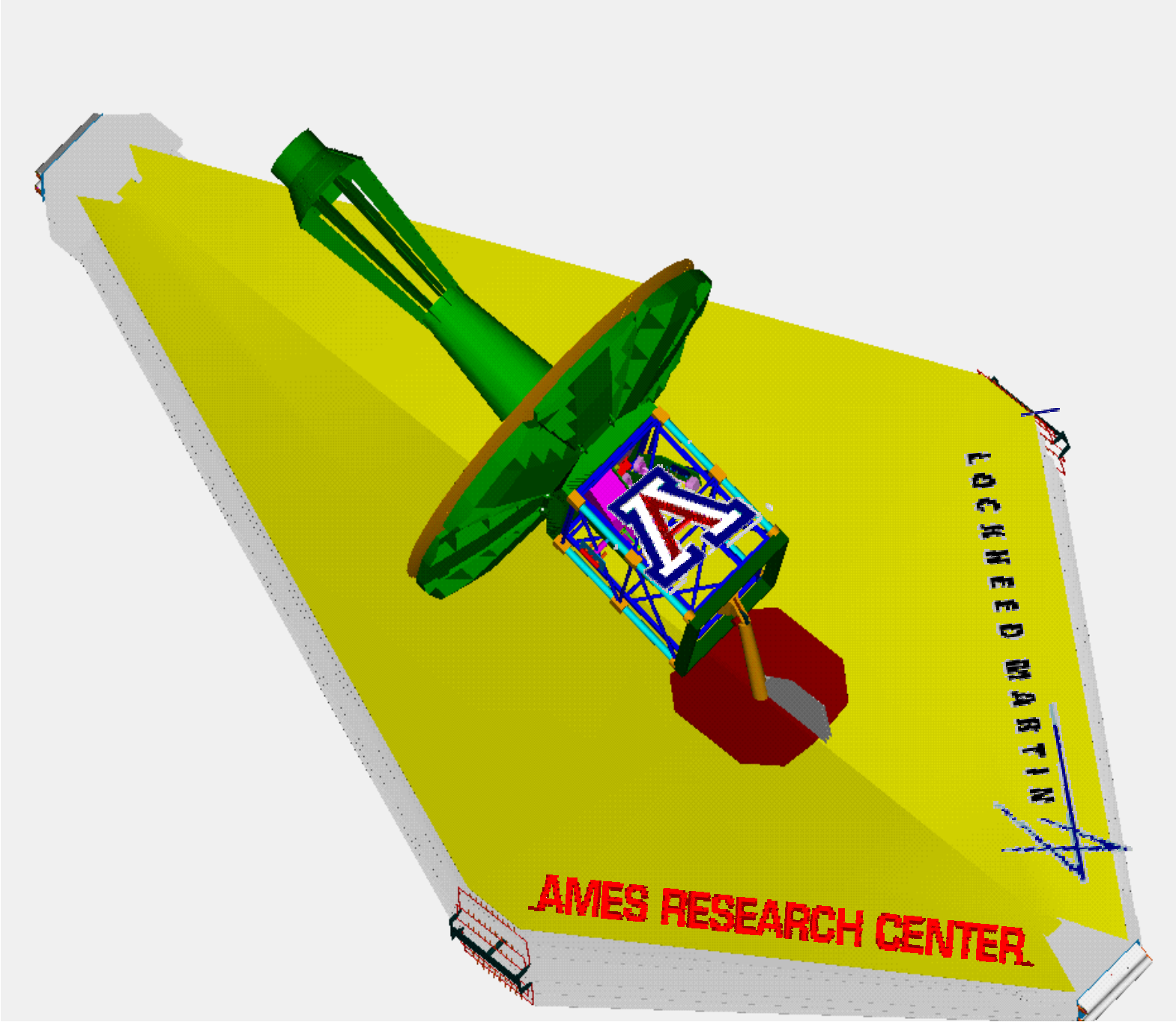
An Integrated Science Instrument Module For the Next Generation Space Telescope



An Integrated Science Instrument Module For the Next Generation Space Telescope

A study conducted by

The University of Arizona,
Ames Research Center, &
Lockheed Martin Corporation
For
NASA Goddard Space Flight Center

Authors:

K. Ennico
J. Bechtold (co-PI), T. Greene (co-PI)
J. Burge, L. Lesyna, M. Menzel, L. Naes, T. Nast
G. Rieke, M. Rieke, R. Sarlot, R. Thompson, J. Uphoff

Edited by G. Rieke
August 26, 1999

Executive Summary

This study has two goals: 1) to explore specific trade issues in sufficient depth both to support development of our ISIM concept and to bring them to the attention of the NGST project and, 2) to develop a comprehensive integrated science instrument module from an independent viewpoint to allow comparison with the GSFC yardstick.

In §1 we list the science goals for NGST and show how they are best supported by extending the wavelength range of the instruments from 0.4µm to 35µm. We also discuss specific trades that guided the architecture of our proposed ISIM. §2 shows that development of Si:P BIB detectors will enable high performance to 35µm without placing additional requirements on the ISIM. It also shows the potential advantages of using HgCdTe arrays in the near infrared and silicon PIN diode arrays in the visible. There are two important conclusions in §3 on coolers: 1) the vibration inherent in linear cooler compressors (e.g., for pulse tube coolers) can be managed so that such devices will not compromise the telescope pointing, and 2) a solid hydrogen cooler of adequate capacity is easily compatible with the other constraints of the ISIM. In §4, we discuss multi-object spectroscopy. First, we show that an imaging Fourier Transform Spectrometer (IFTS) has little or no inherent advantage for near infrared spectral mapping of the sky at low spectral resolution (R<300), compared with an optimized dispersive spectrometer (DS). The dispersive spectrometer is far superior in single source spectroscopy and hence would provide superior capabilities overall. Second, we derive constraints on the performance of slitlets for a multi-object DS. In the near infrared, the slitlets must reject off-slit radiation at a level of 1000:1 or greater to prevent contamination of source spectra by spectra of nearby objects. For wavelengths longer than 5µm, the required rejection becomes greater because of the increase of background flux across the spectral range of the DS. In §5, we show that speeding up the NGST Cassegrain f/ratio from f/24 to f/12-f/16 would probably allow an improved ISIM.

In §6, we propose a specific ISIM. Our philosophy is to create an instrument package composed of multiple, simple sub-modules. To the greatest extent possible, we limit the number of observing modes and capabilities of each sub-module to optimize scientific performance, and at the same time, minimize complexity, cost and risk. §7 and §8 discuss the advantages of the modularity we have proposed for the ISIM and the readiness of the key technical items in the design, both of which are important boundary conditions for estimating the cost. We estimate the cost of our ISIM in §9.

The conclusions from this study are summarized in §10. The scientific capabilities of our ISIM are significantly greater than those of the yardstick, particularly in providing extended wavelength coverage down to 0.4µm and out to 35µm and improved spectroscopy in the mid and longwave infrared. In addition, our ISIM could complete the NGST Design Reference Mission in only 2/3 the time required by the yardstick. With suitable optimization of the telescope as well as of the instruments, the cost of our proposed ISIM is similar to the estimated cost of the yardstick. We believe that adoption of the concepts developed in this study can therefore substantially improve the performance of NGST.

Table of Contents

***** Contents of these sections may be ITAR-sensitive.

1.0 ISIM defining science programs

This section addresses the scientific motivations for the NGST integrated science instrument module (ISIM).

1.1 Prioritized list of NGST science goals

The Ad-Hoc Science Working Group (ASWG) was formed by the NGST project to explore the range of science that could be achieved with NGST. A set of sample projects, the Design Reference Mission (DRM), has been developed to test the capabilities of NGST against probable actual observing programs. Some of these programs grew out of the science goals in "HST and Beyond" (the Dressler report) while others are newly proposed by the ASWG. The ASWG identified five science themes for NGST:

I. Cosmology and Structure of the Universe
II. Origin and Evolution of Galaxies
III. History of the Milky Way and its Neighbors
IV. Birth and Formation of Stars
V. Origin and Evolution of Planetary Systems

All of the DRM programs address at least one of these themes. These programs expand on the science outlined in the Dressler report and illustrate that NGST potentially can impact much of modern astronomy. Of course, the potential for unanticipated discoveries cannot be included explicitly but provides additional excitement given the advance in capabilities that NGST would make possible.

The DRM is intended to be illustrative of how much could be accomplished in the first 2.5 years of NGST observing, with the remainder of the time spent on general observer programs. However, in designing NGST it must be borne in mind that the DRM may not actually be what the telescope does nor is it necessarily the only set of observations that would be done in the early part of the mission. The launch of NGST is sufficiently far in the future and significant advances will have occurred beyond our current understanding of the Universe, for example through both Chandra and SIRTF. The NGST science program must be kept flexible to address the new science goals that will emerge.

1.1.1 Core DRM

Because the full DRM covers a broad range of topics, the ASWG classified the DRM into three broad categories during its 1999 June 4 Chicago meeting. Proposals were deemed to be either: 1) core (meaning essential to mission success), 2) highly desirable, or 3) programs which were scientifically compelling but which should not drive the design of the telescope and instruments. Categories 2 and 3 were not fully statistically distinguishable. The following proposals constitute the core science mission for NGST:

1. The Formation and Evolution of Galaxies I: the Deep Imaging Survey (Lilly et al.) (Theme II)
2. The Formation and Evolution of Galaxies II: the Deep Spectroscopic Survey (Lilly et al.) (Theme II)

3. Mapping the Dark Matter Distribution at High Redshift with NGST (Schneider et al.) (Theme I)
4. Probing the Intergalactic Medium out to the Reionization Epoch (Jakobsen et al.) (Theme I)
5. Measuring Cosmological Parameters with High-Z Supernovae (Kirshner & Madau) (Theme I)
6. The Formation and Evolution of Galaxies V: Obscured Star Formation at High Redshifts (Rieke et al.) (Theme II)
7. The Physics of Star Formation: Understanding the Youngest Protostars (Greene et al.) (Theme IV)

1.1.2 Additional high priority DRM

At this meeting the ASWG determined that a number of other science programs were also scientifically compelling and should be considered but not overly drive the design of NGST. These programs include:

8. The Age of the Oldest Stars from the Faint End Slope of the White Dwarf LF in Globular Clusters (Rich et al.) (Theme III)
9. Detection and Characterization of Jovian Planets and Brown Dwarfs in the Solar Neighborhood (Trauger et al.)(Theme V)
10. Evolution of Circumstellar Disks Around Young Stars: The Search for Gas and Remnant Dust (Meyer et al.) (Theme V)
11. The Evolution of the Cosmic Supernova Rates (Madau et al.) (Theme II)
12. The Origins of sub-stellar Mass Objects: Probing Brown Dwarfs and Extra-Solar Planets in Star Forming Regions (Meyer) (Themes IV and V)
13. The Formation and Evolution of Galaxies III: Cluster Galaxies (Lilly et al.) (Theme II)

The remainder of the DRM proposals were determined to be scientifically compelling but should not influence the design of NGST. These proposals are

14. Observing the IR Transients of Gamma-Ray Bursts and their Host Galaxies (Stockman et al.) (Theme I)
15. Microlensing in the Virgo Cluster and the Role of Baryonic Dark Matter in the Universe (Rix) (Theme I)
16. The Formation and Evolution of Galaxies IV: The Relation between Galaxy Evolution and AGN (Lilly et al.) (Theme II)
17. A Complete Initial Mass Function for Old Stellar Populations (Rich and Margon) (Theme III)
18. The Ages and Chemistry of the Oldest Stellar Halo Populations (Rich and Margon) (Theme III)
19. A Study of Cool, Field Brown Dwarf Neighbors (Liebert et al.) (Theme III)
20. Dynamics and Evolution of the Interstellar Medium: Cosmic Recycling (Bally et al.) (Theme IV)
21. A Survey of the Trans-Neptunian Region (Nicholson and Gladman) (Theme V)

22. Measuring the Physical Properties of Kuiper Belt Objects (Rieke and Nicholson) (Theme V)
23. Explorations in Astrobiology: Evolution of Organic Matter from the ISM to Planetary Systems (Greene et al.) (Theme IV and V)

This classification was arrived at by a subset of the ASWG and does not reflect how interconnected some of these proposals are. For example, the study of circumstellar disks in "The Physics of Star Formation: Understanding the Youngest Protostars" and in "Evolution of Circumstellar Disks Around Young Stars: The Search for Gas and Remnant Dust" are closely connected to the study of the Solar System's disk proposed in "A Survey of the Trans-Neptunian Region."

1.1.3 Non-DRM (7.5 / 10 years)

Because the scientific planning for NGST cannot yet reflect what will be discovered by either Chandra or SIRTF, or in other ways, many uses of NGST cannot be anticipated now. NGST will present such an increase in capabilities over previous missions and facilities operating in the 0.6 to 30μm region that most of the return from NGST will come from areas not covered by the DRM. A guest observer program similar in scope to those executed from HST, Chandra, and SIRTF should be assumed.

1.2 ISIM science goals

Table 1.2-1 lists the core DRM programs and their required instrument capabilities:

DRM Program	Major Required Instrument Capabilities	Comments
1. Deep Galaxy Imaging	Wide Field NIR Imaging (1/2) Wide Field vis and 10μm imaging (1/2)	112 days with 4'×4' FOV
2. Deep Galaxy Spectroscopy	R=100, 1000, multi-object and R>3000 long slit at λ=3.5μm R=5000 at λ=10μm	98 days 20 days
3. Dark Matter	Wide Field NIR Imaging	192 days for 4'×4'
4. Probing the IGM	R=100 NIR spectroscopy	10 days
5. High z Supernovae	WF NIR Imaging	some spect; 1 week follow-up
6. High z Obscured Galaxies	λ=8-36 μm Wide Field Imaging R=300 Multi-object MIR spectroscopy	54 days / at least 2'×2' FOV 10 days
7. Physics of Protostars	λ=15-35μm Imaging R≥3000 spectroscopy λ=6-30μm	70 days, single object observations

Table 1.2-1 Core DRM programs and their major required instrument capabilities.

Programs 1, 5, 6, and 7 all require diffraction-limited imaging with well-sampled PSFs. Note that some of these programs will use the same datasets for different goals. For example, the imaging for Program 1 can be the same as for Program 5 if executed with correct time spacing, and much of the survey for Program 2 can also be the same as for Program 1.

Following the lead of the ASWG, we weigh the observing requirements of these core DRM program programs most heavily. The second group of DRM programs (numbers 8-13 in §1.1.2) place the following extra requirements beyond those imposed by the core programs as summarized in Table 1.2-2.

DRM Program	Major Required Instrument Capabilities	Comments
8. The Age of the Oldest Stars	Optical imaging, suppression of Diffraction spikes	
9. Detection and Characterization of Jovian Planets and Brown Dwarfs	Near-IR coronagraph	
10. Evolution of Circumstellar Disks	R~30000 spectroscopy from 4.5-12µm	
11. Measuring the Rates of Supernovae		Requires reasonably rapid followup capability
12. Origins of Substellar Mass Objects	NIR imaging	
13. Form. & Evol. of Galaxies: Clusters	Multiplexed near-IR spectroscopy at R~1000 and R~5000	

Table 1.2-2 Major required instrument capabilities for additional high priority DRM programs.

Again, many of the proposals in this category need diffraction-limited imaging over the entire 2-30µm range (especially Programs 9 and 12).

The only additional requirements imposed by the third category of proposals are the need for imaging at V with as good a PSF as possible including suppression of diffraction spikes (Program 18) and R~30000 to 100000 over 1-30µm for Program 20. Program 20 would also require a 10'×10' imaging field of view (FOV) and enhanced data downlink rates to support a large area survey.

Finally, we expect non-DRM programs to require more high spatial resolution observations of individual objects than identified in the DRM. They will also probably place greater emphasis on the single-source capabilities of the ISIM than will the large surveys in the DRM. We assign lower weight to the requirements of these programs than to those of the core DRM.

We now discuss in more detail the rationales for the choices of observing modes and optical parameters for which have driven our ISIM concept.

1.2.1 Imaging trades

By dividing the instrument into a number of modules, and generally limiting the spectral range of each module compared to the yardstick ISIM design, we can provide pixel scales that sample the NGST Airy disk more optimally than in the yardstick. A separate mosaic of arrays has been dedicated to large area surveys, rather than having a single pixel scale at each wavelength try to serve the conflicting purposes of high angular resolution and large areal coverage. The issues in this compromise are discussed in the following section.

1.2.1.1 Field of view vs. PSF sampling

Sampling of the point spread function (PSF) always involves a compromise between 1) obtaining a large field of view, and hence tending to undersample and, 2) trying to preserve all the spatial frequencies delivered by the telescope to get the highest possible angular resolution, and hence sampling at least at the Nyquist frequency (i.e., pixels of size $\lambda/2D$ or smaller). In our ISIM study, we have shown the possibilities for both approaches, although it seems likely that more difficult choices would be required for a real instrument design. Further discussion of the design can be found in §6.1.

In designs that require a deformable mirror (DM), the well-corrected FOV is limited by a second order coupling between the DM and field. Assuming an ideal system, with 8-m aperture and 92 mm DM, the DM magnification is 0.0115. At a 5 arcmin field, the cosine effect is 1-cos(.083°/0.0115) = 0.008, so if the primary mirror has errors of 5 μm rms, corrected on axis by the DM, images at 5 arc min will see mirror errors of 0.04 μm rms, or wavefront errors of 0.08 μm rms. That is, the system will be diffraction limited only to about 1.1μm even if there are no other error terms, and realistically the system would not be diffraction limited short of 2μm. This behavior scales with $(\text{field})^2$ times (mirror error).

A very large field of view can be provided by the telescope if all the wavefront correction is achieved with the primary (and/or secondary) mirror, so there is no requirement for a DM (§5.1). In this case, a significant portion of the focal plane can be occupied directly by a mosaic of detector arrays. One or more mosaics of detector arrays can achieve wide field imaging over fields of up to ~6×8 arcmin (in our design). In a direct mode such as for this wide field option, the projected pixel scale is determined purely by the Cassegrain f/ratio of the telescope (and primary diameter) and the pixel size of the array. With an 8-m telescope at f/24, the scale is about 1.1 marcsec/μm, and the image diameter (diffraction limited at 2μm) will be about 50 marcsec, so reasonable pixel scales can be achieved with available arrays that have pixels in the 10-30μm range. This direct mode is limited in flexibility since it does not allow for insertion of grisms or other optical elements into the beam, and hence a steerable focal plane would be needed to feed other cameras with these features.

Another advantage of a fully corrected telescope is that it is possible to place additional optics near the focal plane to form a pupil that is occupied by a steering mirror that directs the beam into different instrument modules. Because of the large corrected field, good imaging can be achieved over a large range of steering at the pupil. This pupil is the logical place for the DM in designs that require one. If the pupil is occupied by a DM, a much

more complex design is required to recover the large field offered by a well corrected telescope with a simple steering mirror at the pupil (if it can be recovered at all). Thus, in general DM-based designs will have difficulties packaging instruments and in sharing the field of view.

The large field imaging is important for many of the high priority programs in the DRM, including: 1) studies of weak lensing to search for dark matter, 2) supernova observations to determine cosmological constants, 3) imaging of early galaxies, 4) searches for hidden IR galaxies and AGNs, and 5) KBO searches. Many of these surveys will produce interesting objects to be observed more extensively at the maximum angular resolution achievable by NGST and also in more filter bands than can be provided in the large field, survey imager. Examples include: 1) interesting weak lensing galaxies, 2) morphology of early galaxies, and 3) morphology of "hidden" IR galaxies with obscured AGNs and starbursts. Other high priority DRM objectives requiring high resolution imaging are: 1) morphology of protostars and their surrounding regions including forming disks, 2) planet searches, and 3) debris disk evolution.

On basic principles, one expects to have to sample at the Nyquist frequency to retain all the information in a diffraction limited image, as would be desirable for these followup studies and others. With larger pixels, some information can be recovered by dithering the image on the array. However, the high spatial frequencies lost by underpixelization cannot be recovered in this way. In fact, simulations of resolution boosting image processing show clearly that the ultimate resolution suffers unless the sampling is at or above the Nyquist level (e.g., Bippert-Plymate et al. 1992). We have therefore included modules in the instrument that project the pixels to $\lambda/2D$ or smaller scales at the spectral midpoint of the module coverage. These units provide also for spectroscopy and multi-band filter imaging. They can be complemented by wide field modules optimized for large field surveying and for which Nyquist sampling is not a high priority.

The high resolution modules are required for maximum angular resolution where the telescope is diffraction limited. At the wavelengths short of the diffraction limit, the image properties will depend on the specific nature of the wavefront delivered by the telescope. In general, though, there will be a significant component at the highest spatial frequencies even with an image at a wavelength a factor of two or three short of the diffraction limited one. Figure 1.2.1.1-1 is illustrative. A series of artificial images was generated by assuming that the surface errors were distributed over the low order Zernike coefficients, as shown in the phase error map. The wavefront error provides the diffraction limit at 2µm. However, the image retains a very sharp core at 1µm.

The image characteristics of NGST short of its diffraction limited wavelength of 2µm are not known at this time. The shortest wavelength at which roughly Nyquist pixel scales are appropriate needs to be determined after more is understood about the imaging properties of the optical system.

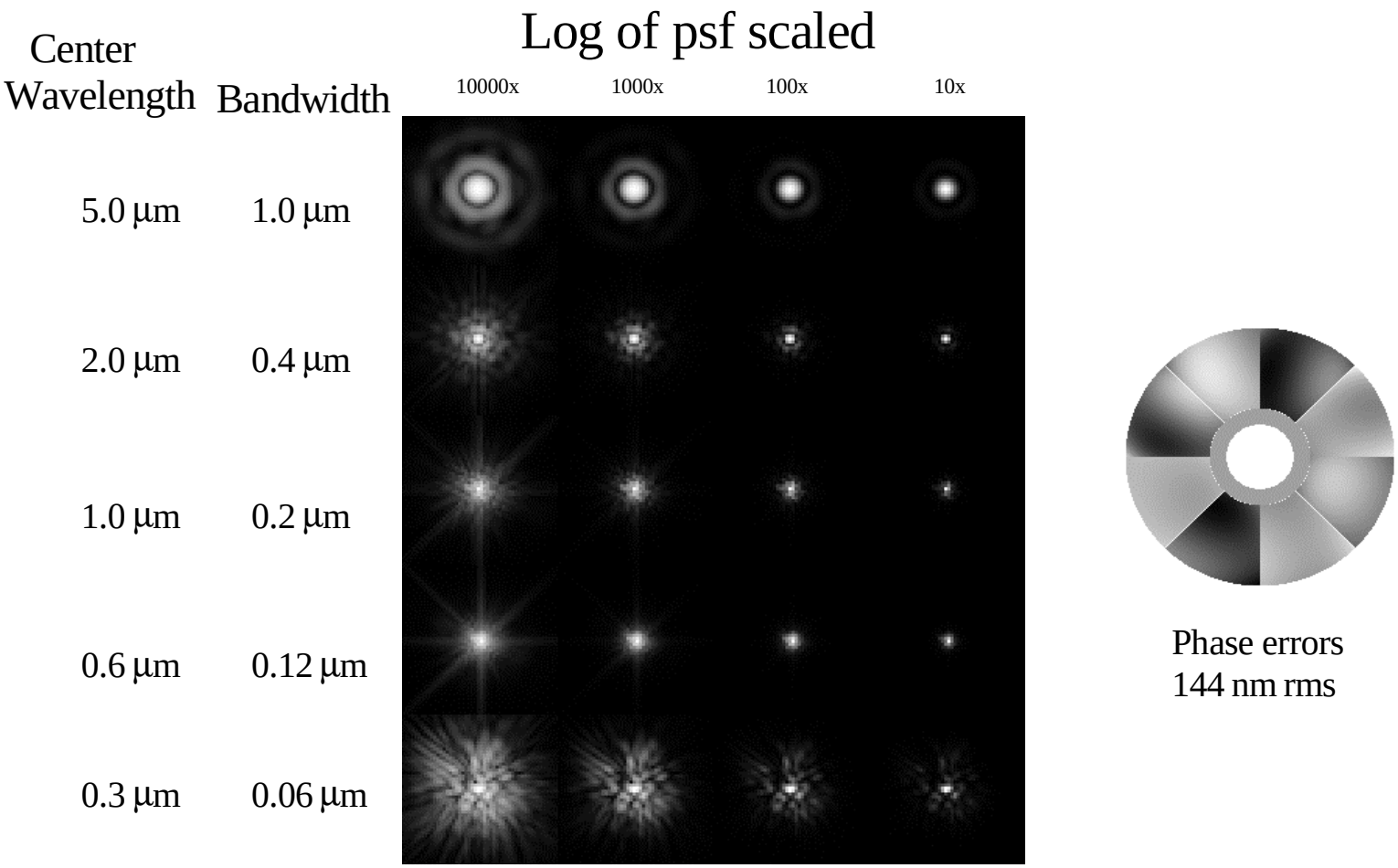


Figure 1.2.1.1-1 Example of possible trend of point spread function with wavelength.

1.2.2 Spectroscopy trades

The DRM is split roughly 50/50 between spectroscopy and imaging. The top seven DRM programs emphasize: 1) multi-object R=100 and 1000, and 5000 single object NIR (1-5μm) spectroscopy, 2) multi-object R=300 MIR (5-10μm) spectroscopy, and 3) single object R>3000 6-30μm spectroscopy. We have responded to these drivers in our ISIM concept.

Because no DRM program calls for optical spectroscopy, we only address 1-35μm spectroscopy in this section.

1.2.2.1 Detector performance

To determine the effect of instrument resolution versus detection limits, we modeled the expected NGST telescope, sky, and instrument backgrounds and computed instrument sensitivities. Details of the components of this model, including assumptions about detector parameters and instrument losses, are given in §12.1. The model is for the MIR compatible architecture, with a 40K telescope primary (ε=0.05) and a 93K sunshade (effective ε=1e-4). The modeled telescope is zodiacal light limited for $\lambda \leq 12\mu m$ for imaging and low resolution (R=100) spectroscopy. At longer wavelengths, the telescope & sunshade thermal backgrounds dominate.

We considered detector performance on the basis of the cosmic ray density near solar minimum, 4 cm^{-2} s^{-1}. Assuming that a typical hit affects four pixels, and that the NIR pixels are 18μm on a side, we concluded that the maximum useful on-chip integration time was 3000s. Larger pixels will lead to shorter maximum integration times. For example, with 30μm pixels the maximum is about 1000 seconds. Thus, arrays with similar nominal specifications (read noise, dark current, quantum efficiency) may lead to significantly different performance in read-noise-limited observations because of pixel size. In general, the smallest pixels are strongly favored for these applications.

With the 3000 second maximum, and assuming a dispersive spectrometer with λ/2D pixels, NGST will be detector (dark current + read noise) limited for $R \geq 300$ in the NIR (1-5μm), even with good detectors, having dark current 0.02e/s and read noise 5e. At this level, the dark current is as serious a limitation as read noise for sky background limited observations. In the short MIR (5-8μm), a similar situation holds if the detectors have dark current 1e/s and read noise 15e.

Thus all proposed moderate and high resolution spectroscopy for NGST at $\lambda < 10\mu m$ is potentially limited by detector performance and further investments in reducing dark current and read noise would yield significant dividends. For example, improving the NIR detector parameters to 0.01e/s dark current with 1.5e effective read noise will decrease the observing time for a 10σ, 10,000s (t_{exp}=1000s) background limited point source detection at 1μm by factors of 1.8-2.5, for R=100-3000, respectively.

It also follows at these wavelengths that the lower the resolution required to meet the science objectives, the better, for it minimizes the detector noise in the continuum. R=100 appears to be adequate for the majority of the proposed programs. Moderate resolutions R=300-500, will come

at a penalty in signal-to-noise unless the detectors are improved from the I_d = 0.02 e/s and read noise = 5e rms level.

For wavelengths 8-12µm, the resolution at which the observatory becomes detector limited increases to R~1000, until the thermal background from the sunshade and primary telescope begin to dominate at longer wavelengths. At these long wavelengths, already achieved dark currents and read noises more than suffice for NGST to be photon noise limited. Additionally, rebinning at high resolution in this photon limited regime recovers the SNR (sensitivity) of low resolution spectra. Therefore, no additional low resolution grisms are required if a long wavelength module possesses high resolution observing modes.

1.2.2.2 Spectrometer implementation

For reasons discussed in §4.1, we assume that the spectrometer implementation will utilize dispersive elements. In a space flight instrument, where complexity must be kept to a minimum, it is attractive to use grisms mounted on a camera filter wheel to add a spectroscopic capability, usually with a slit that can be placed at the telescope focal plane near the entrance to the instrument. Where the free spectral range is smaller than the width of the array and grisms are relatively easy to make with modest angle prisms, we prefer this approach. We will also show advantages from using cross dispersed grisms for high resolution spectroscopy.

For the higher resolution programs, R≥3000, various problems emerge with grism systems. For example, increasing the resolution from R=3000 to R=5000 over the 1-5µm region, will require 8 grisms of very high index material (n>2.4) to cover the entire region, compared to 2-3, as proposed in our strawman ISIM (§6.1). For KRS-5 (n(1-5µm)=2.4), the prisms required to match R=5000, will be too steep, so higher index material such as silicon (n=3.4) or germanium (n=3.9) will be required (or the designs would have to grow to allow for larger collimated beams). These materials are more difficult to rule than KRS-5, although it may be possible to machine them with asymmetric etches. §12.5 gives more information about grism restrictions.

Thus, for high resolution (R>3000) studies, the instrument design becomes rapidly more complex, either requiring very high index grisms with steep angles, larger pupils, or a switch to a reflective grating type spectrometer. Higher resolution will also require that the instrument either be increased in size to allow for the larger detector area for the full spectrum, or for the inclusion of cross dispersing elements. We have therefore avoided these high spectral resolutions wherever possible to simplify the instrument and minimize cost. In particular, the very high resolution case R>10,000 desired for some of the second tier DRM programs was not considered because of their lower priority and the degree to which satisfying this goal drove the instrument designs.

1.2.2.3 Spectral resolution vs. spectral range

The free spectral response of a grism/grating based spectrometer will be the ultimate limiting factor in wavelength coverage, except in the case at the highest resolution where the detector size forms the limit. At the lowest order (m=1), the largest wavelength coverage is one octave. Grisms/gratings operating at higher orders will have smaller spectral coverage. Therefore, several grisms/gratings will be required to cover the spectral range. This point is illustrated by Figure 1.2.2.3-1, which shows the number of first order grisms it would take to cover the 1-5μm range for a series of resolutions. Two detector sizes, 1024×1024 and 2048×2048 pixels are compared.

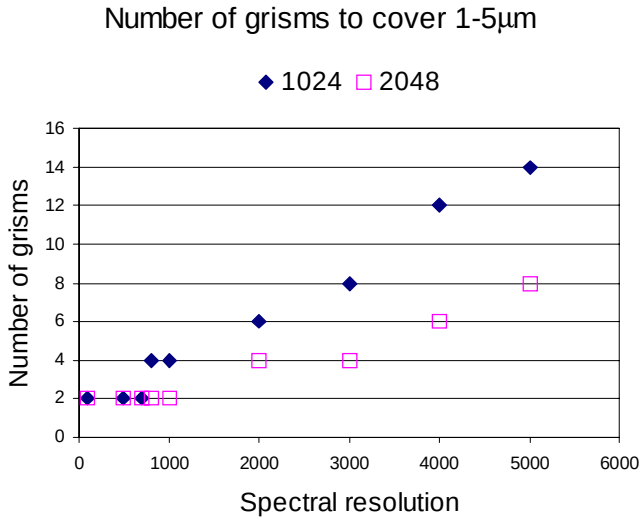


Figure 1.2.2.3-1 Number of grisms required to cover the entire 1-5μm range working in first order, where the maximum spectral coverage is one octave.

Since the faint galaxy DRM program asks for R=100,1000, and ~3000, its objectives can be met with a single 2048×2048 detector over the 1-5μm region for an acceptable number of grisms. NIR instruments composed of mosaics of 1024×1024 devices will suffer from gaps in recorded spectra for the desired highest resolutions. When operated in multi-object mode, when objects will be scattered across the FOV, the full FOV cannot be used without losing parts of spectra.

1.2.2.4 Multi vs. single object spectroscopy

Regimes for single slit and multi-object spectroscopy is illustrated by Figure 1.2.2.4-1. Single object long slit spectroscopy will be the preferred choice for: 1) spectroscopy of unclustered extended objects, 2) spectroscopy of unclustered isolated objects, and 3) high resolution spectroscopy of objects requiring a large spectral coverage. For the last of these options, designs move toward image slicers (maximum accuracy) and cross- dispersed echelles (maximum wavelength coverage).

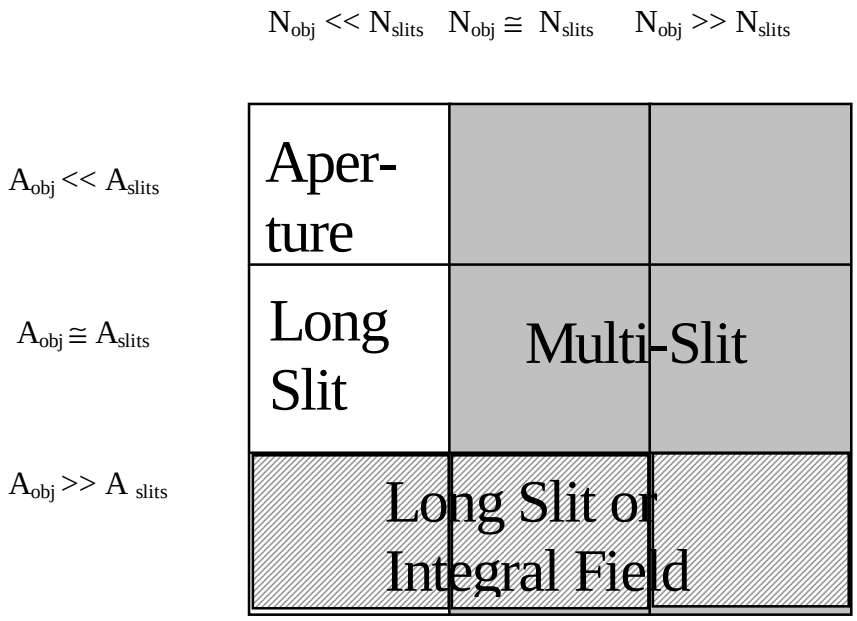


Figure 1.2.2.4-1 Slit spectroscopy regimes. The most challenging case will occur when there are roughly as many slits as objects and the objects fill the slit field of view (central box). N is the number of sources in the field of view, or the number of slits provided by the instrument. A is the typical area of a source on the sky, or the area of a slit.

Several key science programs for NGST will require spectroscopy of many objects, relatively close on the sky. As indicated in Table 1.2-1, these programs are concentrated in the NIR, and we believe the application for this type of observation may be underestimated in the DRM. The instrument design response to these programs is large field of view detectors with multi-object spectroscopic (MOS) capabilities. A slitless spectrometer, as currently used on the NICMOS HST instrument, will have a significant noise penalty at R<10,000 compared to multi-slit schemes, and therefore is not included in our ISIM concept.

Since several MOS schemes, both mechanical configurations for dispersive spectrographs and the IFTS, are being investigated by other study groups, we have not focused on the details of these different approaches. However, in §4 we address important performance trades for these MOS approaches.

Over the 5-35µm range, where the NGST background is increasing rapidly, the technical requirements to provide multi-object spectroscopy become more difficult (see §4.3). In addition, the DRM does not depend strongly on this capability. In general, the lower density of objects on the sky within the range of NGST spectroscopy tends to put requirements in the left third of Figure 1.2.2.4-1, often in the upper left corner. With anticipated maximal detector sizes of 1024×1024 pixels over this wavelength region, a pair of cross-dispersed grating spectrometers will best provide the desired spectral coverage. Details of the proposed modules are found in §6.1.6 and §6.1.8.

1.2.3 Short wavelength extension

An 8 m space observatory offers tremendous, order-of-magnitude improvements beyond HST in sensitivity and spatial resolution for imaging at wavelengths less than 1 μm. Simulations have shown that NGST is expected to have a good PSF with a near-diffraction-limited core in the 0.5 – 1 μm range (see Figure 1.2.1.1-1). Obtaining similar performance using adaptive optics from the ground is technically daunting. In any case, groundbased AO is unlikely to yield a sufficiently stable PSF for many investigations and will probably operate over only a limited FOV. Both aluminum and overcoated silver are candidate mirror coating materials that have good reflection from below 0.5 microns to the far-IR. High performance visible wavelength Si PIN diode detector arrays will be available in large formats (2k x 2k pixels) and have operating temperatures and readout electronics compatible with near-IR arrays (see §2.1).

Thus it is technically feasible for NGST to have good performance at wavelengths below 1 μm. HST will operate for only a couple years after the NGST launch (until ~ 2010), so it is prudent to equip NGST with visible light instruments. NGST will have large visible light performance advantages over ground-based 8-10 m telescopes in the realms of wide field imaging and high resolution / high contrast imaging where a stable PSF is required. We now summarize two compelling scientific programs which would be enabled by these capabilities.

1.2.3.1 The UV properties of the youngest objects

One of the key science drivers for NGST is the searching for and understanding the very first generation of objects, proto-galactic fragments (PGFs) and the first galaxies. Figure 1.2.3.1-1 shows a composite spectrum of a star-forming galaxy as a function of age, and identifies the wavelengths of key spectral features. Initially, the hot UV emitting stars and nebular emission dominate the observed spectral energy distribution (SED), but as the galaxy fades, the 4000Å break becomes stronger and within 10^7 years red supergiants emerge and the SED shows a prominent near-infrared bump with strong CO absorption (Bruzual & Charlot, 1993). A mid- and far-infrared excess develops as dust is produced and heated in the ISM.

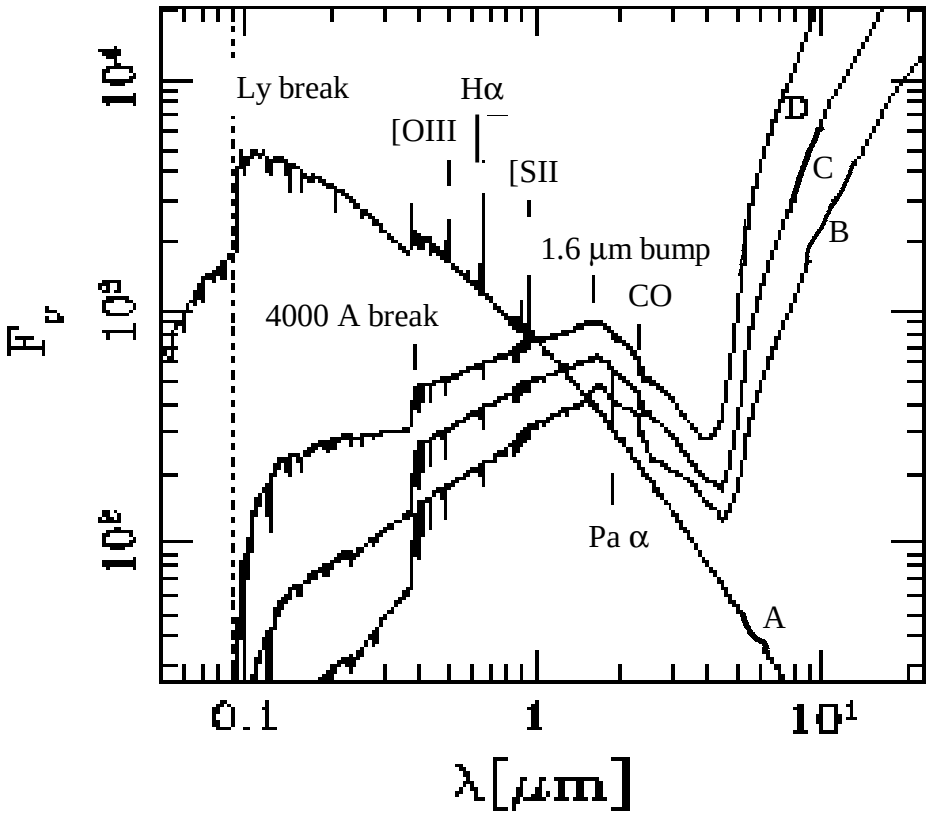


Figure 1.2.3.1-1 Galaxy SED as a function of age. A is a Bruzual-Charlot model at t=3×10^6 years (metallicity 0.004 solar, single burst) with no dust extinction and therefore no far-infrared flux (Bruzual & Sharlot, 1993). B shows the same model at t=1×10^7 years with A_v=1 and far-infrared flux scaled to equal the energy absorbed at short wavelengths. The Calzetti extinction law has been used (Calzetti et al. 1994). C is the same as B, but for t=5×10^7 years. D is the same for B, but for t=1.3×10^8 years. Flux scale is arbitrary.

The first task for a PGF survey is to estimate the redshift of the detected objects. Figure 1.2.3.1-1 shows that these young objects will exhibit a dramatic "break" at the Lyman limit over a range of ages. This break is detected via multi-band photometry. Since the Lyman limit (912Å) falls short of 1μm for z<10, imaging in the visible must complement longer wavelength observations.

In addition to searching for Lyman breaks in the 4 < z < 10 realm, visible light imaging is required to study the UV energy density of galaxies in the 2 < z < 5 range. Ground-based and HDF studies suggest that the energy density of the universe reaches a maximum in the rest frame wavelength region near 2000 Å, so wide field visible imaging with NGST will be required to investigate this further. The high priority DRM proposal "The formation and evolution of galaxies I: The deep imaging survey(s)" (Lilly et al.) also identifies the need for these observations.

1.2.3.2 The ages of the oldest stars

Globular clusters are the oldest objects in the Galaxy. The ages of their stars sets a lower limit to the age of the universe, yielding an age estimate for the universe that is independent of cosmological observations and parameters. Globular cluster ages have been determined thus far by main-sequence turnoff fitting, but this technique is somewhat imprecise due to uncertainties in hydrogen burning stellar models and metallicity effects. NGST will have the sensitivity, FOV, resolution, and stable PSF required to observe white dwarfs in globular clusters, and this will allow estimating their ages by comparison to white dwarf cooling curve models. Thus this technique will offer yet another rare independent method of dating the oldest stars and hence the universe. This concept is developed further in DRM proposal 8, "The age of the oldest stars from the faint end slope of the white dwarf LF in globular clusters" by Rich et al.

1.2.4 Long wavelength extension

We have evaluated a number of approaches to extending the wavelength response of NGST toward 40μm. Based on advantages in operating temperature and manufacturability, we recommend that the response be set by using Si:P BIB detectors, as discussed in more detail in §2.3.2. Here, we summarize the science some of benefits that can be achieved with these devices, which will provide useful response to beyond 34μm. In addition to the examples given below, the long-wavelength extension will also be extremely useful in a variety of star formation studies.

1.2.4.1 "Hidden" IR galaxies

The complex of fine structure lines in the mid infrared is illustrated in the ISO spectrum of the Circinus Seyfert galaxy shown in Figure 1.2.4.1-1. They allow fundamental determination of nebular properties in heavily obscured environments, such as ultraluminous infrared galaxies (e.g., Genzel et al. 1998), starburst galaxies (e.g., Roche et al. 1991) or compact HII regions (e.g., Megeath et al. 1990). The ratio of [SIII] lines at 33.5 and 18.7μm is very useful in studying the densities in the interstellar medium. In addition, the [NeII], [NeIII], and [NeV] lines at 12.8, 15.6, and 14.3μm provide a virtually extinction-free probe of the nature of the UV spectrum responsible for ionizing the gas in a dusty source. For example, they have been used with ISO to probe whether ultraluminous galaxies are excited by AGNs (Genzel et al. 1998). With Si:As, this set of lines can be studied to $z \sim 0.8$. Since galaxy luminosity evolution proceeds at least as fast as $(1+z)^2$, and AGN evolution as $\sim(1+z)^4$, the ability with Si:P to study these lines to $z \sim 1.2$ extends our exploration to 2 – 5 times greater luminosities (assuming pure luminosity evolution). Si:P arrays will therefore give access to a significantly more interesting phase of evolution for starbursts and the nuclear activity in galaxies than is accessible with Si:As.

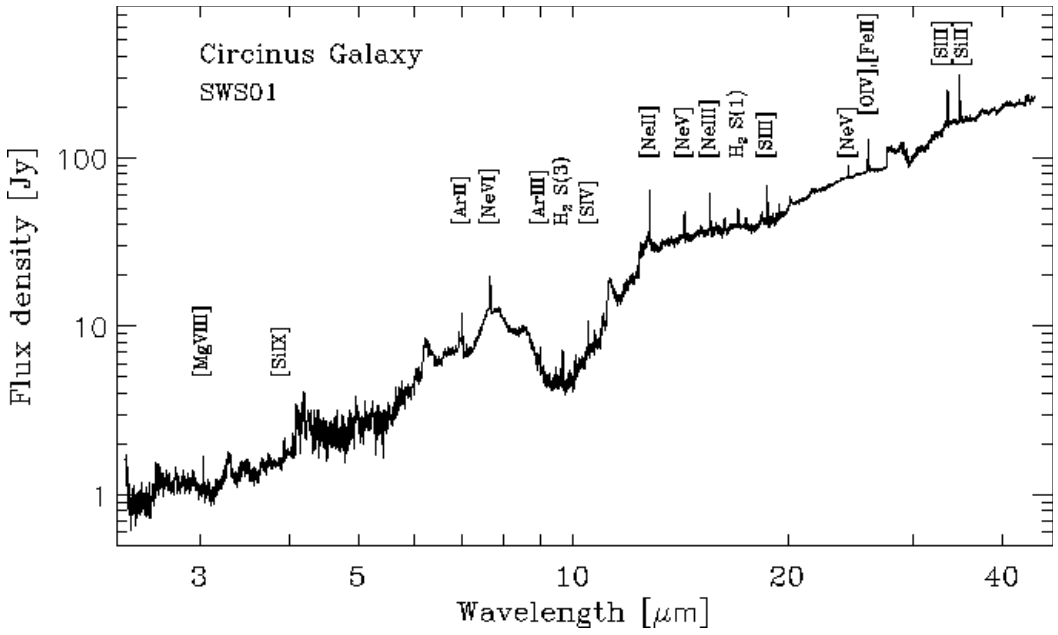


Figure 1.2.4.1-1 ISO-SWS spectrum of the Circinus Seyfert galaxy, showing the prominent mid infrared fine structure lines. The [SIII] line at 33.5μm and [SII] at 34.86μm are just in range for Si:P. There are no additional significant fine structure lines before 50μm, except for a weak [NeIII] line at 36μm.

1.2.4.2 Mineralogy of planetary debris disks

An interesting discovery in the ISO spectroscopy of planetary debris disks is the series of mineralogical emission features extending to 34μm. The bands at the longest wavelength end of this range are important because their relative strengths can distinguish different forms of crystalline silicates in the debris disk grains. For example, olivines have a peak near 34μm, whereas pyroxenes have one near 33μm; ISO data show sources with either or both peaks and find variations in the shape of the band from source to source (Waters 1998). Another potentially critical advantage is that the stellar photosphere is relatively faint at the long wavelengths, giving a favorable contrast ratio of disk to stellar emission. These points are illustrated in Figure 1.2.4.2-1. As shown there, Si:P detectors enable study of the mineralogy of debris disks around evolved stars.

1.2.4.3 Radiometry of Kuiper Belt Objects (KBOs)

The large number of Kuiper Belt Objects being discovered implies that they constitute an important constituent of the solar system. In addition, as small bodies outside the zone cleared by planets, they probably represent the largest members of the debris system of the sun, analogous to the more dramatic debris systems that lie around some nearby stars such as Vega and β Pic. Although it is now known that KBOs have a range of colors in the visible, the albedos can only be guessed and hence the diameters and masses of the individual objects are very uncertain. Radiometry provides a measure of albedos when combined with measurements of reflected light

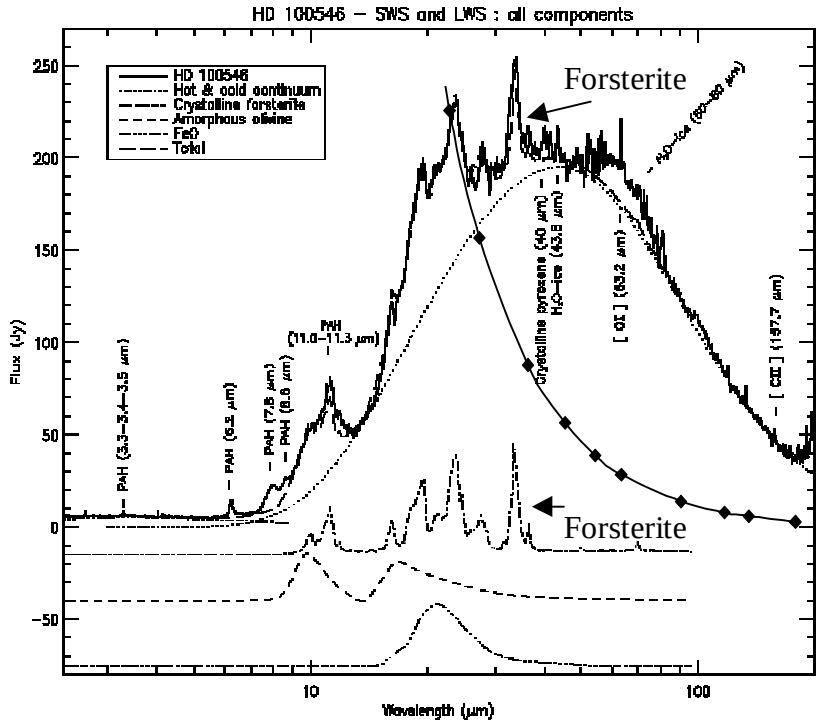


Figure 1.2.4.2-1 Spectrum of the extreme AeBe star HD 100546, which has a very prominent infrared excess with the spectral feature at 33-34μm indicative of forsterite-type crystalline silicates (Malfait et al. 1998). The response of Si:P detectors extends just far enough to detect this feature; there are no additional diagnostic features to 40μm that would be accessible to Si:Sb. The line with diamond markers is the Rayleigh-Jeans spectral energy distribution of a typical primary star. For extreme objects like HD 100546, the full suite of features is accessible. However, for a much less prominent disk typical of those around evolved stars, the 33-34μm features have by far the best contrast to study the mineralogy of the disk.

such as in the visible. Figure 1.2.4.3-1 shows the predicted results for KBOs at the inner edge of the Kuiper Belt (~40AU) and of various combinations of diameter and albedo that give identical signatures in reflected light. It also shows the predicted flux for a low albedo object in the middle of the Kuiper Belt. Fluxes are plotted as νF_ν to indicate where the maximum energy emerges. Although the brightest examples are detectable on NGST with a Si:As array, the Si:P detector has significantly greater capability to study objects over a range of albedo and distance from the sun.

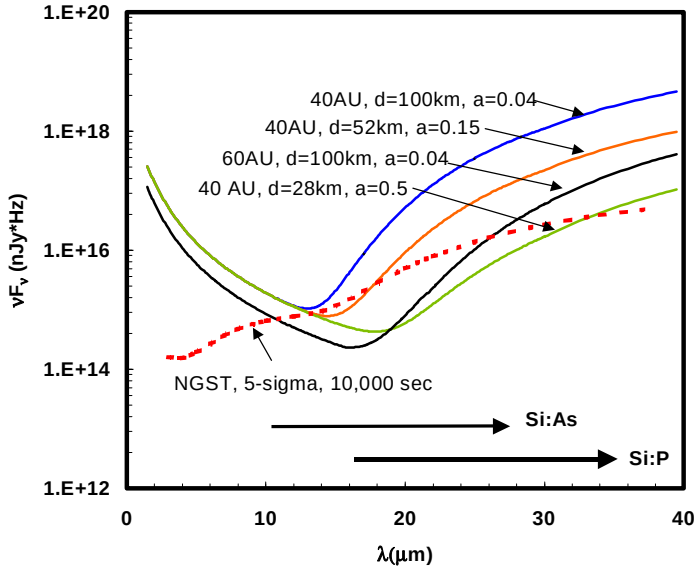


Figure 1.2.4.3-1 Detectability of KBOs as a function of albedo and distance from the sun.

1.2.4.4 Debris disk systems

Following on the results in Figure 1.2.4.3-1, typical debris disk systems have central holes that greatly reduce the surface areas radiating in the middle infrared, so these remnants of planetary system formation are most detectable in the far infrared. The very cold temperatures of the constituent objects make a small longward shift in wavelength potentially important in their detection and imaging (the beam of NGST would be 0.8" at 30μm, corresponding to 8AU at 10 pc). The extension from 27 to 34μm provided by Si:P could be critical in studying low density systems around nearby stars (analogous to the Kuiper Belt) and in probing systems with large inner clearings, where one could deduce that planetary systems had formed over a range of distances from the star.

1.2.4.5 H_2 and the ISM

The 28.2μm line of H_2 is virtually the only way to determine the amount of gas in a cold, T < 100K, dense cloud such as a collapsing pre-protostar, a circumstellar pre-planetary nebula, or even a typical giant molecular cloud. Search for and study of this line is prominently featured in the NGST design reference mission for this reason. Unfortunately, Si:As BIB detectors have lost virtually all their response at the rest wavelength of the line. However, Si:P is near its peak response.

Figure 1.2.4.5-1 illustrates one application of this line. It is one of the very few ways to detect and measure the properties of very cold gas in the interstellar medium. The conversion of indirect measurements of molecular hydrogen densities to total masses has been the source of controversy for decades. With measurement of the 17 and 28.2μm lines, it can be constrained directly. In the example in Figure 1.2.4.5-1, the ratio of line strengths between 17 and 28.2μm requires a molecular gas temperature of about 100K and a density of about 3000 cm^{-3} (Thi et al. 1998a).

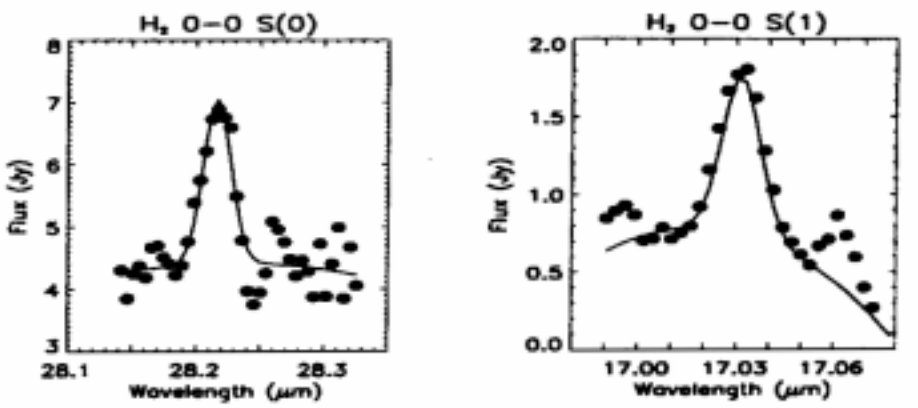


Figure 1.2.4.5-1 Pure rotational H_2 emission lines from the peripheral regions of S140, from Thi et al. (1998a)

Figure 1.2.4.5-2 illustrates another, very different application. It shows H_2 from the circumstellar disk of the very young star GG Tau. The molecular hydrogen originates in about 0.007 solar masses of relatively dense and cold material at distances up to 100 AU from the central star. These lines provide a unique probe of this regime, within which planetary systems are likely to be forming.

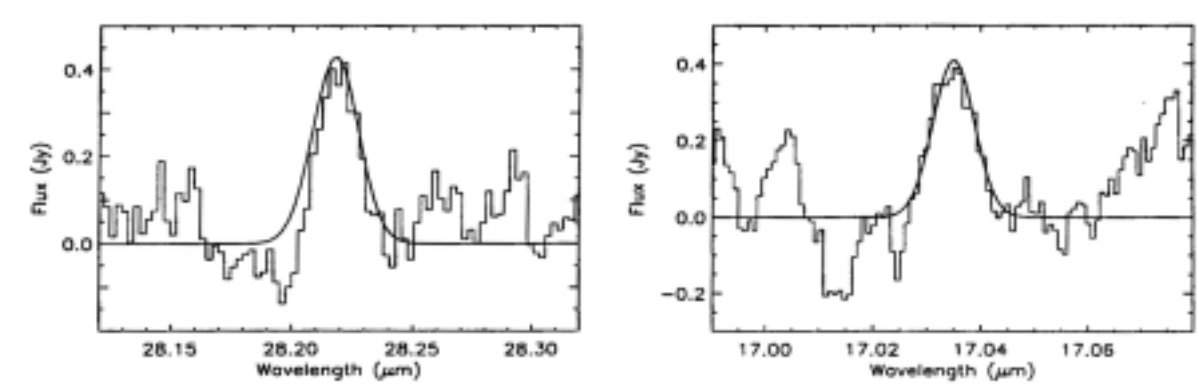


Figure 1.2.4.5-2 Pure rotational molecular hydrogen in GG Tau, from Thi et al. 1998b.

2.0 Detectors

Our ISIM concept uses a suite of several detector types. We feel that suitable detector choices can increase the science yield within a fixed budget relative to the yardstick. This section summarizes the results of our detector trade studies and our preferred detector choices.

2.1 Visible detectors

The yardstick instrument uses an infrared array to provide its visible range sensor. Although this approach has the advantage of minimizing the number of array types, it has a number of disadvantages.

First, to provide reasonable matching of the pixel scale to the telescope image, it is possible that the pixels as projected onto the sky should be smaller than would be appropriate in an infrared band. This change in pixel scale is likely to be needed even given that the telescope is not diffraction limited short of 1 micron, since the image will contain significant information at high spatial frequencies that otherwise would be lost (see §1.2.1). Implementing the smaller pixel scale would require either a scale change in the optics (requiring a mechanism), or a dedicated optical train and array.

Second, the infrared array may not provide equivalent performance to an optimized visible-wavelength array. Although InSb can have high quantum efficiency in the visible region, it will probably have higher capacitance than an equivalent optical detector, degrading read noise. Finally, if the yields on the InSb arrays are lower than those on visible arrays, it is possible that it will be more economical to use a suitable optimized visible array, even if it introduces a new array type into the NGST mix. Some of these arguments will become clearer from the following discussion.

We believe the tradeoff between the yardstick approach and the use of an optimized visible-range array must begin by selecting the best type of array for the visible, since it may be possible to retain most of the advantages of the "infrared-array-only" approach of the yardstick without accepting its disadvantages. In addition, the constraint of good visible performance is removed from the infrared array, which may allow a more optimum choice for it also.

There are three candidate array technologies for optimized visible detectors: CCDs, CMOS detectors and Si PIN diodes.

2.1.1 CCDs

CCDs are the most familiar of the candidates. They offer very large formats (e.g., 4k×4k pixels or larger), low noise, high quantum efficiency in the visible and blue, and high fill factor. Disadvantages include their susceptibility to radiation damage, poor red quantum efficiency (although deep depletion devices offer excellent potential in this area), the fact that they must operate near 150K (posing a significant challenge for the thermal design of the ISIM), and the fact that their controller requirements are quite different from those of infrared arrays, potentially increasing warm electronics cost.

2.1.2 CMOS

CMOS detectors are a rapidly emerging technology in which silicon diodes are grown side-by-side on a single chip with readouts. Compared with CCDs, they should have greater radiation hardness, should operate colder, and with greater red quantum efficiency.

The largest existing CMOS imagers have 2k×2k pixels, and in the NGST time frame will conceivably have 4k×4k pixels. Their main drawbacks are poor fill factors (50%) and relatively high noise (>10 electrons). Thus, although they can be more red-biased in response than most CCDs, they may actually have lower net sensitivity. In some cases, these devices are equipped with microlenses to achieve better fill factors. However, it is not established that the microlenses will survive thermal cycling.

CMOS detectors that would have much more promise for NGST are under development at JPL (G. Domingo, private communication). The initial goals are to supply quantum efficiency of 75%, read noise < 5 electrons, and a format of at least 1k×1k. The conflict between space for the readout amplifier and the detector itself will be resolved by back illuminating the devices. If this development is successful, these arrays could become the visible light sensor of choice for NGST, since they should have high yields and hence relatively low prices, good radiation tolerance, and control requirements compatible with the same electronics that would be needed for infrared arrays.

2.1.3 Si PIN diodes

Direct hybrid PIN diodes are a technology being pursued by Rockwell Science Center, Raytheon SBRC, and other detector manufacturers. Their construction is closely analogous to an infrared array, except the detectors are PIN diodes. The detector array is bump bonded to a silicon readout and back illuminated. These devices have excellent red quantum efficiency (see Figure 2.1.3-1) and high fill factors, their readouts can be identical to those for infrared arrays, they are resistant to radiation damage, and they can operate at the low temperatures optimum for infrared arrays. Their disadvantage is that control of the forces required for hybridization may limit their formats to 2k×2k, similar to the largest formats for infrared arrays. Pixels would be about 20µm in size.

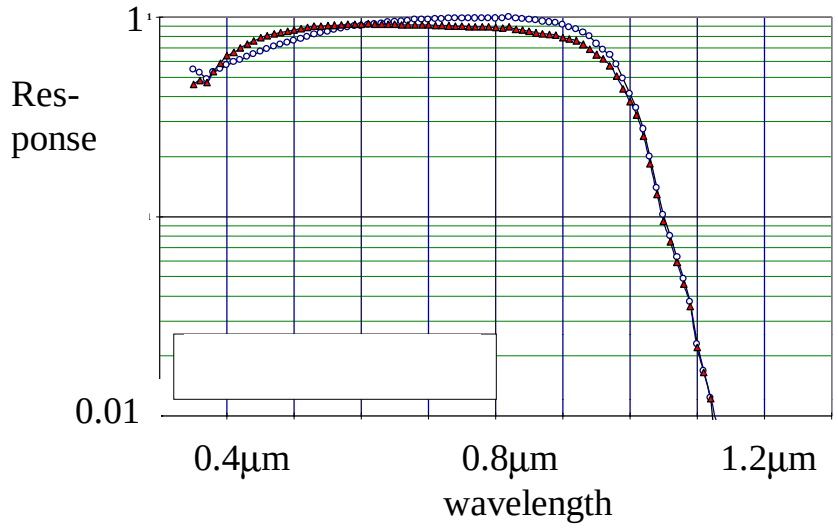


Figure 2.1.3-1. Silicon PIN diode arrays can have excellent red response (from Raytheon SBRC).

2.1.4. Conclusion

Comparing these detector array types, the best choice appears to be the direct hybrid PIN diodes. Next generation CMOS imagers might be a better choice if current developments are successful. We now compare use of these dedicated visible imagers with use of an infrared array for the visible detector. The readouts and control electronics could be identical, so the incremental cost of the PIN diode array would be qualifying its detector material. Fortunately, silicon PIN diodes are a very well developed and conservative approach, so these costs should be relatively minor. The PIN diode array would probably have better performance, with lower read noise and at least as high quantum efficiency. Because of the built-in field in a PIN diode, cross talk would be reduced. We expect with this solution overall cost would be reduced, since the control electronics could be identical for the infrared and visible array and the manufacturing cost for the PIN diode array is likely to be less than for an equivalent number of infrared arrays to be used in the visible. Therefore, we recommend that the detector suite for the ISIM include direct hybrid PIN diode arrays.

2.2 Near infrared detectors

Two general detector types deserve consideration for the near infrared: InSb, as in the yardstick concept, and HgCdTe. The InSb is better established for the 1-5µm in terms of single-pixel performance. It also has broader spectral response, with good quantum efficiency from 5µm down into the visible. However, there has been recent rapid progress on HgCdTe in this regard. HgCdTe is more advanced than InSb in terms of large array formats, demonstrated read noise, and array manufacturing yields. In addition, it may provide adequate performance at higher operating temperature than InSb, a potential advantage in system design. Another advantage of HgCdTe is its cutoff wavelength, and hence dark current, which can be tuned for specific applications. Thus the NIR modules could use detectors with 2.5-3µm cutoffs, having much lower dark current than InSb. If a dedicated detector type is used in the visible, then HgCdTe deserves serious consideration for the near infrared.

A substantial advance in HgCdTe detectors has been achieved at Rockwell Science Center with molecular beam epitaxy double layer crystal growth. The best performance is achieved with a CdZnTe substrate, which is a good match to the HgCdTe crystal structure. The pixel yields in the MBE HgCdTe tend to be relatively low but are improving as experience is gained in process control and optimization. The first generation of the devices with 5.2µm cutoff has dark currents in 40µm pixels operated at 50K of 0.4e/sec (Bailey et al. 1998). We have independently confirmed dark currents <1e/s at this temperature with additional devices (D. McCarthy, private communication). A 1024×1024 array with 4.2µm cutoff on a silicon substrate (expected to result in elevated dark currents) had mean quantum efficiency of 58%, and 99.6% operable pixels, with dark current <1e/s per 18µm pixel (K. Vural, private communication). Recent data on a second generation 5.2µm cutoff array with 18µm pixels have found dark currents of 0.04 e/sec at 63K; this device had 75% quantum efficiency from 1 to 4µm (D. Hall, private communication). These values are close both to the goals for NGST and to the per-pixel performance of InSb. The demonstrated manufacturability of HgCdTe in 2k×2k formats, higher operating temperature, and lower read noise all are important advantages that make it potentially the detector of choice for

NGST given the demonstrated performance and rate of improvement in per-pixel performance and yields in the past three years.

2.3 Mid infrared detectors

Several DRMs call for imaging and spectroscopy longward of 5µm. The detector choices available, however, all have to be cooled below that of the passively cooled temperature of the ISIM (~35-40K). We present here the results of our choices for >5µm response detectors. Coolers are discussed in §3.

2.3.1 Si:As BIB arrays

Si:As BIB is by far the most common IBC detector. It has response to about 27.7µm (here and in the following we define the limit of useful response to be 10% of the peak). It requires cooling only to about 10K, and can be built in conventional semiconductor facilities because arsenic is a common dopant for integrated circuit construction. A competitor for these detectors between 6 and about 12µm is MBE HgCdTe. These detectors have a potential advantage in higher operating temperature over Si:As BIB devices, but they do not cover the whole spectral range and their demonstrated performance in terms of dark current and in producability is not as good. At present, we believe Si:As BIB detectors are the best candidate for the NGST detectors working at wavelengths longer than 6µm out to its cutoff wavelength just short of 28µm. With further development of long wavelength HgCdTe and of Si:P BIB detectors (see below) there could be an interesting option to use only these two detector types, if it eased overall thermal design and provided useful performance from 6 to 34µm.

2.3.2 Si:P BIB arrays

For use in SIRTF, where superfluid helium cooling is available, Si:Sb BIB detectors have been developed (van Cleve et al. 1995) that provide response to about 39.3µm, but require cooling to <4K and depart from conventional semiconductor processing. Battelle Institut (1988) demonstrated a Si:P device and single detectors using this same dopant have been built at Boeing. Si:P BIB detectors have a cutoff wavelength of about 34.5µm and can be built through conventional semiconductor processing, since phosphorus is another commonly used dopant. The response of these detector types is compared in Figure 2.3.2-1.

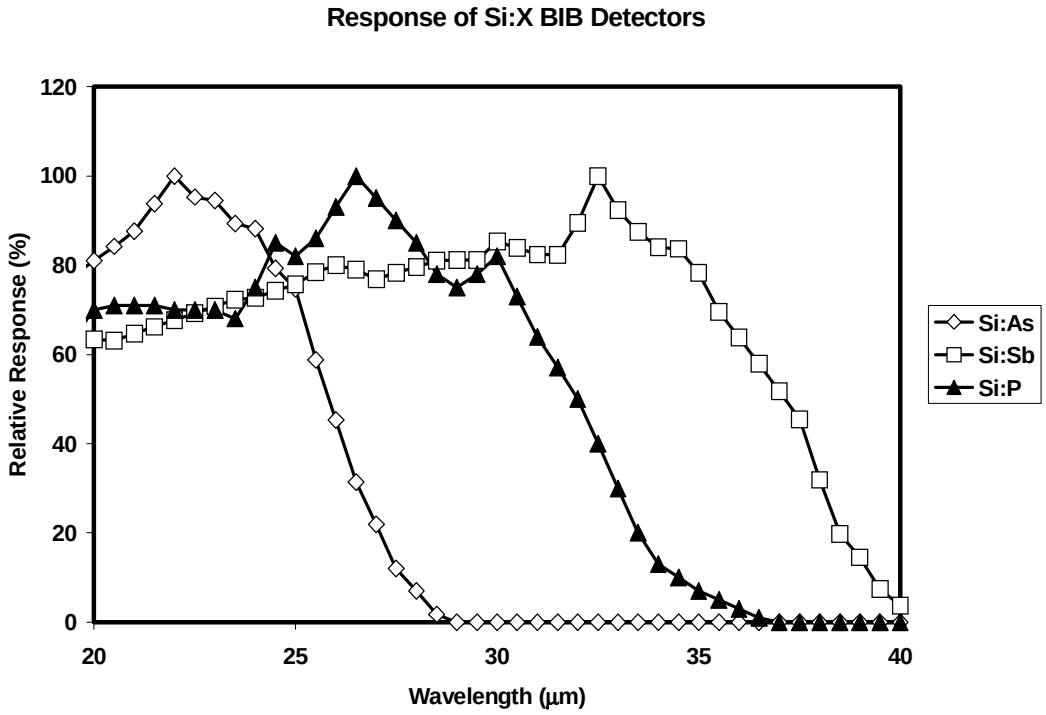


Figure 2.3.2-1 Response of Si:As, Si:P, and Si:Sb BIB detectors. The Si:As and Si:Sb devices were built and measured at Boeing; the Si:As detector has an anti-reflection coating to enhance its long wavelength response as much as possible. The Si:P device was manufactured and measured at Battelle Institut. Modest further enhancement of the long wavelength responses of both the Si:P and Si:Sb could be achieved with AR coatings.

2.3.2.1 Technical and system level advantages of Si:P

Since NGST has the potential for spectacular performance at wavelengths well beyond the zodiacal background limit performance (λ~12μm) (§1.2.4, §12.1.2 (Figure 12.1.2-1), our ISIM concept includes these long wavelength detectors. Although Figure 2.3.2-1 implies that Si:Sb BIB detectors would be the most attractive because of their response 39μm, for a number of reasons associated with manufacturability, we recommend that Si:P be used instead.

2.3.2.1.1 Semiconductor fabrication mainstream

The chemistry used for doping silicon with phosphorus is generally available because this doping is used commonly in the semiconductor industry. Thus, the raw material for detector fabrication can be obtained from a number of suppliers. Specialty houses such as Lawrence Semiconductor Research Laboratory (LSRL) can make these materials with excellent control of background impurities, opening up new possibilities for very high quality detectors.

As an example, Boeing has shifted some of the fabrication of Si:As wafers to LSRL. The background impurity levels are three times lower than Boeing was able to achieve with in-house material fabrication, making possible a new generation of detectors that probe new levels of performance. By lowering the As concentration and increasing the thickness of the IR-active

layer, it has been demonstrated that detectors can be built with very low dark currents (~1e/s) at much higher temperatures than previously (11K). The ultra-high purity levels achieved at LSRL allow full depletion of the thicker IR-active layers, so high quantum efficiencies can be achieved with these devices. Eventually a limit in elevated temperature performance is set by the increase in device response time.

In addition to the improved yields and performance made possible by use of mainstream semiconductor capabilities, there is a second benefit. There can be a continuing struggle to maintain the capability for production of relatively low volume, specialized devices such as Si:X BIB arrays. For example, so far as possible the fabrication of the readouts for such arrays has been shifted to large volume, commercial foundries which maintain a stable business base. Experience has shown that smaller, specialized foundries (e.g., Hughes Carlsbad) are hard to maintain, and when they close the processes need to be re-established elsewhere – an expensive and potentially risky procedure. By moving the detector wafer fabrication into the commercial arena, the maintenance of the long wave detector capability will be made easier than with Si:Sb.

2.3.2.1.2 Better controlled processing

The Si:Sb detector material is difficult to grow because there is no good gaseous source for chemical vapor deposition with silicon sources. Instead, Boeing "bubbles" a liquid source and with this process there can be difficulties in dopant control and uniformity. In addition, the processing must be at relatively high temperature. These requirements on the processing also add to the difficulties in growing ultra pure devices. Thus, Si:Sb cannot be advanced to the next generation of performance, which requires higher purities and thicker layers than used at present.

In contrast, the Si:P material can be prepared with the same, relatively low temperature, silicon deposition process that is used successfully with Si:As and which has made possible the improved performance next-generation devices.

2.3.2.1.3 Operating temperature

The latest generation of Si:As BIB detectors, best suited to NGST requirements, can be operated at temperatures of 9 to 10K. Assuming the first order relation that the operating temperature should be proportional to the excitation energy, one would expect to be able to operate Si:P up to 8K. We expect with high purities and optimization of the detector design, this limit can actually be achieved. The same excitation energy argument suggests that Si:Sb BIBs could be operated up to 7K. However, a more typical limit is only about 4K (van Cleve et al. 1995). Although further optimization may bring some improvements for this material, the processing issues discussed in §2.3.2.1.2 represent a fundamental barrier to development of Si:Sb BIBs that work well at 7K.

The difference between cooling to 8K and to 4K can have an immense effect on system design. For example, as illustrated by the development for the WIRE mission, it allows use of liquid hydrogen with much greater cooling capacity, rather than liquid helium. Alternately, it opens the possibility for a broad variety of closed cycle coolers of high reliability. Although none of these advantages are fundamental, they promise to yield lower cost and more reliable overall sensor systems.

2.3.2.1.4 Radiation damage

Si:X BIB detectors are subject to damage when struck by ionizing particles, which can produce crystal defects, reduced impedance, and a "hot" pixel with increased dark current. This behavior is greatly reduced if the detector can be operated at low bias voltage. However, to deplete the infrared-active layer and maximize the detector quantum efficiency requires a bias voltage proportional to the minority impurity concentration in the infrared active layer. Thus, for previous generation detectors with relatively high minority impurity levels, a tradeoff must be made between minimizing the radiation-induced dark current and maximizing the radiometric performance of the detector. This tradeoff became widely apparent when the Si:As BIB detectors in the Short Wavelength Spectrometer (SWS) of ISO exhibited increasing dark currents in flight, until the bias was reduced with a resulting downward revision of the instrument performance.

The solution to this class of problem is to produce very high purity infrared-active layers. This approach has been implemented very effectively for Si:As BIB detectors, with the best behavior obtained with the super-pure detector material from Lawrence Semiconductor Research Laboratories (Raytheon IBCs also use very high purity processing and have relatively little radiation damage). Thus, at the typical exposure levels for a space astronomy mission, radiation damage is not a serious issue for Si:As detectors. However, the reduced control of impurities in the processing required for Si:Sb has made a similar solution for it impossible. The arrays of this material to be flown in the Infrared Spectrograph (IRS) on SIRTF will degrade over the life of the mission.

Because they can be produced with equivalently high purities to the Si:As detectors, we expect Si:P detectors to be much more radiation hard than Si:Sb at the dose levels of interest.

2.3.3 Conclusion

Extended long wavelength capabilities beyond 5µm are highly desirable in achieving several NGST science goals, as well as opening up new scientific objectives (§1.2.4). Over the 5-25µm range, we feel the highly developed Si:As IBC detectors are the most sensible choice. Longward of 25µm, when the Si:As response begins to deteriorate, Si:P IBC detectors are the most promising technology due to 1) their higher operating temperature, 2) more radiation robust, and 3) ease of manufacture.

3.0 Coolers

The minimum temperature required for Si:As and Si:P BIB detectors is about 8K. We have considered four ways to maintain this temperature: 1) turbo-brayton, 2) sorption, 3) pulse tube; and 4) solid hydrogen. The turbo-Brayton and sorption approaches are described in detail in other project literature. We summarize briefly the discussion of them by Castles (1998) and DiPirro (1999). We have concentrated our attention on the latter two types. Each of the four cooling approaches has its own unique advantages and disadvantages. The major results of this section of our report are: 1) we find that the vibration inherent in the linear compressor used with pulse tubes can probably be managed so as not to disturb the pointing of NGST, and 2) with the compact packaging of our instrument modules and a demonstrated qualifiable heat switch, a solid hydrogen cryostat is a viable option. The advantage of either of these two approaches is that they are fully demonstrated with relatively little further technology development required.

3.1 Turbo-brayton

Because of their turbo compressors, turbo-brayton coolers are basically vibration free. Thermal design with these coolers can be eased because the compressor and its associated power dissipation can be placed a significant distance from the cold head. A turbo-brayton cooling system permits 8 K temperature regions to be placed with great freedom in the ISIM, as plumbing lines can direct the fluid coolant to appropriate locations. A space rated version is planned for use in reviving NICMOS, which will be cooled to the ~70K range. Similar devices are used down to ~30K. The major issue with turbo-brayton coolers is the need for further technical development (for example, miniaturization of critical turbine components) and demonstration to satisfy the NGST requirement for cooling to ~8K. At present, only a single company, Creare, makes space qualified turbo-brayton coolers.

3.2 Sorption coolers

Space qualifiable sorption coolers are also basically vibration free. System thermal design can be eased with these units as with turbo-braytons because the warm (~400ºC) compressor unit can be remote from the cold head. Space qualifiable units are under development for the Planck mission, but they will only reach ~20K. Sorption coolers use more power that the other three approaches. Further development will be required to satisfy the NGST requirement for cooling to ~8K.

3.3 Pulse tube coolers

Pulse tube coolers require a linear compressor which can be a source of vibration. As a result, they have been generally not considered for use in NGST. They do however have major advantages: existing technology for single stage systems, compact integration, low mass and good power efficiency. Although commercial pulse tube coolers have been operated as low as 2.5 K, some development is required to optimize and demonstrate compact pulse tube systems at 8 K. We discuss pulse tubes at some length in this report as they have not been extensively investigated for NGST applications. LMMS has demonstrated performance on two stage pulse tube designs and tested the third stage to 8 K, where cooling power of 50 mW was achieved with reasonable input power. Space qualified pulse tubes are also available from TRW.

Relatively little technical development is required for NGST, as discussed in more detail in §8.2. The pulse tube has no cryogenic moving parts and the reliability of the compressor assembly is very high. There is extensive heritage for the compressors from previous cryogenic applications. Based on these previous designs, a failure modes and criticality analysis (FMECA) and fault tree analysis at Lockheed Martin shows that a compressor reliability of 0.99 can be achieved for a 10 year mission.

A system issue with pulse tubes is that the compressor must be placed within ~1 meter of the cold head, so careful thermal design is required to permit radiating the power dissipated in the compressor. In addition, the pulse tube length is limited to a similar size. However, these limitations of pulse tubes are overcome by designing the ISIM to place the parts requiring the lowest temperature operation near an outside edge.

We have analyzed the vibration issue in a number of ways. First, assuming a typical uncompensated vibration level of 0.1 N at the frequency of the compressor operation (~40Hz), one can do a simple rigid body analysis to understand the magnitude of the pointing influence. This calculation places an upper limit of about 1 milliarcsec on the effect on the NGST pointing. A more complete analysis, which models the mechanical structure of the entire LM reference observatory architecture, predicts a pointing error of 10 milliarcsec. This is somewhat higher than the pointing error needed for a diffraction limited telescope at 2μm (5 mas). However, this calculation describes the worst case in which nothing is done to mitigate the vibration.

The uncompensated vibration has a broad frequency spectrum extending above the frequency of compressor operation. Both the fundamental and the higher frequencies can be damped significantly by soft mounting the unit. Figure 3.3-1 shows the results of modeling these effects. A mount with stiffness corresponding to 16Hz is sufficient to reduce the observatory pointing error induced by vibrations to 1 mas. Even with a relatively stiff mount (32Hz resonance), the high frequencies are sufficiently damped that excitation of resonances by them becomes unlikely.

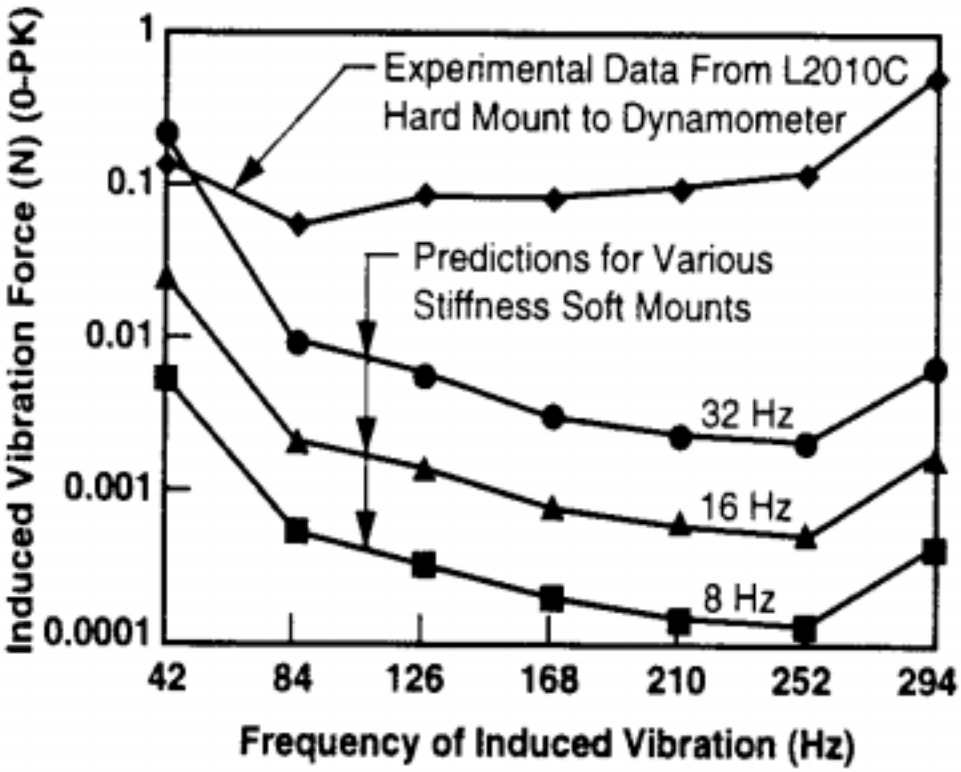


Figure 3.3-1 Vibration of a linear compressor can be substantially reduced by a soft mount.

With a soft mounted pulse tube compressor, then, the issue on vibration becomes resonances excited at the fundamental frequency of the compressor operation. By tuning this fundamental frequency away from mechanical resonances, a variety of observatory architectures may be accommodated. Therefore, we believe that pulse tube vibrations do not pose a fundamental obstacle for deployment on NGST.

Pulse tubes are also expected to be similar or more power efficient than turbo-brayton coolers at or near 8 K. Table 3.3-1 compares the efficiency of LM pulse tubes with that expected from the Creare turbo-brayton cooler (Swift et al. 1999). Direct comparisons are difficult to make because of the different assumptions relating to thermal grounding temperatures and cooling loads, but the predicted power requirement is similar between the two and the system weight is much lower for the pulse tube system.

Item	Turbo-Brayton (expected, from Swift et al.)	LM Pulse Tube for NGST (expected)	LM Pulse Tube (Lab Data)
Heat Rejection Assumptions	No staging at lower temperatures. 300 K radiator	Radiator at 250 K for compressor	Radiator at 230 K for compressor
Cooling Capability	20 mW at 6 K	4 mW at 8 K 127 mW at 15 K	26 mW at 8 K
Compressor Power Consumption	>= 100 W	50 –100 W	50 – 100 W
System Weight	40 kg	16 kg including electronics	8 kg (utilizing existing 2010 compressor)

Table 3.3-1 Comparison of Pulse Tube and Turbo-Brayton Coolers

Pulse tubes are also much more mature than turbo-brayton coolers. Figure 3.3-3 illustrates laboratory data from the third stage of a LM developed pulse tube in which the upper stage temperature is maintained by helium coolant. The cold stage load versus the cold stage temperature is plotted for selected upper stage temperatures of 20 K, 30 K, and 35 K. The cold stage load for a 20 K upper stage far exceeds our ISIM requirement of 5 mW cooling at 8 K. To adapt the LM pulse tube for NGST, the upper stage of cooling would also need to be provided by the pulse tube, thus a multi-stage pulse tube is required. Pulse tube cooling at these higher temperatures is very mature and therefore low risk. Models for predicting and enhancing pulse tube performance are also well established, enabling flexible and realistic design. Electronic control systems are more mature than for turbo-brayton coolers. We also note that space missions within NASA and DoD favor pulse tubes over turbo-brayton coolers in the ratio of about 9 to 1. We expect that pulse tubes will continue to maintain an advantage in flight heritage, reliability, performance and maturity for the next ten years.

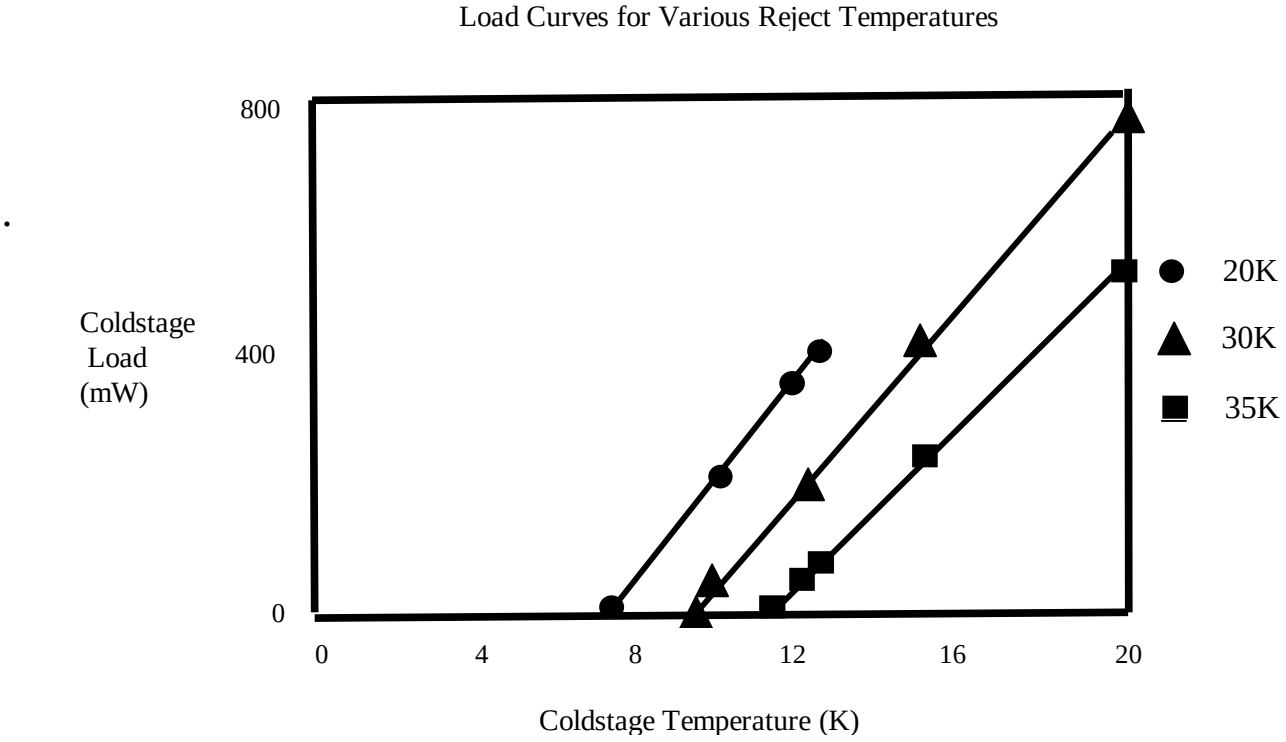


Figure 3.3-3 Performance of a LM Pulse Tube for the indicated upper stage temperatures. Adequate cooling for NGST is provided with an upper stage temperature of 20 K.

3.4 Passive (solid hydrogen) cooling

The temperature requirements can also be met by solid hydrogen. Scaling from the power requirements for the Boeing Si:As BIB arrays in SIRTF, operating the larger format arrays we propose for the ISIM would require only about 2mW (time averaged to 0.5mW if the arrays are used 25% of the time). We estimate other parasitics for a small solid hydrogen dewar to be about another 2mW, given that the dewar will operate in a ~30K environment. If we provide a 50% contingency, an upper limit to the total power into the cryogen is 6mW. One hundred liters of solid hydrogen could provide this cooling power for 10 years. A round cryostat of diameter 0.8m and mass loaded of about 50kg can provide the required amount of solid hydrogen. The efficient packaging we propose for the ISIM, obtained by moving the feed mirrors up and out of the nominal ISIM volume, provides ample room for such a cryostat as shown in Figure 3.4-1.

To avoid having to provide a vacuum shell around the mid infrared instruments, they would be launched warm. A thermal strap will penetrate the dewar shell with a hermetic, thermally insulating feedthrough and will connect to the heat sink(s) for the mid infrared arrays after the instruments are radiatively cooled to the ISIM cavity temperature of 40K. If ultrahigh purity copper is annealed in an oxygen atmosphere, the oxygen getters the residual iron in the copper and RRR values >>1000 can result. Copper treated in this manner produces very high conductivity thermal straps at the low temperatures required by the NGST instruments. Within the

cryostat, the strap will be attached to a heat switch. The switch will remain open until the ISIM has cooled on orbit below 40K. At this point the specific heat of the array mounts (and most other items in the ISIM) will be very low, and the heat switch can be closed to cool the detectors to < 8K with very little load on the cryogen.

A variety of approaches could be used for the heat switch. LMMS has built, tested and developed flight type gas gap and differential contraction thermal switches for a number of different programs. LMMS has tested these for launch vibration and operated them for over 3000 cycles, so their thermal switch technology is very extensive. Another approach was developed by Paul Richards for the adiabatic demagnetization refrigerator originally in the MIPS instrument for SIRTF. It uses a solenoid and a two-armed lever arrangement to close a clamp on the heat strap, establishing a high thermal conductivity between the strap and the clamp jaws. A flight prototype was built at Ball Aerospace and subjected both to vibration testing and to an accelerated life test of 50,000 cycles.

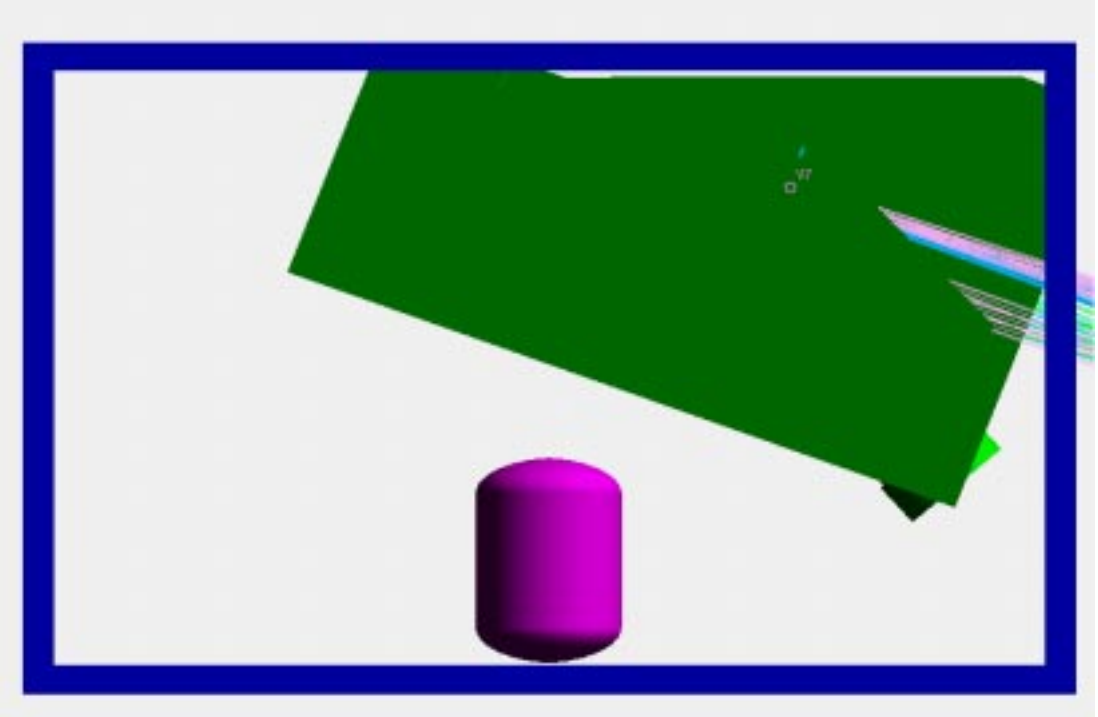


Figure 3.4-1 100 liter solid H_2 cryostat compared to the frame of the ISIM (blue) and the proposed instrument complement mounted in the ISIM (green). Light from the telescope is incident from the right. There is ample space for a solid hydrogen cooler for the mid infrared detectors.

4.0 Multi-object spectroscopy options

Several core DRM programs require multi-object spectroscopy (MOS). A number of study groups are investigating various implementations for MOS on NGST, from multi-aperture dispersive spectrometers to Imaging Fourier Transform Spectrometers (IFTSs). We have therefore concentrated on conceptual trade studies of issues related to MOS options for NGST.

4.1 FTS vs. dispersive spectroscopy

Although discussions of the tradeoff between Imaging Fourier Transform Spectrometers (IFTSs) and dispersive spectrometers (DSs) can become highly mathematical, it is useful to begin with a simple but rigorous treatment based on first principles.

4.1.1 Single object spectroscopy

It is generally agreed that, for spectra of isolated single objects, a DS using state of the art detector arrays is faster to a given result across all the NGST spectral range and envisioned spectral resolutions. The inherent advantage of the DS arises because it need only be read out once for the entire spectrum, whereas the FTS must be read out 2×S times where S is the number of spectral elements. Thus, in the read noise limit the DS has potentially a significant advantage. In the photon noise limit, each spectral element for the FTS is degraded by the noise associated with the photon flux over the entire free spectral range of the instrument, whereas the DS exposes each spectral element only to the photon flux appropriate to that element. An example, calculated with the NGST backgrounds (§12.1) is given in Figure 4.1.1-1. Throughout this section, we quote performance in Johnson magnitudes; at K, $m_{AB} \sim m_{Johnson}+2$.

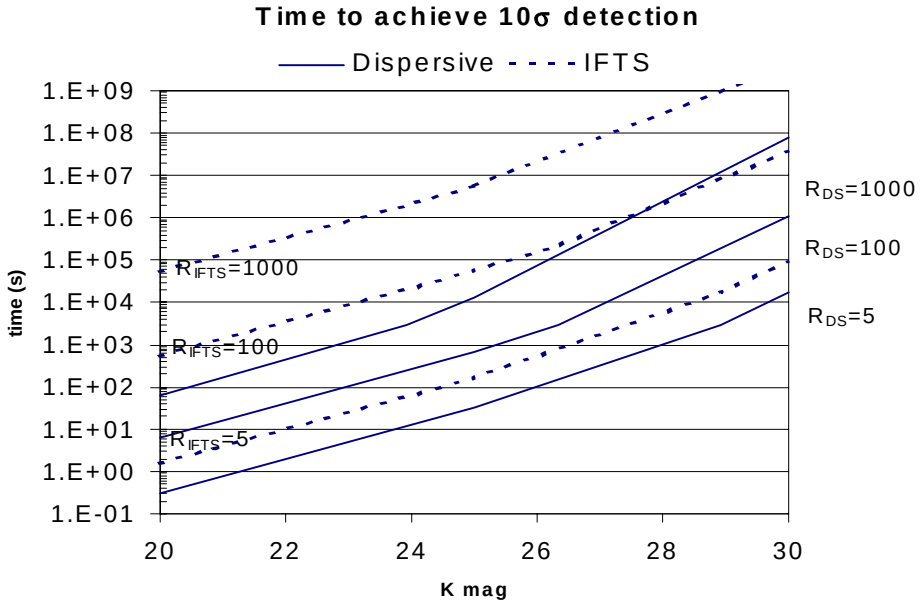


Figure 4.1.1-1 Comparison of exposure times required to achieve a 10σ detection of an object for a dispersive and IFTS spectrometers in the 1.25-2.5μm region, sampled at 2μm. The dispersive spectrograph always is faster for a given result across all resolutions. The maximum allowed integration on each object was 3000s, with longer integrations achieved by co-adding. Details of the calculations used are given in §12.2.

4.1.2 Spectral Mapping

We now turn to one of the applications where the IFTS is at its strongest, that is, spectral mapping of the sky. We first consider the situation where both instruments are photon noise limited on the background. To simplify the discussion, we will assume that all the efficiency factors for each instrument – beam splitter performance vs. grating efficiency, optical throughput, etc. can be taken to be the same. We assume that each uses an array of size Y times Y pixels and that an area on the sky at least as large as the projected area of this array is to be mapped to provide S spectral elements for each point on the sky. We further assume that each instrument can derive an adequate flat field frame from the suite of measurements (likely to be true for the dispersive instrument through use of the time when pixels view sky with only very faint sources to generate a superflat) or from other observations (e.g., flat fields obtained at other times). The IFTS is taken to use both outputs, i.e., to have two detector arrays. In this case, the signal to noise on one spectral element on a single position on the sky is $S^{1/2}$ higher for the DS than for the IFTS, or equivalently the DS is S times faster to a given result. It is this factor that is at the foundation of the advantage of the DS for single sources. However, in this example the IFTS gathers useful spectra simultaneously on Y^2 pixels, so we need to determine whether this gain compensates for its per-pixel disadvantages.

A DS is often designed with a single entrance slit for single object spectroscopy. For such an instrument, useful spectra are obtained over the slit length, Y, so the relative speed for mapping to the same signal to noise is

$$\frac{t_{DS}}{t_{IFTS}} = \frac{1/(YS)}{1/(Y^2)} = \frac{Y}{S}.$$

However, for mapping the DS should have an optimum number of slits. If S << Y, then it needs Y/S slits to make use of the entire array. This point is illustrated in Figure 4.1.2-1, where Y/S ~ 7. The speed of mapping is then increased by the factor Y/S:

$$\frac{t_{DS}}{t_{IFT}} = \frac{1/(YS)}{1/(Y^2)} \times \frac{S}{Y} = 1.$$

Thus, if both are photon noise limited, a properly designed DS will be able to map sky to a given sensitivity limit and spectral resolution at the same speed as an IFTS.

There are many second order effects that need to be taken into account, given this simple result using first order arguments. The DS would map by taking an exposure, then adjusting the telescope pointing one slit width perpendicular to the slit and taking another, and so forth. Thus, the IFTS has a potential advantage in more accurate pixel registration, since telescope pointing errors would affect all the pixels equally. The IFTS also gains an advantage if one or both instruments is read noise limited, a subject we discuss in greater detail below. Although it might be thought that the IFTS has an advantage because of its greater free spectral range, so long as

one is willing to change blocking filters and gratings to map the equivalent range with the DS, the derivation above remains valid and there is no intrinsic advantage.

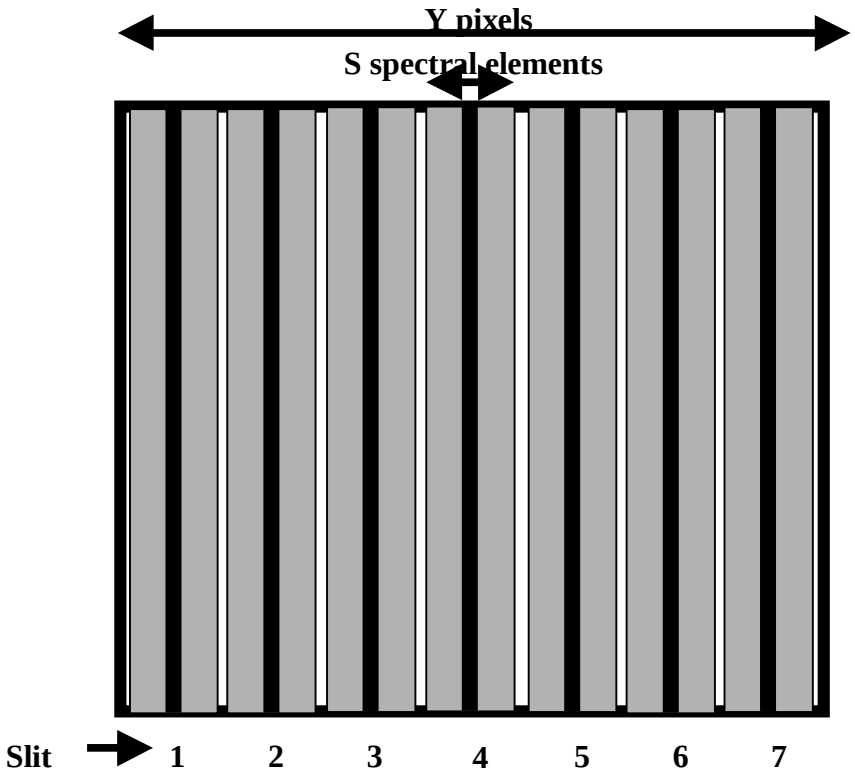


Figure 4.1.2-1 For mapping, a dispersive spectrometer should have Y/S slits.

A significant disadvantage for the IFTS is the higher data rate. Each spectral point requires measurement of a frequency and an amplitude in the transform. If both outputs of the instrument are used, there are then four data points per spectral point, compared with one for the DS. A Fourier Transform requires a large dynamic range, so the advantage in data handling with the DS may even be understated by a simple factor of four. Another fundamental issue is that we have implicitly assumed a spectral shape optimal for the IFTS. Its performance will be degraded where the flux density is low by the energy received from regions in its free spectral range where the flux density is high. To counteract this tendency, it has been proposed that one use a composite instrument that passes the light through a slit and disperses it at low resolution over a modest number of pixels, so the IFTS can operate over limited spectral range for each pixel (Bennett 1999). Such arrangements add complexity, and cannot overcome the basic result that intrinsically in the photon noise limit a DS and an IFTS are equally fast for spectral mapping, assuming the most favorable possible input spectrum for the IFTS.

The other limiting case in any comparison is when both the IFTS and the DS are detector noise limited. Because the IFTS generates 4 data points for every one data point with the DS, it pays a penalty of $4^{1/2} = 2$ in signal to noise. It does achieve a multiplex advantage, however, so that this noise suffices for spectra of S points. The net effect is that, for a single source, $t_{DS}/t_{IFTS} = ¼$, and for imaging spectroscopy:

$$\frac{t_{DS}}{t_{IFTS}} = \frac{(1/Y)\times 1/4}{1/Y^2}\times\frac{S}{Y} = \frac{S}{4}.$$

In general, one expects an IFTS to be operating in the photon noise limit for NGST. However, the result just derived emphasizes that instruments using pre-dispersion to limit the spectral range on each pixel will gain little advantage even in the most favorable case for the IFTS unless the number of spectral elements per pixel is substantially more than four.

A more general relation allows us to consider intermediate cases where the DS is read noise limited and the IFTS is background limited. Assume that I_{ph} is the photocurrent for a single spectral element and t_{int} is the integration time. In making a map, each pixel of the IFTS must be read out 2S times. Each pixel of the optimized DS must be read out $Y/N_{slit} = Y/(Y/S) = S$ times. We then find that:

$$\frac{t_{DS}}{t_{IFTS}} = \frac{I_{ph}t_{\mathrm{int}} + I_d t_{\mathrm{int}} + N_R^2}{2[\sum_1^S I_{ph}(t_{\mathrm{int}}/2) + 2I_d(t_{\mathrm{int}}/2) + 2N_R^2]}\times\frac{Y^2}{Y(Y/S)}\times \mathrm{O}(1) \approx \frac{SI_{ph}t_{\mathrm{int}} + SI_d t_{\mathrm{int}} + SN_R^2}{SI_{ph}t_{\mathrm{int}} + 2I_d t_{\mathrm{int}} + 4N_R^2}\times \mathrm{O}(1),$$

where N_R is the read noise and I_d is the dark current. All the efficiency terms (instrument throughput, beamsplitter performance, grating efficiency, etc.) have been collected into a term of order 1 (O(1)). Such terms enter only linearly in the time to a given result, so it would take a major difference between the two spectrometer approaches to have a strong effect on the relative times. There are many complex considerations in estimating this term, and given the likelihood that it is not significantly different from 1 we have set it to that value. We have assumed both outputs of the IFTS are used; otherwise, it is two times slower than indicated. In the approximation, we have also made the favorable assumption for the IFTS that the contributions to the photocurrent from all the spectral elements are the same, so the sum over photocurrent can be replaced by SI_{ph}. This latter expression easily reduces to the two limiting cases already considered. It is accompanied by the same second order effects already discussed: four times the data rate for the IFTS, etc.

Using the above relations, we calculated the mapping speed of this specific design of DS containing S slits across a detector compared to a FTS. Both are assumed to have the same efficiency. The NGST backgrounds and detector parameters used are those summarized in §12.1. We evaluated the background and source contributions for a flat spectrum, characterized at a magnitude measured in the center of the band, and an integrated background over the band. The hypothetical NGST IFTS is background limited at all resolutions and wavelengths, whereas the

DS is only background limited for bright objects and/or low resolutions. Figure 4.1.2-2 illustrates these regions, where as expected, the mapping time ratio is near unity when both instruments are photon noise limited. The exposure times were set to give a 10-σ detection of the point source, with a maximum of t=3000s for an individual exposure. For objects requiring longer exposures, 3000 sec ones were co-added as discussed in §12.1.2. A maximum integration time of 10 hours per point in the field with the DS determined the maximum magnitude plotted. This limit corresponds to ~ 1000 hours total integration at R = 100 to cover the field of a 2Kx2K array.

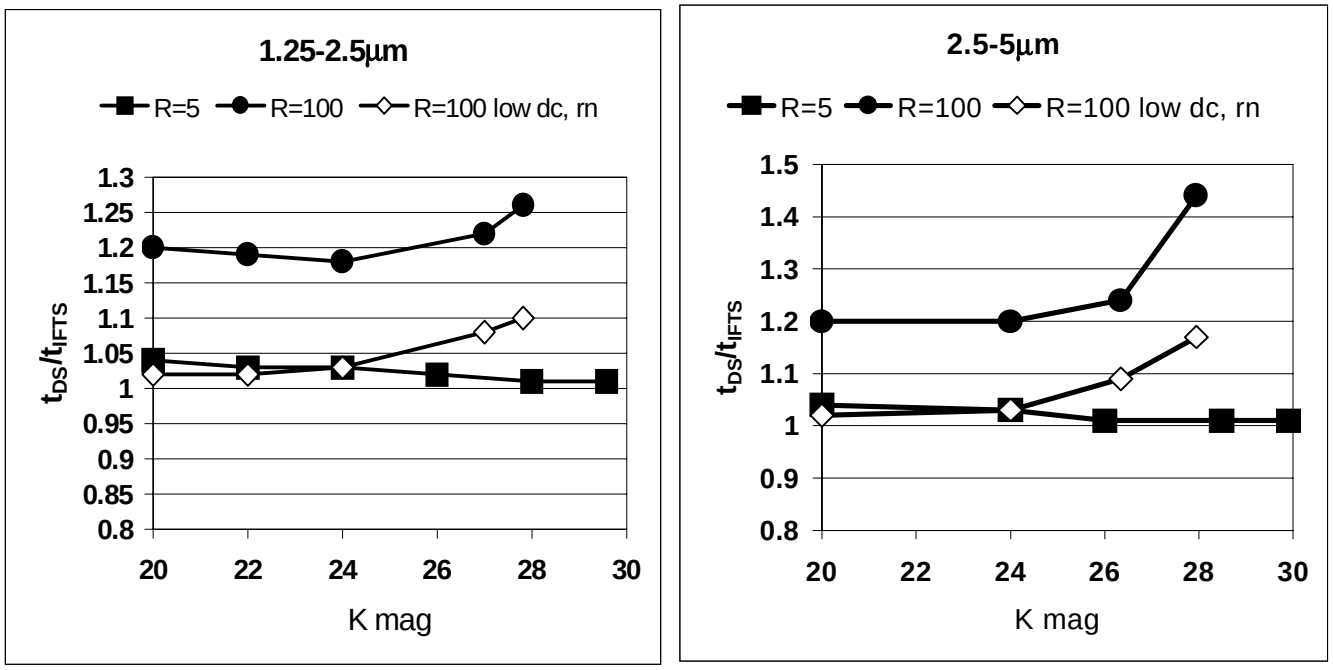


Figure 4.1.2-2 Comparison of mapping speeds over the 1.25-2.5μm and 2.5-5μm bands (0.05" pixels). At R=5, both the DS and IFTS are photon noise limited, and the ratio is unity. At R=100 (solid diamond), the DS becomes detector limited and its relative mapping speed decreases comparable to the IFTS, but far from the extreme of S/4=25 for even the faintest magnitudes. Decreasing the detector contributions (open diamond), will decrease the IFTS advantage.

Figure 4.1.2-2 demonstrates how small the potential advantage of the IFTS would be even in this relatively favorable application, so long as it is compared with a dispersive spectrometer with an optimum number of slits rather than only one.

4.1.3 High Resolution

The IFTS achieves a traditional multiplex advantage only when the DS is read noise limited. Exactly where this transition occurs depends on the way the DS is optimized. To provide an explicit example, we have considered a spectrometer optimized for the minimum in natural background at 3.5μm, have sized its pixels to λ/D at this wavelength, and assumed an instrumental efficiency of 50% and a detector quantum efficiency of 80%. The transition then occurs near R = 300. For sky mapping at R = 1000, the IFTS is about 4.4 times faster than the DS. Of course, this ratio depends on the detector parameters, particularly the dark current, and could be significantly smaller if advances occur from our assumption of a read noise of 5 electrons and dark current of 0.02 e/s.

In general, however, the density of sources of interest decreases as one goes to higher spectral resolution, because of signal to noise limitations. Assuming the source positions are known in advance, we can make a comparison based on Figure 4.1.1-1 for speed in single source spectroscopy. For example, at R = 1000 and $K_{mag} \leq 25$, the speed advantage for the DS is about a factor of 500. If there are 500 objects of interest in the field, then the two spectrometer types can complete the program in the same time; the IFTS with a single long exposure of the whole field and the DS by doing the objects sequentially through a single slit.

The time for a program will scale as the number of the faintest objects in the program – the integrations on the brighter objects will be sufficiently shorter that they will not drive the total time. At 24 < K < 25, there will be roughly 300 objects within a 2.5arcmin x 2.5 arcmin field of view (Thompson et al. 1999), so a DS with a single slit would be slightly faster than an IFTS. With either instrument, the total time for the program would be about one month. Similarly for 26 < K < 27, there are about 800 objects (Thompson et al. 1999), and the DS is about 200 times faster on a single object. Therefore, a multi-object DS with 4 settable slits would be roughly as fast as the IFTS if all the objects in the field were of interest. Obtaining spectra of 800 objects this faint is at or even beyond the plausible limits for NGST, yet only a modest multislit capability makes the DS competitive with the IFTS. We conclude that a DS with only a few settable slits would be competitive with or superior to the IFTS for almost any program at R = 1000.

4.1.4 Other issues

In addition, a dispersive spectrometer, either used in spectral single-object, multi-object, or spectral mapping mode, will have several practical advantages over the IFTS, especially in the areas of reliability and ease of operation.

4.1.4.1 Reliability and cost

The IFTS must maintain alignment to tolerances on the order of 0.1 microns for the entire effective lifetime of the instrument. In addition, a positioning system of similar accuracy must be maintained to move the moving mirror of the multiplex system to produce the spectrum. This requires very precise bearing and stable servo systems for space operation. Errors in the sample position of the interferogram have a similar impact as inaccuracies in the ruling of a grating. They produce ghost spectral images and place power in other spectral locations. Gratings can easily be

manufactured without these misrulings. The IFTS drive system has to reproduce the ruling engine accuracy during every observation.

The grism spectrometers we propose below use dispersive elements set on a relatively low precision filter wheel. Rather than depending on a precise grating mechanism to extend spectral coverage, we do so with multiple grisms, each fixed in the filter wheel. The filter wheel mechanisms should be of lower cost than the IFTS drive, because of its lower precision and because it will be similar, probably identical, to filter wheel mechanisms that will be used for other functions in the ISIM.

Most interferometer systems depend on frequency and mode stabilized lasers for servo control of the moving mirror, maintenance of interferometric alignment, and for determination of the mirror positions for data sampling. Failure of the laser system is catastrophic for the operation of the interferometer.

4.1.4.2 Detector properties

Since dispersive spectrometers place smaller amounts flux on each pixel they gain speed from a reduction of read noise for objects up to S times brighter than multiplex spectrometers. This means that the observation times will be reduced as the detector read noise improves for a much larger range of objects.

With a read noise of 5 electrons and a dark current of 0.02 electrons per second, for exposure times on the order of 1300 seconds the dispersive dark current noise begins to exceed the read-noise and becomes a significant contributor to the total noise. Dispersive spectrometers will therefore benefit from reductions in the dark current. The IFTS system does not share this benefit since the source noise overwhelms the dark current for reasonable observing times. Because 5 electrons read noise and 0.02 e/s dark currents are close to demonstration now, it is likely that the comparisons in performance will tilt more toward the DS by the time the detectors are manufactured for NGST. IFTS spectroscopy, however, is very near the limits imposed by the high photon flux on the detectors and cannot expect to achieve significant gains in performance through better detectors.

4.2 Sliding slits with Dispersive Spectrometers

Given the selection of dispersive spectrometers, we can now consider various implementations of multi-object capability.

4.2.1 Number of slits

The expected number densities for a subset of DRM programs requiring NIR and MIR multi-object spectroscopy are schematically shown in Figures 4.2.1-1&2 scaled for our ISIM concept's fields of view (see §6).

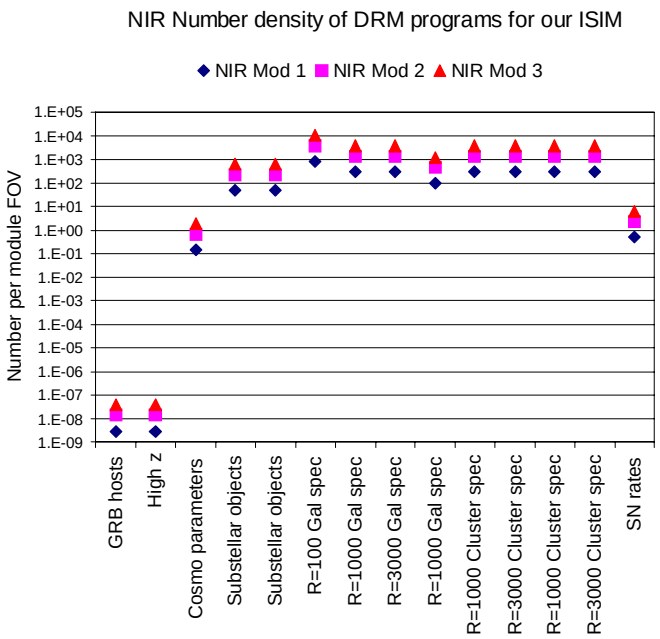


Figure 4.2.1-1 The near infrared (1-5μm) number density of the faintest limits for DRM programs requesting NIR multi-object spectroscopy. These have been scaled to numbers per 0.7'×0.7', 1.5'×1.5', and 2.5'×2.5' for NIR modules 1, 2, and 3, respectively. The most demanding programs are the high redshift galaxy spectroscopy.

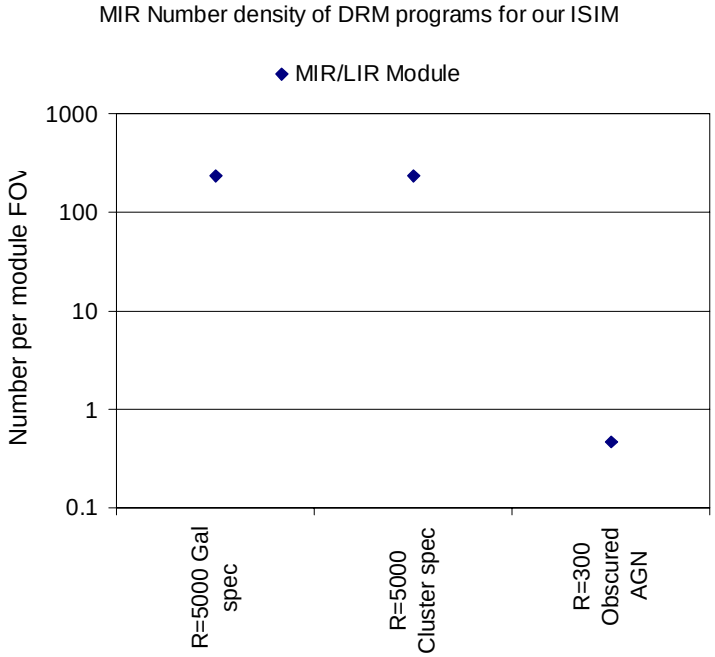


Figure 4.2.1-2 The mid infrared (>5μm) number density of DRM programs desiring MIR multi-object spectroscopy. Again, these have been scaled to number per 3.4'×3.4' for our MIR and LIR echellette imaging/spectroscopic modules (§6.1.6, §6.1.8).

4.2.2 Sliding Slits

One way to provide variable multi-object slit configuration is to have physical slits which can be moved into position for each object. The performance and degree of complication can be judged from two ground based multi-object instruments presently planned to use such devices. The HET low resolution spectrograph will use thirteen individually configurable slits (Hill, et al. 1998).The units are configured out of the beam and then inserted, allowing for an imaging mode for setup on faint objects without disturbing the slits at all. A more versatile mechanism is that used in FORS1 for the VLT, which has nineteen slitlet jaws (Bohnhardt, et al. 1995). The width of each jaw is made adjustable so the instrument can adjust the slit widths independently for each object. Either multi-slitlet configuration will require motorized components.

No cryogenically cooled slitlet mechanism exists. Although there appears to be no fundamental difficulty in adapting a room temperature device for cryogenic use, the overall mechanical complexity, tight tolerances, and multiple mechanisms would give one pause in taking this approach for a high visibility space mission. One could envision alternate approaches that might make a spaceflight slitlet device more palatable, such as flexures as an alternative to sliding bearings for slit configuration, and perhaps a linear motor with a servo for positioning. With such modifications (basically, no traditional bearings required), a slitlet design might be feasible, and we will use it as a baseline against which we measure other approaches.

The sliding slit approach will only provide one slit per dispersion direction, as compared to the 2D mapping of either micromirror (§4.3) or microshutter (§4.2.3) arrays. This implies the maximal slit number will be limited by the slit length and detector size. For example, if there are three Airy FWHM disks per slit length, and each pixel is $\lambda/(2D)$, then each slit will map to 6 pixels long, so a 2048×2048 array can accommodate 340 objects. The smaller 1024×1024 arrays for the mid infrared modules would allow 170 slits.

4.2.3 Microshutters

Microshutters are an implementation of multiple slits that can be more flexible than slitlets, since more than one shutter can be opened along one line of dispersion and more shutters can be opened at once than would be feasible with slitlets. They potentially provide an efficient, high contrast individual aperture selection for MOS instruments.

They will be arranged in a two dimensional array, made up of individual shutters of size ~100µm square with expected filling factor up to 80%. This reduced filling factor is caused by the presence of a mask to which each Si shutter's torsion flexure is connected. Since they work in transmission, there is no loss of contrast due to diffraction or scattering off the edges of adjacent unactuated pixels. They would be operated cryogenically. However, they are not yet fully demonstrated.

4.3 Micromirror arrays with dispersive spectrometers

It has been proposed that versatile multi-object spectroscopy in a space-borne instrument can be best performed by cryogenic micro-mirror arrays, hereafter MMA. These devices are key to the implementation of multi-object spectroscopy in the yardstick NIR spectrometer

Micromirror arrays were originally developed for commercial projection devices. It is uncertain whether they can be adapted to operation at cryogenic temperatures and with the performance demanded for an astronomical spectrometer. To probe these issues, NASA has funded two studies of the feasibility of the MMA devices for the NGST ISIM: 1) silicon approach (P.I. MacKenty); and 2) GSFC all aluminum design, 30K operation temperature, 100µm pixel pitch, 2048×2048 format (Greenhouse).

In brief, the MMA is based on a large number of small silicon or aluminum mirrors. Each of them can be placed by an electric field in either of two positions, "ON" or "OFF". If the mirror is in the "ON" position, the light selected by the micromirror is sent into the spectrograph, whereas in the "OFF" position, the light is reflected in another direction, which could either be an imager or a light trap (e.g., black surface to absorb light and control scattering). This is shown schematically in Figure 4.3-1. As with any reflective spectrometer slit, some light will be scattered into the spectrograph by structure around an "ON" mirror; in this case the structure is a field of "OFF" micromirrors and hence may not be optimum for control of this stray light. This issue is avoided with a conventional slit or sliding slitlets, since the light passes through the slit in these cases and the opaque slit jaws block all off-slit stray light.

We investigated the issue of stray light for two cases: 1) the NGST diffuse IR background sky and self-emission and 2) individual contaminating sources. We define an efficiency parameter, named “rejection ratio,” which is simply the ratio between the reflection of a MMA mirror in its “ON” position and its “OFF” position. A rejection ratio of 100:1 or 0.01 equates to 1% scattering off a MMA mirror in its “OFF” position.

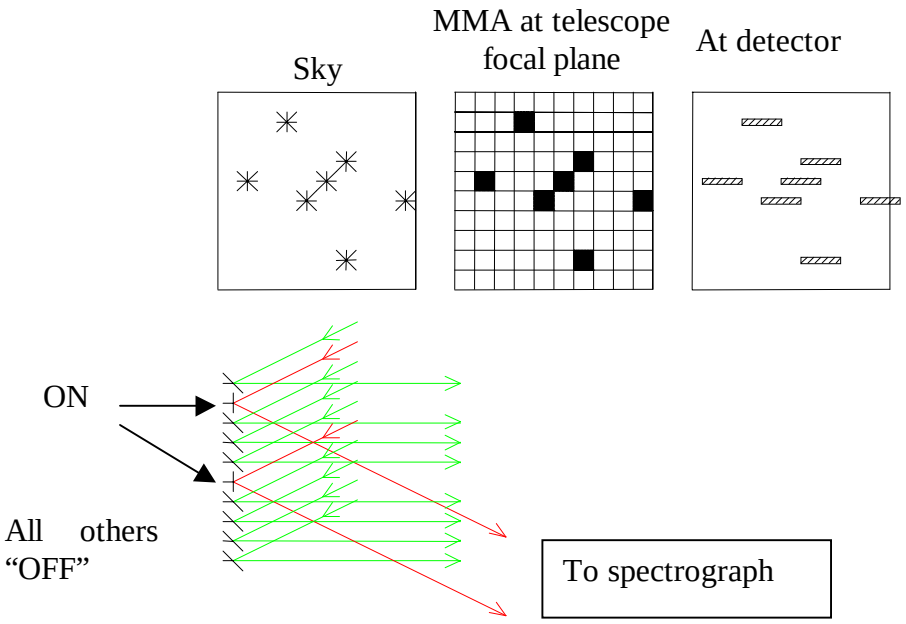


Figure 4.3-1 Schematic of a micromirror array device. Mirrors tilted in their "ON" position deliver light to the spectrograph (red lines), whereas, in principle, adjacent mirrors which are "OFF" (no object), redirect the light far from the spectrograph's input aperture.

4.3.1 Rejection of the NGST background sky and self emission

We calculated what rejection ratio criteria must be met to keep the degradation in signal to noise for a 10σ, 10,000s exposure to less than 10% compared to the case for a slit, where there would be no contribution of scattered light from adjacent overlapping sky spectra. Details of the background model and sensitivity calculations and the micromirror rejection calculations are given in §12.1.1-2 & §12.3.1, respectively. The 10,000 second exposure was broken into ten 1000s exposures. Results of these calculations are summarized with Figures 4.3.1-1 to 3.

Figure 4.3.1-1 shows the calculations for a near infrared module. Two spectrometer configurations were studied, each covering an octave of the spectrum, one from 1.25 to 2.5μm and the second from 2.5 to 5μm. We also assumed a read noise of 5 electrons, dark current of 0.02e/s, quantum efficiency of 80%, and instrument efficiency of 50%. We assumed a pixel size of λ/2D at 1.6μm = 19 marcsec for the 1.25 – 2.5μm spectrometer and of λ/2D at 3.5μm for the 2.5 - 5μm one. The micromirrors were used to place a source on a 2×2 pixel region of the array (in the absence of dispersion). We found that rejection ratios better than 300:1 are required for the 1.25 – 2.5μm spectrometer and better than 500:1 for the 2.5 – 5μm one for R=100-1000. Lower dark currents or larger pixels will impose tighter constraints. The degradation of signal to noise scales as the square root of the rejection ratio, so a ratio of 100:1 would degrade the sensitivity by 17% for 1.25 – 2.5μm and by 22% for 2.5 - 5μm. In this case, the relatively flat spectrum of the contaminating light across the spectrometer passbands results in a moderately low rejection ratio requirement.

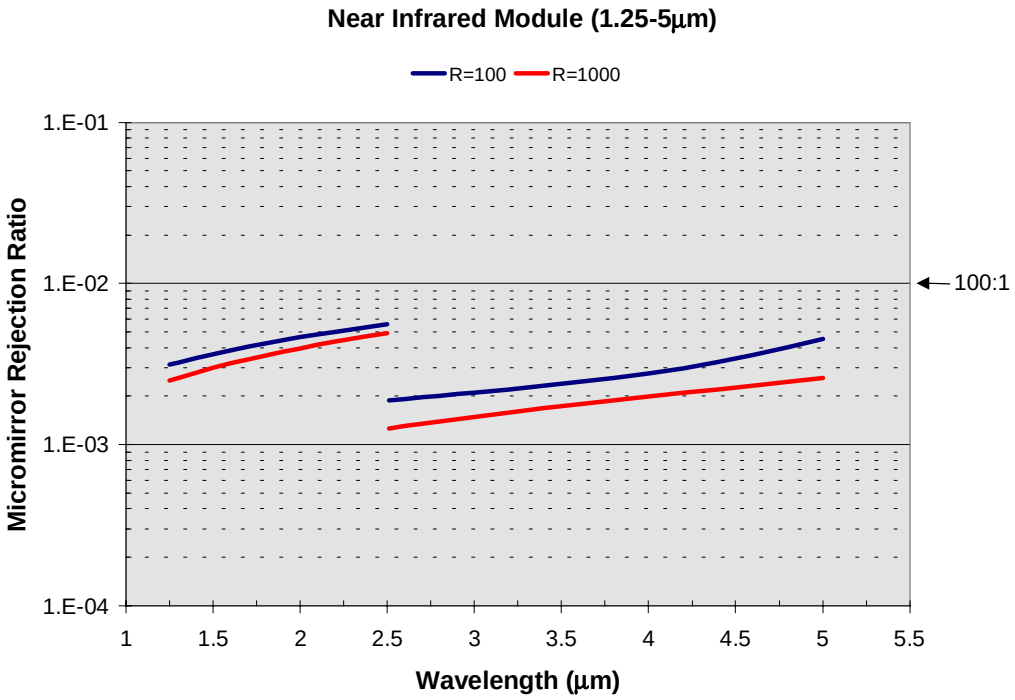


Figure 4.3.1-1 MMA rejection ratios required for background rejection in the near infrared.

Figure 4.3.1-2 shows a similar calculation for a spectrometer operating between 5 and 10µm. We assumed a read noise of 15 electrons, dark current of 1e/s, quantum efficiency of 25% to 7.5µm and 55% from there to 10µm, and instrument efficiency of 50%. The pixel size was λ/2D at 9.5µm = .103 arcsec, and the area for the spectrum was a 2×2 set of pixels in the absence of dispersion. Rejection ratios better than 1000:1 are required. The R=100 case for higher dark current (10e/s) was also modeled. Rejection ratios better than 500:1 are required. NGST detectors, however, are expected to meet and exceed the lower dark current requirements (<1e/s).

Figure 4.3.1-2 also shows the results if the spectral response of the above module is extended with a second spectrometer to the octave from 10 to 20µm. Rejection ratios of ~ 30,000:1 are needed in this case. This progression toward tighter requirements on the rejection of off-slit light results from the steep rise of the spectrum of the contaminating light toward the long wavelength end of the octave over which the spectrometer works. Only a small proportion of the high flux of photons near 20µm need to be scattered onto the array to provide a significant increase in flux at 10µm. In general, making micromirrors work well in such a situation will be difficult.

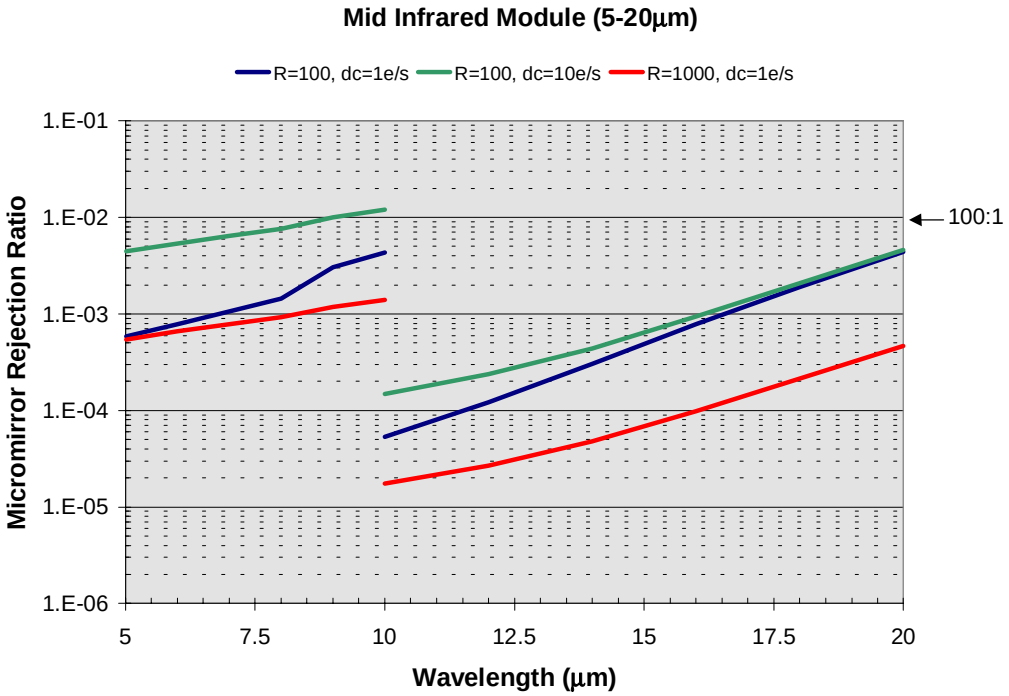


Figure 4.3.1-2. Rejection ratios required for MMAs in the mid infrared.

Figure 4.3.1-3 shows another example of this behavior. As in the case in Figure 4.3.1-2 for λ>10μm, there will be tight constraints to block the large NGST background over the 20-40μm region.

In summary, rejection of the diffuse sky background and telescope emission poses constraints for micromirror performance. These constraints grow rapidly with wavelength due to the NGST background shape, most likely ruling out a MMA version of a long wavelength multi-object spectrometer. Additionally, the higher the spectral resolution, the tighter the constraints on reflection. Thirdly, the lower the dark currents and larger the pixels in detectors for λ<10μm, the better the micromirrors need to perform.

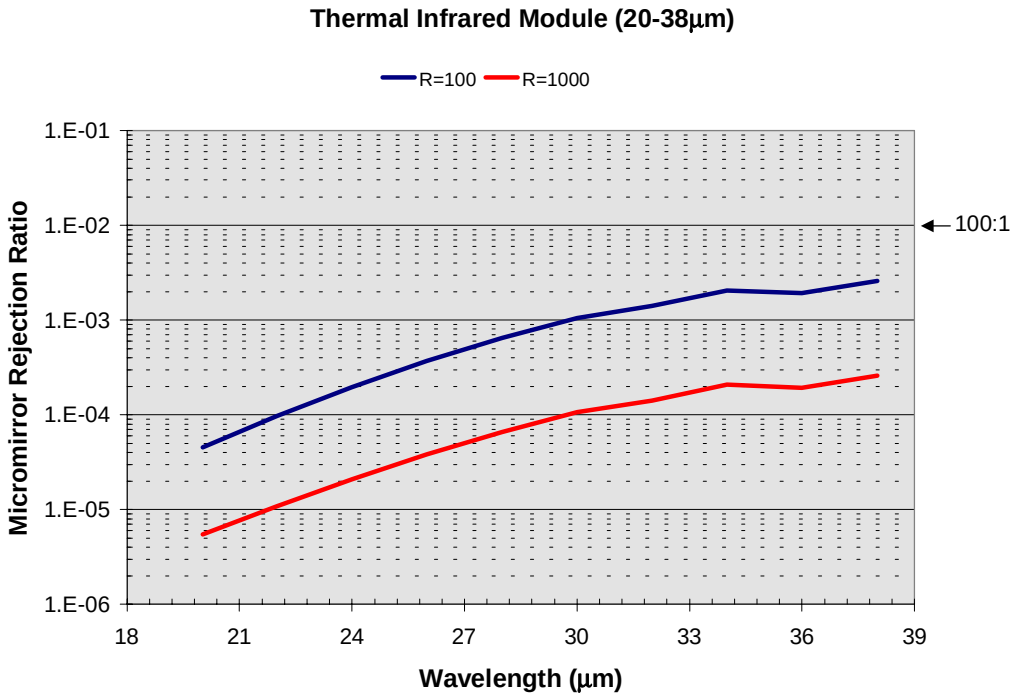


Figure 4.3.1-3. Rejection ratios required for MMAs in the far infrared.

4.3.2 Rejection of contaminating objects

We also investigated the effects of imperfect rejection of neighboring sources from scattering off a MMA mirror (i.e., when it is in its "OFF" position) in a hypothetical NGST spectrometer. We modeled this effect through a Monte Carlo simulation which computes a distribution of confusing sources over a fairly large area around a target object's slit.

We chose the DRM of evolution of faint field galaxies at high redshift as our example. For this DRM, many spectra are required in the near infrared where the MMA works the best. This problem is actually a fairly general test of the performance of MMAs, since they only provide an important advantage when there are many faint sources of interest within the field of the instrument. If, in comparison, the sources are relatively bright, then the penalty in telescope time to observe them one at a time is modest.

Simulated instrument parameters, background source distribution, and criteria for contamination are given in §12.3.2, and were summarized briefly in the preceding section. We calculated the number of times the target's spectra was deemed invalid, due to the confusion of an overlapping source contributing 10% of its flux, as a function of resolution and survey limiting magnitude for several micromirror rejection ratios.

Figures 4.3.2-1 to 4.3.2-3 illustrate the effect of the confusing sources on the sky for R=100, 1000, 3000. Three rejection ratios of the micromirrors, namely 0.01, 0.001 and 0.0001 are plotted. The source density is defined as the number of confusing sources brighter than the target

per given area. They have been scaled for a 4'×4' field of view. Equivalently, one can read off the source density per 4'×4' FOV and relate that value to a target brightness, since the source density is that defined for integrated objects brighter than the target. This is shown in the upper axis where the target's AB magnitude is plotted.

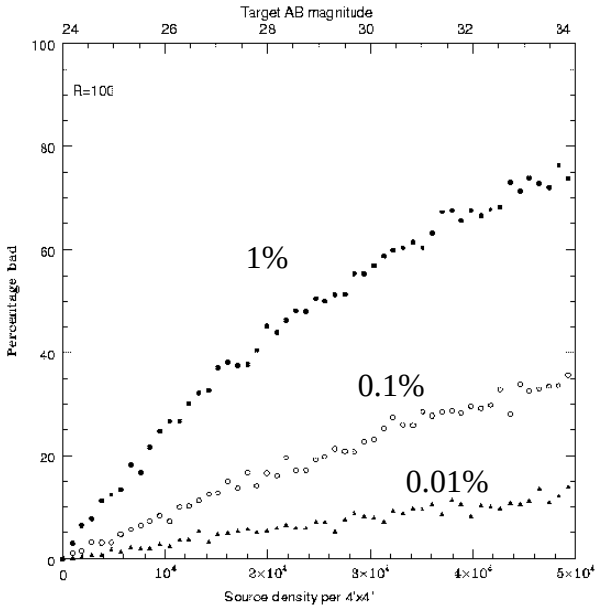


Figure 4.3.2-1 shows what happens at R=100. 50% contamination is reached for target magnitudes of AB=29.4 for 1% rejection by the micromirrors. This is unacceptable for NGST limiting sensitivity predictions of AB=34. High rejection ratios are required, approaching 1000:1. Thus, the requirement to avoid overlapping spectra results in a similar, or slightly stricter, requirement on the rejection ratio in the 1-5µm range compared with the need to reject diffuse background.

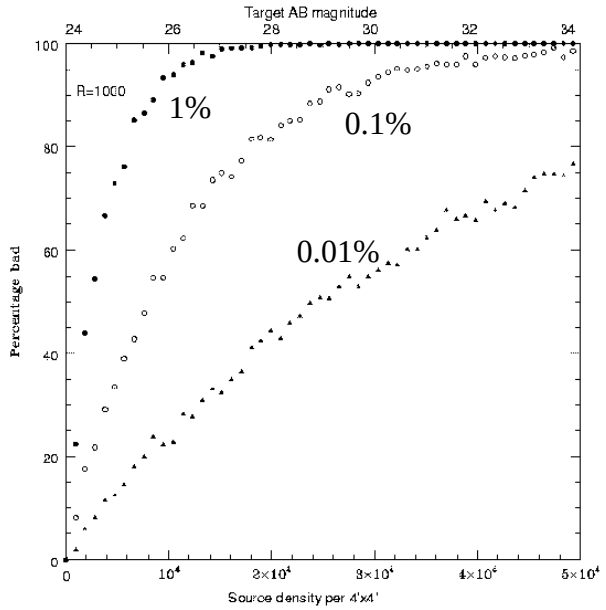


Figure 4.3.2-2 shows a similar calculation for R=1000. 50% contamination is met for target magnitudes of AB=24.5, 25.8, and 29.1 for 1%, 0.1% and 0.01% rejection by the micromirrors, respectively. Even at 0.01% rejection, the level at which this high level of contamination occurs is above the detection limit for deep exposures at AB ~ 31.

The trend in these figures continues toward higher resolution. For example, at R=3000, 50% contamination is reached for target magnitudes of AB=24.1, 24.5, and 25.9 for 1%, 0.1% and 0.01% rejection by the micromirrors, respectively. These limits are far above the detection limits at this resolution (at K(2.2μm), $m_{AB} \sim m_{Johnson} + 2$).

Figures 4.3.2.-3 and 4 illustrate the effect of spectral resolution for two moderately deep spectroscopic surveys.

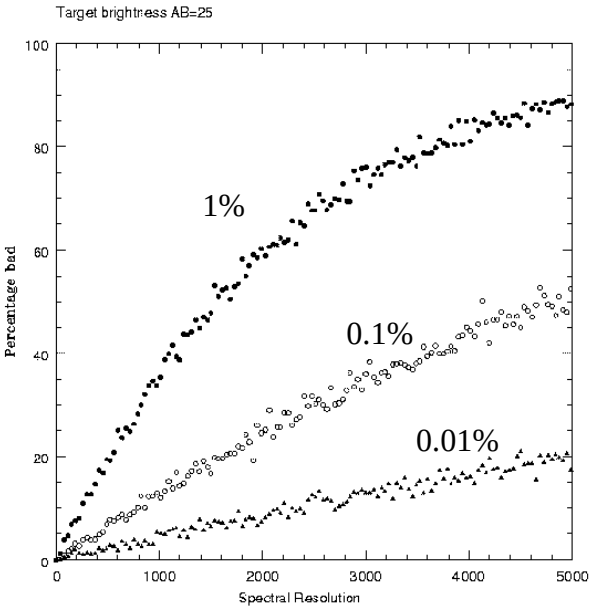


Figure 4.3.2-3 is based on the expected number of sources brighter than AB=24 (~1700 galaxies/4'x4'). 50% contamination is reached for R≥1500 and R≥4700 for rejection ratios 1% and 0.1%, respectively.

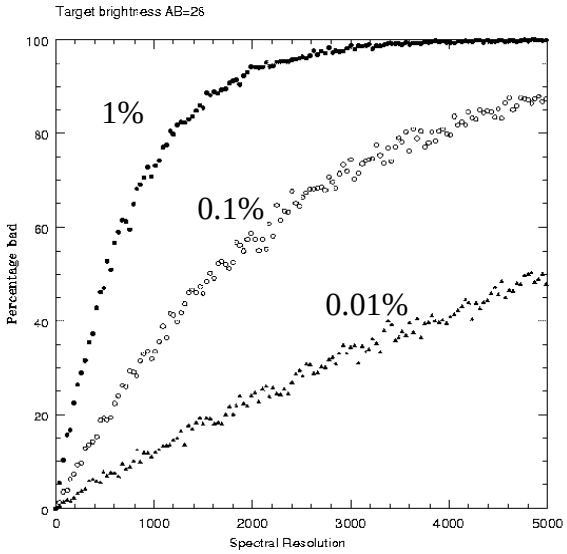


Figure 4.3.2-4 is based on the expected number of sources brighter than AB=28 (~5000 galaxies/4'x4'). 50% contamination is reached for R≥500, R≥1600, and R≥5000 for rejection ratios 1%, 0.1%, and 0.01%, respectively.

In summary, rejection of confusing sources will be a major concern for a MMA fed MOS instrument with rejection ratios near 1%. Again, as concluded by independent calculations for rejection of the diffuse sky background, the higher the spectral resolution, the tighter the constraints on the MMA performance (i.e., need lower rejection ratios). In particular, at R=1000, we suggest that the rejection ratios need to be greater than 10,000:1 to reach targeted NGST sensitivities at AB=34. In general, the rejection ratios to prevent serious overlap of the source spectra are higher than those required to preserve the intrinsic signal to noise without significant degradation due to extra background emission reaching the detector.

5.0 Effects of telescope concept on the ISIM

We discuss some issues related to the telescope design that can strongly affect the scientific and potential operation of the ISIM instruments.

5.1 Impact of a non-deformable-mirror telescope architecture

Due to the envisioned deployable nature of NGST, the observatory will require active control to align and maintain alignment of the segments. In the yardstick telescope design, coarse alignment and adjustment are provided by actuators behind the primary, while fine wavefront control is provided by a separate deformable mirror (DM).

There are several lightweight optical materials being tested for NGST. The concept developed by Jim Burge at the University of Arizona allows a high density packing of cryogenic actuators behind a glass membrane mirror (~2000 over the primary area), which removes the need for a DM to meet the wavefront control requirements. The concept is to correct the primary mirror itself, rather to attempt to compensate for its errors. Primary mirror (PM) actuators are intrinsically less risky than deformable mirror because PM actuators nominally apply no corrective force and there is more space to engineer redundancy. Although Burge's development is for a glass membrane mirror, it would appear feasible to use this aspect of his concept with other mirror materials.

The simulations and analysis of the surface errors in the primary mirror segment support this approach. The optical design is shown in Figure 5.1-1. It consists of a 8.4m primary, a convex secondary, a concave tertiary, and a flat fast steering mirror. The basic telescope parameters are listed in Table 5.1-1. This design yields a well corrected FOV (Figure 5.1-2) and simulations and lab test data show that it is diffraction limited to 200nm (Fig 5.1-3). For visible and infrared science, no gain would be made from using even an ideal deformable mirror. This telescope architecture allows wide field imagers to be placed in the telescope focal plane for large field surveys, as discussed in more detail in §5.2.

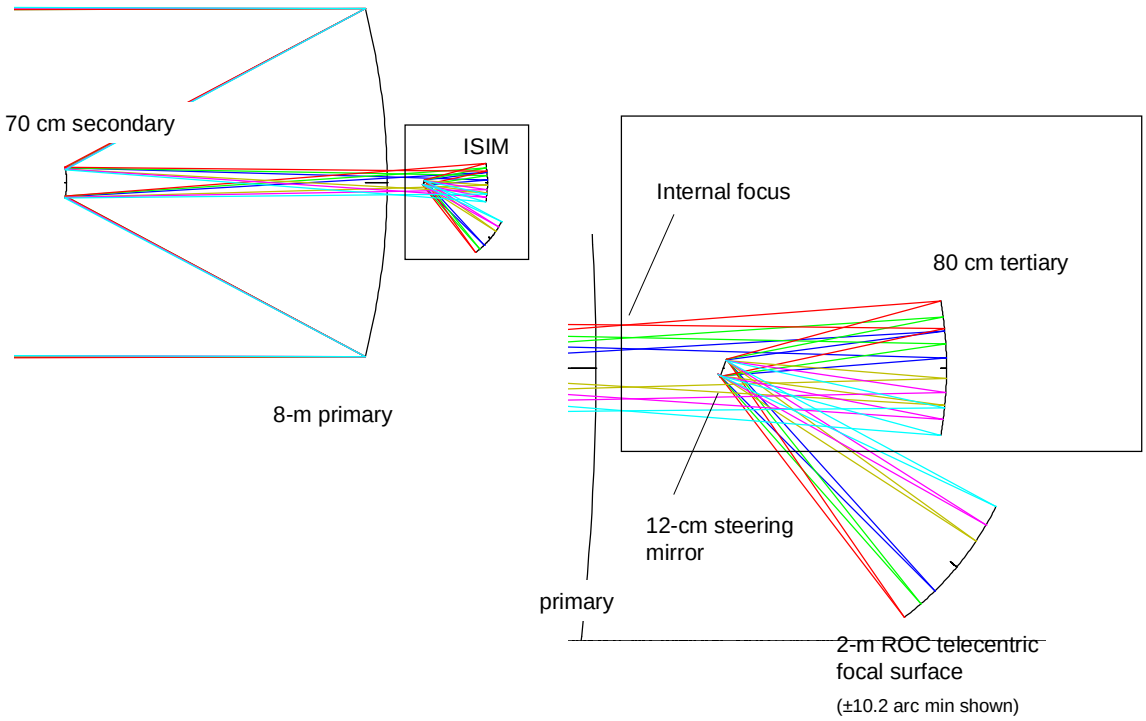


Figure 5.1-1 Design of NGST telescope giving 20 arcmin field. The dense packing of the primary mirror actuators allows for a well corrected field of view. A steering mirror at a pupil allows the field to be placed on various instrument modules.

Primary mirror	8.4m diameter, f/1.14
Secondary mirror	73 cm convex hyperboloid
Tertiary mirror	86 cm concave ellipsoid
Steering mirror	10×12cm flat, tilted 15º, located at pupil and at center of focal surface
Focal surface	Spherical, concentric with steering mirror

Table 5.1-1 Basic telescope parameters for non-DM telescope architecture.

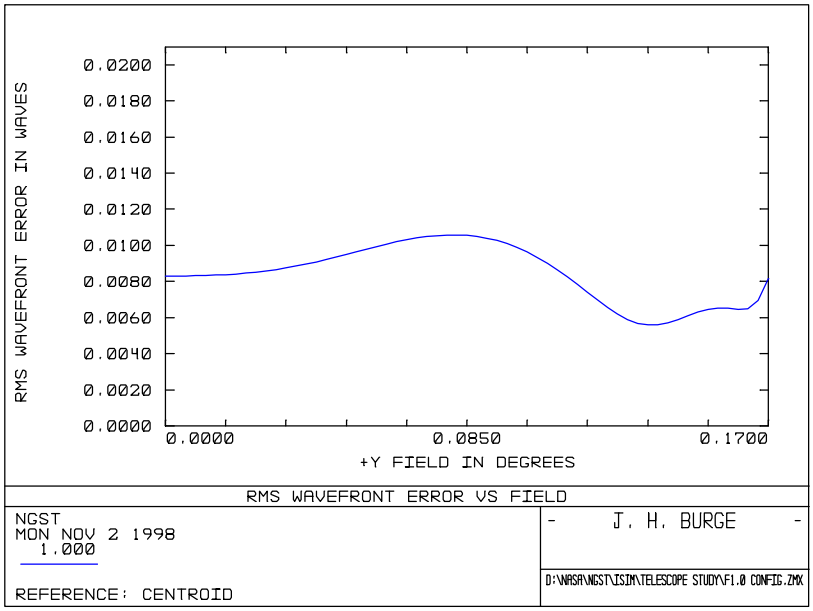


Figure 5.1-2 RMS wavefront error for the telescope design shown above. It is well corrected over a 20 arcmin diameter field of view.

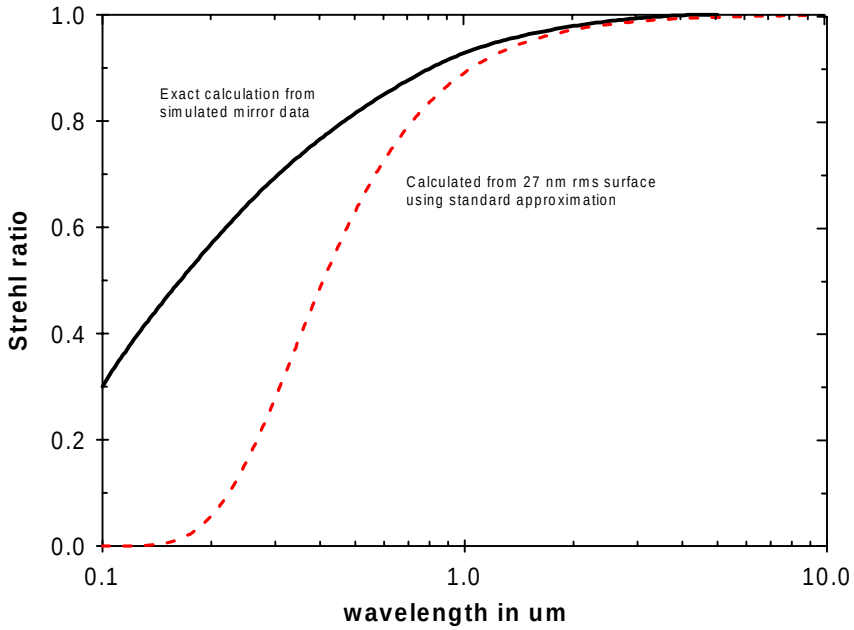


Figure 5.1-3 The telescope design shown in Figure 5.1-1 provides diffraction limited imaging down to 200nm.

The steering mirror coincides with an image of the stop, otherwise occupied by the DM, and it also lies at the center of the spherical focal plane. This allows the line of sight to be steered by extremely large angles by moving this mirror, with the telescope fixed in space.

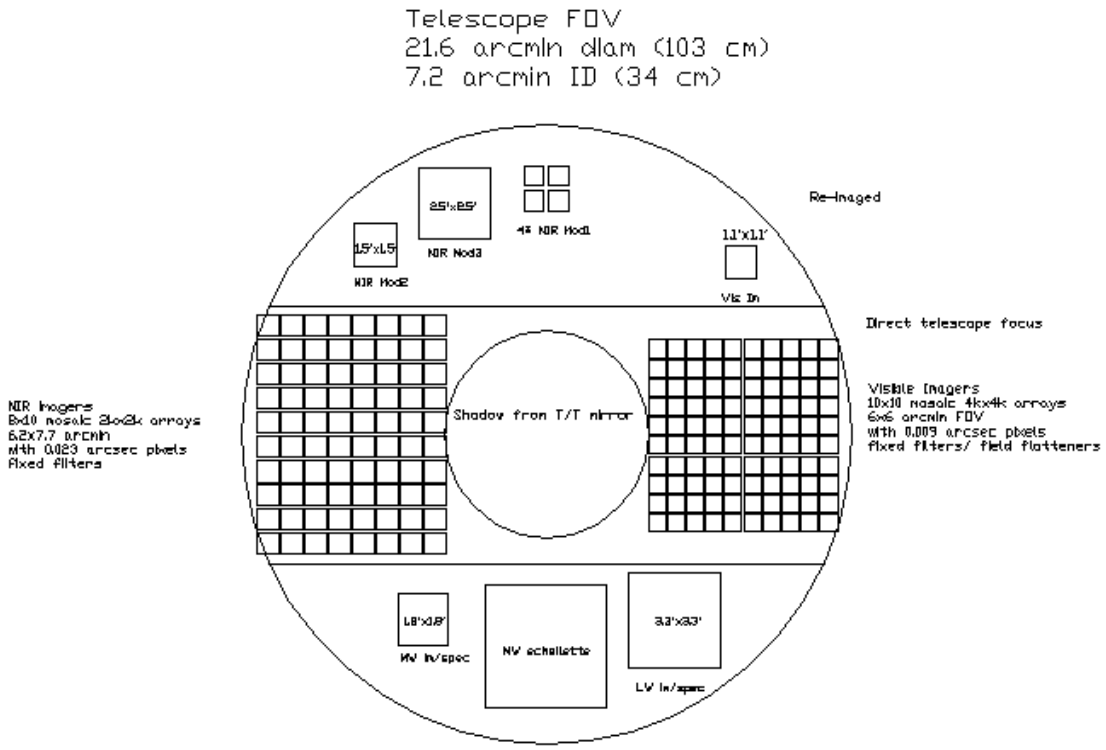


Figure 5.1-4 Illustrative schematic of the NGST telescope FOV for the non-DM telescope design. This design gives a large (21.6 arcmin diameter) well-corrected FOV. There is ample room for wide field imagers in addition to a collection of dedicated modules for high resolution imaging and spectroscopy. The latter FOVs are shown re-imaged after their respective relay optics. Other telescope architectures will give different FOVs which will determine the largest field for these direct imager mosaics.

Figure 5.1-4 illustrates the flexibility in instrument placement around the focal plane that can be achieved with such a steering mirror.

5.2 Large field imagers

As summarized in §1.2, more than 200 days of the core DRM is spent in deep surveying at near infrared and visible wavelengths. In fact more than half the total observing time for this part of the DRM is used in these surveys. Therefore, it is a high priority to provide a large imaging field of view to survey efficiently.

In §5.1, we show that a large and well corrected field of view can be obtained with NGST if the full wavefront correction is carried out with the primary (or primary plus secondary) mirror. The key to straightforward implementation of a wide field option is to place a simple steering mirror at a pupil (where otherwise one might need to place a deformable mirror) and to use this mirror to

direct the beam to an appropriate wide field instrument module (§5.1). The focal plane scale in this module is just that for the telescope itself. For an 8m telescope at f/24, the scale is about 1.1 marcsec/μm, so the 2μm diffraction limited image FWHM is 47μm in diameter.

We have assumed that the detector arrays that would be used in the wide field modules would have pixels of about 20μm size. This size is based on current manufacturing practice, as well as our desire to limit the number of cosmic ray hits on a pixel so that long integrations can be obtained. However, the projected size on the sky of ~ 22 marcsec is significantly smaller than would be ideal for a large field imager for NGST. Unfortunately, as shown in §9, the cost of these modules is also quite high because of the large number of arrays that must be purchased.

5.3 Faster Cassegrain designs

We have carried out only a brief exploration of the effect of a faster telescope on the packaging of the rest of the ISIM. Based on two previous case studies, we believe that a faster f/ratio may improve the packaging of these instruments. For example, in the case of SIRTF, it was found that the instruments could be made more compact when the telescope design was changed from f/24 to f/12. The current instruments could probably not be packaged in the existing volume at f/24. A second example is large (6.5-10m) ground based telescopes. We (Rieke 1987) conducted a careful optimization of the configuration for the Magellan Telescope that showed that f/15 was superior to f/24. Independently, the Gemini project has decided that f/16 was more optimum than slower f/numbers. One of the important considerations in both studies was the delivered image quality in realistic instruments.

A faster telescope design also provides a more satisfactory pixel scale for the large field imagers (§5.2) such that they would reach the background limit quickly. Thus, for this application, telescope f/numbers of 12 to 16 would be better than the f/24 baseline. In addition, with a faster telescope fewer arrays would be needed to provide the equivalent field of view, substantially reducing the price of the wide field modules (§9.4). Thus, we believe that a faster telescope is virtually a requirement if NGST is to be well optimized for large field surveying.

6.0 Our ISIM concept

Our philosophy is to design the ISIM composed of a number of modules, each of which is relatively simple. To the greatest extent possible, we limit the number of observing modes and capabilities for each individual module to optimize performance, and at the same time minimize complexity, cost and risk. We also strive for redundancy in capabilities among the individual modules, when feasible. The modularity allows relatively easy modifications at later stages, and incremental descopes, if necessary. We have, in fact, intentionally considered more modules than would probably be appropriate for the final ISIM design, so that we could evaluate alternatives and bring them to the attention of the project. After a description of all the individual modules studied, we will identify the specific ones we have selected for an ISIM that provides the best compromise between cost and capabilities.

6.1 General description of sub modules

A summary of the capabilities and formats of the modules we have studied is given in Tables 6.1-1 & 2. Not all of these modules would be included in the final ISIM: the ultimate choice of which of these modules would be best to build will be based on a trade of scientific priorities, detector and cooler technology, spacecraft and telescope performance, and cost.

	Visible Imager	NIR Module 1	NIR Module 2	NIR Module 3	MIR im/spec	MIR/LIR Echellette*		LIR im/spec
Imaging λ (μm)	0.4-1	1-2.3	2.3-5	2.3-5	5-27	5-25, 18-35		20-34
Spectroscopy λ (μm)		1.25-5	1.25-5	1.25-5	5-27	5-18, 17-35		20-34
Nyquist λ (μm)	0.5	1.5	3.4	5.7	8	15.5		27
Detector	Si PIN hybrid	HgCdTe	HgCdTe	HgCdTe	Si:As	Si:As, Si:P		Si:P
Detector Temperature	≤60K	≤60K	≤60K	≤60K	≤10K	≤8K		8K
Pixel size	18μm	18μm	18μm	18μm	27μm	27μm		50μm
Format	10240^2	2048^2	2048^2	2048^2	1024^2	1024^2		512^2
Field ('×')	1.1×1.1	0.7×0.7	1.5×1.5	2.5×2.5	1.8×1.8	3.4×3.4		3×3
Spectroscopic Resolution	Imaging	100, 1000, 3000	100, 1000, 3000	100, 1000, 3000	100, 500, 1000	100 500 1000 3000	500 1000 3000	500, 1000

Table 6.1-1 Summary of optical design parameters for ISIM modules under study requiring relay optics. At the Nyquist wavelength, pixels are λ/(2D) in size. Modules not retained in the final selection are shaded.

*****Two identical modules except for detector type, filters, and spectral elements.

	Wide field visible imager	Wide field NIR imager
λ Imaging (μm)	0.4-0.8	1-2.5
Nyquist λ (μm)	0.7	1.8
Detector	8×10 mosaic 2k×2k Si PIN detectors or 10×10 mosaic 4k×4k CMOS detectors	8×10 mosaic 2k×2k HgCdTe FPAs
Pixel size	18μm or 7μm	18μm
FOV ('×')	6.2×7.7 or 6×6	6.2×7.7
Filters	Broad band	Broad band

Table 6.1-2 Parameters for proposed wide field imagers.

In all the optical designs shown below, light is collected by the telescope and brought to a focus from the left. The vertical line indicates the size and position of this telescope focal plane. Light then proceeds to the right to a pupil, which is where all of the rays intersect at another line, often at an angle from vertical. This is where a filter wheel (FW) (or wheels) is located, cooled to an appropriate operating temperature. The rays then proceed, through mirrors or lenses to a final focus, where the focal plane array (FPA) is located. An enclosing light tight box (not shown), also cooled, surrounds all optics after the filter wheel.

For the multi-object modules, the mechanisms for providing multi-object spectroscopy are not explicitly shown. Transmissive multi-slit or microshutter devices can be placed at the re-imaged telescope focal plane with no additional elements. A reflective micromirror device would require additional flats for relaying the field from the telescope into the module.

We now discuss the individual modules. Further details can be found in §12.4.

6.1.1 Visible light imager

Although a telescope that is only diffraction limited to 2μm may not need Nyquist sampling in the visible, some telescope concepts are capable of achieving the diffraction limit to much shorter wavelengths (e.g., §5.1). We have therefore designed a visible light module that would take advantage of this imaging performance. We designed relay optics assuming 18 micron pixels. This value is the likely to be the smallest possible for non-CCD detectors (e.g., Si PIN diodes, see §2.1.3), and is easily attainable with CCDs. We therefore require re-imaging optics to f/72 for Nyquist sampling at 0.5μm. Reflective designs are possible but the resulting module is large. We chose instead refractive optics. The overall design is shown in Figure 6.1.1-1. It has been optimized to yield diffraction limited images across the full field down to 0.5μm.

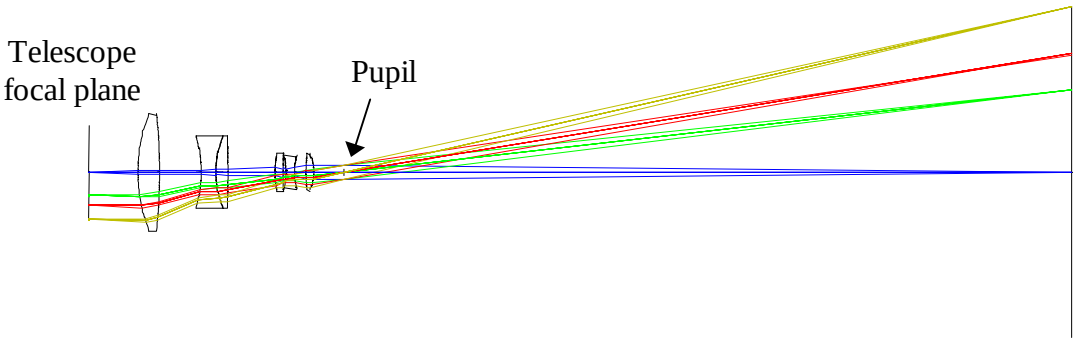


Figure 6.1.1-1 Visible Light Imager. The system is a six lens refractive design composed of simple lenses. From left to right, the components are PSK53A, F7, LASF36A, LASF36A, K2FSN5,and PK51A. These lenses form a pupil (~10mm diameter) where a filter wheel would be placed.

6.1.2 Wide field visible light imager

For a DM-less design (§5.1), one can envision additional wide field imagers placed directly in the telescope focal plane with no re-imaging optics. Thus, a mosaic of visible light detectors could achieve a very wide field-of-view. This module, primarily designed for survey science, would possess a limited filter assortment. One can envision mounting a pattern of broadband filters to the arrays or a filter selector similar to a juke box mechanism. To Nyquist sample at 0.7microns, the pixel size should be 7 microns at f/24. However, the telescope images will probably not be at the diffraction limit (see Figure 1.2.1.1-1), so 20μm pixels are likely to be adequate.

CCDs can easily be fabricated with either pixel size, and a very wide field (up to 6×8 arcmin) obtained. Baffling is critical so that the CCDs can be run warm (150K) without swamping the IR and mid-IR modules.

Hybrid Si PIN diode detectors run at 60K, would be similar operationally to the near-IR FPAs, and can be fabricated with 18 micron pixels (see §2.1.3). A single 2k×2k chip would provide a field of about 40 arcsec on a side, and might be very well matched to the intrinsic telescope resolution. These detectors would operate at low enough temperature that they would pose no unique problems for baffling.

CMOS detectors are also now only fabricated with larger than optimal pixels but could in principle be made with 7-9 micron pixels However, the fill factors would suffer even more than at present as the pixels are decreased in size.

The telemetry requirements of this module exceed the yardstick by a large factor. The most likely solution is to use this module in a very limited set of observing modes, with a great deal of on-board processing. For example, it could be used for very long exposures only, with most of the cosmic ray rejection done on board before downloading.

A critical design issue for this module is the ability to baffle scattered light from the sky. First order baffling would be accomplished by placing a stop at the pupil on the fast steering mirror in the UofA telescope design (§5.1), and similar measures can be taken for other telescope designs. However, baffle design details depend on the spacecraft and telescope architecture adopted.

We decided therefore that it was premature to invest our study resources in generating a specific baffle design. However, we anticipate that there might be real differences among architectures being considered in the ability to baffle scattered light and therefore in the ultimate performance of this module, and the proposed NIR wide field imager (§6.1.4).

6.1.3 Near infrared modules

We investigated two different designs for the near-IR modules: a refractive design (NIR module 1), and a fully reflective design (NIR modules 2 & 3) The refractive design is significantly smaller. We envisage several NIR modules would be built with different plate scales, in order to Nyquist sample the NGST images over the full 1-5μm wavelength range. The modules' characteristics are listed in Table 6.1-1. Each will provide high resolution imaging and multi-object spectroscopy via grisms with a corresponding slit mask.

The NIR module 1 five element refractive optical system (Figure 6.1.3-1), composed of simple BaF_2 and SF6 infrared glasses, provides diffraction limited images across the whole detector FOV, illustrated with sample spot diagrams (Figure 6.1.3-2) and the calculated RMS WFE (wavefront error) across the field (Figure 6.1.3-3). The reflective design of NIR modules 2 &3 is shown in Figure 6.1.3-4.

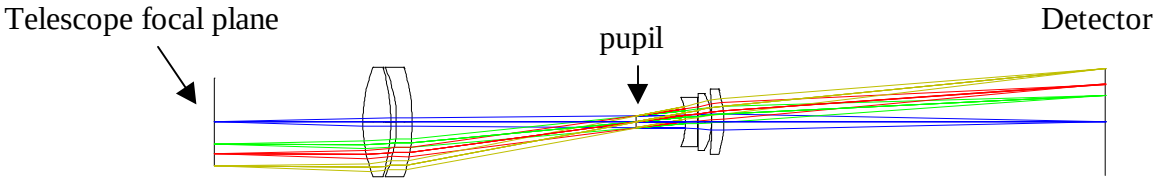


Figure 6.1.3-1 Design of NIR module 1. This module is used for high spatial resolution imaging and multi-object spectroscopy over the 1-2.3μm region. The overall length of this module from re-imaged telescope focal plane to detector is 450mm, compared to an optimal folded 900mm effective length for the refractive NIR module 2 (Figure 6.1.3-4).

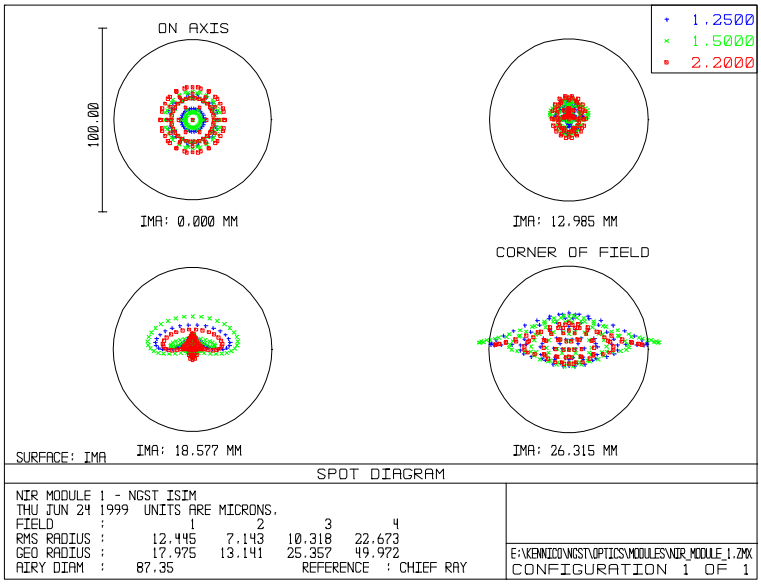


Figure 6.1.3-2 Spot diagrams for NIR module 1 for on axis, intermediate and corner of field. A 4 pixel diameter circle indicating the Airy disk diameter at 1.5µm is shown in each spot diagram. The FOV of this module is 0.7'×0.7'.

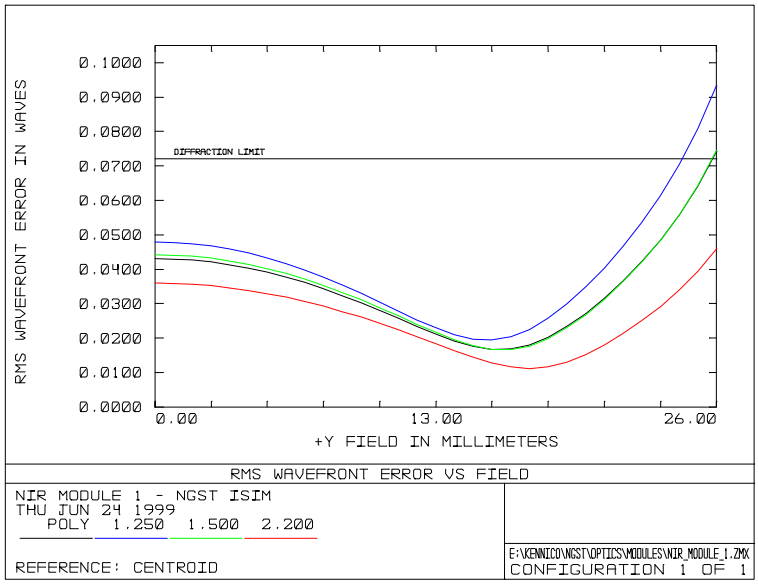


Figure 6.1.3-3 Calculated RMS wavefront error for NIR module 1. The system is diffraction limited across the field (+Y field 26.00mm is at the corner of a 2k×2k detector) for wavelengths 1.25-2.3µm.

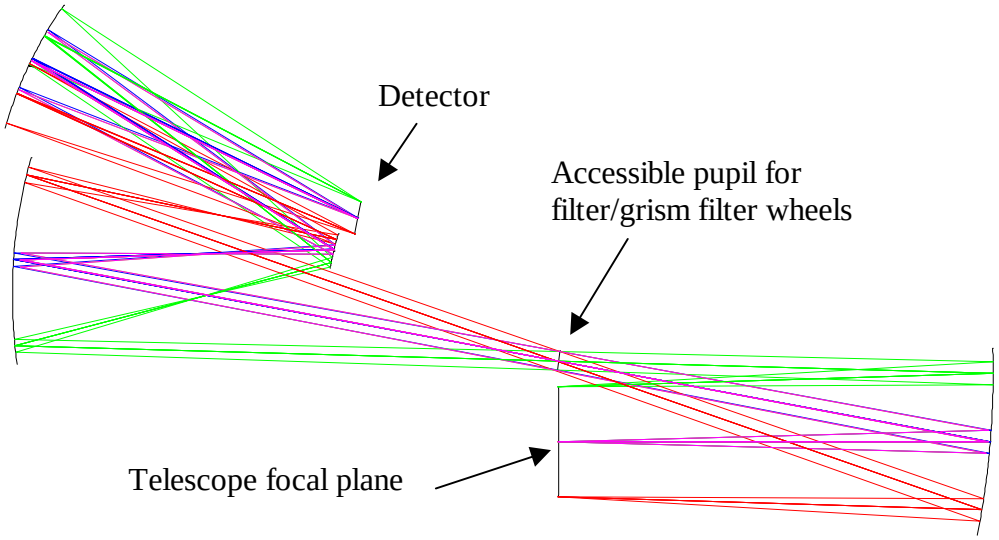


Figure 6.1.3-4 Design of NIR modules 2 and 3 (two nearly identical units). These modules will provide high spatial resolution from 2.3-5µm and wide field imaging over 2.3-5µm with pixels of λ/2D at 3.4µm and 5.7µm, respectively. The larger pixel size for NIR module 3 helps optimize this module for spectroscopy.

The reflective design, shown above in Figure 6.1.3-4 for NIR modules 2 & 3 is replicated in the optical designs of the longer wavelength modules. Each system contains four mirrors, which together provide a better collimated space (systems work at high f/#) and an accessible pupil for placement of filter wheels. The system also provides more degrees of freedom to optimize the image quality over a larger field of view. For multi-slit work, where large fields and good imagery across the fields are important, this design accomplishes both. Additionally, in principle, it is potentially easier to align. The drawback is having an additional component over the more common three mirror (the simplest system to correct all basic aberrations) approach.

All three NIR modules will provide R=100, 300, 3000 multi-object spectroscopy with grisms and a corresponding slit mask. Details of a possible suite of grisms summarizing resolutions and spectral coverage are given in §12.4 Table 12.4-2.

Since much of the core science requires multi-object spectroscopy of faint objects, maximizing the field-of-view of the spectrometers is important. We note that if we want to increase the field-of-view of an individual module by say a factor of 4, the volume of that module increases by roughly the cube of 2, or 8.

However, with the large focal plane of the DM-less design (§5.1), it is possible to have 4 copies of a module and achieve the factor of 4 increase in FOV, with only a factor of 4 increase in volume. This is a most attractive approach for a refractive instrument, whose size is more compact.

6.1.4 Wide field near infrared imager.

A very interesting possibility is to pave the telescope focal plane with 2k×2k HgCdTe arrays (18 µm pixels) in the DM-less design. This could give 6×8 arcmin FOV, Nyquist sampled at 1.8µm wavelength (0.023 arcsec pixels). As with the wide field visible imager (§6.1.2), this camera would have essentially no re-imaging optics, and limited filter assortment.

The cost and lead-time for producing the 100 or so flight qualified FPAs required for this module may be prohibitive. Similar concerns exist already for the yardstick ISIM which requires 80 InSb arrays, which may be more difficult to produce than HgCdTe. However, if the Cassegrain f/ratio is decreased, then a similar field of view can be imaged with slightly greater than Nyquist-sized pixels and at much less cost, because fewer arrays are required. This module also raises telemetry bandwidth issues similar to those already discussed for the wide field visible option.

6.1.5 Mid-IR wavelength imager/spectrometer

We design for a 1024×1024 Si:As BiB array with 27µm pixels. The operating temperature for these detectors is 8-10K so that this module requires active cryogenic cooling. The QE of the Si:As detector is good between 8 and 27 microns. Between 5 and 8µm, the achievable sensitivity will be somewhat compromised by the drop in QE for Si:As IBC detectors, and the fact that InSb and high performance HgCdTe detectors do not work beyond about 5.5µm.

For the direct imager, we chose Nyquist sampling at λ=8µm, resulting in 0.103 arcsec per pixel and a FOV of 1.8×1.8 arcmin^2. This results in acceptable sampling over the 5-27 micron range, oversampling by a factor of 3.4 at the extreme end, 27 microns, and undersampling by a factor of 1.6 at the blue end. The overall design is shown in Figure 6.1.5-1. It is a relatively compact design, comparable in size to NIR Modules 2 & 3.

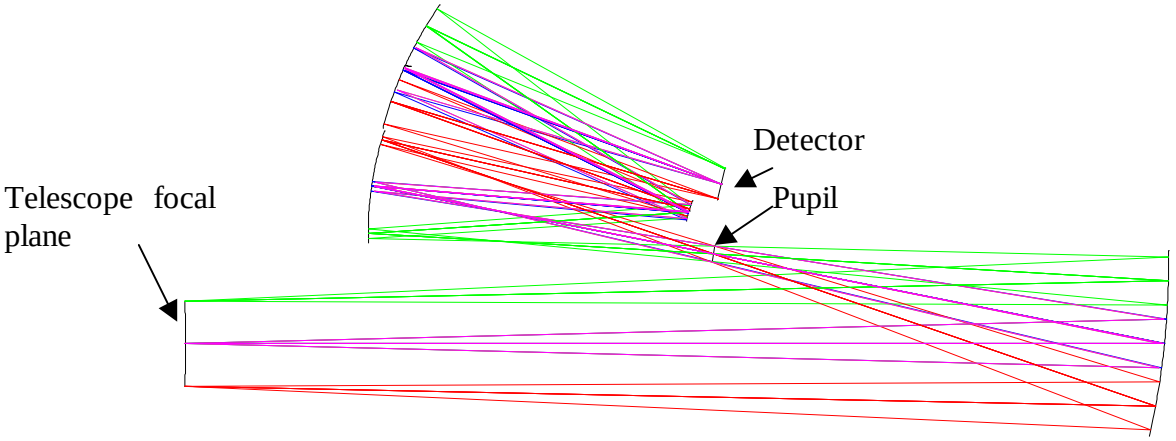


Figure 6.1.5-1 Design of the mid-IR imager and spectrometer.

A selection of broad band filters is required to give color information over this broad spectral range. A limited number of narrow band imaging (1% filters) can also be provided. Shortward of

30 microns, the SIRTF experience has shown that interference filters are available from a number of commercial vendors.

This module will also feature grisms for single and/or multi-object R=100, 500, and 1000 spectroscopy, with a corresponding slit mask at the re-imaged telescope focal plane. R=100 will be best for spectrophotometry of the faintest objects. R=500 is required for some applications, such as the nebular diagnostic emission line studies in high redshift galaxies. R=1000 will aid in detailed studies of bright objects, primarily Galactic sources. Details of a possible suite of grisms summarizing resolutions and spectral coverage are given in §12.4, Table 12.4-2.

At present, the design is not fully optimized, as we decided the mid-IR cross-dispersed echellette module (§6.1.6) could serve both for its original design for R=3000 cross-dispersed spectroscopy and also provide the imaging and low to moderate resolution capabilities of this mid-IR imager & spectrometer.

6.1.6 Mid-IR echellette spectrometer/imager

The proto-star DRM program (one of the seven highest priority programs), would highly benefit from R≥3000 spectroscopy over the λ=6-30μm region, in particular with spectroscopy covering the 12, 17, and 28μm H_2 lines. To cover this wavelength range at this resolution, an instrument module would require many high index grisms with steep angles. A better alternative is to use a cross-dispersed spectrograph to provide the high resolution spectra with large wavelength coverage but only a few grisms. We show now the designs for four echellettes, the third and fourth discussed as part of the long wavelength imager & spectrometer (§6.1.8).

The term "echellette" refers to dispersing elements where $\tan(\theta_B) < 1$, where θ_B is the grating/grism blaze angle. Echellettes are used in orders ~5-20, that is, moderate dispersion. In comparison, high resolution echelles are used in orders >20 or so and typically have $\tan(\theta_B)=2$. These echellette designs are for single slit.

We design for a 1024×1024 Si:As BiB array with 27μm pixels, operated at 8K. Our designs feature a multi-order echellette grism spectrograph with grism cross-disperser. Since this design also relies on grisms placed in the filter wheel, this module can replace the Mid-IR imager & spectrometer (§6.1.5) after populating the filter wheel with lower resolution grisms. This has attractive cost reductions as summarized in §9.4. The optical design of this module is shown in Figure 6.1.6-1. Spot diagrams are virtually identical to the similar design for a long wavelength imager and shown in Figures 6.1.7-2 and 6.1.7-3.

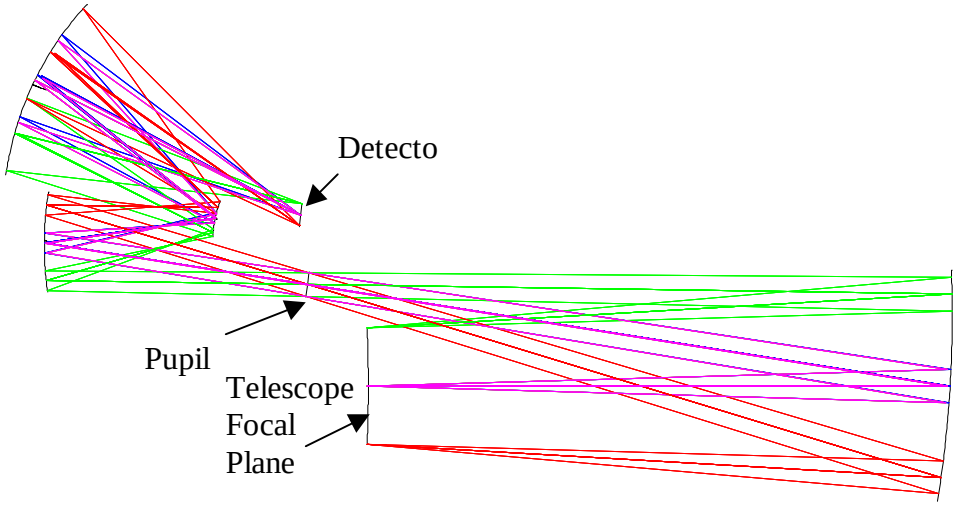


Figure 6.1.6-1 Mid-IR echellette spectrometer optical layout. A 38mm diameter pupil can locate the grisms for R=100-1000 spectroscopy, as well as a grism plus grism cross-disperser for R=3000 coverage over the 5-17µm region, using two combinations. The module is the largest of them all, measuring approximately 1.7m in length in this folded configuration.

We undersample the spatial resolution, and use 0.4 arcsec slit width projected onto 2-3 pixels. Using grisms provides a compact form, previously only attainable with lower dispersion prism cross-dispersers, as well as enough dispersion in the mid-IR to give adequate order separation. A similar instrument, built for the NIR (1-2.5µm), has been successful at implementing this approach to an all-grism based cross-dispersed spectrograph (McGregor, et al. 1994). The grisms required will necessitate the need for high index material, if we opt to keep the pupil size the same. For our 38mm diameter pupil, we have found that KRS-5 grisms will work adequately, although moving to higher index material (e.g., Si or Ge) might relax the steep angles required for the >10µm wavelength echellettes.

The 1024×1024 pixel format of the detector then allows the 5.2-18.75 micron range to be covered with 2 echellette/cross-disperser choices. Two additional echellettes, part of the long IR imager & spectrometer (§6.1.7), would extend this coverage to 35µm for use with a Si:P array. Figures 6.1.6-2 illustrates the two dimensional spectral layout at detector. Details of the grism parameters used in generating these plots are given in §12.4, Table 12.4-3.

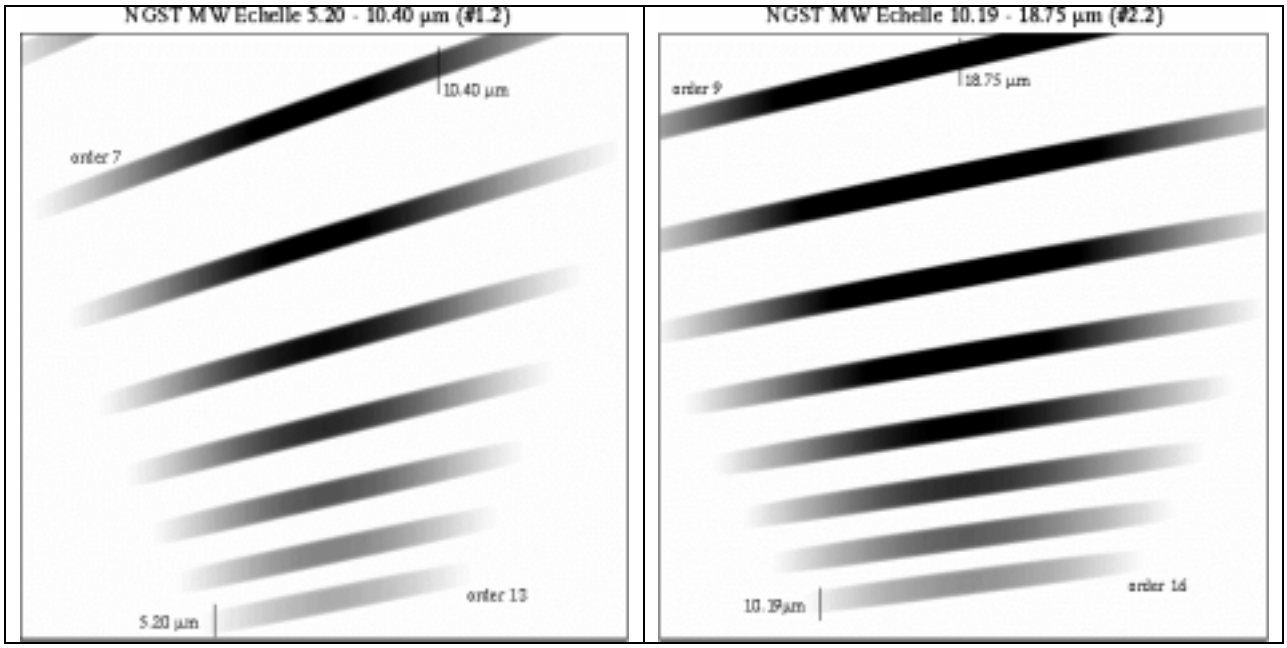


Figure 6.1.6-2 Illustration of proposed two dimension spectra for R=3000 cross-dispersed echellette. The left image shows orders 7-13 covering 5.20-10.4µm on a 1024×1024 pixel array. The right image shows orders 10-16 covering 10.19-18.75µm. These fixed formats each require one echelle grism and one cross-dispersing grism, both optimized for the wavelength regions. They can be placed in parallel with two filter wheels at the pupil position.

6.1.7 Long wavelength imager/spectrometer

We design for a 512×512 pixel BIB/IBC array with 50µm pixels, using Si:P BIB detectors. The design would be unchanged with a 1024x1024 array with 27µm pixels. The short wavelength limit is 20 microns. This provides substantial overlap and redundancy in the 20-27 micron region with the Si:As mid-IR modules, but with pixels sized for the longer wavelengths. The long wavelength module cut-off will be limited by the chosen detector. Since we recommend using an Si:P IBC array (§2.3.2), this will be around 33-34µm.

We choose Nyquist sampled images at λ=27µm, resulting in a plate scale of 0.387 arcsec per pixel and total FOV of 3×3 arcmin^2. The design is shown in Figure 6.1.7-1. Examples of spot diagrams for rays at center and edges of the field as well as the WFE across the detector are illustrated by Figures 6.1.7-2 & 3.

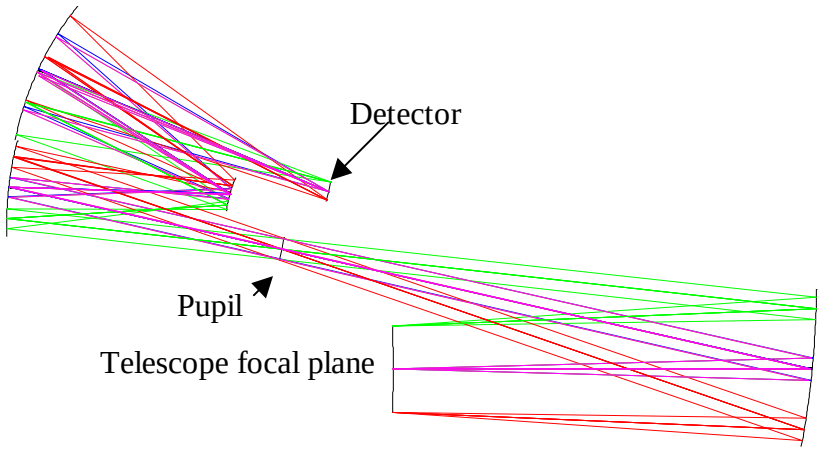


Figure 6.1.7-1 Optical layout for the long wavelength imager/spectrometer. This module is the second largest, measuring about 1.5m in length in its folded position. It has a 40mm diameter pupil.

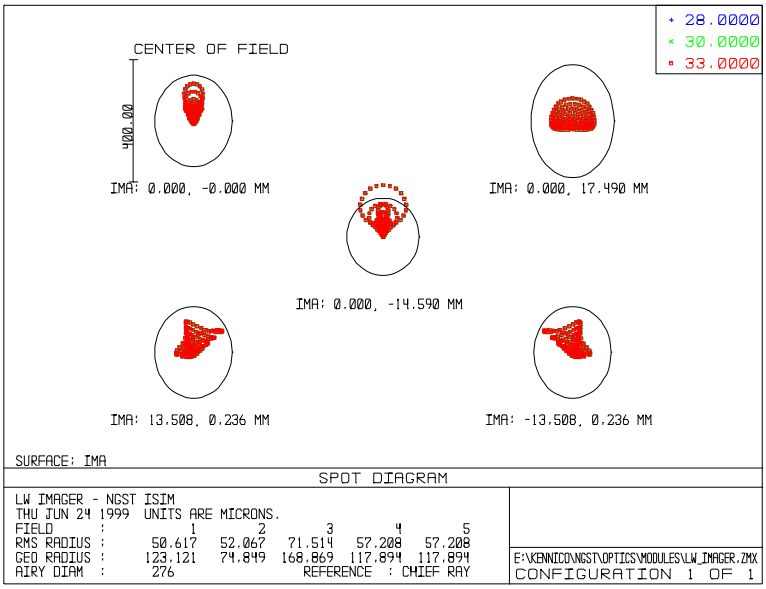


Figure 6.1.7-2 Spot diagrams for optimized optical design for center and edges of field on a 512×512 detector. A 4.8 pixel diameter circle indicating the Airy disk diameter at 27μm is shown in each spot diagram.

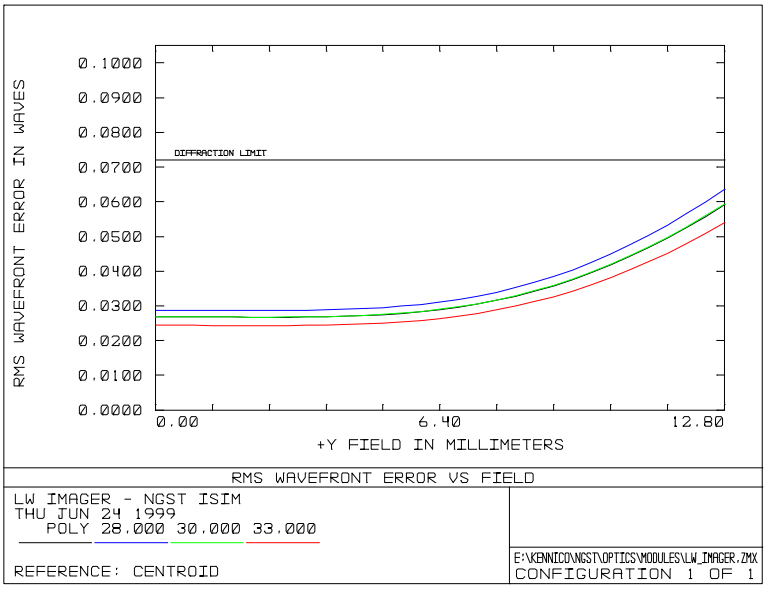


Figure 6.1.7-3 RMS wavefront error across the 512×512 array. The system is diffraction limited at all wavelengths.

The module will provide single or multi-object grism spectroscopy with a slit mask for R=500 and R=1000. With our design (40mm diameter pupil), the highest resolution suggested for a KRS-5 grism would be R~3000, with prism angle near 60º and 42 l/mm. Higher spectral resolution grisms would be possible with a higher index material, but we found no other promising candidates that transmit longward of 20 microns (§12.5, Figure §12.5.1-1). The suggested grism complement for this module, along with resolutions and wavelength ranges is given in §12.4 (Table 12.4-2).

6.1.8 Long wavelength echellette spectrometer /imager

Stemming from our preliminary work on a conceptual design for a cross-dispersed grism echellette for the mid-IR (§6.1.6), we realized that this identical design could be adapted to the requirements of the long wavelength imager and spectrometer. We present a design of a R=3000 dual echellette covering the 17-35µm range, to complement the two Mid-IR echellette configurations. Figure 6.1.8-1 illustrates the two dimensional spectral layout at detector. Details of the grism parameters used in generating these plots are given in §12.4, Table 12.4-3. With our design (38mm diameter pupil), the highest resolution suggested for a KRS-5 grism would be R~3000, with prism angle near 60º and 42 l/mm. Higher spectral resolution grisms would be possible with a higher index material, but we found no other promising candidates that transmit longward of 20 microns (§12.5, Figure §12.5.1-1).

We design for a 512×512 pixel BIB/IBC array with 50µm pixels, using Si:P BIB detectors, or a 1024x1024 array with 27µm pixels. Either type of array is satisfactory. The 1024x1024 array might be of lower cost because it has a high degree of commonality with the 1024x1024 Si:As

BIB/IBC array. The optical design can be identical with that in Figure 6.1.6-1. Spot diagrams are very similar to those described in Figure 6.1.7-2 and 6.1.7-3.

As for the LIR module, this module will provide single or multi-object grism spectroscopy as well as a high resolution mode. The lower spectral resolution gives a single exposure coverage of the entire wavelength region of the module. KRS-5 grisms can be fabricated with acceptable ruling requirements and prism angles. This "spectral energy distribution" mode will be used to measure the overall thermal dust emission SED, and has adequate spectral resolution to resolve broad dust and PAH features in the mid-IR. Lower resolution (R<500) grisms are not necessary in this module, since at these wavelengths the thermal background from the telescope and sunshade dominates the detector noise. Hence lower dispersion SEDs for faint sources can be obtained by binning, at no penalty in S/N. The R=1000 mode is desired for detailed study of spectra lines, notable H_2. Multi-object spectroscopy could be implemented with microshutters or sliding lsits (as shown in §4.3, micromirrors probably will have inadequate performance). The suggested grism complement for this module, along with resolutions and wavelength ranges is given in §12.5 (Table 12.5-2). The module can also serve all the imaging programs by removing the grisms and slits from the beam.

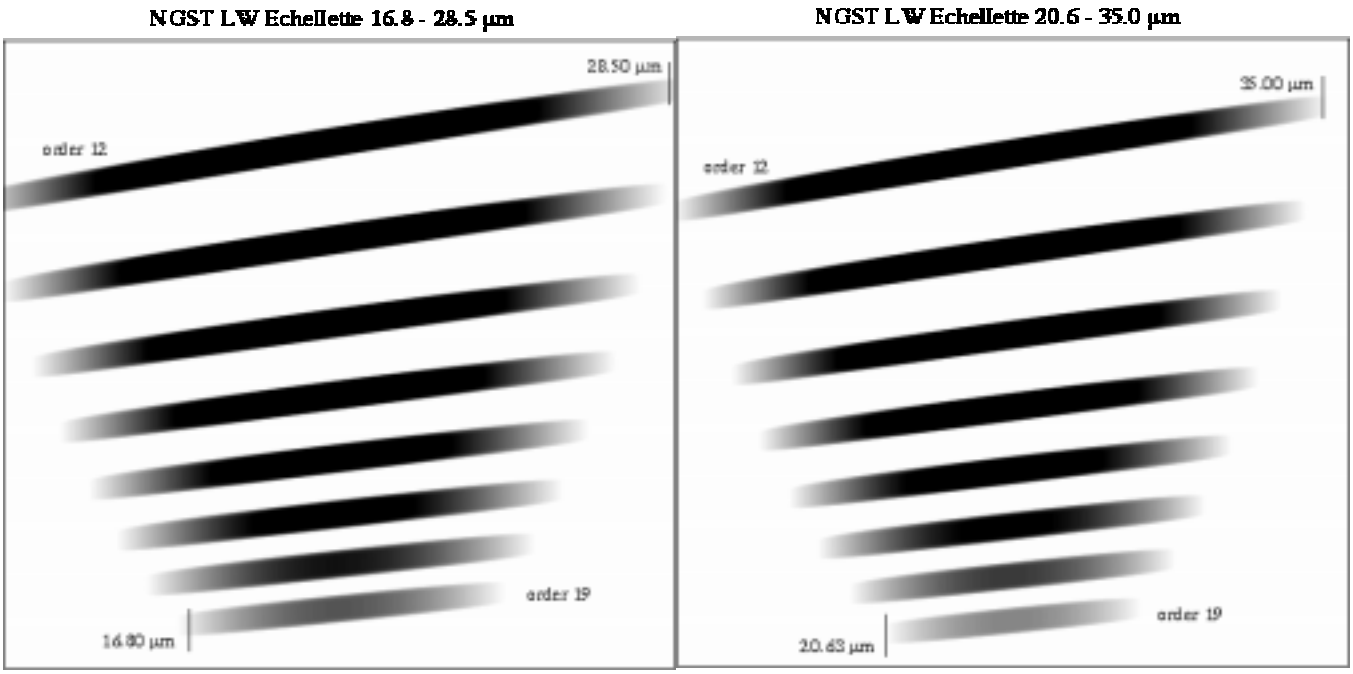


Figure 6.1.8-1 Illustration of proposed two dimension spectra for R=3000 cross-dispersed mode of the long wavelength imager & spectrometer. Two grisms are used to cover 16.8 to 18.5 and 20.6 to 35.0μm respectively, using either a 512x512 50μm pixel or a 1024x1024 27μm pixel array.

6.1.9 Summary

The various module designs explore a variety of possibilities, and each is optimized for somewhat different applications. However, for reasons of cost as well as overall ISIM complexity it would be inadvisable to build all of them. We have selected a subset that we believe provide NGST with the best compromise between cost and capabilities. A consideration in these selections is that two identical or nearly identical modules will cost less than two modules of different design.

We would recommend an ISIM with:

- a visible light imager, if the telescope provides adequately good images;
- NIR modules 1-3 to provide optimized capabilities over the prime wavelength range of NGST; and
- two identical (except for detector type) echellettes to provide imaging and spectroscopy from 5 to 35μm.

Further optimization could consider whether the NIR modules could all have similar designs, to reduce their overall cost. We would also include both

- visible and
- near infrared wide field imagers,

and we assume that the telescope Cassegrain f/ratio would be adjusted to bring the costs of these imagers down within the affordable range.

6.1.10 Comparison of DRM integration time with the yardstick.

We have estimated the relative integration times to achieve the DRM with our ISIM and with the yardstick. The basic result was that our ISIM required only about 2/3 the time as the yardstick. In this particular comparison, it was predicted that improving the sampling in the MIR echellette/imager would increase the speed in completing the DRM despite the penalty in field of view. This gain depends critically on the exact mix of science programs and further study would be advisable before implementing it. However, a much larger gain resulted from out inclusion of a wide field near infrared imager. A more modest gain also resulted from the wide field visible imager. We conclude that the project should consider seriously both of these capabilities, and particularly the near infrared one.

In addition to its improved performance on the DRM, our proposed ISIM has greater wavelength coverage compared with the yardstick, extending down to 0.4μm and out to 35μm. Its spectrometers for the MIR and LIR are also much more capable than the one proposed for the yardstick.

6.2 Opto-mechanical design

Drawings showing how our ISIM fits into the NGST observatory are shown in Figures 6.2-1 and 6.2-2. We selected the LM OTA reference design, but packaging our ISIM with the Goddard yardstick OTA would result in a similarly looking observatory. We omitted the radiators that cool the ISIM in all of the diagrams in this section in order to enable the components to be seen. Radiator geometry is discussed in § 6.3.

We selected the ISIM structure appropriate for the Lockheed Martin OTA design, but note that other candidate telescope designs are compatible with our ISIM. The advantages of the Lockheed Martin OTA include a large $22m^3$ ISIM volume. Figures 6.2-3 and 4 show the path of rays from the secondary telescope mirror (from the right) as they travel to the tertiary mirror, next to a fold mirror, to the fast steering mirror, to another fold mirror, and then into the ISIM. Note that only the tertiary mirror intrudes into the ISIM volume, and even so only at the front of the ISIM. This allows us to flexibly place ISIM instrument modules within the ISIM volume, and to satisfy other constraints imposed by mechanical and thermal interfaces.

For our packaging concept, we included the six modules plus two wide field imagers designated in Table 6.1-1. We oversized the tertiary and subsequent OTA mirrors to enable an even larger field of view, from 4 x 10 arcmin to 6 x 30 arcmin. These changes are expected to have a minimal system impact. As each module has its own optical bench, they may be optically aligned and tested for mechanical integrity independently. A disadvantage is that the overall mass can be greater than an integrated design which allows instruments to share a common bench. Modularity issues are discussed in more detail in §7.1.

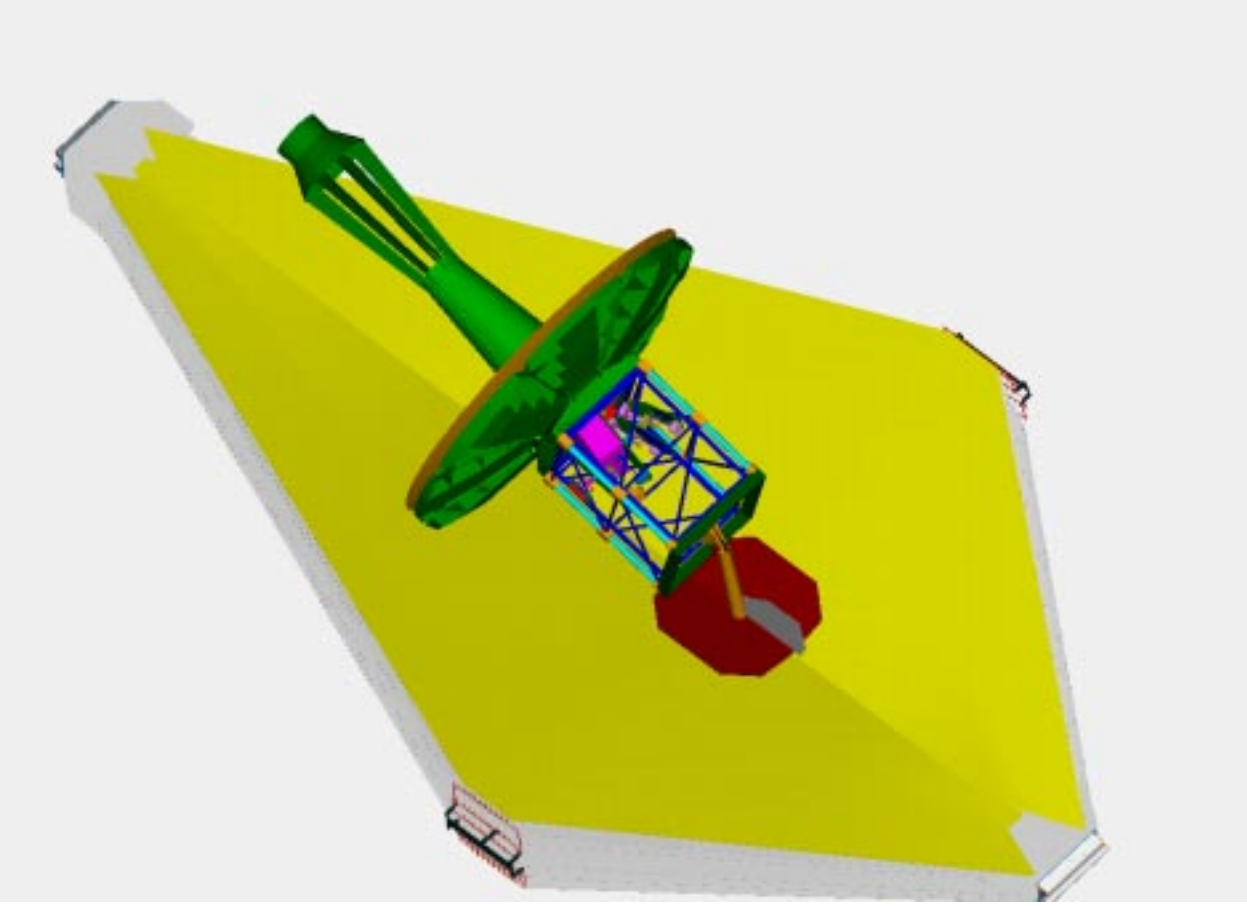


Figure 6.2-1 Our ISIM, occupying the space within the framed structure, integrated with the entire NGST observatory. The ISIM radiators are suppressed for clarity.

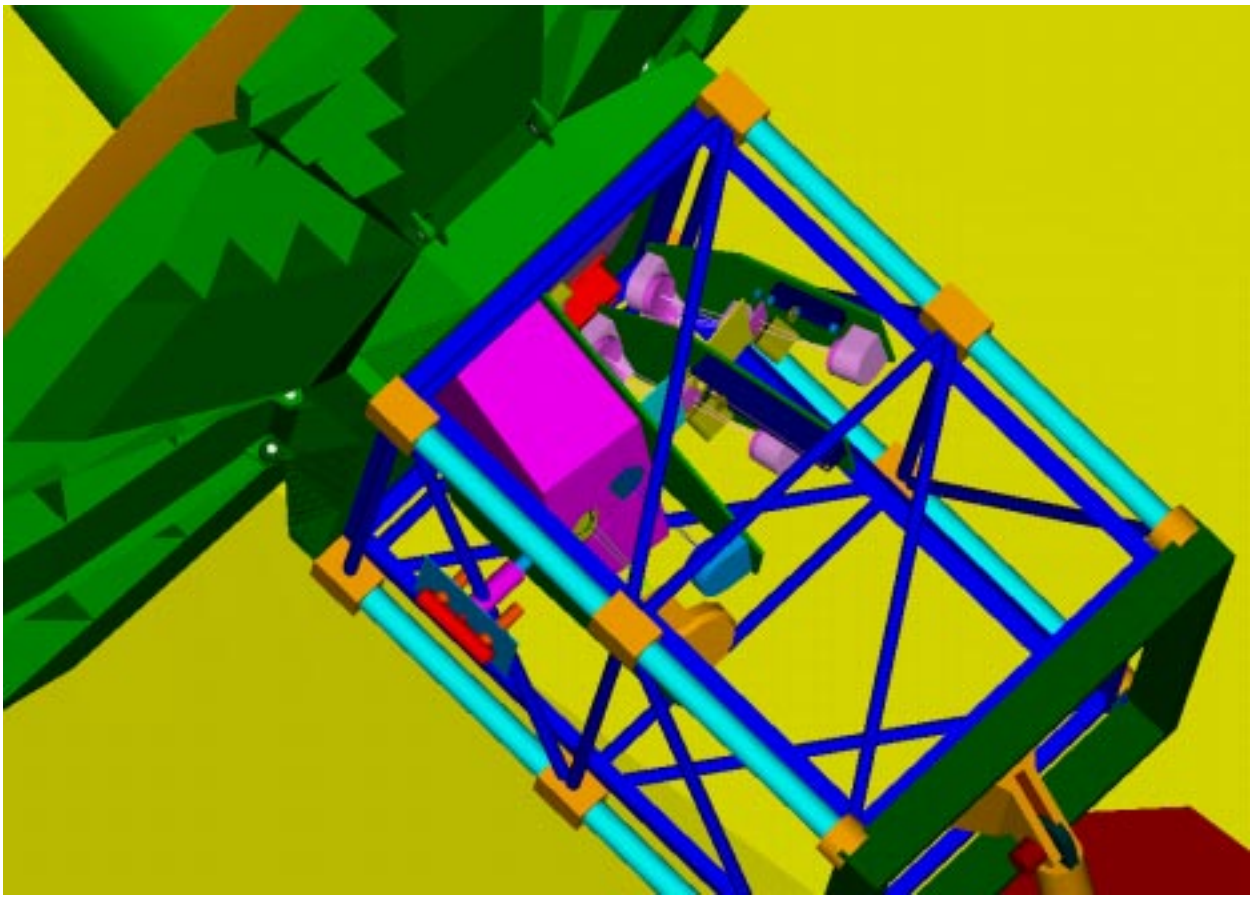


Figure 6.2-2 Expanded view of Figure 6.2-1, centered on telescope focal plane. Note the magenta colored box cooled to 15 K, enclosing the MIR and LIR detectors.

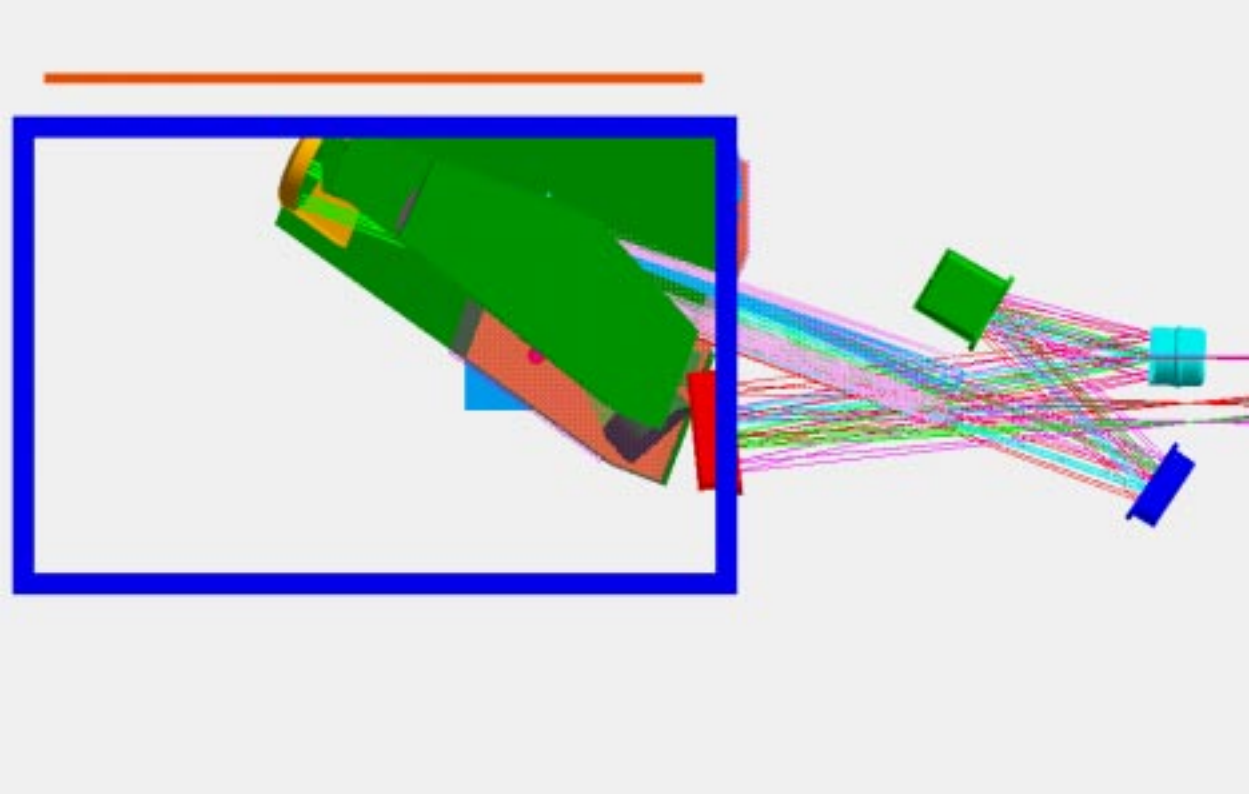


Figure 6.2-3 ISIM structure (blue) in the LM OTA design where only the tertiary mirror (red) intrudes into the ISIM volume. Our module concept (green) fits nicely into the spacious ISIM volume.

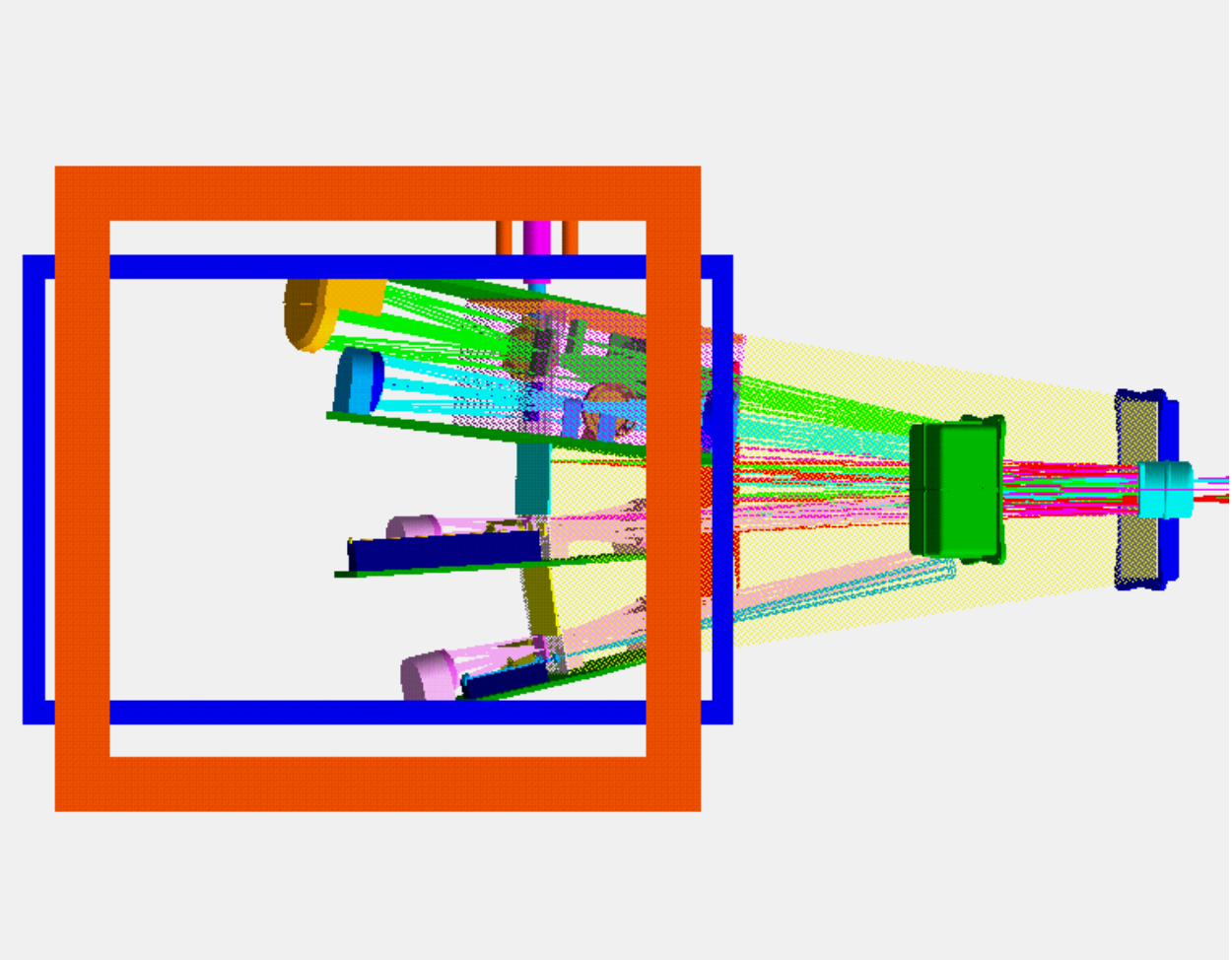


Figure 6.2-4. Figure 6.2-3 has been rotated by 90 degrees along the telescope optical axis. This allows identification of rays intercepted by each module from the curved telescope focal plane. The boundary of a radiator (in orange) is shown for reference.

Although we have not done a complete packaging assessment, a preliminary examination indicates that the entire complement could also be fit into the smaller $11m^3$ ISIM volume provided by OTA configurations such as the drop-leaf telescope. However, substantially less room is left over for other functions.

The telescope focal plane is 6×30 $arcmin^2$ and is underpopulated by the instrument suite. This is shown in Figure 6.2-5, where the color coded fields of view of each instrument module are shown to scale. This figure also indicates the relative placement of the FOVs. The long wavelength and mid-IR wavelength modules are placed on the left side of the diagram to provide a continuous volume for the 15K and lower temperatures they require. The boundary of the 15 K region is shown in a translucent pink color. The pulse tube extends into this region to provide necessary cooling. Our modular design gives us the flexibility to move FOVs about the telescope focal plane, to allow more instruments, or arrays at the focal plane for other OTA designs. Alternatively, we may repackage our strawman complement in a telescope with a similar field of

view down to the equivalent of 6 arcmin by 18 arcmin. The instrument packaging is shown from the side in Figure 6.2-6.

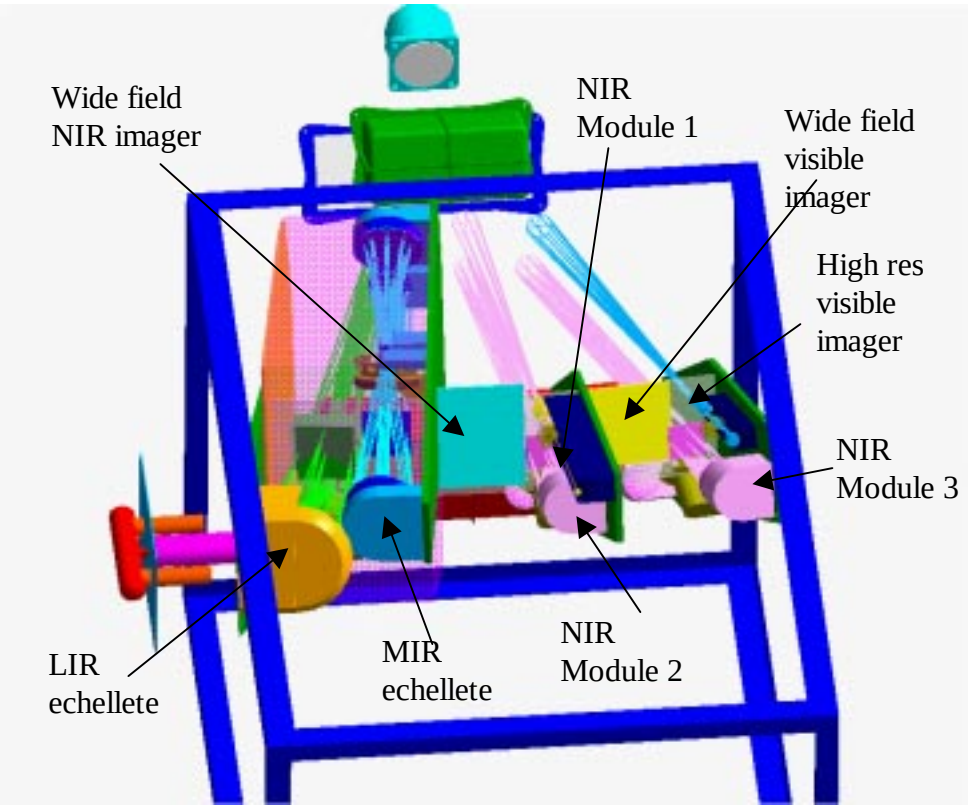


Figure 6.2-5 Overview of the ISIM package with the instrument modules indicated.

We have gone to some effort to package the instrument complement in a realistic way. Fold mirrors have been inserted into the instrument modules to provide a compact volume and to co-locate detectors requiring cooling below 40K. This reduces the cryo-cooler cooling load and eases thermal management as discussed in §6.3.

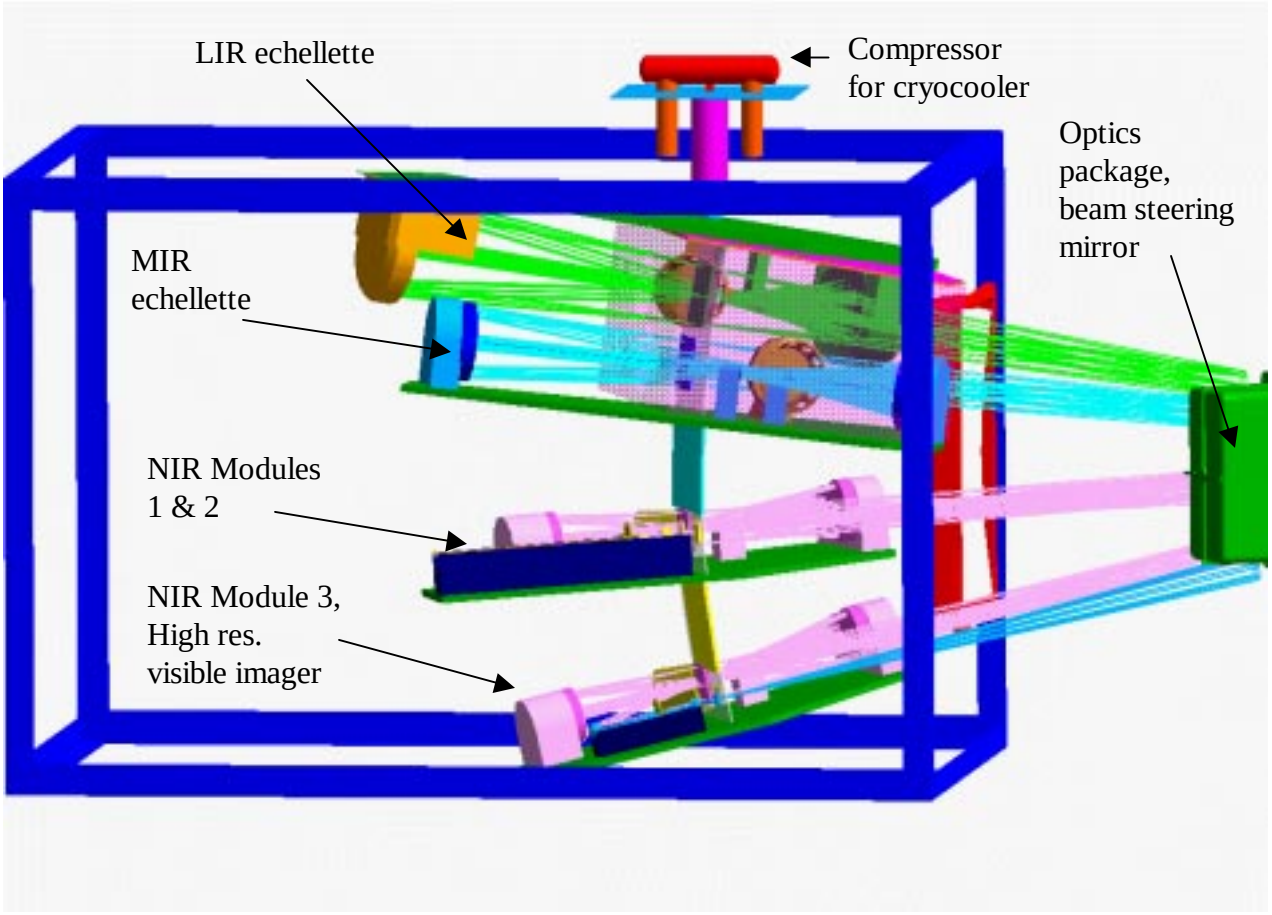


Figure 6.2-6. Side view of the ISIM package.

We present a close-up view of the long wavelength imager and spectrometer module (or the MIR and LIR echellettes, which are similar) in Figure 6.2-7. The diagram illustrates the physical layout of the optical bench and the mounting structure for the mirrors and the filter wheel. The telescope focal plane corresponding to this module's 3×3 arcminute^2 FOV is shown as a dark blue square. The rays strike a collimating mirror, and are reflected back to a pupil where a 12 position filter wheel is located. Not shown is the enclosing box cooled to 15K, which surrounds this filter wheel and all succeeding optics. The function of this box is to prevent stray light from other instruments providing excessive background photons from reaching the Si:P detector. The field of view of the detector is limited by a snout cooled to 8K, to match the cone of light emanating from the last fold mirror.

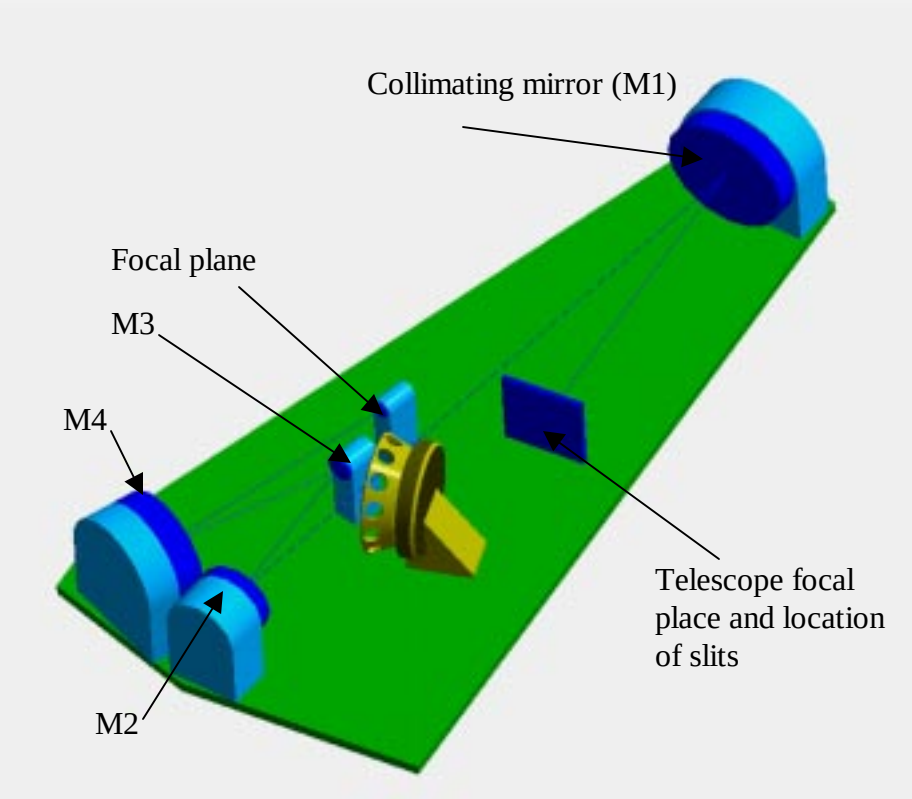


Figure 6.2-7 Mechanical packaging of long wavelength imager/spectrometer or the MIR and LIR echellettes (all are similar).

One possible use for the available space remaining in the ISIM module is shown in Figure 3.4-1 (see §3.4). Although our baseline is to use pulse tube cooler, a solid hydrogen cryostat can easily fit into the ISIM structure with no impact to theoretical and mechanical design of the ISIM. The size of the hydrogen cryostat will easily accommodate a 10 year mission lifetime, yet it occupies a very small fraction of the ISIM volume. The pulse tube or turbo-brayton coolers occupy even smaller volumes.

The total mass of the strawman complement is 450 kg. This estimate assumes solid body construction of aluminum but that a significant level of relieving for light-weighting is used for the optical bench and mirror mounts. Aluminum is a good choice for these functions as it has mechanical stability and established heritage for similar instruments used in space. A contingency of 35% should be added for a total projected mass of 608kg. Alternative materials, such as graphite used in a honeycomb structure can substantially lower the ISIM mass. If such materials are employed, the mass could drop to 270 kg or less (plus contingency). However, cost is likely to increase substantially over an all-aluminum design.

6.3 Thermal assessment

Our baseline instrument configuration calls for a total of eight modules (including the wide field imagers) to be cooled to temperatures ranging from 8 K to the ISIM cavity temperature of 38K. Table 6.3-1 presents the specific temperature and cooling requirements for each instrument.

INSTRUMENT	TEMPERATURE REQUIREMENT	POWER DISSIPATION	COMMENTS
Visible Imager	<40K	FPA: 25-mW Filter Wheel: 1.4-mW	•Cooled by coupling with ISIM radiator.
Wide Field Visible Imager	<40K	80-mW	•Cooled by coupling with ISIM radiator. No optics required.
NIR Module 1	<40K	FPA: 4-mW Filter Wheel: 1.4-mW	•Cooled by coupling with ISIM radiator.
NIR Module 2 (Same layout as for NIR Module 3).	<40K	FPA: 4-mW Filter Wheel: 1.4-mW	•Cooled by coupling with ISIM radiator.
NIR Module 3 (Same layout as for NIR Module 2).	<40K	FPA: 4-mW Filter Wheel: 1.4-mW	•Cooled by coupling with ISIM radiator.
Wide Field NIR Imager	<40K	80-mW	•Cooled by coupling with ISIM radiator. No optics required.
MIR Spectrometer/ Imager	FPA: <10K Filter Wheel: <18K Optics, after FW: <18K	FPA: 1-mW Filter Wheel: 1.4-mW	•Filter Wheel and optics cooled by coupling with pulse tube 2rd stage. FPA cooled by 3rd stage pulse tube cryocooler.
LIR Spectrometer/ Imager	FPA: 8K Filter Wheel: 15K Optics, after FW: 15K	FPA: 0.25-mW Filter Wheel: 1.4-mW	•Filter Wheel and optics cooled by coupling with pulse tube 2rd stage. FPA cooled by 3rd stage pulse tube cryocooler.

Table 6.3-1 Summary of NGST instrument cooling requirements.

Three unique temperature zones have been identified by combining similar temperature requirements: (1) an 8 K temperature capability for cooling the MIR and MIR FPA's, (2) a 15 K temperature zone for cooling of the low temperature filter wheels and re-imaging optics, and (3) a <40 K temperature zone within the ISIM enclosure for cooling of all remaining modules containing FPA's, re-imaging optics, and other elements. To take advantage of this commonality

in temperature zones, we have co-located each of these temperature zones into a common corner of the ISIM. A single cooling system (regardless of the type selected) can be easily accommodated for cooling at each temperature zone. This offers a further benefit to the design as reducing the extent of the cooled volumes will reduce surface area and supported mass; this leads to reduced system parasitic heat load requirements in the colder zones as well. This is very important for NGST since even small additional heat loads at the colder temperature zones will have a significant impact on the overall system power and mass. We have also assumed that the ISIM radiator will be used to dissipate the 0.100-W from the telescope fast steering mirror (FSM). In the yardstick OTA, the fast steering mirror is located in the ISIM space. Other architectures, such as the LM OTA, place the FSM outside of the ISIM. For these architectures, alternative paths of removing heat from the FSM might reduce the ISIM radiator size substantially.

In Figure 6.3-1, the cooling schematic is shown for the full complement of the eight NGST instruments. Our baseline thermal design is to cool the ISIM (and each component tied to it) to less than 40 K (38K) by a single stage radiator operating at 35K. A guard radiator may be added to further reduce the radiation heat leak from the sunshade, but our preliminary analysis suggests that this may not be necessary. The ISIM will be insulated with a 0.25-inch thick blanket of MLI to reduce the radiation coupling between the sunshade and ISIM. The remaining five (5) surfaces can be used for radiating the ISIM parasitic heat load; 22.1-m^2 is required to provide the necessary cooling. This area represents only a fraction of the total surface area made available by the ISIM surfaces so it is reasonable to expect the final design to double as panel stiffeners to the ISIM structure. The radiator geometry is shown in Figure 6.3-2.

An evaluation of system heat loads was made for each of the three temperature zones (8 K, 15 K, and the 38K ISIM cavity). Calculated heat loads include the effects of support structure, MLI blankets, instrument FPA leads, component power dissipation, and direct radiation. We assumed that the detector power dissipation was 1 mW per 1024 x 1024 pixel FPA for the low background NIR modules. We read the other FPAs at 1/4 the NIR rate to reduce power dissipation, and accept a small penalty in effective read noise. Background photon noise, rather than detector noise becomes significant for these modules. In calculating the support heat load, the elements were sized to survive buckling with a 10-g quasi-static load imposed, while simultaneously maintaining a frequency of >35 Hz. The frequency constraint assures that the preliminary tube sizing yields a realistic support conductance, subject to the accuracy of the mass estimate.

Given the preliminary nature of the design, we guaranteed a conservative thermal analysis in two ways. First, we doubled the expected instrument mass in sizing the support structure. Specifically, for each module we assumed 2.3 kg each for the FPA's and their associated support structure (6.8 kg total) and 45 kg for each of the re-imaging optics assemblies (140 kg total). Second, we imposed an extra 50% margin on the lowest temperature cooling stage, formally by assuming it was used for three modules rather than the recommended two. In calculating the direct radiation to the 8K and 15K temperature zones, a low emissivity (0.035) surface was assumed. Studies have shown that a low emissivity surface will outperform an MLI blanket at these boundary temperatures - so long as localized regions of high effective emissivity (e.g. cavity effects or

exposed surfaces with high emittance) can be avoided. Should the detailed design yield such effects, localized blanketing of low emittance will be used as appropriate.

FPA and instrumentation lead wire heat leaks were derived assuming that flat ribbon cables with manganin conductors, each having a cross sectional area equivalent to a 0.005-inch diameter wire, would be used. A total of 100-leads (50-signal, 50-ground) were assumed for each instrument. Effective lead lengths for heat leak calculations were taken to be as follows: 1.25-m (100K to 38K), 0.3-m (38K to 15K), and 0.2-m (15K to 8K). This partitioning is consistent with a total lead length of 3.0-m from the warm electronics box (250K) to FPA. It is assumed that 0.5-m of length between the 250K and 100K section will be required, and that some reduction in the effective thermal length internal to the ISIM will be necessary as mechanical support of the cabling to adjacent structure will probably be required inside the ISIM. Total mechanism power dissipation is 10 mW for the filter wheel motors and focal plane mask mechanisms. A heat map describing the ISIM heat flow is summarized in Table 6.3-2.

Our baseline for cooling the 8K and 15K temperature zones is to use a 3-stage pulse tube mechanical cryocooler (See discussion in Section 3). The compressor and ambient heat exchanger of the pulse tube are radiatively cooled to less than 250K by a passive radiator. Fiberglass thermal isolators off the ISIM structure support this radiator and cryocooler hardware (compressor and pulse tube). Location of the 250K radiator is such that neither the NGST telescope optical components nor ISIM 35K radiator surfaces view the warm 250K cryocooler radiator. Soft mounts would not introduce significant heat increases into the system, should they be needed to reduce cryocooler vibrations to acceptable levels.

The second stage of the cryocooler is used to cool the 15 K MIR and LIR re-imaging optics; the third stage cools the MIR and LIR FPA's. No cooling of the ISIM instrument is performed by the pulse tube first stage, as it only serves to improve the net cooling to the lower temperature stages. We did explore the possibility of using this first stage for 40 K cooling, but rejected this because it would have required substantially higher electrical power.

LM has three compressor assemblies which could be used to provide the requisite PV input to the pulse tube. These include a high capacity (35 mm diameter piston) compressor developed as part of the Low Cost cryocooler program; the L-1710 compressor that was developed for use with our split-Stirling cryocooler; and the Mini-cryocooler compressor currently under development for NASA/GSFC. Preliminary sizing suggests that the L-1710 compressor assembly would be the optimal choice for NGST, providing over a factor of three margin in the required PV. Estimated mass and power requirements with the L-1710 compressor driving the three-stage pulse tube are included in Table 6.3-3, along with the mass of the ISIM radiator. Our conclusion is that our baseline thermal system adequately conforms to all of our ISIM requirements with ample engineering margins. In addition, relevant technology required for instrument cooling exists and has been demonstrated in the laboratory giving us high confidence in achieving the design goals.

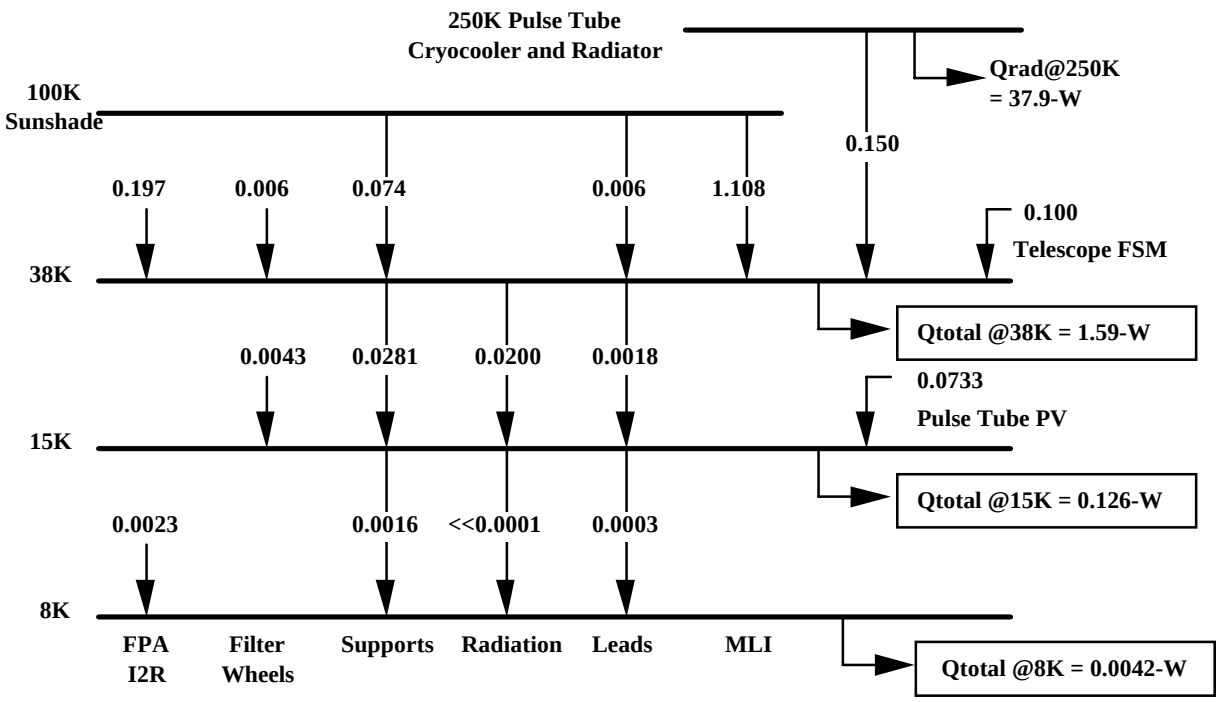


Table 6.3-2 NGST ISIM Heat Map

Performance Characteristics	
Mass (kg)	
Cryocooler	10.0
38K Radiator	39.2
Totals (kg)	**49.2**
Power (W)	**52.3**

Table 6.3-3 Summary of NGST Cooling Characteristics

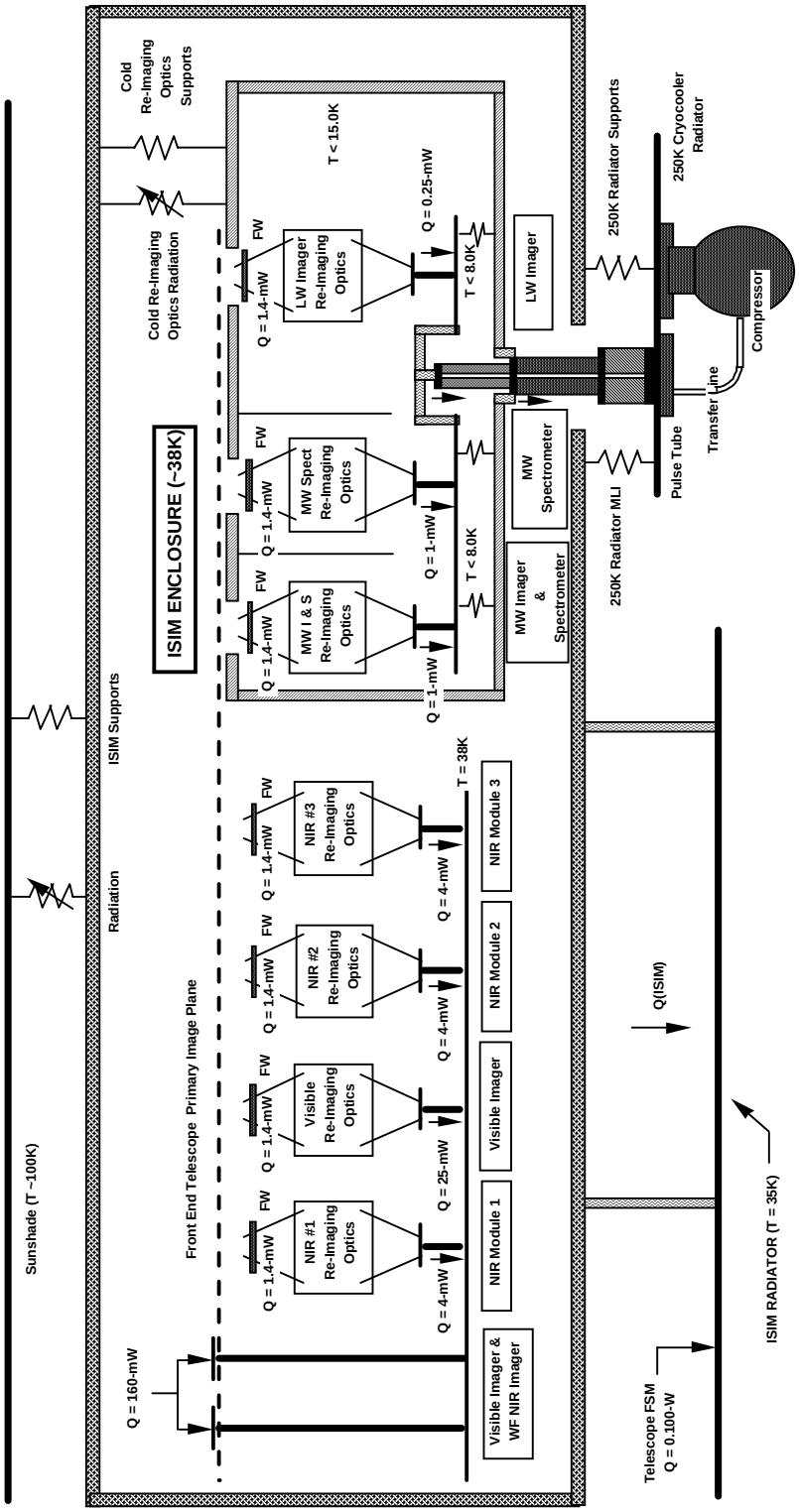


Figure 6.3-1 Cooling configuration schematic detailing heat sources, significant thermal loads and conduction paths.

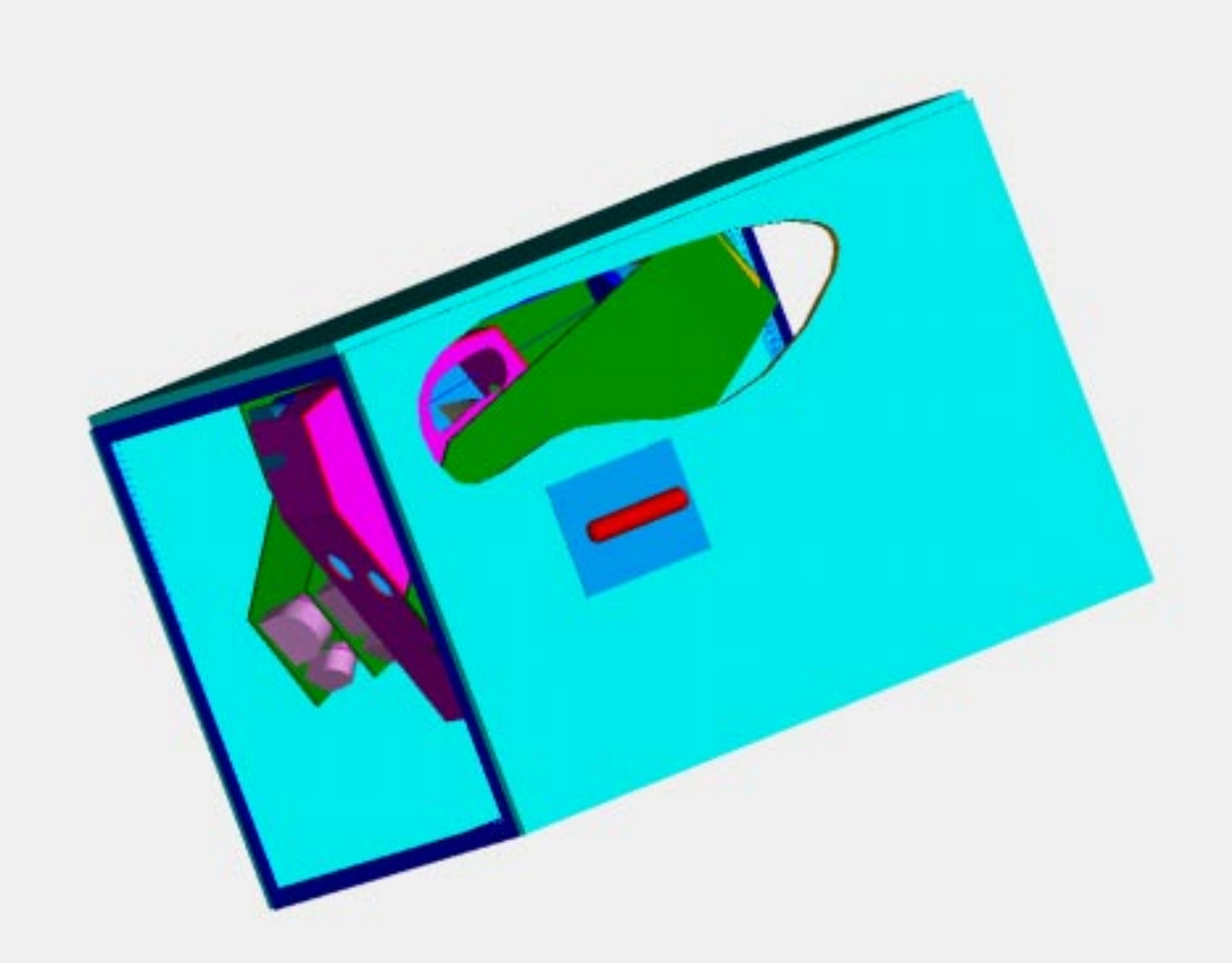


Figure 6.3.-2. Radiator Geometry for our ISIM. 35 K radiator surfaces, shown in turquoise, cover three surface of the ISIM structure. The pulse tube compressor has a separate 250 K radiator, shown in blue. The cutaway in the radiator reveals the green ISIM optical benches cooled to 38 K. The cutaway in the magenta box reveals the MIR and LIR components inside. The box and its contents are cooled to 15 K.

6.3.1 Comparison with yardstick thermal model

Both our ISIM concept and the yardstick ISIM (Parrish et al. 1999) need to cool a complement of NIR and MIR detectors and so their thermal designs have similarities. One significant difference is the NIR focal plane temperature and its effect on radiator design. The yardstick has its InSb FPAs operating at 30 K with a total detector power dissipation of 160 mW. Three stages of radiative cooling, each with its own independent radiator, are needed to define separate temperature zones for the ISIM optical bench, parasitics and the detector. This results in a complex radiator design. Our ISIM has more pixels than the yardstick and dissipates 200 mW, but operates at a significantly higher temperature of 38 K. Only a single temperature zone is required,

which simplifies ISIM construction and installation. The overall radiator area is smaller than for the yardstick, which also results in mass savings.

The total mechanism dissipation in both systems is dominated by the 100 mW of fast steering mirror power; after subtracting this component, the remaining mechanism dissipation for filter wheels and focal plane masks is 48 mW for the yardstick and 10 mW for our ISIM. Our mechanism disspation could be further reduced by using superconducting electrical components, especially in the MIR and LIR instruments operating at 18 K. The yardstick uses a turbo brayton cryocooler capable of providing temperatures of 6 K, whereas we have chosen a pulse tube cooler operating at 8 K. Both cryocoolers require a compressor operating at temperatures in the 220 – 250 K range. The electrical power required by the turbo-brayton compressor is 100 W for the yardstick. The pulse tube cooler requires slightly more than half of this amount; an important difference should spacecraft electrical power be limited.

6.4 Comparison with Yardstick

We summarize a comparison of the major features of the yardstick and our baseline ISIM in Table 6.4-1. Primary discriminators among the two systems are 1) extended wavelength range of our ISIM to include the detection of the astrophysically important H_2 line; 2) better sampling of the telescope focal plane via various plate scales, which provides diffraction limited imaging at a broader range of wavelengths; 3) much larger fields in the visible and NIR for efficient mapping; and 4) more spectroscopic capability, particularly at MIR wavelengths.

ISIM Configuration	Our ISIM[a]	Yardstick
Spectral Range	0.4–34 µm	0.6–28 µm
Number of modules	6 plus 2 wide field	6
Number of mechanisms	13[b]	28
Minimum λ for Nyquist sampled image	0.5 µm	2 µm
Total imaging FOV	129 arcmin^2	20 arcmin^2
Imaging 0.4-5µm	106 arcmin^2	16 arcmin^2
Imaging 5-34µm	23 arcmin^2	4 arcmin^2
Total spectroscopy FOV	32 arcmin^2	9 arcmin^2
Spectroscopy 0.4-5µm	9 arcmin^2	9 arcmin^2
Spectroscopy 5-34µm	23 arcmin^2	Single object
NIR spectroscopy pixel size	λ/2D, λ/D or 3λ/D	λ/2D
Resolutions and λ coverage		
0.4-5µm	R=100 (1.25-2.5µm, 2.5-5µm) R=1000 (1.25-2.5µm, 2.5-µm) R=3000 (1.25-2.25, 2.25-3.25µm, 3.25-4.25µm)	R=100, 1000 (0.6-1.2µm, 1-2µm, 1.8-3.6µm, 3.4-5.4µm)
5-34µm	R=100 (5-10µm, 10-20µm) R=500 (5-10µm, 10-20µm, 20-34µm) R=1000 (5-10µm, 10-20µm, 20-27µm, 27-34µm) R=3000 (5.20-10.4µm, 10.19-18.75µm, 16.8-34.1µm)	R=1000 (5.8-13.1µm,12.75-21.2µm)
Mass	450 kg[c]	460 kg
Cryo-coolers	Pulse Tube	Turbo-Brayton
Cooler Electrical Power(W)	50	100

Table 6.4.1 Comparison of our ISIM concept and the GSFC yardstick.

[a]Our ISIM complement is composed of two wide field imagers, each of 80 detectors (f/24 design), one visible imager, three NIR modules, and two MIR/LIR modules, unless otherwise mentioned. [b]One filter wheel per modules, with exception of two for the modules supporting cross-dispersed echelle spectra, plus one generic multi-mirror/slit mechanism for modules supporting spectroscopy. The wide field imagers are assumed to have a fixed filter pattern. [c] A contingency of 35% is recommended on top of this estimate.

6.5 Operations

Although we have not shown it explicitly in our ISIM concept, we believe that NGST would benefit from having a dedicated camera for guiding and image analysis. In the yardstick, it is assumed that these functions can be performed by a subarray within one of the arrays in a science camera. There are a number of arguments against this approach:

- Failure of the guide camera would probably be classified as a mission critical single point failure. However, failure of a science instrument module would not have this level of impact. Designing a science instrument to avoid single point failures could significantly impact its intended function (e.g., by making use of filter wheels questionable).

- Other problems would occur when trying to take imaging data and guide on the same array. It is conventional to operate guide cameras with out of focus images to reduce the effects of "bias" errors (meaning pointing bias due to effects such as small layout errors in the array geometry); such an approach would not be acceptable if science data were being obtained in parallel.

- A simple and optimized guide camera (single filter, optimized for fast readout) would be relatively inexpensive compared with a science instrument and would have superior performance.

- Interfacing a science instrument to the spacecraft pointing system and/or to the telescope control would be complex, crossing lines of responsibility within the NGST project, and would also diffuse responsibility for the guiding performance

For these reasons, we recommend that NGST include an optimized guide camera that was the responsibility of the supplier with which it has the most demanding interface (telescope, if fine guiding uses moving optics).

Our ISIM provides the capability for wavefront sensing for primary or deformable mirror control. One method would be the phase diversity technique where a prism is inserted into the beam to produce two images corresponding to differing optical path lengths. Such a capability could be achieved by placing a prism in one of the filter wheel positions of a visible or near infrared imager. To provide redundant capability for this essential function, we recommend that more than one module be so equipped. The other details of implementing this plan would need to be coordinated with the prime contractor to insure mirror actuator control requirements are met. However, the detector sensitivities and read rates provided by our imaging systems appear to be more than adequate to meet the wavefront control requirements.

7.0 Development

7.1 Modular design

Selection of an optimum overall architecture for the instruments can make a substantial difference in the ability to control cost and schedule. We recommend that the instrument complement be put together in a number of specialized modules, each to be as simple as possible and focused on a limited set of capabilities. Responsibility for these modules should be divided among a smaller number of investigator teams, and along lines that ally similar technical problems under a single team. These general statements will be illustrated by discussing the possible organizations in more detail.

7.1.1 Large consolidated instruments

A traditional approach to major astronomical missions like NGST is to build a small number large and complex instruments; in the case of NGST, three instruments would be roughly in line with this tradition, perhaps two of which would be built in the United States. Some advantages are:

- The efforts would be naturally concentrated into only two teams, with the overheads associated with many teams eliminated. Consolidating further to a single team might not be a good idea, since it might tend to reduce the breadth of experience brought to general instrument issues for NGST.
- A high degree of commonality naturally results in areas like warm electronics design, use of common mechanisms, and so forth, minimizing money wasted through parallel development of items to perform the same functions.
- Savings can be achieved in instrument mass, through more integrated mechanical design, in electronics packaging; in power consumption, and so forth, all through sharing resources among a large portion of the total instrument complement.

Some disadvantages include:

- Large instruments usually have many operating modes, complicating test and operations after launch.
- Test of a large instrument must generally be delayed until it has all been integrated.
- A failure in a large instrument during test and integration usually requires halting the entire integration process until a repair has been made and certified for continuation.
- Large instruments generally demand larger and more complex test facilities.
- A small number of large instruments may result in more single point failures for the net instrument complement of NGST.

7.1.2 Small modular instruments

Rather than two large instruments, the U.S. instrument complement could be organized into about half a dozen small modular instruments. The advantages and disadvantages roughly reverse compared with the situation with two large instruments. Advantages include:

- Each instrument module will have only a small number of operating modes, simplifying test and post launch operations.

- Test can begin with the first completed module, accelerating the integration schedule.
- A variety of strategies may allow continuation of integration and test while a failed instrument is taken off line for repair.
- Test facilities can be tailored to the smaller size of smaller instruments.
- Generally, a larger number of instruments will have fewer possible single point failures.

Disadvantages are:

- The efforts would be spread over many (six?) teams, increasing overheads associated with instrument construction.
- There will be a tendency to develop individual solutions to technical problems, thus resulting in expensive parallel developments in many areas.
- Total instrument mass, power, packaging, and so forth will tend to be less optimized because of the inability to develop highly integrated designs.

7.1.3 Large instruments organized into modules

A better solution than either extreme would be to organize two instrument teams, but with each instrument to be built in a modular fashion with about three major components, each focused on a specific function. In this approach, all the advantages of the extreme modular approach can be realized. In addition, the first two listed advantages of the large instrument approach will also result. The only disadvantage will be that the modular design of each instrument will have some penalties in use of resources. For example, additional structure will be required for separate modules compared with a consolidated instrument, so the result will have larger weight. However, the gains in flexibility in integration and test, and in simplicity in operations and in reduction of single point failures, in particular, will easily outweigh this single disadvantage.

Figure 7.1.3-1 shows our design for the ISIM from a perspective that highlights the accessibility of its modules. Although we have not laid out the details of the mounting arrangement, it appears that it would be feasible to remove and install most of the modules independently and with minimal disturbance of the interfaces established for the other modules. Such an arrangement would increase the potential gains from modularity with regard to flexibility in the integration and test sequence.

These gains could be critical in holding to the envisioned instrument development schedule, as discussed below.

7.1.4 Collaborations

A further advantage of a modular ISIM is that it could allow increased flexibility in laying out collaborations. For example, the visible and near infrared wide field modules make a logical pair for collaboration, or they could be given to two separate collaborators. The three near infrared imagers/spectrometers are another logical piece; the mid and long wave modules are a sensible common development; and high resolution visible imager is yet another. Alternately, the NIR modules and NIR wide field imager could be developed together and the high resolution and wide

field visible imagers would make another logical package. Thus, our proposed ISIM can be broken into a number of separate development projects.

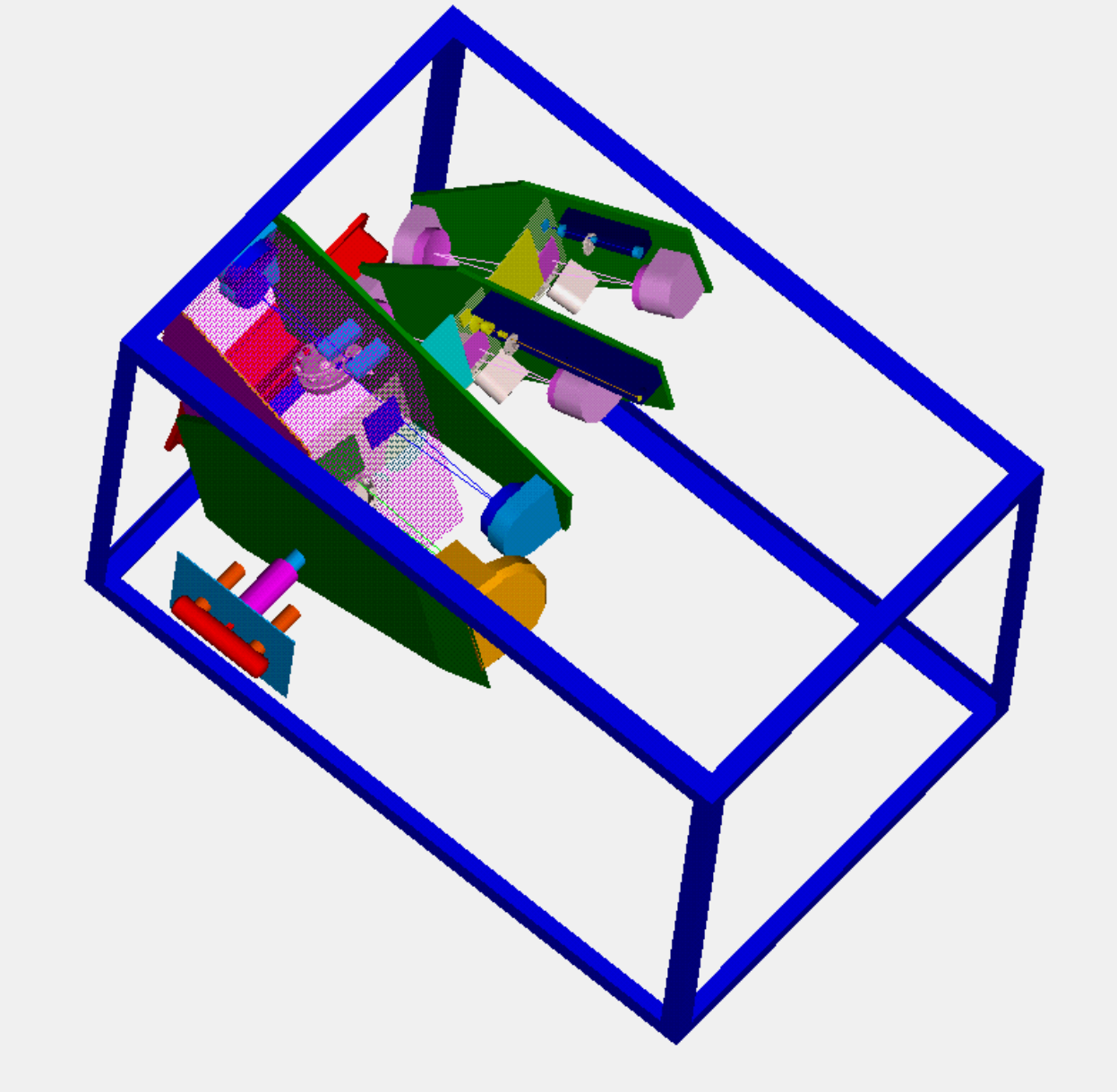


Figure 7.1.3-1 Overall layout of the ISIM, showing the inherent accessibility of the instrument modules.

8.0 Technology readiness

We favor the use of established technologies wherever possible to keep cost and risk low. However, we have identified several emerging and evolving technologies that are critical to the proper functioning of our ISIM baseline (or its options). We believe that the expected benefit of scientific return is worth the additional development effort needed and its associated risk. While not ready for immediate deployment on NGST, these technologies are expected to become mature within a few years, if actively pursued. These technologies include:

- Micromirrors/shutters/slits
- IR Detectors (including Si:P)
- Turbo-Brayton Coolers
- Pulse Tube Coolers
- Grisms

We have chosen to address technology readiness assessment for detectors, pulse tube coolers, and grisms only. Micromirrors are being investigated by another study effort and are also part of the yardstick ISIM. Turbo-Brayton coolers have also been assessed for the yardstick ISIM.

8.1 Detectors

We have indicated at appropriate points in this report where detector technology advances would reduce risk and/or improve the performance of NGST. A partial list is:

- Development of PIN diode hybrid arrays will allow optimized optical detection with readouts identical to those used in the infrared (and hence little cost impact).
- Science grade CMOS imagers, although farther from fruition, could have similar benefits as PIN hybrids, but with larger formats and lower cost.
- Near infrared detectors with very small pixels will benefit the spectroscopic applications since they allow long integrations with acceptable levels of hits by cosmic rays.
- Improvements in dark current as well as read noise will benefit near infrared spectroscopy
- HgCdTe arrays are competitive with InSb for the near infrared: smaller pixels, lower demonstrated read noise, larger demonstrated format. Detector development should consider significant effort for these devices
- Si:As IBC arrays are needed in a 1024x1024 format
- Si:P IBC arrays would provide wavelength extension to 35μm without significantly increasing cooling requirements. In addition, these detectors should be readily manufacturable and should be radiation damage resistant.

8.2 Pulse tube coolers

Lockheed Martin has conducted extensive research and development of pulse tube coolers, and has demonstrated that pulse tube cooling to temperatures of 8K is feasible for space applications. NGST imposes some unique requirements on pulse tube coolers that must be addressed. These special requirements include extending the tube length to reach into the 8K areas of the ISIM, managing the vibration levels, and adapting the compressor to work at reduced temperatures. In addition, overall pulse tube reliability must be consistent with a mission lifetime of 10 years.

The component limiting the reliability of current pulse tube cryo-coolers is the controller drive electronics. LM's standard flight controller package used successfully on a number of flight missions has a large part count that limits the single string reliability figure of the system to 0.85 for ten years. Currently, LM is redesigning the controller, incorporating state-of-the-art technology that greatly reduces the envelope and parts count, increasing the single string box reliability to over 0.95 for ten years operational lifetime. Implementing wholly redundant boxes further increases the controller reliability to over 0.999. A pulse tube cryo-cooler system is wholly compatible with high reliable, long life operation for NGST.

Currently, pulse tubes are somewhat shorter than the estimated 1–2 meter lengths we envision for our ISIM. However, these short pulse tubes also have measured cooling powers of up to 10 W; this is substantially more than the 3.9 W that needed by our ISIM. LM has accurate models that predict pulse tube performance as a function of design parameters. By seeking an optimum pulse tube design for the overall length, an effective system tailored for the ISIM can be implemented.

Present pulse tube compressors function at temperatures of about 200 K or greater. For NGST, we would probably want to operate the compressor at a reduced temperature between the ISIM cavity temperature of 40 K and 200 K. A low temperature compressor may require changes in piston diameters and its linear flexures, as well as different construction materials to control the effects of thermal expansion in its mechanical clearances. There appear to be no fundamental reasons precluding reliable compressor operation at temperatures down to 40 K, but this needs to be demonstrated and a reliability analysis performed.

Implementing a soft mounted pulse tube compressor will still require adequate thermal coupling for dissipation of the compressor waste heat. This can be achieved by attaching flexible copper ropes to the compressor mounts that provide a soft mount approach (<10-Hz possible), yet a high thermal resistance coupling for waste heat power dissipation. LMMS has extensive experience in the development of flexible thermal couplings with low thermal resistance. For launch, a hard connection will be required, but this can be mechanically decoupled (for example, with nut releases) once in orbit. Final pulse tube force levels ranging from 0.01 to 0.001 N appear achievable.

In summary, we have identified these key areas that must be addressed before pulse tubes may be used in our ISIM. While substantial work needs to be done, they may be investigated and addressed before a final cryo-cooler selection for NGST is scheduled. A decision based upon the performance of pulse tubes, Turbo-Brayton or other technology ready cryo-coolers should then be made.

8.3 Grisms

The use of IR grisms is crucial to the designs of all our imager/spectrometer modules. We have selected KRS-5 as the baseline material of choice for these grisms since it has good transmission (greater than 70%) from 1 – 35 μm (Figure 12.5.2-1). Its moderate index (n=2.4) is a good compromise between lower index materials which require steeper prism angles and higher index materials which require higher density rulings (see § 12.5 for more grism details). KRS-5 grisms with moderate ruling densities (~ 100 lines mm^{-1}) and prism angles (~ 25°) have been manufactured and have demonstrated good performance with low scattering and good efficiency (~ 55% in second order) in near-IR astronomical spectrographs (Rayner 1998).

In order to minimize overall instrument size, our instrument designs use the smallest reasonable pupil sizes. These small pupils (6 – 40 mm diameters) require that our high resolution (R ~ 3000) KRS-5 grisms have high ruling densities (up to 400 lines mm^{-1}) in the near-IR and large prism angles (α ~ 55°) in the mid-IR. KRS-5 grisms with these high ruling densities or wedge angles have not yet been manufactured or performance tested. The higher groove densities may cause high scattering / low efficiencies, and the large prism angles may produce dispersions which are different from those predicted by the simple geometric optics formulae. We are optimistic that KRS-5 will still yield low scattering grisms at high ruling densities due to its low hardness, and other high index grisms of similar angles have been measured to have negligible dispersion deviations from the geometric formulae (Kaufl 1998). However, we need to manufacture and test KRS-5 grisms with ruling densities and prism angles similar to those required in our designs in order to verify this. The scattering performance of low ruling density, high prism angle grisms must also be measured in moderately high orders (m ~ 10) to ensure that these optics are suitable for use in the echellette modules. We do not anticipate problems with obtaining KRS-5 grisms in the required diameters.

This manufacture and test of grisms can be done relatively rapidly for reasonable cost (approximately $20 K manufacturing cost per grism) before the formal NGST instrument call. If the KRS-5 grisms show scattering problems, there would be time to consider and test alternatives before submitting an instrument proposal. For example, Si is a promising material which has already been manufactured into immersion gratings which show low scatter in high orders at high groove densities (Keller 1998). We consider Si a secondary material choice because its ruling techniques are not as well developed and it does not transmit light at all needed wavelengths (see section 12.5).

9.0 Cost

Our recommended configuration for the ISIM – two wide field imagers and six additional modules – could be built for virtually the same cost as the yardstick ISIM. To establish this comparison, we begin with cost modeling of all the modules we have studied. This first run takes no advantage of savings through common designs or an optimized telescope configuration. In § 9.5, we discuss our recommendations to contain cost while providing a scientifically powerful ISIM. We show that the cost of our optimized ISIM should be $186M (no fee, FY 96 dollars), within the errors the same cost as the yardstick ISIM. This version of the ISIM is the one which we compared in speed with the yardstick in §6.1.10, finding it could complete the DRM in only 2/3 the time required by the yardstick

Cost estimates for the proposed ISIM were estimated in FY96 dollars. This estimates were arrived through a combination of grass roots cost estimates, engineering estimates and parametric cost models. The estimate has the following specific assumptions:

- The ISIM is comprised of three science instruments, a Wide Field Visible / Near Infrared Instrument, a Near Infrared / Visible Imager (NIR / VIS) and Spectrometer Instrument and a Mid Infrared/ Long Wave (MIR/LIR) Instrument. The design and development of each of these instruments are managed by separate SI program managers.
- Development and risk mitigation costs of the focal planes and special devices such as microshutter arrays and the large format focal plane arrays are not included. It is assumed that such efforts are successfully completed by NASA funded development activities in these areas. Development engineering to design the individual instrument modules incorporating these devices is included and estimates for learning curve efficiencies are included in the estimates of modules with similar designs such as the NIR modules.
- The cost of a fine guidance sensor is not included in this estimate. It is assumed the NGST Prime Contractor will provided a separate instrument module for this function. We estimate that the cost for such a module is approximately $10M in FY96 dollars.

9.1 Costing methodology

The majority of the costs of this estimate were computed using the Automated Cost Estimating and Integrating Tool (ACEIT). ACEIT has been commercially developed and is available to the government with no licensing fee. It is an automated architecture and framework for Life Cycle Cost (LCC) estimating that is used extensively in the government and contractor community for top level, multiple case trade studies. The software package employs parametric Cost Estimation Relationships (CERs) for the LCC of hardware elements. Among the CERs used is the Strategic Experimental IR Sensor Cost Model II. Software costs were arrived at from engineering estimates from similar instrument architectures. The cryocooler costs were also derived from engineering estimates from similar architectures.

The results of the ACEIT cost estimates compare favorably with costs for similar instrument architectures such as SIRTF, where an average of $10M per instrument module was found to hold. The exceptions to this for our ISIM are those modules with large focal plane arrays.

9.2 Work Breakdown Structure

Table 9.2-1 gives the WBS which was used to arrive at our cost estimate. The WBS is first organized ISIM activities and Science Instruments which are then further broken down into individual instrument modules. The WBS includes the cost of a common processor shared by the NIR / VIS and MIR/LIR SIs under the NIR/VIS cost element. Software is also budgeted under this element. The Wide Field Instrument is assumed to have its own processor and software.

Table 9.2-1 Work Breakdown Structure

ISIM Work Breakdown Structure and Dictionary	
WBS Element	Definition
1.0 ISIM Program Management	Program management, cost / schedule tracking, contract control
2.0 ISIM Systems Engineering	Resource allocations, interface definitions,
3.0 ISIM Structure and Thermal	ISIM Common structures, mechanisms, radiators, blankets
4.0 ISIM Power Distribution and Harness	ISIM power harnesses and power distribution devices
5.0 Wide Field Science Instrument	NIR and VIS Wide Field Imagers
5.1 Program Management	Program management for Wide Field SI
5.2 Wide Field NIR Module	
5.2.1 Development Engineering	Development engineering, design and analysis
5.2.2 I&T	Integration and test (calibration) of NIR Wide Field Module
5.2.3 Detectors	Detector hardware costs
5.2.4 Optics	Not Applicable
5.2.5 Mechanisms and Structure	Structure, mechanism, and thermal interface hardware costs
5.3 Visible Wide Field Imager	
5.3.1 Development Engineering	Development engineering, design and analysis
5.3.2 I&T	Integration and test (calibration) of VIS Wide Field Module
5.3.3 Detectors	Detector hardware costs
5.3.4 Optics	Not Applicable
5.3.5 Mechanisms and Structure	Structure, mechanism, and thermal interface hardware costs
5.4 Wide Field SI Processor	Hardware costs for a Wide Field processor
5.5 Wide Field SI Software	Software development costs for Wide Field software
6.0 NIR / VIS Science Instrument	NIR/ VIS Imagers and Spectrometers
6.1 Program Management	Program management for NIR/VIS SI
6.2 Visible Imager Module	
6.2.1 Development Engineering	Development engineering, design and analysis
6.2.2 I&T	Integration and test (calibration) of VIS Module
6.2.3 Detectors	Detector hardware costs
6.2.4 Optics	Optics hardware costs
6.2.5 Mechanisms and Structure	Structure, mechanism, and thermal interface hardware costs
6.3 NIR Module 1	
6.3.1 Development Engineering	Development engineering, design and analysis
6.3.2 I&T	Integration and test (calibration) of NIR Module 1
6.3.3 Detectors	Detector hardware costs
6.3.4 Optics	Optics hardware costs
6.3.5 Mechanisms and Structure	Structure, mechanism, and thermal interface hardware costs

6.4 NIR Module 2	
6.4.1 Development Engineering	Development engineering, design and analysis
6.4.2 I&T	Integration and test (calibration) of NIR Module 2
6.4.3 Detectors	Detector hardware costs
6.4.4 Optics	Optics hardware costs
6.4.5 Mechanisms and Structure	Structure, mechanism, and thermal interface hardware costs
6.5 NIR Module 3	
6.5.1 Development Engineering	Development engineering, design and analysis
6.5.2 I&T	Integration and test (calibration) of NIR Module 2
6.5.3 Detectors	Detector hardware costs
6.5.4 Optics	Optics hardware costs
6.5.5 Mechanisms and Structure	Structure, mechanism, and thermal interface hardware costs
6.6 NIR/VIS/MIR/LIR Processor	Hardware costs for a processor shared between NIR/VIS and MIR/LIR SIs
6.7 NIR/VIS/MIR/LIR Software	Software development costs for NIR/VIS/MIR/LIR software
7.0 MIR / LIR Science Instrument	MIR/LIR Imagers and Spectrometers
7.1 Program Management	Program management for presumed U.S. delivered NIR SI
7.2 MIR Imager/Spectrometer	
7.2.1 Development Engineering	Development engineering, design and analysis
7.2.2 I&T	Integration and test (calibration) of MIR Imager/Spectrometer
7.2.3 Detectors	Detector hardware costs
7.2.4 Optics	Optics hardware costs
7.2.5 Mechanisms and Structure	Structure, mechanism, and thermal interface hardware costs
7.3 MIR Spectrometer	
7.3.1 Development Engineering	Development engineering, design and analysis
7.3.2 I&T	Integration and test (calibration) of MIR Spectrometer
7.3.3 Detectors	Detector hardware costs
7.3.4 Optics	Optics hardware costs
7.3.5 Mechanisms and Structure	Structure, mechanism, and thermal interface hardware costs
7.4 LIR Imager	
7.4.1 Development Engineering	Development engineering, design and analysis
7.4.2 I&T	Integration and test (calibration) of LIR Imager/Spectrometer
7.4.3 Detectors	Detector hardware costs
7.4.4 Optics	Optics hardware costs
7.4.5 Mechanisms and Structure	Structure, mechanism, and thermal interface hardware costs
7.5 Cryo-Cooler	Hardware costs for cryo-cooler system used for MIR and LIR modules
8.0 ISIM I&T	Integration and alignment, test and calibration of SIs into ISIM

9.3 Schedule

The cost estimate assumes ISIM design, development, integration and testing to roughly follow the time table shown below:

SI Team Selection	Jan-02
SI Preliminary Design Reviews	Jan-03
SI Critical Design Reviews	Jan-04
SI Delivery to ISIM Integrator	Apr-06
ISIM Delivery to NGST Prime Integrator	Apr-07

NGST Launch Date June-08

A top level program schedule broken out by WBS is shown in Figure 9.3-1.

9.4 Cost Breakdown

Table 9.4-1 gives the cost estimates to the module level for the ISIM. Table 9.4-2 sorts these costs by functional discipline to enable a comparison with the WBS used by GSFC in its Yardstick ISIM study as presented at the March 99 Quarterly.

Table 9.4-1 Cost estimate to the module level.

WBS Element	Cost ($M FY96)
1.0 ISIM Program Management	2.622
2.0 ISIM Systems Engineering	11.739
3.0 ISIM Structure and Thermal	5.054
4.0 ISIM Power Distribution and Harness	0.756
5.0 Wide Field Science Instrument	125.392
5.1 Program Management	2.622
5.2 Wide Field NIR Module	59.435
5.3 Visible Wide Field Imager	59.435
5.4 Wide Field SI Processor	1.500
5.5 Wide Field SI Software	2.400
6.0 NIR / VIS Science Instrument	67.860
6.1 Program Management	2.622
6.2 Visible Imager Module	19.994
6.3 NIR Module 1	10.327
6.4 NIR Module 2	11.487
6.5 NIR Module 3	13.330
6.6 NIR/VIS/MIR/LIR Processor	1.500
6.7 NIR/VIS/MIR/LIR Software	8.600
7.0 MIR / LIR Science Instrument	44.574
7.1 Program Management	2.622
7.2 MIR Imager/Spectrometer	9.872
7.3 MIR Spectrometer	13.960
7.4 LIR Imager/Spectrometer	10.120
7.5 Cryo-Cooler	8.000
8.0 ISIM I&T	11.455

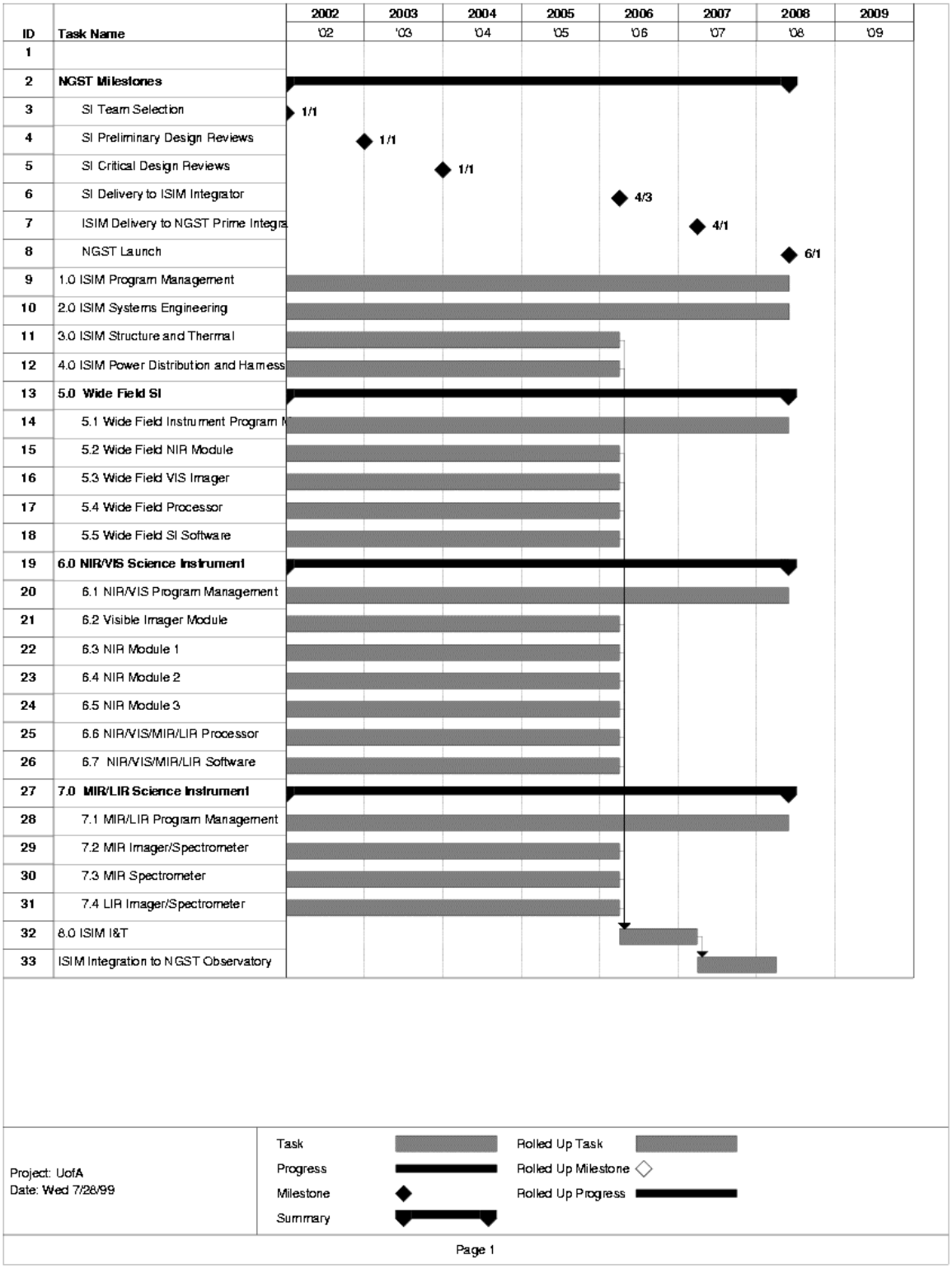


Figure 9.3-1 Top level schedule by WBS element.

Table 9.4-2 Cost Breakout by Functional Area

Functional Areas	Cost
Management	10.488
Science / Development Engineering	51.121
Systems	11.739
SR&QA / I&T	34.365
Structure	10.531
Optics	20.183
Electronics	3.756
Operations	
Thermal	4.513
Software	11.000
Detectors	103.756
Cryocooler	8.000

9.5 Cost of the optimized ISIM

We have used the detailed costing in the preceding section to derive a cost for the specific combination of modules we recommend for a complete ISIM: 1.) high resolution visible imager; 2.) wide field visible imager; 3.) three NIR imager/spectrometers; 4.) wide field NIR imager; and 5.) two echellette spectrometers/imagers for the 5 to 35μm region.

9.5.1 Wide field imager costs

The cost model shows $118.9M for the visible and near infrared wide field imagers. The great majority of this price is for the large number of focal plane arrays required. As shown in §5, if the telescope had an f/number of 12 to 16, equivalent surveying capability would still be possible (and possibly the instrument packaging would benefit overall). However, equivalent fields of view could be provided with the faster telescope using one half to one quarter the number of arrays we have had to cost for the baseline telescope.

We assume a more optimum telescope would reduce the cost of the wide field imagers by a factor of two, saving $60M directly. We estimate that another $8.6M could be saved through proportionate adjustments in management, systems engineering, structure and thermal, and I&T, where we have assumed the savings are proportional to the overall cost reduction.

9.5.2 MIR and LIR echellette spectrometers/imagers

From our study, it appears that a single module could be built that would carry out imaging and spectroscopy between 6 and 20μm, based on the echelle spectrometer design. Modest compromises might be required compared with use of two modules for these goals, for example in high resolution spectroscopic throughput (use of grisms rather than reflection gratings) and in imaging throughput (more optical elements than an imager). However, it appears that these compromises are only in the 10-20% range of sensitivity. A further economy could be made by making the 20 – 35μm virtually identical in design to the 6 – 20μm one. A further benefit is that moderately high spectral resolution could be provided at the long waves (important for emission line observations that are featured in parts of the DRM). We would consider carrying this

similarity even to detector array format, since it is probably the case that SiP BIB arrays can probably be made with high enough performance that only a minor performance penalty would result from vastly oversampling the PSF at the long wavelengths. Since the processing and testing of the Si:As and Si:P arrays could then be identical other than doping the material, the cost of the flight arrays would be reduced.

In estimating the cost savings, we have eliminated the MIR imager and have assumed that the LIR echellette spectrometer/imager would cost half of the price of the MIR echellette module, since it would make extensive use of the designs, drawings, and test programs for the MIR echellette. This estimate is supported by experience reported to us for the Infrared Spectrograph for SIRTF, which is built of modules with strong design similarity. The direct savings are then $13M. An additional $1.9M would be saved through reductions in areas like management, systems engineering, and integration and test, where we have assumed the savings are proportional to the overall cost reduction.

9.5.3 NIR Modules 2 & 3

NIR Modules 2 & 3 have been costed individually. In fact, they can have virtually identical designs, except for a small change in the final optical elements to change the scale on the focal plane. We assume that this commonality would save ~30% on the lower cost unit, or with proportionate reductions in management and systems engineering too, about $3.5M.

9.5.4 Electronics

The estimated cost of electronics seems low. A comparison has been made with the IRS/MIPS electronics for SIRTF. The SIRTF unit controls two different array multiplexer types and a mechanism. It is block redundant, i.e., two identical units with suitable cross strapping. The total for these electronics in FY96 dollars is $6M. The ISIM electronics are of somewhat greater complexity so we have added $3.5M to the estimate for a total of $7.2M. This adjustment offsets the potential reduction in §9.5.3.

9.5.5 Cost summary

Table 9.5.5-1 contains the information from Table 9.4-1, adjusted for the optimized ISIM.

Table 9.5.5-1. Cost of the optimized ISIM by WBS element

WBS Element	Cost ($M FY96)
1.0 ISIM Program Management	1.912
2.0 ISIM Systems Engineering	8.559
3.0 ISIM Structure and Thermal	3.685
4.0 ISIM Power Distribution and Harness	0.756
5.0 Wide Field Science Instrument	66.182
5.1 Program Management	1.912
5.2 Wide Field NIR Module	29.435
5.3 Visible Wide Field Imager	29.435
5.4 Wide Field SI Processor	3.000
5.5 Wide Field SI Software	2.400
6.0 NIR / VIS Science Instrument	65.704
6.1 Program Management	1.912
6.2 Visible Imager Module	19.994
6.3 NIR Module 1	10.327
6.4 NIR Module 2	8.041
6.5 NIR Module 3	13.330
6.6 NIR/VIS/MIR/LIR Processor	3.500
6.7 NIR/VIS/MIR/LIR Software	8.600
7.0 MIR / LIR Science Instrument	30.852
7.1 Program Management	1.912
7.2 MIR Imager/Spectrometer	---
7.3 MIR Spectrometer	13.960
7.4 LIR Spectrometer	6.980
7.5 Cryo-Cooler	8.000
8.0 ISIM I&T	8.352
Total	186.002

This form of ISIM could be provided for $186M (no fee), or $213.9M (15% fixed fee). The costs of instrument science team activities and of NASA oversight of the ISIM development need to be added to this estimate to obtain a total cost of the ISIM.

We have run the same cost model on the Goddard yardstick ISIM, which it estimates at $167M (no fee). Thus, our proposed ISIM cost is within 10% of the comparable cost of the yardstick, a difference which is within the estimation errors. We believe that either form of ISIM could be procured for similar cost and that the choice between them can be made on the basis of performance.

10.0 Conclusion

In this study, we have found that a suite of instruments can be supplied to NGST that is significantly more powerful than the yardstick proposal, but which does not cost significantly more. Our instrumentation includes a much larger field of view in the visible and near infrared for quicker deep surveying, as well as more versatile spectrometers and an extension of the wavelength coverage down to 0.4μm and out to 35μm. In addition to the greater scientific capabilities provided by greater wavelength coverage, the instrument could complete the DRM in only 2/3 the time required by the yardstick ISIM. Costs are controlled by:

- Optimizing the telescope as well as the instruments in a systems approach. In particular, we recommend that a somewhat faster Cassegrain f/ratio than f/24 be considered.
- Designing an instrument complement around a number of simple, optimized modules that closely resemble each other so that design, fabrication, test, and qualification costs can be shared.
- Selecting very large format near infrared arrays to minimize device count
- By maintaining simple modules, also reducing the number of potential operating modes and hence operational costs for the mission.

In arriving at the proposed suite of instruments, we have considered a number of important trades:

- We recommend that optimized detectors such as Si PIN hybrid detector arrays be developed for the visible and Si:P IBC arrays for the 28 to 35μm region.
- We find that either pulse tube coolers with soft mounted compressors, or passive solid hydrogen coolers can be used without the expense of development and qualification required by more developmental approaches.
- We show that imaging Fourier Transform Spectrometers are not competitive with optimized dispersive spectrometers. This comparison is particularly favorable for dispersive instruments if detector arrays with small physical pixels are considered (to minimize cosmic ray hit rates) and if future likely improvements in dark current and read noise materialize.
- We derive the degree of background rejection required for multi-object, multi-slit dispersive spectrometers, finding that rejection at the level of 1000:1 is needed in the near infrared for spectral purity, that is, to avoid interference by spectra of neighboring sources. In the mid infrared, the rejection of the telescope background is so demanding that a reflective type of multi-slit is probably not feasible.

Although our suite of instruments differs in important ways from the yardstick, it also has an important resemblance: we find that very powerful and versatile scientific capabilties can be supplied to NGST for a total instrumentation cost of < $200 million in FY96 dollars (and not including fee, science team support, etc.). Thus, our study tends to confirm the underlying assumption of the project, but at the same time it suggests directions for further optimization of the facility.

11.0 References

Bailey, R. et al., "Prospects for Large Format IR Astronomy FPAs Using MBE-Grown HgCdTe Detectors with Cutoff Wavelength >4µm," SPIE 3354:77-86 (1998)

Battelle Institut, 1988, Final Report, ESTEC Contract No 6939/86/NL/PB(SC)

Beckers, J.M., Gatley, I. "Grisms for Infrared Observations," ESO Conference VLT Telescopes and their Instrumentation, p. 1093-1101 (1988).

Bely, R., et al. "Implications of the Mid-Infrared capability for NGST," NGST Project Office Memorandum, 9 November 1998.

Bennett, C. L. 1999, preprint, submitted to "Imaging the Universe in Three Dimensions: Astrophysics with Advanced Multi-Wavelength Imaging Devices," to appear in ASP Conference Series, ed. W. van Breugel & J. Bland-Hawthorn

Bippert-Plymate, T., Paul, A. E., and Rieke, G. H 1992, in "Astronomical Data Analysis Software and Systems I", ASP Conf. Series, 25, 205.

Bohnhardt, H. et al., "Design benchmarks for the FORS instrument for the EDO VLT," in Scientific and engineering frontiers for 8-10m telescopes. Instrumentation for large telescopes in the 21st century, eds. M. Iye & T. Nishimura, 1995.

Bruzual, A., Sharlot, S. 1993, ApJ, 405, 438.

Calzetti, D., Kinney, A., Storchi-Bergmann, T. 1994, ApJ, 429, 582.

Castles, S. 1998, "NGST Cooler Trades," NRA PI Meeting, October 14, 1998

Dekker, H., D'Odorico, S., Arsenault, R. "First results with a transmission echelle grating on the ESO Faint Object Spectrograph: observations of the SN 1986a in NGC 3367 and the nucleus of the galaxy," A&A 189:353-360 (1988).

DiPirro, M. 1999, "Mid IR Cooler Options," NGST Quarterly, March 8-9, 1999

Dwek, E. & Arendt, R. G. in "After the Dark Ages: When Galaxies were Young", ed. S. Holt & E. Wmith, AIP, p354 (1999)

Estrada, A.D., et al. "Si:As IBC IR Focal Plane Arrays for Ground-Based And Space-Based Astronomy," SPIE 3354:99-108 (1998)

Genzel, R., et al. 1998, ApJ, 498, 579

Gillett, F. C., & Mountain, M. in "Science with the NGST", ed. E. Smith & A. Koratkar, ASP Conf. Ser., 133, 42 (1998)

Herter, T. et al., "Mid- and Far-Infrared Hybrid Focal Plane Arrays for Astronomy," SPIE 3354:109-115 (1998)

Hill et al., "The Hobby-Eberly telescope low resolution spectrograph mechanical design," SPIE 1998 Kona Proceedings.

Käufl, H.U., Kühl, K., Vogel, S. "Grisms from germanium/silicon for astronomical instruments," SPIE 3354:151-158 (1998).

Keller, L. D., Jaffe, D. T., & Doppmann, G. W. "Design for a near-infrared immersion echelle spectrograph: breaking the R=100,000 barrier from 1.5 to 5 • m," Proc.
SPIE 3354: 295-304 (1998).

Klocek, P. *Handbook of IR optical materials.* Optical Engineering, Vol. 30, M. Dekker: New York, 1991.

Kozlowski, L.J. et al., "HgCdTe 2048^2 FPS for Infrared Astronomy: Development Status," SPIE 3354:66-76 (1998)

Malfait, K., Waelkens, C., Waters, L. B. F. M., Vandenbussche, B., Huygen, E., and De

Graauw, M. S. A&A, 332, L25 (1998)
McGregor, P., Hart, J., Downing, M., Hoadley, D., & Bloxham, G., Experimental Astronomy, 3, 139 (1994)
Megeath, S. T., Herter, T., Gull, G. E., and Houck, J. R. 1990, ApJ, 356, 534
Neviere, M. "Electromagnetic study of transmission gratings," Applied Optics 30, 31:4540-4547 (1991).
NGST ASWG DRM study, meeting minutes and discussions
Parrish et al. Thermal Design and Analysis of NASA's NGST Yardstick ISIM, April 1999 NGST Detector Workshop
Rayner, J.T. "Evaluation of a solid KRS-5 grism for infrared astronomy," SPIE 3354:289-294 (1998).
Rieke, G. H. "Infrared configurations for 8m telescopes with fast primaries," Report to the Magellan Project Science Advisory Committee, September 11, 1987.
Roche, P. F., Aitken, D. K., Smith, C. H., Ward, M. J. 1991, MNRAS, 248, 606
Swift, W. et al. 1999 "Developments in Turbo-Brayton Technology for Low Temperature Applications" presented at the 16th Space Cryogenics Workshop, July.
Thompson, R. I., Storrie-Lombardi, L. J., Weymann, R. J., Rieke, M. J., Schneider, G., Stobie, E., & Lytle, D. AJ, 117, 17 (1999)
Thi, W. F., van Dishoek, E. F., Black, J. H., Jansen, D. J., Evans, N. J., and Jaffe, D. T. 1998a, to appear in proceedings of "The Universe as Seen by ISO", eds. P. Cox et al. (ESTEC, ESA-SP)
Thi, W. F., et al. 1998b, ibid.
Van Cleve, J. E., Herter, T. L., Butturini, R., Gull, G. E., Houck, J. R., Pirger, B., and Schoenwald, J.1995, Proc. SPIE, 2553, 502
Waters, R. 1998, invited talk at "The Universe as Seen by ISO", eds. P. Cox et al. (ESTEC, ESA-SP)

12.0 Appendices

12.1 NGST Background Models & Sensitivities

To investigate various performance trade studies and to establish our proposed instrument criteria, we desired an acceptable NGST sensitivity model. This appendix deals with the details and assumptions of our adopted model.

12.1.1 Background Model

Figure 12.1.1-1 illustrates the several component which comprise the expected NGST background. They are given in units of photons per second per pixel for a R=100 sampling. Details of each of these components are outlined in the following paragraphs.

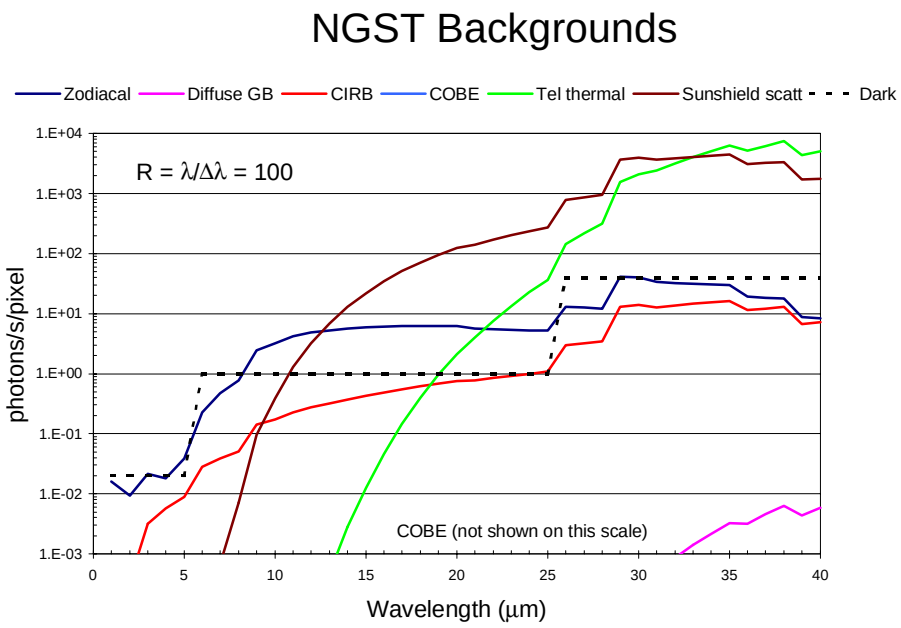


Figure 12.1.1-1 NGST background model components for R=100.

The NGST telescope temperature was modeled as a T=40K blackbody spectral curve with emissivity, ε=0.05. A scattered light component from the sunshield was modeled as a T=93K blackbody with an effective emissivity, ε=1e-4. These values represent the proposed MIR-compatible NGST architecture (Bely, et al. 1998). A breakdown of the diffuse sky background is shown with the aid of Table 12.1.1-1.

Component	Model	Reference/Comments
Zodiacal scattered	B(5500K), ε=1.9e-13	
Zodiacal thermal	B(278.5), ε=2.7e-8	
Diffuse galactic background	B(24.45), ε=1.78e-5	Significant for λ>20μm
Cosmic IR background	1e-9 ergs s^{-1} cm^{-2} sr^{-1} $Å^{-1}$	Thompson, *et al*. 1998; Dwek & Arendt, 1999
COBE T=3K background	B(3), ε=1	Significant for λ>40μm

Table 12.1.1-1 Breakdown of the constituents of the diffuse sky background for NGST.

The conversion from flux units (e.g., Jansky/sr) to photons/s/pixel assumed a D=8m diameter telescope and an overall telescope + instrument efficiency of 50%.

The model is based on breaking the ISIM wavelength coverage into separate instrument modules, each optimized for a particular sampling. Hence, the actual size of a pixel in arcseconds on the sky is a function of the designed sampling parameters for each module, show in Table 12.1.1-2. Additionally, the detector characteristics (e.g., dark current, read noise) for each module are summarized in Table 12.1.1-2. Adopted QE values for the selected detectors are illustrated by Figure 12.1.1-2.

As illustrated for R=100 (and imaging) in Figure 12.1.1-1, the telescope is zodiacal light limited for λ≤12μm. The 40K telescope primary will dominate the background when λ≥33μm. At higher resolutions (R=3000, not shown), the system is detector limited for λ≤15μm, with longer wavelength backgrounds dominated by the sunshield emission.

Detector	λ_{min} (μm)	λ_{max} (μm)	λ_{Ny} (μm)	Pixel size (arcsec)	Dark current (e/s)	Read Noise (e- RMS)	Full well (e-)	Reference
Visible	0.4	1	0.5	0.006	0.001	1	1e5	
HgCdTe	1	2.5	1.5	0.019	0.02	5	1e5	Kozlowski, *et al.* 1998
HgCdTe	2.5	5	3	0.039	0.02	5	1e5	Bailey *et al* 1998
Si:As	5	20	12	0.155	1	15	2.5e5	Estrada et al. 1998, Herter *et al.* 1998
Si:Sb/Si:P	20	34	27	0.348	40	40	2.5e5	Herter, *et al*. 1998

Table 12.1.1-2 Default instrument module parameters for sampling.

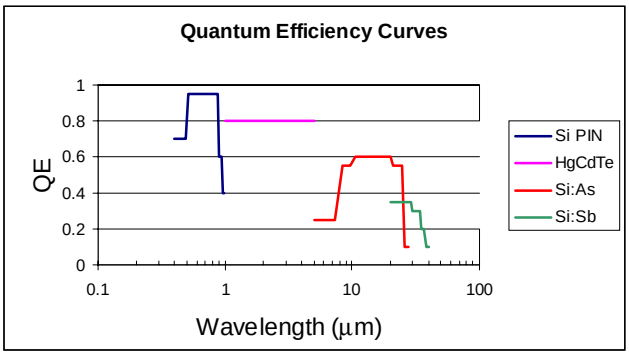


Figure 12.1.1-2 Adopted detector quantum efficiencies for visible detector (Si PIN diodes, see §2.1.3), HgCdTe, Si:As, and Si:Sb. These have taken from the reference articles mentioned in Table 12.1.1-2.

12.1.2 Sensitivity Calculations

The signal to noise calculations for a point source object were calculated along similar lines to the model used in Gillett & Mountain (1998).

Namely, we defined the SNR of an observation with exposure time, *t*, to be

$$\frac{S}{N} = \frac{I_{obj}(t) * t}{N(t)},$$

where

$$I_{obj}(t) = F * A_{tel} * \varepsilon * DQE * \frac{\lambda}{R} * \frac{\lambda}{(h * c)} \text{ electrons/s/SRE}$$

is the object's photocurrent per spectral resolution element (SRE). A SRE is defined as a 2×2 pixel region. F is the source strength in Janskys, A_{tel} is the telescope collecting area; ε is the total efficiency of the system including telescope & instrument efficiency and taking into account slit losses; DQE is the detective quantum efficiency of the detector; and λ/R is the size of a spectral resolution element in microns.

The measurement noise is given by

$$N(t) = \sqrt{(I_{obj}(t) + I_{sky(t)} + n * I_{dark}(t)) * t + n * N_r^2},$$

where

$$I_{sky}(t) = \beta * A_{tel} * \varepsilon * DQE * \Omega * \frac{\lambda}{R} * \frac{\lambda}{(h * c)} \text{ electrons/s/SRE}$$

is the detected sky background photocurrent per SRE. Here β is the sky background surface brightness in Janskys per square arcsec; Ω is the solid angle in square arcsec that makes up a SRE; n is the number of pixels composing a SRE; $I_{dark}(t)$ is the detector dark current; and N_r^2 is the detector read noise per pixel.

A total of *k* observations of time *t* were combined for a total exposure time of $T_{\text{int}} = k * t$.

The resultant SNR was calculated to be

$$\frac{S}{N}(T_{\text{int}}) = \frac{S}{N}(k * t) = \sqrt{k} * \frac{S}{N}(t).$$

For all modules, we assumed a total telescope + instrument efficiency to be 0.5 times the respective detector quantum efficiency (Figure 12.1.1-2). The blaze profiles of gratings were not included in these first order sensitivity calculations.

5-σ, point sources sensitivities were calculated for T_{int} = 10,000s, composed of several shorter individual integration times. Individual exposure times were set to 1000s over the 1-5μm to be sufficiently short enough to limit cosmic ray damage, and continually shorter towards longer wavelengths to avoid overfilling the detector full wells, down to 1s at 35μm (R=5).

The results are shown in Figures 12.1.2-1 & 2 for imaging (R=5) and R=3000. Also plotted on these graphs are the equivalent sensitivities for other ground and space observatories covering our proposed ISIM wavelength range, 0.4-34μm.

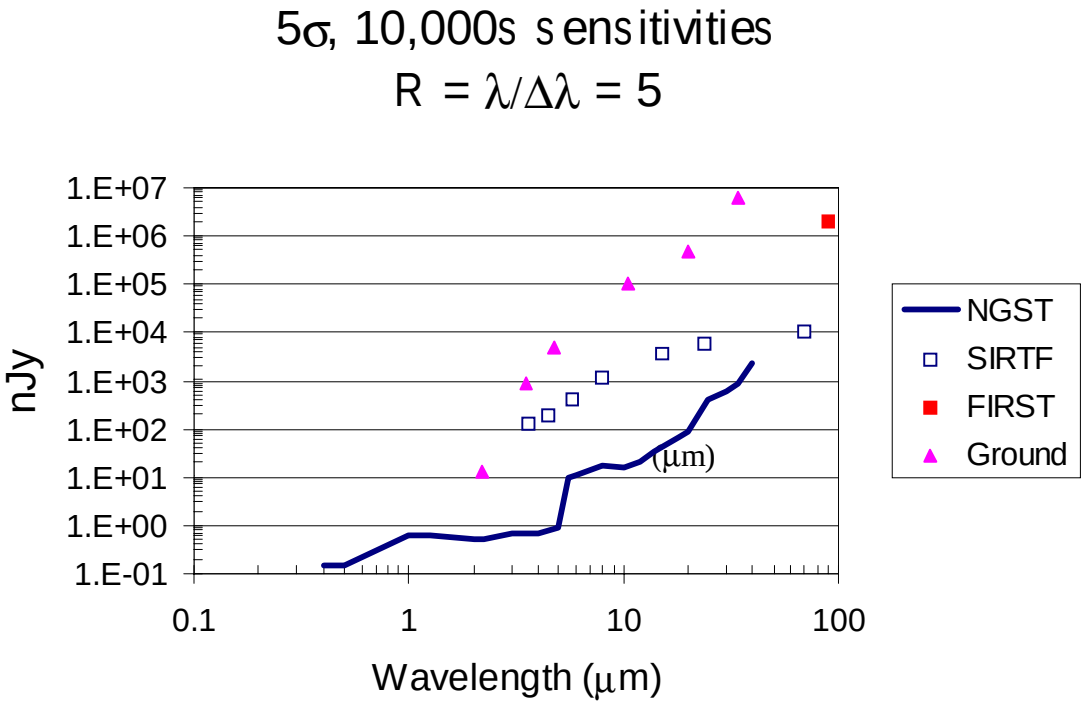


Figure 12.1.2-1 R=5 (imaging) 5-σ detection limits for point source objects in a 10,000s total integration. NGST will be the most sensitive observatory out to 40μm.

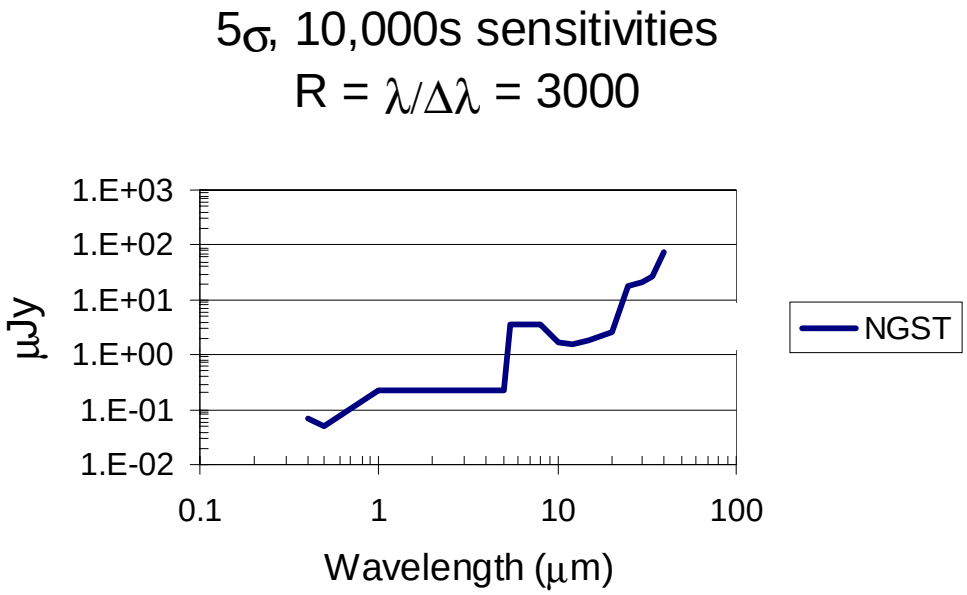


Figure 12.1.2-2 R=3000 5-σ detection limits for point source objects in a 10,000s total integration. Over the scientifically important 1-5μm region, the observatory is detector dark current limited for the values used in Table 12.1.1-2. Lowering the detector dark current will yield greater sensitivities than suggested here. Note change of scale to μJy.

12.2 IFTS calculations

All of the spectrometers chosen for our ISIM concept are dispersive rather than multiplex or Imaging Fourier Transform Spectrometers (IFTS). This choice ensures the highest possible signal-to-noise for NGST spectroscopic observations. Dispersive spectrographs (DS) are also capable of exploiting the future improvements in both detector read-noise and dark current.

This appendix summarizes the relationships and assumptions used in evaluating the performance of an IFTS in §4.1.1-2.

12.2.1 Basic concepts

The principal difference between the dispersive and FTS spectroscopic methods is the way in which wavelength information is coded. Dispersive spectroscopy codes wavelengths by position along an array of detectors. Each detector receives only the radiation from the spectral element that it is recording. FTS spectroscopy codes wavelengths by frequency, either in time or in spatial position of a moving mirror in an interferometer. Radiation from all of the spectral elements falls on the two output detectors. For a given wavelength the input power is sinusoidally switched between the two output detectors by the motion of the moving mirror. Wavelengths are decoded at a later time by a Fourier transform of the signal. In both instruments all of the signal is observed all of the time. The dispersive spectrometer does it with many pixels, the IFTS with only two for a point like object.

12.2.1.1 Dispersive spectroscopy

Signal-to-noise analysis for a dispersive spectrometer is similar to that for photometry. A single pixel integrates the electrons produced by the source and background photon flux in a single spectral element for a given period of time. The pixel is then read out with a given read-noise. The total noise is the square root of the sum of the electrons produced by the source, background, and the square of the read-noise. The equations used are the same as those summarized in §12.1.2.

12.2.1.2 FTS Spectroscopy

Signal-to-noise ratio analysis for a FTS spectrometer is more complicated but still straight forward. The spectrum is recovered from the output signal by performing a Fourier transform. For signal-to-noise purposes the Fourier transform is best thought of as a series of cross correlations of the output with sinusoids that correspond to the modulation frequency for each wavelength. Each spectral element is the result of the cross correlation with its specific modulation frequency. The output signal is produced by reading out the two pixels at equally spaced mirror positions. For S spectral elements the sampling theorem requires a minimum of 2S output points. This results in 4S total readouts for the two pixels.

In this case each product is between the appropriate sinusoid and the output sample which contains the detector read-noise. If the read-noise is N_R, the square of the noise for the cross correlation is the sum of the squares of the individual reads, $4SN_R^2$. The square root of this is the total read-noise for the cross correlation and applies to the single spectral element. The read-noise per spectral element for a multiplex spectrometer is therefore $(4S)^{1/2}$ higher than for a dispersive spectrometer, where S is the number of spectral elements. If we take into account the IFTS's multiplex advantage, this noise suffices for spectra of S points, and hence the penalty is $4^{1/2}$. For our comparisons shown in §4.1.1, we have adopted the former since we compare SNR at a particular point in the object's spectrum.

A similar situation applies to the photon noise for FTS spectroscopy. Since photons from all of the spectral elements fall on the two detectors the photon noise for any spectral element is equal to the total noise from all of the spectral elements. If all S spectral elements contain the same power there are S more photons falling on the detectors with an increase in the noise of $(S)^{1/2}$. The photon noise for each IFTS spectral element is therefore increased by $(S)^{1/2}$ over the noise for a dispersive spectral element.

The remaining noise contributor is dark current in the detectors. Both methods of spectroscopy are sensitive to dark current in the same way. The only difference is that two detectors used in multiplex spectroscopy doubles the contribution of the dark current.

12.2.2 Sensitivity calculations

We calculate the signal-to-noise of the IFTS similar to the sensitivity equations described in §12.1.2. A spectral resolution element is defined to be one pixel for our calculations. The signal terms for the dispersive spectrometer and the IFTS, are

$$I_{obj}(T) = F * A_{tel} * \varepsilon_{DS} * DQE * \frac{\lambda}{R} * \frac{\lambda}{(h*c)}, \text{ and}$$

$$I'_{obj}(T) = F * A_{tel} * \varepsilon_{IFTS} * \tau * DQE * \frac{\lambda}{R} * \frac{\lambda}{(h*c)} \text{ electrons/s}$$

respectively. The sky terms, $I_{sky}(T)$ and $I'_{sky}(T)$, are related in a similar way, additionally each being multiplied by Ω, the solid angle seen by a pixel. The actual values used for the sky and detector quantum efficiency are from the NGST background model described in §12.1.1. The other terms, including detector parameters and pixel scales, are described in §12.1.1-2.

The efficiency terms (ε_{DS}, ε_{IFTS}) include instrument transmission, grating efficiency, and slit losses. The extra term τ in the IFTS equations is the modulation efficiency of the beam splitter. If the transmission and reflection of a beam splitter are not exactly equal, some photons which fall on the detector will not produce a signal. These photons do contribute to the photon noise, however. For this reason the multiplex signal is multiplied by τ but the photon noise contribution is not. For our calculations, whose results are shown in §4.1, we have assumed that ε_{DS} and ε_{IFTS} are each 50%, and τ is unity. τ is often 90% for existing IFTSs.

The measurement noise is given by

$$N(T) = \sqrt{(I_{obj}(T) + I_{sky}(T) + I_d(T)) * T + N_r^2} \text{ , and}$$

$$N'(T) = \sqrt{S * (I'_{obj}(T) + I'_{sky}(T)) * T + 2 * I_d(T) * T + 4 * S * N_r^2} \text{ ,}$$

respectively. Here, for the IFTS, the background and object are assumed to contribute equal amounts of flux to each spectral element, and thus increase the photon noise by a factor of S. These should be replaced by integrating these signals over the spectral range.

The resultant signal-to-noise ratios are given for a total integration time, T, which in practice might be composed of several shorter exposures, with maximum exposure time t=3000s. We treated the dispersive spectrograph in the same way as in §12.1.2. For the IFTS, we replaced each T in the above equations with T/(2S), the time per spectral element, because our illustrated examples in §4.1.1-2 compare the SNR of the IFTS to a spectral element in the dispersive spectrometer.

12.3 Micromirror calculations

This appendix summarizes the calculations used in evaluating performance criteria for a micromirror array (MMA) based multi-object spectrograph, the results of which are found in §4.3.1-2.

12.3.1 Rejection of diffuse IR background sky

Micromirrors provide a method of configurable multi-object spectroscopy, in reflection. As with any reflective spectrometer slit, some light will be scattered into the spectrograph by the structure around an “ON” mirror, namely a field of “OFF” micromirrors. These regions of adjacent sky which, if scattered from mirrors in their “OFF” position, will contribute an additional unwanted signal which will degrade the spectrometer’s sensitivity. Our simple model did not address the source of the scattering, and is based on the assumption that the calculated rejection ratios for the

mirrors are based for light which is scattered on axis, and hence could be underestimating the unwanted contamination. Thus, these results serve only as a figure of merit in ascertaining the required performance of the MMA option.

The "OFF" mirrors need to reject the diffuse sky background sufficiently over a region of sky for which sky spectra will overlap that of the object. This is shown in Figure 12.3.1-1. If *x* is the spectrum's length on the detector (i.e., number of pixels, function of spectral resolution), the sky background's contributing region extends *2x* in length. At each wavelength, the additional contribution of scattered sky light is the sum of all these overlapping diffuse sky spectra.

That is,

$$S_{\substack{scattered \\ sky}}(\lambda_i) = \varepsilon * \int_{\lambda_{\min}}^{\lambda_{\max}} S_{sky}(\lambda_i) d\lambda$$

for all $\lambda_{min} < \lambda_i < \lambda_{max}$.

The efficiency, ε, is defined as the percentage reflection of the micromirrors in the "OFF" position (e.g., 100%="ON", 0%=slit where metal blocks the sky adjacent to the object's spectra in the spectral direction). Similarly, we can define a rejection ratio as the ratio of the efficiency of the mirrors in the "ON" position to the "OFF" position. Thus, a scattering of 1% would have a rejection ratio of 0.01.

We ran sensitivity calculations (§12.1.2) to compare performance between a micromirror and a conventional slit fed spectrograph, the latter which is not affected by off-target scattering. The sensitivity model calculated the number of photons per spectral resolution element, adopted to be 2×2 pixels, for the contributions from 1) the object itself, 2) the sky background, 3) detector dark current and 4) read noise contributions, and in the MMA model, 5) an additional scattered sky light component from adjacent MMA mirrors in their "OFF" position.

A signal-to-noise measurement for a standard observation of total exposure time of T_{int} = 10,000s, composed of several shorter individual exposure times. Individual exposure times were set to 1000s over the 1-5μm to be sufficiently short enough to limit cosmic ray damage, and continually shorter towards longer wavelengths to avoid overfilling the detector full wells, down to 1s at 35μm (R=5).

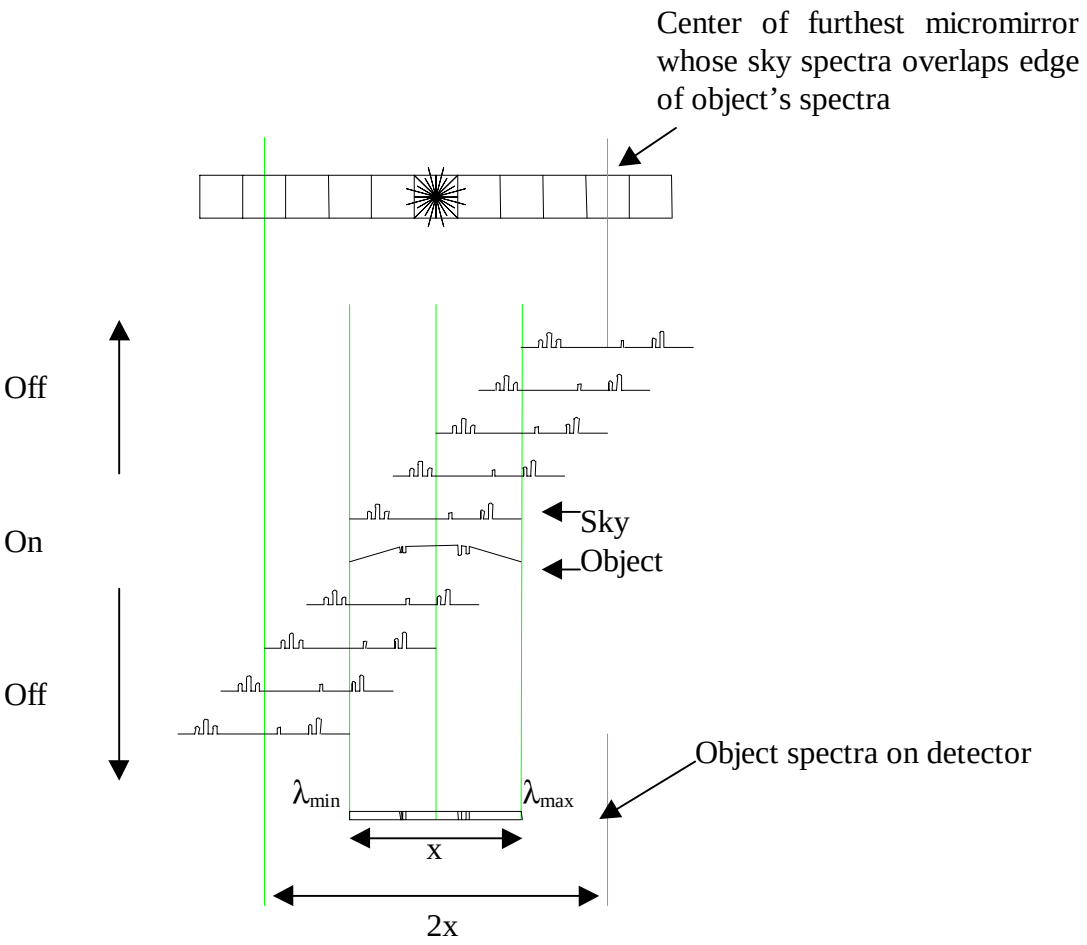


Figure 12.3.1-1 Schematic of the "avoidance zone" for over-lapping sky spectra from sky adjacent to target. At top is a row of a micromirror arrays with a star positioned in the center. Below is a projection of the spectra of the object and sky on the detector, the latter from both a mirror in the "ON" position, and from adjacent mirrors in their "OFF" positions. The contaminating region on the detector extends *2x* pixels, where *x* is the length of the object's spectrum.

The model calculated the target brightness which would give a SNR=10 with zero scattering (i.e., the slit case). Introducing the scattered sky background would naturally degrade the instrument sensitivity. We calculated what the efficiency parameter would be to compute a SNR=9, or a 10% performance from the zero scattering case.

The contribution from the scattered diffuse sky background is dependent on the filter bandpass ($\lambda_{min} < \lambda < \lambda_{max}$). In practice, the maximal spectral bandpass per one grism is set by its free spectral range. For a grism in first order, it is maximal at one octave. Therefore, the bandpasses chosen for the calculations represented first order grism coverage.

12.3.2 Rejection of contaminating objects

We also investigated the criteria for rejecting background contaminating objects in order to obtain suitable SNR spectra of faint targets. This was parameterized by running several Monte Carlo simulations and calculating the percentage of contaminated object spectra by a random distribution of background sources.

A background source's spectrum contaminated our target object if all the following criteria were met:

1) it was located within the slit width,
2) it was located within 2 * slit length, and
3) its contributing flux was above a "threshold brightness."

This threshold brightness takes into account the rejection ratio of the micromirrors as well as the percentage contribution of the source's brightness is contributed to the target's spectrum. For example, if we judge a source which fulfills criteria 1) & 2) to be invalid if it contributes 10% or more of its brightness with scattering 1% from the micromirrors, the faintest the source can be is brightness 100 if the target brightness is 10.

In these calculations, the slit width was defined to be four Airy disk FWHMs, where one Airy disk FWHM is λ/D in size. The slit length was calculated from the dispersion, $\lambda/(2R)$, in µm/pixel, and a bandpass, $\lambda_{max} - \lambda_{min}$, as

$$\text{slit length} = \frac{(\lambda_{max} - \lambda_{min})}{\text{dispersion}} = 2 * R * \frac{(\lambda_{max} - \lambda_{min})}{\lambda} \text{ pixels} .$$

The examples illustrated in §4.3.2 were calculated with λ_{min}=1µm, λ_{max}=5µm, sampled at λ=3µm.

The number of sources distributed randomly in the simulation were calculated from a normalized power spectrum of the form,

$$N(> x) = C * x^{-\alpha} ,$$

where N(>x) is the number of sources brighter than brightness *x*, an integrated form of the number magnitude count distribution and, α is the negative of the index of this power law integral. Values of $\alpha \sim 0.5$, agree with the present day predictions used for deep galaxy counts for NGST. The simulations were run for 1000 Monte Carlo trials.

12.4 Further details of ISIM modules

This section contains additional information about the modules, not mentioned in §6.

The modularity of our ISIM concept allows for individually simple and small modules. Figure 12.4-1 shows all our proposed modules on the same scale. For comparison, we show the yardstick NIRCAM as well.

Table 12.4-1 summarizes the first order optical design parameters for the modules discussed in §6.

Table 12.4-2 lists possible KRS-5 grism combinations for the modules.

Table 12.4-3 summarizes the model parameters used to address the formatting and coverage of the echellette modes for the Mid-IR and long-IR spectrographs (§6.1.5-7).

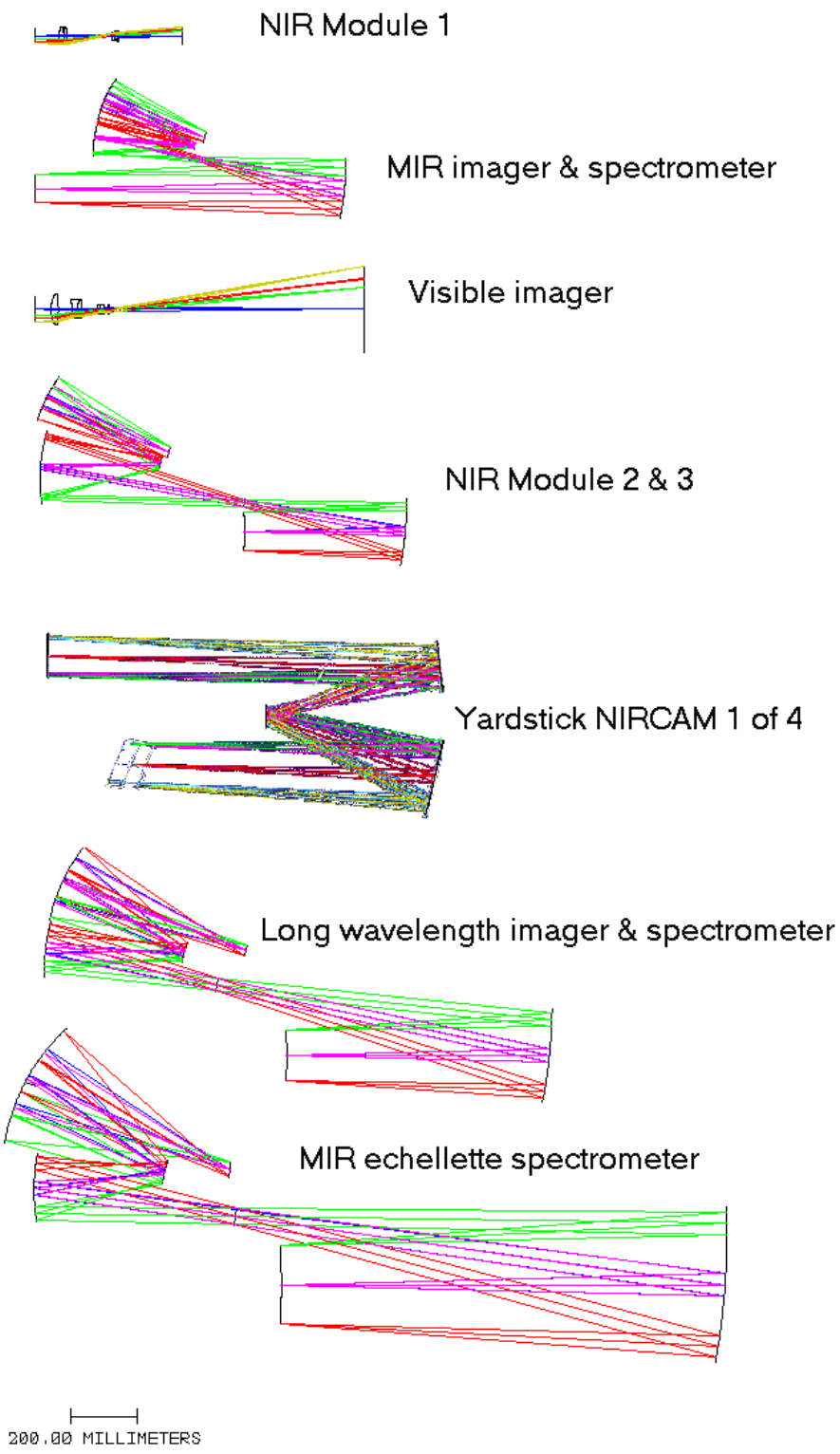


Figure 12.4-1 Relative sizes of the modules in our ISIM, compared with the size of the yardstick NIRCAM (1 of 4 modules), all on the same scale.

	Module	Module	Module	Module	Module	Module	Module	Module	Wide Field	Wide Field
Telescope diameter	8000	8000	8000	8000	8000	8000	8000	8000	8000	8000
Telescope f/num	20	20	20	20	20	20	20	20	20	20
Report Section	**§ 6.1.1**	**§ 6.1.3**	**§ 6.1.3**	**§ 6.1.3**	**§ 6.1.5**	**§ 6.1.6**	**§ 6.1.7**	**§ 6.1.8**	**§ 6.1.2**	**§ 6.1.4**
	Visible Imager	NIR module 1	NIR module 2	NIR module 3	MIR im/spec	MIR echellette	LIR im/spec	LIR echellette	WF Vis imager	WF NIR imager
Im min wavelength	0.4	1	2.3	2.3	5	5	20	18	0.4	1
Im max wavelength	1	2.3	5	5	27	27	34	35	1	5
Spec min wavelength		1.25	1.25	1.25	5	5.2	20	17		
Spec max wavelength		2.5	5	5	20	18.75	34	34		
Nyquist wavelength	0.5	1.5	3.4	5.7	8	15.5	27	15.5	0.7	1.8
array size in pixels	10240	2048	2048	2048	1024	1024	512	1024	40960	18432
pixel size (um)	18	18	18	18	27	27	50	27	7	18
array size in mm	184.32	36.864	36.864	36.864	27.648	27.648	25.6	27.648	286.72	331.776
Magnification	0.3	0.8	1.9	3.2	3.0	5.7	5.4	5.7	1.0	1.0
Final focal ratio	72.00	24.00	10.59	6.32	6.75	3.48	3.70	3.48	20.00	20.00
Pixel size in arcsec	0.006	0.019	0.044	0.074	0.103	0.200	0.349	0.200	0.009	0.023
Full FOV arc min	1.102	0.661	1.499	2.512	1.763	3.416	2.975	3.416	6.171	7.140
telescope image size in mm	51.2	30.72	69.632	116.736	81.92	158.72	138.24	158.72	286.72	331.776
Coll pupil diam in mm	10.2	6	10	20	15	38	40	38		
collimator focal length mm	204	120	200	400	300	760	800	760		
nominal size of collimator (mm)	61.4	36.72	79.632	136.736	96.92	196.72	178.24	196.72		
camera FOV degrees	14.4	14.7	20.0	16.7	15.7	12.0	9.9	12.0		
camera focal length mm	734.40	144.00	105.88	126.32	101.25	132.39	148.15	132.39		
collimation for 0.1 wave astig (f/num)	25.3	24.4	13.2	18.8	14.0	20.8	8.9	13.0		
Spectral shift due to field		0.37%	0.68%	0.47%	0.42%	0.24%	0.17%	0.24%		
Spectral broadening due to Coll. requirement for 0.1 wave astig		0.01%	0.03%	0.02%	0.03%	0.01%	0.07%	0.03%		

Table 12.4-1 Module optical parameters. Pixels are 2/(λ*D) in size at the Nyquist wavelength.

Module	Grism No.	λ_{blaze} (μm)	R	Bandpass (μm)	Prism angle (degrees)	l/mm
NIR Module 1 (6mm pupil)	1	2	100	1.25-2.5	1.0	12
	2	2	1000	1.25-2.5	10.1	123
	3	1.6	3000	1.25-2	28.2	413
	4	2.2	3000	2-2.5	28.2	301
NIR Module 2 & 3 (20mm pupil)	1	2	100	1.2-2.5	1.4	17
	2	3.7	100	2.5-5	1.4	9
	3	2	1000	1.25-2.5	1.4	165
	4	3.7	1000	2.5-5	13.7	89
	5	2	3000	1.25-2.25	13.7	412
	6	2.7	3000	2.25-3.25	36.1	305
	7	3.7	3000	3.25-4.25	36.1	223
MIR im/spec (15mm pupil)	1	8	100	5-10	2.2	7
	2	15	100	10-20	2.2	4
	3	8	500	5-10	10.8	33
	4	15	500	10-20	10.8	17
	5	8	1000	5-9	20.9	62
	6	15	1000	9-16	20.9	33
	7	22	1000	16-20	20.9	23
MIR echellette low resolution modes (38mm pupil)	1	8	100	5-10	1.7	5
	2	15	100	10-20	1.7	3
	3	8	500	5-10	8.3	25
	4	15	500	10-20	8.3	13
	5	8	1000	5-9	16.2	49
	6	15	1000	9-16	16.2	26
	7	22	1000	16-20	16.2	18
LIR im/spec (40mm pupil)	1	27	500	20-34	10.8	13
	2	23	1000	20-27	28.3	29
	3	30	1000	27-34	28.3	22
LIR echellette low resolution modes (38mm pupil)	1	27	500	20-34	8.3	7
	2	23	1000	20-27	16.2	17
	3	30	1000	27-34	16.2	13

Table 12.4-2 Spectral coverage versus resolution for the modules. The grism parameters assume KRS-5 material and designed for first order.

	Mid-IR Echellette #1 5.20-10.4-µm, orders 7-13		Mid-IR Echellette #2 10.19-17.70µm, orders 10-16	
Pixel size (µm)	27µm = 0.20 arcsec		27µm = 0.20 arcsec	
Nyquist wavelength (µm)	15.52 (λ/(2D) pixels)		15.52 (λ/(2D) pixels)	
Array size (pixels)	1024×1024		1024×1024	
Pupil diameter (mm)	38		38	
Slit width (pixels)	2		2	
Slit length (arcsec)	9 (45.7 pixels)		9 (45.7 pixels)	
Minimum y separation (pixels)	20		25	
Camera focal length	132.3		132.3	
	Echelle grism	Cross-disperser grism	Echelle grism	Cross-disperser grism
Resolution	3000	700	3000	480
Prism index	2.36	2.36	2.36	2.36
Prism angle (degrees)	39.48	11.8	51.85	13.42
Blaze angle (degrees)	31.4	8.345	42.5	9.5
l/mm	12.3	36.1	6.3	21.6
Reference order	9	1	12	1
Blaze λ of reference order (µm)	7.82	7.71	14.03	14.64
Blaze dispersion of reference order (Å/mm)	484.3	2047.8	578.9	3408.3

	Long-IR Echellette #1 16.8-28.5µm, orders 12-19		Long-IR Echellette #2 20.6-35µm, orders 12-19	
Pixel size (µm)	27µm = 0.20 arcsec		27µm = 0.20 arcsec	
Nyquist wavelength (µm)	15.5µm (λ/(2D) pixels)		15.5µm (λ/(2D) pixels)	
Array size (pixels)	1024×1024		1024×1024	
Pupil diameter (mm)	38		38	
Slit width (pixels)	4 (2 if 512×512 array)		4 (2 if 512×512 array)	
Slit length (arcsec)	10 (48 pixels)		10 (48 pixels)	
Minimum y separation (pixels)	20		20	
Camera focal length	132.3		132.3	
	Echelle grism	Cross-disperser grism	Echelle grism	Cross-disperser grism
Resolution	3000	425	3000	425
Prism index	2.36	2.36	2.36	2.36
Prism angle (degrees)	59.7	14.1	59.7	14.1
Blaze angle (degrees)	50.51	10.01	50.51	10.01
l/mm	3.6	14.9	2.9	12.1
Reference order	15	1	15	1
Blaze λ of reference order (µm)	21.85	22.32	26.79	27.39
Blaze dispersion of reference order (Å/mm)	675.3	4887.2	828.3	6001.5

Table 12.4-3 Parameters used for evaluating resolution, wavelength coverage, and format for the high resolution cross-dispersed echellettes mentioned in §6.1.6-7.

12.5 Infrared grisms

All of our designs for the spectroscopic modes in our ISIM concept rely on grisms. The most attractive aspect of using grisms as the dispersing element is that they can be placed in a filter wheel without deviating the desired central wavelength and can thus be treated like a filter. It allows a simple conversion between an imager and spectrograph with only a filter wheel. The largest drawback for a grism alternative over a grating, is their lowered dispersion, by a factor of (n-1)/2 as compared to a grating with identical groove parameters when n is the refractive index of the grism substrate.

This section addresses some of the design parameters used for designing grism spectrometers, with particular emphasis on materials to be used to achieve high dispersion in the near and mid infrared. The grism choices featured in our strawman ISIM (§6.1, §12.4) are based on what can presently be made, but these should evolve to pace future development. Continued work on the manufacture of high quality IR grisms will be required for NGST.

The spectroscopic properties of grisms are derived from geometric optics (Beckers, et al. 1987). It can be shown that the resolution of a grism (normal incidence, matching two pixels to slit width) blazed for a particular wavelength λ is proportional to (n-1), where n is the index of refraction of the grism material, to the tangent of the prism angle, and to the diameter of the collimated beam incident on the grism,

$$R = \frac{\Delta\lambda}{\lambda} = \tan(\alpha) * (n-1) * \frac{f_{cam}}{(2 * d_{pix})} = \tan(\alpha) * (n-1) * \frac{D_{pupil}}{\lambda}.$$

Similarly, ruling requirements (i.e., a, the spacing between facets) for the grism can be calculated, being proportional to (n-1) and the sine of the prism angle, at a given wavelength for a grating order m,

$$\frac{1}{a} = (n-1) * \frac{\sin(\alpha)}{m * \lambda}.$$

Line densities a > 300l/mm are practical with a variety of different materials and technologies.

Conventional optical grisms be manufactured in two ways: replicated or directly ruled. For the former, a master grating is pressed into a transparent resin which has been deposited onto the hypotenuse of a prism. For the latter, a grating is directly ruled into the hypotenuse of a prism. Resin grisms work well out to 2.5µm, beyond which the resins absorb strongly and can no longer be used. At these longer wavelengths, a directly ruled grism is the only current manufacturing solution. Hybrid grisms, ones which have a grating manufactured into a different optical material (e.g., ruled resin on glass), are possible, but care is needed to choose materials to prevent large reflection losses at the interface surface and internal reflection.

Two techniques for ruling long wavelength (λ>2.5μm) grisms are under development: 1) solid state micro-machining/anisotropic etching into germanium or silicon and, 2) direct ruling into materials such as CaF_2, ZnSe, and KRS-5 (Rayner, 1998; Käufl et al. 1998).

Direct ruling under oil of KRS-5 (n~2.4 at 2μm) has proved to be a successful way of producing moderately efficient (>50%) grisms with low scattering. Experiments with ZnSe (n~ 2.4 at 2μm) have led to cracking of the material when it is diamond machined, causing a lot of scattered light. CaF_2 rules well and successful CaF_2 grisms have been produced which are currently being used in the NICMOS instrument on HST. However, the low refractive index of CaF_2 (n~1.4 at 2μm) results in relatively low spectral resolutions.

12.5.1 Moderate resolution grisms

All dispersions in the near-IR modules and the cross-dispersion in the mid-IR modules are done with first order grisms which avoid large prism angles and excessively high ruling densities (Tables 12.4-2 and 12.4-3). Grisms of similar design have been demonstrated to perform well over a number of years in various ground-based instruments at the IRTF (Rayner et al. 1998) and are also being used at other observatories. Thus we expect that this existing technology will perform well for NGST.

12.5.2 High resolution issues

The spectral resolution of a grism spectrometer can be increased by increasing the collimated pupil size. For a cryogenic instrument, the pupil size (cold stop) drives the size and weight, and hence, complexity of the instrument. In practice this number cannot be increased easily in a design. Thus to achieve higher spectral resolutions, materials of high index of refraction must be used. Figures 12.5.2.1-2 summarize the infrared transmissions and indices of refraction for a selected list of high refractive index materials, taken from the *Handbook of IR optical materials.*

The spectral resolution can also be increased, by increasing the prism angle. However, it has been shown in practice that it is desirable to keep the prism angle $\alpha \leq 60^o$ to avoid complicated optimization of the grism (Käufl et al. 1998). At such steep angles the above approximating equations for relating resolutions and line densities to the central wavelength no longer hold, and spectrographs may suffer anamorphic effects and lateral displacement of the pupil (Neviere 1991; Käufl et al. 1998). High ruling densities (a > 300 lines mm^{-1}) increase polarization, and the grism efficiencies of different polarizations can differ substantially in grisms with large prism angles.

The lack of transmissive materials for wavelengths longer than 28μm creates limited choices for grism materials for the ISIM long wavelength module (§6.1.7). Only KRS-5 has been found to show acceptable transmission at these wavelengths. This restricts the maximum achievable spectral resolution mode to about R=3000, with a 58.8º, 43 l/mm grism, for a pupil diameter of 40mm for an echelle-type grism operated in order m=11 at a central blaze wavelength of 24 μm (for operation over $16.8 < \lambda < 33.6$ μm).

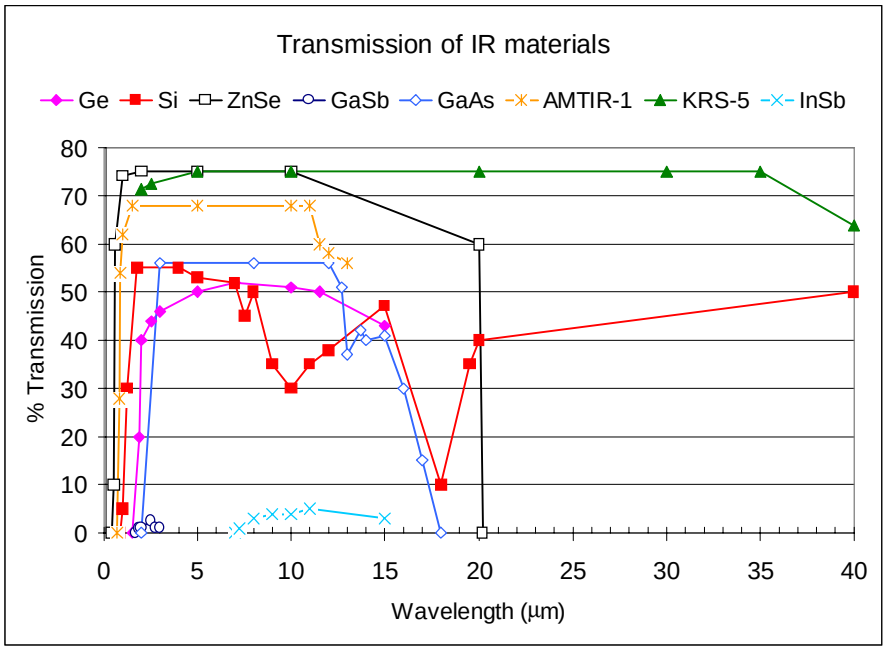


Figure 12.5.2-1 Transmission of several high refractive index materials. The best material for high index grism choices for λ>2.5µm will be among KRS-5 (solid triangle), Si (solid square) and Ge (solid diamond). (Klocek 1991, *Handbook of IR optical materials*)

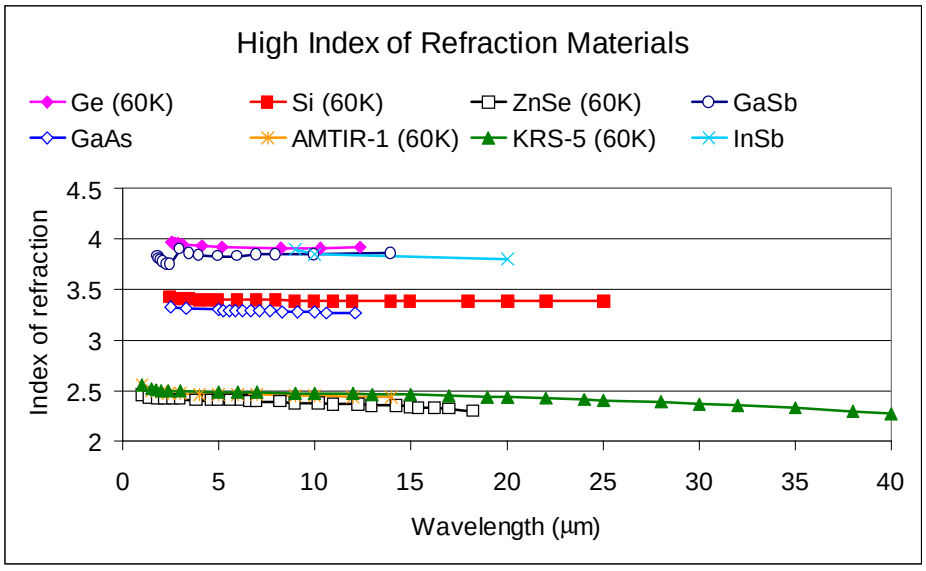


Figure 12.5.2-2 Indices of refraction at 60K for selected list of high refractive list materials. Entries without specified temperature were measured at 25ºC. Most materials' refractive indices will tend to decrease with decreasing temperature, with the exception of KRS-5. (Klocek 1991, *Handbook of IR optical materials*)

Similar steep-angled echelle-type grisms have been used successfully in both optical (Dekker et al. 1998) and mid-IR (Käufl et al. 1998) astronomical spectrographic cameras. Any reduction in grism efficiency resulting from a large prism angle (caused by shifting the blaze peak away from

the central undeviated wavelength) can be largely restored by either tilting the grism slightly (Dekker et al. 1998) or else modeling these effects in the grism design (Neviere 1991). Grisms generally do scatter more light when used in higher orders, but we do not anticipate that this will be a problem with our echelle-type designs because they require such low groove densities (a ~ 10 lines mm^{-1}) and KRS-5 has proven itself as a low-scatter grism material (Rayner 1998). Thus we see no serious problems with manufacturing or achieving good performance from our echelle grism designs.

The long wavelength response for silicon (n~3.4) might increase the achievable resolution in the LW module to R~5000, with a 57.4º, 77 l/mm grism for the same above pupil size. A number of groups have demonstrated basic steps toward silicon grisms by means of anisotropic etching techniques, although they have not yet been manufactured routinely (Käufl et al. 1998). If they can be manufactured, these silicon grisms will require anti-reflection coatings, because with such a high refractive index material, the Fresnel reflection losses amount to ~30%.

12.6 Details of cost model components

Table 12.6-1 details the work breakdown structure elements used in evaluating the cost for our ISIM concept. The summaries of the final cost, arranged by module and by functional area are given in §9.

WBS Element	Cost ($M FY96)	Definition
1.0 ISIM Program Management	2.622	Program management, cost / schedule tracking, contract control
2.0 ISIM Systems Engineering	11.739	Resource allocations, interface definitions,
3.0 ISIM Structure and Thermal	5.054	ISIM Common structures, mechanisms, radiators, blankets
4.0 ISIM Power Distribution and Harness	0.756	ISIM power harnesses and power distribution devices
5.0 Wide Field Science Instrument	125.392	
5.1 Program Management	2.622	Program management
5.2 WF NIR Module	59.435	
5.2.1 Development Engineering	12.699	Development engineering, design and analysis
5.2.2 I&T	7.604	Integration and test (calibration) of NIR WF Module
5.2.3 Detectors	38.022	Detector hardware costs
5.2.4 Optics	0.000	Not Applicable
5.2.5 Mechanisms and Structure	1.110	Structure, mechanism, and thermal interface hardware costs
5.3 Visible Wide Field Imager	59.435	
5.3.1 Development Engineering	12.699	Development engineering, design and analysis
5.3.2 I&T	7.604	Integration and test (calibration) of VIS Wide Field Module
5.3.3 Detectors	38.022	Detector hardware costs
5.3.4 Optics	0.000	Not Applicable
5.3.5 Mechanisms and Structure	1.110	Structure, mechanism, and thermal interface hardware costs
5.4 WF Processor	1.500	Hardware costs for a processor for the wide field cameras
5.5 WF VIS/NIR Software	2.400	Software development costs for wide field cameras
6.0 NIR/VIS Science Instrument	67.860	

6.1 Program Management	2.622	Program management for presumed U.S. delivered NIR SI
6.2 Visible Imager Module	19.994	
6.2.1 Development Engineering	3.396	Development engineering, design and analysis
6.2.2 I&T	1.017	Integration and test (calibration) of VIS Module
6.2.3 Detectors	11.725	Detector hardware costs
6.2.4 Optics	2.746	Optics hardware costs
6.2.5 Mechanisms and Structure	1.110	Structure, mechanism, and thermal interface hardware costs
6.3 NIR Module 1	10.327	
6.3.1 Development Engineering	3.296	Development engineering, design and analysis
6.3.2 I&T	0.987	Integration and test (calibration) of NIR Module 1
6.3.3 Detectors	3.292	Detector hardware costs
6.3.4 Optics	1.642	Optics hardware costs
6.3.5 Mechanisms and Structure	1.110	Structure, mechanism, and thermal interface hardware costs
6.4 NIR Module 2	11.487	
6.4.1 Development Engineering	3.711	Development engineering, design and analysis
6.4.2 I&T	1.111	Integration and test (calibration) of NIR Module 2
6.4.3 Detectors	3.292	Detector hardware costs
6.4.4 Optics	2.263	Optics hardware costs
6.4.5 Mechanisms and Structure	1.110	Structure, mechanism, and thermal interface hardware costs
6.5 NIR Module 3	13.330	
6.5.1 Development Engineering	4.370	Development engineering, design and analysis
6.5.2 I&T	1.308	Integration and test (calibration) of NIR Module 2
6.5.3 Detectors	3.250	Detector hardware costs
6.5.4 Optics	3.292	Optics hardware costs
6.5.5 Mechanisms and Structure	1.110	Structure, mechanism, and thermal interface hardware costs
6.6 NIR/MIR/LIR Processor	1.500	Hardware costs for a processor shared between NIR and MIR/LW Sis (US Sis)
6.7 NIR/MIR/LIR Software	8.600	Software development costs for NIR/MIR/LW software
7.0 MIR / LIR Science Instrument	44.574	
7.1 Program Management	2.622	Program management for presumed U.S. delivered NIR SI
7.2 MIR Imager/Spectrometer	9.872	
7.2.1 Development Engineering	3.133	Development engineering, design and analysis
7.2.2 I&T	0.938	Integration and test (calibration) of MIR Imager/Spectrometer
7.2.3 Detectors	2.189	Detector hardware costs
7.2.4 Optics	2.502	Optics hardware costs
7.2.5 Mechanisms and Structure	1.110	Structure, mechanism, and thermal interface hardware costs
7.3 MIR Spectrometer	13.960	
7.3.1 Development Engineering	4.595	Development engineering, design and analysis
7.3.2 I&T	1.376	Integration and test (calibration) of MIR Spectrometer
7.3.3 Detectors	2.189	Detector hardware costs
7.3.4 Optics	4.690	Optics hardware costs
7.3.5 Mechanisms and Structure	1.110	Structure, mechanism, and thermal interface hardware costs
7.4 LIR Imager	10.120	
7.4.1 Development Engineering	3.222	Development engineering, design and analysis
7.4.2 I&T	0.965	Integration and test (calibration) of LW Imager
7.4.3 Detectors	1.775	Detector hardware costs
7.4.4 Optics	3.048	Optics hardware costs

7.4.5 Mechanisms and Structure	1.110	Structure, mechanism, and thermal interface hardware costs
7.5 Cryo-Cooler	8.000	Hardware costs for cryo-cooler system used for MIR and LW modules
8.0 ISIM I&T	11.455	Integration and alignment, test and calibration of Sis into ISIM
Functional Areas		
1.0 Management	10.488	
2.0 Science / Development Engineering	51.121	
3.0 Systems	11.739	
4.0 SR&QA / I&T	34.365	
5.0 Structure	10.531	
6.0 Optics	20.183	
7.0 Electronics	3.756	
8.0 Operations		
9.0 Thermal	4.513	
10.0 Software	11.000	
11.0 Detectors	103.756	
12.0 Cryo-cooler	8.000	
ISIM Elements		
ISIM Integrator	31.626	
VIS SI	85.951	
NIR SI	107.301	
MIR/LW SI	44.574	

Table 12.6-1 Details of components for ISIM cost breakdown structure.