Next Generation Space Telescope

Visible wavelength instruments

Summary report

Reference : DSO/940/DD/039.99

21 September 1999

Report prepared by:

Joël BOYADJIAN **Dominique DUBET** **Olivier SAINT PE**	**Matra Marconi Space (MMS)**	
Cornelius SCHALINSKI	**Dornier SatellitenSystem (DSS)**	
Martin J. WARD **Paul O'BRIEN** **Andrew HOLLAND**	**University of Leicester (UL)**	
Mark CROPPER **Richard BINGHAM** **Franscisco DIEGO**	**University College London (UCL)**	
Denis BURGARELLA **Jean Luc GACH**	**Institut Gassendi pour la Recherche Astronomique en Provence (IGRAP)**	
Emmanuel PECONTAL	**Centre de Recherche Astronomique de Lyon**	
Slimane BENSAMMAR	**Observatoire de Paris (OP)**	

ESA contract 13328/98/NL/MS

ESA manager: J.W. CORNELISSE

Table of contents

1. SCIENCE

1.1 EXECUTIVE SUMMARY

Unlike HST the NGST will not be a serviceable observatory - what you launch is what you get. For this reason it is even more important to consider very carefully the composition of the instrument suite. Another boundary condition is that there will be a stringent cost cap, and we cannot reasonably expect to afford a comprehensive payload of instruments covering all of the observational parameter space potentially accessible to the telescope. Recognizing these constraints we must therefore set priorities.

The scientific rationale for near infrared instruments (1-5 microns) aboard the NGST has been well rehearsed from the very inception of the project. For example the huge gains in sensitivity by virtue of efficient passive cooling, lack of atmospheric OH airglow and thermal emission. The main science drivers are also obvious and compelling, for example galaxy evolution via study of the optical region red-shifted into the IR, and the importance of dust attenuation at high redshift and for studies of star formation locally.

Having identified the infrared region as the raison d'être for NGST, other important issues need to be considered. By the time of launch of NGST, HST operations will be close to an end. Are we really willing to give up access to large aperture space-borne observations at optical wavelengths? On the other hand, by 2008 there will be about ten 8m class ground-based telescopes. However we should note that the background at L2 should be from 5-10 times less than from the ground in the range 0.5-0.8 microns. In addition assuming that adaptive optics fulfils its promise, the isoplanatic patch will be only around 10 arcsecs in diameter, and observations will suffer from time variable PSFs. At optical wavelengths it is unlikely that routine AO observations will have spatial resolutions much better than a few tenths of an arcsec. Clearly a wide field visible camera covering several arcmins with pixels less than 0.1 arcsec would have capability unparalleled from the ground. Field censored spectroscopy over the whole field would be highly desirable, but needs to be technically viable, and lie within complexity and budgetary limits. Another powerful instrument would be a high resolution camera, making use of the best spatial resolution provided at optical wavelengths. This could have a field of view between one and two arcmins. Finally, an integral field spectroscopic (IFS) mode is desirable. Since the gain factors over ground-based instruments are less for high resolution spectroscopy, a medium resolution of about R ~ 1000, is considered. This would permit emission line diagnostics and kinematic studies over a complete 10x10 arcsec field of view sampled at 0.1 arcsecs.

Although these technical points support the provision of an optical camera for the NGST, the main purpose of the material in this section is to put forward a convincing scientific case. A visible camera cannot be justified as an 'add-on' instrument for old times sake. It must provide highly relevant and complementary data which would enhance the value of the IR observations, as well as undertaking unique and important programmes in its own right. The selection and study of targets on the basis of their broad-band energy distribution has proven crucial over a vast range of astronomical applications, ranging from distant galaxies to cool white dwarfs. The usefulness of such broad-band studies increases greatly with the breadth of the wavelength-baseline. Extending this wavelength-baseline by nearly a factor of two (~0.5 microns to 1.0 microns) and providing NGST with an optical imaging capability that matches its near-IR capabilities in depth, spatial resolution and field-of-view, will greatly enhance the scientific efficiency of the observatory. A visible camera would not be a niche instrument, but would be used as the instrument of choice for projects ranging from studies of distant galaxies, to Kuiper belt objects. The NGST Design Reference Mission currently lists 19 suggested programmes, of which 8 require observations shortward of 1 micron.

The topics described below have been suggested by members of our Science Study Group, whose combined astrophysical expertise covers a broad range. In addition we have received valuable inputs

from a much wider community of interested astronomers, and benefited from the dissemination of ideas submitted as DRM program elements.

1.2 SCIENCE CAPABILITY

1.2.1 The origin and evolution of galaxies

Studying the formation and evolution of galaxies is one of the main tasks for NGST in the context of the 'Origins' program and has been identified as the top-ranked science area in the NGST Design Deference Mission (DRM). In order to enable a derivation of the galaxy luminosity function and the star formation rate (SFR) of the Universe as a function of redshift (i.e. look-back time or age) between the redshift range currently probed by HST ($z < 3$-4) and that at which galaxies formed (possibly $z = 10$-20), it is essential that visible capability be provided on NGST. Having only IR capability would leave a gap in the redshift coverage and significantly limit our understanding.

In this section we discuss the need for access to the rest-frame UV. Examples of recent ground-based and HST studies are then presented, followed by the implication of these studies for surface number-density estimates of likely targets for NGST and required sensitivities. The visible instrument proposed here is capable of meeting all the requirements, and would provide a very powerful tool with which to study galaxies in the Universe.

1.2.1.1 Why look in the rest-frame ultraviolet?

Although the study of galaxy evolution has been revolutionised by recent HST and large-aperture ground-based telescope observations, it is clear that these observations only sample the tip-of-the-(luminosity/mass)-iceberg. Also, simulations show that high redshift galaxies tend to be mis-classified based on their apparent morphology, unless one is able to detect the evolved (fainter) stellar population which defines their true morphology. Thus, in order to study the full assembly of galaxies, and hence properly test galaxy evolution theories, we need to detect galaxies at much fainter flux levels and at higher redshifts. The sensitivity of NGST in both the optical and IR will make it a uniquely powerful facility for such studies. Figure 1.2.1-a shows examples of faint blue galaxies detected with the HST. These are mostly at low redshifts (typically $z = 1$-2). The proposed visible instrument for NGST will enable us to detect the building blocks of galaxies at much greater distances, as well as determining their morphology and redshifts.

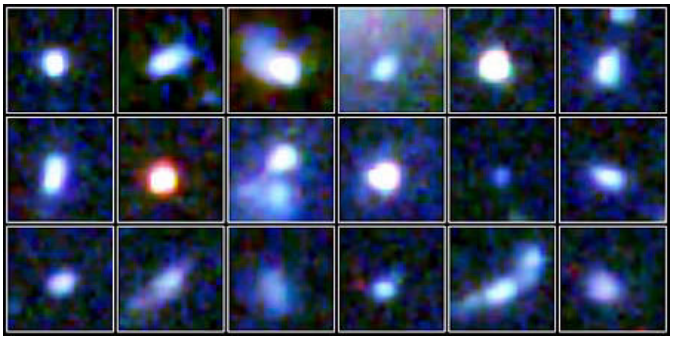


Figure 1.2.1-a: Examples of faint blue galaxies detected with the HST Windhorst & Pascarelle (1996, PRC96-29b). Composite blue, green and far red (optical) images

Based on data from HDF-N, Madau et al. (1996, MNRAS, 283, 1388; 1998, ApJ, 498, 106) claimed that the SFR increases from z=0 to z=1 and then declines beyond z=2 (the 'Madau plot'). These data imply that the SFR is strongly peaked, with more than 80% of stars made around z=2-3. However, more recent analyses (e.g. Pozzetti et al. 1998, MNRAS, 298, 1133; Steidel et al. 1999, ApJ, 519, 1) suggests that the SFR may be rather more constant over the range $0.8 < z < 3$ and possibly out to z=5.

At even higher redshifts the SFR is also of considerable interest with regard to re-ionization of the intergalactic medium. The sources responsible must have UV bright (unreddened) continua (shifted into the optical band) and could be starforming galaxies at z > 5. Whatever the exact dependence of SFR with redshift, it is clearly crucial to be able to extend the present studies to identify and study starforming galaxies at medium-high redshifts.

The spectra of young, star-forming galaxies should be dominated by hot O and B stars, and therefore should appear similar to those of nearby, giant star-forming regions (see, for example, the recent review by Thommes (1998, Ringberg Conference, 'From stars to galaxies in the Universe', astro-ph/9812223). Detailed models predict three key features all of which appear in the rest-frame UV:

(1) Strong Lyman-alpha emission, with rest-frame EW = 50 - 250Å.

The detection of Lyman-alpha provides a clear redshift indicator. Lyman-alpha carries about 3-6% of the bolometric luminosity, and recombination theory states it should be 8-11 x H-alpha. But, as Lyman-alpha photons can be resonantly scattered many times before escaping, they can be absorbed more easily by dust. Nevertheless, recent detections (see below) have shown strong Lyman-alpha emission is seen with rest-frame EW up to 100Å (corresponding to an observed EW up to 600Å) (section 1.2.1.2). NGST will be capable of detecting large numbers of such objects per square-arcmin (section 1.2.1.4). We note that the first generation of stars should form in a dust-free environment. Thus, the detection of strong Lyman-alpha would be a sensitive indicator of the very first star-forming regions. In such systems Lyman-alpha will be much stronger, and therefore easier to detect, than H-alpha (e.g. Dey et al., 1998, ApJ, 498, L93).

Once high-redshift galaxies have been identified, most likely through filter imaging (either broad-band or comb filters), follow-up spectroscopy using either the proposed visible integral-field spectrograph or a low-resolution prism will allow more detailed studies. For example, the Lyman-alpha profile can be obtained at moderate resolution (R=1200) using the IFS for point-source magnitudes down to the detection limit in the HDFs. The Lyman-alpha line profile can be used to estimate the gas outflow velocity Heckman (1998, in 'The most distant Radio Galaxies', Reidel, astro-ph/9801155).

Based primarily on Keck spectra, strong evidence is seen for outflowing gas in high-redshift galaxies (i.e. P-cygni profiles, Figure 1.2.1-b). The observed velocity shift enhances the possibility of detecting Lyman-alpha if the velocity shift is between the neutral gas, containing most of the dust, and the HII regions. These outflows may be a way to solve the 'metallicity problem' at high redshifts (Pettini, 1999, in 'Chemical evolution from zero to high redshift' (Berlin: Springer), astro-ph/9902173), in which the observed metal content of intervening absorption systems (such as damped Lyman-alpha absorbers) is too small compared to predictions. Outflows may place the metals into hot gas in halos and proto-clusters (Renzini, 1998, in 'The birth of galaxies' (Gif-sur-Yvette, Edition Fronteres), astro-ph/9810304) where it would be difficult to detect.

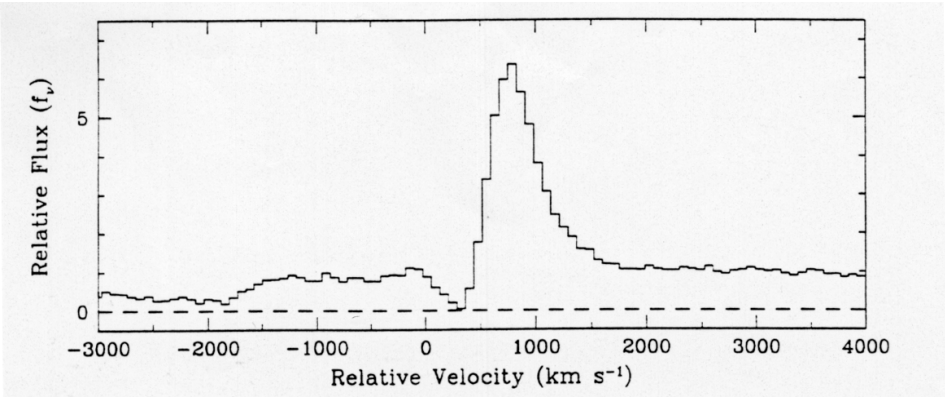


Figure 1.2.1-b: An example P-cygni Lyman-α profile for a high-redshift galaxy (Pettini et al 1997)

(2) Flat UV continua from Lyman-limit to Balmer-limit.

To determine the SFR the IR studies will primarily focus on redshifted H-alpha. However, for the predominately faint sources observed in the NGST deep-fields, moderate-resolution IR spectroscopy (R~100) will provide limited information (Kennicut, 1998, NGST Liege Colloquium, p.81). In contrast, the rest-frame UV continuum which arises mainly from massive, hot, short-lived stars will be much more easily observed. The detailed shape of the UV continuum is a measure of the SFR and the age of the starburst. Star-formation models suggest an intrinsically flat UV continuum (e.g. Charlot & Fall, 1993, ApJ, 415, 580), particularly for a constant SFR, while the near-IR continuum shape varies by only minor amounts with the SFR. The proposed visible instrument can provide the UV continuum shape using individual broad-band filter images (S/N=10 for point-sources down to m_{AB}~31.5 in 100ksecs per filter), simultaneous multiband (comb-filter) images (S/N=10 for m_{AB}~29 in 100ksecs) or by a prism (R~100) exposure (S/N=10 for m_{AB}~27 in 100 ksecs) - all of these magnitudes are per spectral-resolution element.

The rest-frame UV spectra will also provide reddening estimates through comparison of the observed UV continuum shape with those observed locally in starburst galaxies (e.g. Calzetti, 1999, in Ringberg Conference, astro-ph/9902107). A good correlation exists between the UV continuum shape and the extinction (e.g. Heckman, 1998; Meurer et al. 1999, astro-ph/0003054). Applying these calibrations to high redshift observations leads to extinctions at 1500Å (rest-frame) of 1-2 magnitudes. Such values are confirmed by a comparison of the stellar and dust emission in galaxies (Buat & Burgarella, 1998, A&A, 334, 72).

(3) The continuum flux drops sharply shortward of the Lyman-limit.

The continuum flux in high-redshift galaxies is expected to drop in the far-UV for three reasons: (a) the effect of intergalactic absorption due to material along the line of sight to the galaxy; (b) absorption in the high-redshift galaxy ISM; and (c) the intrinsic continuum shape of the O and B stars. The detailed shape of the latter depends on the age of the starburst. These features can be used to determine the redshift of the object using the 'dropout' or 'Lyman-break' technique.

Most of the recent work on galaxy evolution has been done using samples selected via the so-called 'Lyman-break' technique. These Lyman-break galaxies (LBGs) have been found using multi-colour imaging photometry (e.g. the HDFs) followed up by ground-based spectroscopy (section 1.2.1.2). These studies suffer from two major limitations: (a) the limiting magnitude of the HDFs (S/N=10 limit is V~28) prevents detection of small, faint galaxies, which may form the building blocks of the large 'Hubble type' galaxies seen at low redshift, and (b) even using 8m-class telescopes one generally cannot reach deeper than continuum magnitudes of V~26 with moderate-resolution spectroscopy in a reasonable time (1 night). Thus the majority (~90% or more) of the faint objects detected in the HDFs let alone those which NGST will detect are beyond the spectroscopic capabilities of ground-based facilities. If the wavelength limit for the NGST is restricted to longward of 1 micron, LBGs will only be detectable for z > 10. To discover LBGs in the redshift range approximately 5-9 requires a visible instrument on NGST (the lower-redshift limit obviously depends on exactly what short-wavelength cutoff NGST has - the lower the better in this context).

1.2.1.2 Examples of recent spectroscopic work on high-redshift galaxies

This and the following subsection give an overview of the current work on high redshift galaxies. As this is a large, fast-moving field of research only representative examples are given.

Most detailed work on high redshift galaxies has been based on Keck spectra (using LRIS), although similar studies will soon be possible with the VLT, Gemini, etc. The Keck/LRIS instrument spectroscopy limit is R~25-26 (in a night) for detection of the continuum, typically using a spectral resolution of 3-4Å (e.g. Lowenthal et al. 1997, ApJ, 481, 673) or 12 angstroms (e.g. Steidel et al. 1996, AJ, 112, 352). In fact, to save exposure time, most of the high redshift galaxies observed to date

have R=23-25 (i.e. they are L>L* galaxies). These objects are preselected using the Lyman-break technique. The Keck groups have currently surveyed over 800 square-arcmins down to R~25.5.

Observing moderately high-luminosity (L>L*), z=3, Lyman-break galaxies (LBG) from HDF-N, Steidel et al. (1996) suggest some clustering - walls of galaxies - at high-redshift. To confirm this important result requires deeper, larger-area/larger-volume surveys. A major problem with the HDFs is that they sample very small sky areas (~10 square arcmins). Such small-area surveys are sensitive to clustering. Indeed, HDF-N and HDF-S show a difference in galaxy number density of ~15% (Gwyn, 1999, astro-ph/9907336). To study the spatial and angular correlation functions in detail requires larger area surveys. The proposed wide-field visible camera would enable a HDF-depth survey (down to m_{AB}~28 at S/N=10) over ~800 square arcmins in a month.

The high-redshift galaxies observed by Steidel et al. (1996) show multiple-structure with half-light radii of 0.2-0.3 arcsecs. The visible wide-field camera would resolve such sub-structure at all redshifts. The high-resolution camera would resolve <100pc at all redshifts, providing a powerful probe of galaxy structure, young AGN and interacting/merging galaxies. Lowenthal et al. (1997, ApJ, 481, 673) present further Keck spectroscopy of luminous LBGs in HDF-N and find similar results to Steidel et al. The galaxy nuclei are quite small, with most of the light within 3 arcsecs, and usually within 1 arcsec. Over 75% of the galaxies show multiple-clumps of emission, which are presumably intense star-forming regions. A diverse range of spectral features are observed, including weak or strong Lyman-alpha emission (rest-frame EW=0-60Å). The Lyman-alpha lines have FWHM = 100-230 km s^{-1} - thus only moderate spectral resolution (R~1000) is required to study their line profiles. The continuum near the Lyman-alpha lines imply SFRs similar to late-type spirals/HII galaxies today, whereas the Lyman-alpha luminosities imply much lower SFR, suggesting some internal reddening. The spectra also reveal strong Lyman-line breaks (i.e. a drop in the continuum flux shortward of Lyman-alpha due to Lyman-alpha absorption along the line of sight). This spectral feature provides an additional redshift indicator to the Lyman-continuum break.

For the brightest high-redshift galaxies studied with Keck, CIV, SiIV and NV absorption lines from their ISMs are detected, all of which have narrow profiles and low EW (<5Å) resembling stars in the SMC, suggestive of low metallicity. High-resolution spectra (R=5000 or so) would be required to study these absorption line profiles in detail, which in general is beyond the capability of either ground-based 8m telescopes or NGST.

Given the current detection limits, the highest redshift galaxies known are all high-luminosity objects (e.g. Weymann et al. 1998, ApJ, 505, L95; Dey et al. (1998, ApJ, 498, L93; Hu et al. 1999, ApJ, in press, astro-ph/9907079). The galaxy found by Dey et al. at z=5.34 has a very strong Lyman-alpha emission line (rest-frame EW = 95Å), which is asymmetric, with FWHM = 280 km s^{-1}. The Lyman-alpha flux is about 3.5×10^{-17} erg cm^{-2} s^{-1}. To produce such a flux would require about 10,000 O5 stars (producing 10^{43} erg s^{-1}). The continuum flux/shape suggests little reddening, so this could be a very young starburst. The host galaxy is larger than most with FWHM ~0.7 arcsecs, although this is in the I-band filter (which includes Lyman-alpha line emission).

Little work has yet been done on comparing IR and optical spectra of high-redshift galaxies, due to the relatively poor IR-flux limits from ground-based telescopes. Pettini et al. (1998, ApJ, 508, 539) show IR spectra of a few z=3 galaxies. Comparing IR and optical spectra they suggest internal reddening causes a reduction in rest-frame UV flux of 1-2 magnitudes, consistent with extinctions deduced from other observations. They also confirm redshift differences between Lyman-alpha, the nebular emission lines and the absorption lines of several hundred km s^{-1}, suggesting large-scale outflows are common.

1.2.1.3 Examples of recent imaging surveys for high-redshift galaxies.

Using narrow-band imaging and slit-spectroscopy (Keck/LRIS) Hu et al. (1996, Nature, 382, 231; 1998, ApJ, 502, L99) detect several high-redshift galaxy candidates. Their imaging Lyman-alpha detection limit is 1.5×10^{-16} erg cm^{-2} s^{-1} (10 sigma), whereas the slit-spectroscopy limit is 10^{-17} erg cm^{-2}

s^{-1} (10 sigma). At their imaging flux limit they find about 15,000 Lyman-alpha emitters per square-degree per unit redshift at z=3.4 and 4.5. For the fainter spectroscopy search they find about 60,000 per square-degree per unit redshift in the range z=3-5.6 (based on an area of about 200 square arcsecs). Thommes warns that the Hu et al. survey may be biased due to clustering over the small survey areas. As noted above NGST will overcome this problem by providing much larger area surveys. Many of these galaxies are too faint for LBG detection by HST but would all be detectable by the proposed NGST visible camera. Although in principle it is harder to convert observed Lyman-alpha flux into the SFR than for H-alpha, these observations suggest rest-frame Lyman-alpha line EWs in the 100-200Å range, close to the maximum predicted from star-forming galaxy models. These data therefore imply little reddening and support the concept of no drop in the SFR at high redshifts. The lack of dust implies that an NGST visible camera would have many possible targets with detectable Lyman-alpha emission at high redshifts.

Thommes (1998) reviews the Calar Alto Deep Imaging Survey (CADIS), which uses an imaging Fabry-Perot combined with broad-band and medium-band photometry. Their 10-sigma limit is $6x10^{-17}$ erg cm^{-2} s^{-1}, similar to the Lyman-alpha flux observed by Dey et al. (1998). The CADIS survey also finds many extremely red objects (R-K >6), which may be galaxies at z=0-2. If so this would imply massive galaxy formation occurs at late epochs. Alternatively they are old ellipticals or mergers/AGN. To detect such galaxies at higher redshift requires NGST as you need to go deeper than can be obtained with HST.

For comparison with the deep optical surveys, Teplitz, Malkan and McLean (1998, ApJ, 506, 519) show data from IR imaging around known QSOs. They obtained narrow-band Keck imaging (R~100) over 2.1-2.3 microns plus broad-band K'. The line detection limit is 2.3 10^{-16} erg cm^{-2} s^{-1} in the narrow-bands and K' ~20.5 (both 10-sigma). At high redshifts the most likely line in their observed wavelength range is H-alpha. Overall, they find several candidate high-redshift galaxies with implied SFR comparable to the optical surveys.

1.2.1.4 Surface density implications for the NGST visible camera

Based on the current surveys we can make some estimate of the likely surface density of targets available to an NGST visible camera. The HDFs give a surface density of about 500,000 galaxies per square degree down to I=29 (from WFPC2; see Figure 1.2.1-c) and double that at I=30 (based on the small STIS field in HDF-S). Thus a reasonable lower-limit is for 1,000,000 objects per square-degree in the visible at m_{AB}~31. Assuming a FWHM=0.2-0.3 arcsecs for high-redshift galaxies, the total sky coverage is likely to be 5-10% (the lower end of that range is consistent with the coverage deduced from the HDF-S STIS images). An NGST visible camera with a 5x5 arcmin field of view could therefore detect in excess of 7000 galaxies per deep-field. Several thousand of these would be bright enough to provide prism spectra or multiband imaging data.

Of course only a fraction of the objects seen in such deep, wide-field images would actually be high-redshift galaxies. Using the Hu et al. (1996, 1998) surveys, and assuming a redshift range for the Lyman-break technique with the NGST visible instrument of 5-10, gives ~100 high-redshift (z>=5) objects per square arcmin. [N.B. Most of these galaxies will have resolved sub-structure, so there will actually be several hundred high-redshift emission clumps per square-arcmin]. Given this distribution, obtaining moderate-resolution (R~1200) spectra of all of them would be practical only if a wide-field field-censoring device, such as a micro-mirror-array, becomes available. The proposed visible IFS will combine reasonable spatial (0.1 arcsec pixels) and spectral (R=1200 resolution to provide integral-field spectroscopy over 100 square-arcsecs on spatial scales unavailable to ground-based telescopes with AO systems. The IFS will have many hundreds of potential targets per NGST deep-field so target pre-selection would be the normal mode of operation. However, even random pointing (perhaps in parallel observing mode) would provide ~2 potential high-redshift targets (probably with sub-structure) and ~5 other targets per IFS field-of-view.

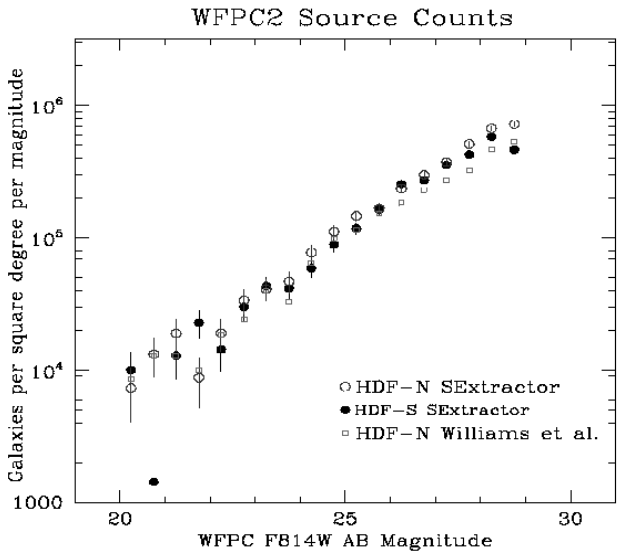


Figure 1.2.1-c: The observed surface-densities derived from HDF-N and HDF-S

The wide range of galaxy redshifts encompassed within any deep or moderately deep imaging by NGST means that pre-selection of galaxies by redshift for targeted spectroscopic observation by NGST optical or IR spectrographs will be essential. Based on detailed modelling of both early- and late-type galaxies, it is clear that having only IR colours will provide insufficient discrimination. While J-H:H-K should distinguish the high redshift 4<z<10 galaxies quite effectively, V-R:R-I will be needed to make any pre-selection of objects in the range 1<z<3.5. This redshift range is thought to include the most active period of Star-Formation in the history of the universe and currently only the brightest galaxies in this range have been probed by ground-based observations. For many of the 1<z<3.5 galaxies too faint for NGST spectroscopy at NIR wavelengths, the optical colours will provide the *only* information on their redshift and star-formation rate. These 1<z<3 galaxies are expected to constitute the vast majority of galaxies detected at faint optical and NIR limits by NGST. In Figure 1.2.1-d, we show the H-K versus J-H colour. Late-types at z=4,5,6 are clearly discriminated, but galaxies at 1<z<3.5 have very similar colours. At IR magnitudes of 27<K<27.5 the redshift distribution peaks at z around 1 with most galaxies expected to be late type. Using I-H versus H-K improves the situation slightly, but it is only by going down to the V-band that a significant improvement occurs (Figure 1.2.1-e). The V-R versus R-I colours allow discrimination at all redshifts for both early and late-type galaxies.

In addition to looking at the highest-redshift galaxies, the IFS could also be used for a spectroscopic survey to test hierarchical models, in which low redshift galaxies are formed by mergers of smaller structures which formed much earlier. The small blue galaxies seen in the HDFs are likely to have narrow emission-line spectra similar to those of low redshift, low metallicity, gas-rich dwarf galaxies. Due to their low metallicity, many of the strongest cooling lines would appear at rest-ultraviolet wavelengths such as [OII] 2470Å, OIII] 1663Å and 2321Å, NIII] 1750Å, CIII] 1909Å and CIV 1549Å. These lines will be redshifted into the visible range for z = 1-5 (i.e. the redshift range predicted for the faintest objects in the HDFs). These optical spectra, when combined with NGST IR spectra, would yield abundances for relatively unprocessed material at intermediate and high redshifts, and hence yield information on how the first heavy elements were produced in the Universe.

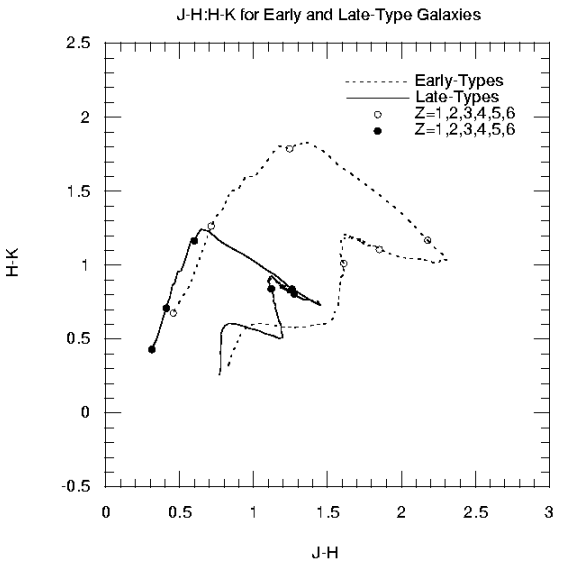


Figure 1.2.1-d: The predicted H-K versus J-H colour for galaxies as a function of redshift, with z=0 as the unmarked start of the track, and the circles correspond to z=1,2,3,4,5,6 in sequence.

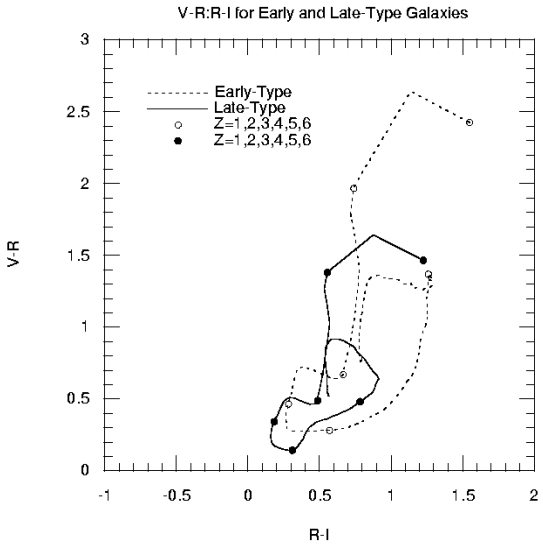


Figure 1.2.1-e: The predicted V-R versus R-I colour for galaxies as a function of redshift

1.2.2 Determining the age of the Universe: White Dwarfs, Cepheids and Supernovae

The age of the Universe is a critical parameter in choosing between possible cosmological models, but it is still uncertain and the subject of much debate. The expansion age derived from the Hubble constant with dependence on the density parameter (omega) and the cosmological constant (lambda), does not agree well with ages derived via stellar evolution models applied to globular clusters. One can force consistency by arbitrarily adjusting the values of omega and lambda. To constrain these values we must know the Hubble constant and Universe age to within about 10% accuracy. The value of H_0 is now converging quite well, but as described above the age parameter is still quite uncertain. An NGST optical camera would make a vital contribution to distance scale determinations, and hence constrain models of the Universe, by detecting white dwarfs, Cepheid variables and supernovae at distances far beyond those accessible to ground- based 8m telescopes or the HST.

A promising new approach is to directly measure the age of evolved stellar associations by using the white dwarf cooling sequence, which is free from the usual set of assumptions. The advantages of this technique are numerous: no metallicity dependence, abundant local calibrators with well determined trigonometric parallaxes, plus a tight (0.03 mag. width) and physically well understood cooling sequence. The only model-dependent quantity is the core mass-luminosity relation, which is the most robust of the theoretical predictions. In practice, then, the uncertainty in age from the cooling sequence should correspond to about a 5% uncertainty in age, which is within the error budget of the other parameters of the problem. These ages can then be compared with the theoretical evolutionary calculations. The lower ages inferred from the recent Hipparcos calibration of subdwarfs would require substantial modification of the basic models, but this remains an area of continuing controversy (see Pont et al, 1998, A&A, 329, 87). A completely independent approach such as using the cooling sequence, would provide crucial input to settle this debate, as well as its obvious relevance to the cosmological age issue.

Despite the clearly promising potential of this method, there are several challenging requirements for its successful application. In order to establish precisely the location and consistency of a cluster WD sequence, at least several hundred bona fide WDs spanning a range of at least 3-4 magnitudes, need to

be reliably identified and properly located in the colour/magnitude diagram. This translates to stringent observing configurations and sensitivities which are summarized below;

(i) Need to be able to detect WDs with accuracy about 0.1 magnitude down to V=30, corresponding to a cooling time of about 15 Gyrs (Hansen, 1999, ApJ. 520, 680), for nearby clusters; (ii) Need to be able to observe individual stars in the dense cores of globular clusters, where we expect the highest density of WDs, and the lowest density of contaminating main sequence red dwarfs; (iii) At the detection limit, need to be able to efficiently discriminate between the cluster WDs and cluster quiescent CVs, soft X-ray transients, and distant field galaxies.

To achieve the above goals we require sufficient field of view that the number density of faint WDs is comfortably larger than the background at V=30 cf. just as the normal main sequence stands out against the background of non-cluster members. We can also use multi-band colours to select them. Contrary to expectations old WDs become bluer as they cool because their stellar atmospheres deviate very significantly from black bodies, due to continuous absorption in the near IR by molecular hydrogen (see Figure 1.2.2-a from Harris et al. 1999, ApJ., in press, astro-ph/9906065). This means that a combination of visible and near-IR colours are necessary to discover them. High spatial resolution images are needed to reject distant background galaxies, which might coincidentally have similar colours.

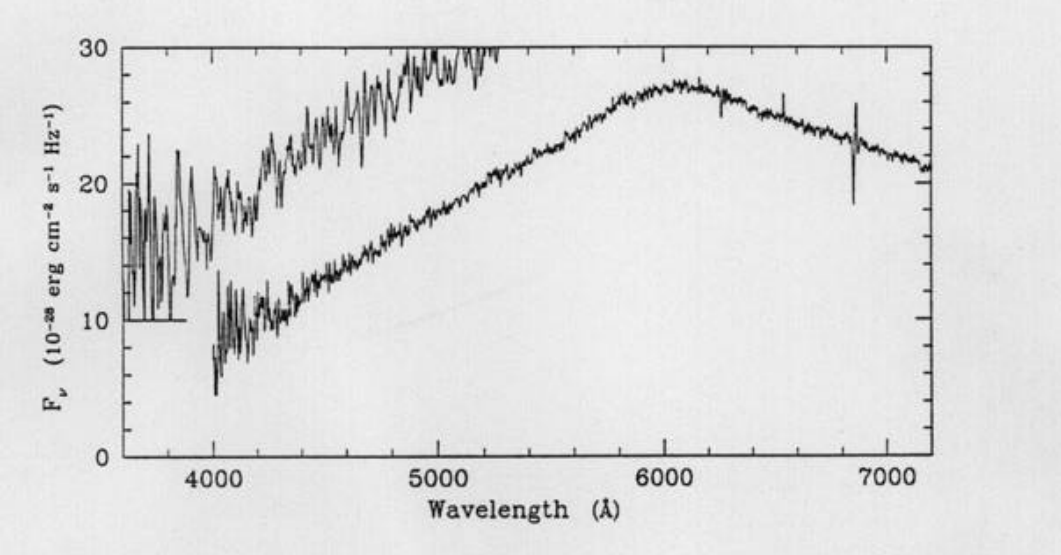


Figure 1.2.2-a: Spectrum of the cool white dwarf LHS 3250 taken with Keck II (lower) and the Steward 2.3m (higher - displaced for clarity)

Finally, images obtained several years apart can be used to separate cluster members on the basis of their proper motions. The combination of these requirements can only be obtained using a visible camera on the NGST.

Traditionally, Cepheids have been regarded as the most reliable distance indicators. Since they only display large amplitude variability in the visible bands, and show virtually no variability in the near-IR (JHK), their detection in distant galaxies depends crucially on the visible rather than IR sensitivity. Predicting the exact capability of the visible instrument is difficult at present primarily due to the unknown PSF of the telescope. The point-source detection limit is sensitive to the wavelength at which the telescope is diffraction limited. Nevertheless, if the telescope maintains a good quality optical PSF (e.g. has a good strehl ratio), it may be possible to detect Cepheids at the distance of the Coma cluster in similar exposure times and at similar S/N as the HST detects Virgo cluster Cepheids.

Since the Coma galaxies are 6 times further away, but the spatial resolution is the same factor (or higher) than for WFPC2, crowding should be no worse of a problem at Coma than at Virgo. Single epoch I band imaging is also required to obtain the Cepheid reddening. A field-of-view of 5 arcmins diameter would be well suited.

The HST Key Project (Freedman et al. 1994, ApJ, 427, 628) has provided Cepheid distances to about 20 spiral galaxies out to Virgo. These distance determinations show that those distances obtained using the Tully-Fisher method are about 20% too small, with a corresponding effect on H_0 estimates (Giovanelli et al., 1997, ApJ, 477, L1; Shanks, 1997, MNRAS, 290, L77). NGST data would quantify the effect of the Malmquist bias on Tully-Fisher distances, and would also provide an important check on other distance indicators, such as the Type Ia supernovae in spiral galaxies. The high accuracy of Cepheid distances will also permit new measurements of bulk motions in the Universe, leading to improved estimates of the mass density and large- scale mass distribution in the Universe.

NGST will discover substantial numbers of supernovae, particularly if regular 'monitoring' of galaxy samples occurs. The high-redshift supernovae rate constrains star-formation rate, galaxy evolution and Cosmology. The exact light-curve depends on supernovae type and metallicity. Dahlen & Fransson (1999, A&A, in press, astro-ph/9905201) recently recalculated the predicted rates for high-redshift supernovae and concluded that previous estimates were underestimates. The highest number-counts will be in near-IR - due to the strong UV cutoff at 4000Å (rest-frame), with ~35 in every 16 square-arcmin field-of-view at J (assuming a 10,000 second exposure, S/N = 10). Most of these will be around z=1, with some up to z=4. However, although the visible number-counts (~10 per 25 square-arcmin field-of-view in the same time) will be lower, as for galaxies, optical data will be required to provide redshift and supernovae type information. For most of the supernovae the clearest spectral feature - the 4000Å cutoff - will be in the visible. The combination of NGST optical photometric data and IR spectra of supernovae Type Ia at high redshifts will therefore provide crucial data for determination of galaxy distances beyond those possible using Cepheid studies.

NGST will also easily detect Type II supernovae in dwarf galaxies, globular clusters, or giant galaxies, providing independent estimates of the global star-formation rate without the usual bias towards 'big' galaxies. For this purpose redshift estimates based on combined optical and IR photometry would be sufficient.

1.2.3 Imaging and Spectroscopy of clusters of galaxies

Clusters of galaxies are the largest bound structures in the Universe. Surveys at X-ray energies are a proven and efficient means to select galaxy clusters, because at these energies the flux is proportional to the square of the electron density. Currently X-ray surveys can detect clusters out to redshifts of about 1 (eg. Stanford et al, 1997, AJ, 114, 2232, and future observations using Chandra and XMM are predicted to detect clusters out to redshifts 2-3. Studies of distant clusters will place strong constraints on cold dark matter models, which predict that such large structures form at a relatively late epoch. Having located high redshift cluster candidates, to progress further we must distinguish between cluster members and field galaxies. To do this we will need to employ a photometric redshift determination. As has already been noted in order to obtain reliable photometric redshifts it is essential to combine colours at visible and near infrared wavelengths. Using this technique from the ground, and using WFPC2 on the HST it has been possible to measure photometric redshifts down to R=24 (Brunner, Connolly and Szalay, 1999, ApJ., 516, 563). When these photometric redshifts are compared with Keck spectra, the redshift accuracy is shown to be better than 0.05, which for high redshift clusters is adequate for most purposes.

The NGST visible camera can be used to obtain photometric redshifts of distant cluster members, and by removing the field galaxy contamination, will permit morphology-density relations to be examined. The highest spatial resolution will be needed to determine the morphological types of the cluster members. Deep observations using visible band colours are essential to avoid discrimination against

early-type galaxies, which would skew the galaxy-type versus cluster location distribution (Lubin and Brunner, 1999, astro-ph/9907404).

Rich galaxy clusters can act as gravitational lenses, producing distorted and magnified images of background galaxies. Careful analysis of the distorted images, including their image separation and position within the cluster potential, is needed to fix the parameters of the lens model and so compare the dark matter and stellar distributions. In particular high spatial accuracy is essential to determine H_0 using gravitational lens techniques. This will require use of the high resolution camera on NGST. Furthermore, by using integral field spectroscopy blindly centred on the optical or X-ray emission centroid, we can obtain a complete spectroscopic inventory of the lensed galaxies out to a significantly greater redshift limit than is possible for field galaxies. A mosaic of IFS observations across the cluster, combined with near IR data, will yield spectral diagnostics which can be used to measure the SFR in high redshift young clusters.

1.2.4 Resolving stellar populations in nearby galaxies

The broad-band colours and spectra of galaxies have traditionally been used to infer the properties of the stellar populations present. A serious problem is that such observations are of the integrated light from large numbers of spatially unresolved stars, which must be deconvolved in some way using evolutionary models.

Observations of resolved stellar populations adds new dimensions to these studies, in terms of estimating the star-formation history, chemical composition hence enrichment history, initial mass function, and when combined with spectroscopy of system, its dynamical history. In principle resolved stellar populations studies of nearby galaxies could be used to test evolutionary theories such as the faded disc model, by identifying the fossil record of past star formation. However there are complications, for example the problem of distinguishing between the effects of age and metallicity. Modelling by Brocato et al. (1999, A&ASS, in press, astro-ph/9812323) show that the colour index V-K appears quite sensitive to the age once the metallicity has been constrained using visible-IR colours (Figure 1.2.4-a).

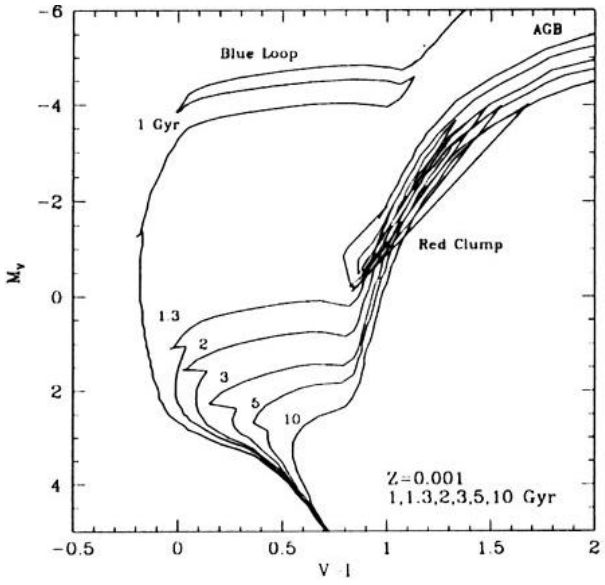


Figure 1.2.4-a: Isochrones for a single metallicity (Z=0.001) and a range of ages (marked in Gyr at the MSTO). This plot illustrates the age sensitivity, although to interpret real cluster data a population synthesis model would be required. (Tolstoy, 1999, astro-phys/9901252)

By using giant stars it should be possible to measure colour indexes of sufficient accuracy to determine metallicities in galaxies out to the distance of the Virgo cluster. The visible band colours are particularly important when dealing with stars hotter than about 8000K, since they have similar colours in the near infrared because at these wavelengths their atmospheres approximate to the Rayleigh-Jeans tail of the distribution.

Unambiguous age information can be obtained from the HR diagram, using the Main Sequence turn-offs (MSTOs), provided the observations sample faint enough down the HR diagram. It should be possible to measure the MSTOs in galaxies out to about 2 Mpc. This information can be used to distinguish between bursting and continuous star formation.

To undertake this programme would require observations at visible wavelengths, using the highest possible spatial resolution in order to overcome the stellar crowding problem. The proposed visible high-resolution camera would be very well suited to this task.

1.2.5 Microlensing in the Virgo Cluster and Dark Matter

The search for microlensing events towards neighbouring galaxies allows us to probe the distribution of compact objects - MACHOs (MAssive Compact Halo Objects) along these lines of sight. As was shown by Gould (1996, ApJ, 470, 201), microlensing in the Virgo cluster towards M87 and other large galaxies, potentially provides a sensitive test of whether the cluster dark matter is comprised of MACHOs. Gould considered the observing capabilities of HST, and concluded that such a project was marginally feasible by using 'pixel' microlensing, i.e. searching for brightening of unresolved stars using image subtraction techniques. However, a slight adjustment in the assumptions would put the project beyond the reach of HST. NGST will be vastly superior to HST for this project, since it gains a factor of 10 from the aperture and another factor of several from the improved diffraction limit assuming a conservative PSF (if the NGST is diffraction limited at 1 micron then the overall sensitivity advantage over HST will be a factor 100). In the visible the project is best carried out at the longer optical wavelengths (e.g. I-band), since giant type stars will dominate the signal.

One of the DRM proposals (submitted by H.-W. Rix) outlined such an observing programme using NGST NIR/CAM. He concluded that 5000f microlensing events should be detectable in each of 10 galaxies in 5 days total observing time, where the f-factor is the cluster MACHO fraction. This would very accurately measure the MACHO fraction; however it would not provide much information on their mass function, of fundamental importance in determining what the MACHOs are. This is because in 'pixel' microlensing (i.e. microlensing of unresolved stars), there is a degeneracy between the (unknown) baseline flux of the lensed star and the 'intrinsic' Einstein ring crossing time t_c, (which provides the information about the lens mass). The actual observed event duration is proportional to the product of t_c and the fraction of the flux from the lensed star within the resolution element. The degeneracy of this arithmetic product cannot be broken, and hence only the total lensing optical depth is measurable.

This degeneracy problem does not occur if the lensed star dominates the flux within a spatial resolution element, (see Melchior et al. 1999, A&A Suppl., 134, 377). For M87, the outer regions approach this regime since the characteristic 'fluctuation magnitude' in the I-band is about m_{AB}=29.8 at 2 arcmins from the centre of M87, and the surface brightness at this wavelength is about m_{AB}=28.5 per NGST spatial resolution element. Thus, a typical lensed star will contribute a substantial fraction of the flux. Clearly, sampling with several pixels per resolution element is optimal for this project, which would require a high resolution camera with pixels about 10mas. Although the overall event detection rate would be lower than for observations using the NIR/CAM or for the proposed visible wide-field camera due to the lower surface brightness and smaller field of view; the additional crucial information gained on the lens masses would be unique to the use of a high resolution camera.

The optimal strategy would be to undertake an initial survey using an NGST wide-field camera, to map the lensing optical depth as a function of projected position on the galaxy image. Follow-up observations of selected regions using a higher spatial resolution would then be obtained. Allowing for the likely field-of-view for the high-resolution imager, several hundred events should be detectable in any image, providing a large sample of event timescales and thus a reliable measure of the MACHO mass function.

1.2.6 Extragalactic Planetary Nebulae in, and Between, Galaxies

Theuns & Warren (1997, MNRAS, 284, L11) identified a population of intergalactic stars in the Fornax cluster of galaxies, detecting 10 planetary nebulae (PNe) in an imaging survey that used a narrow-band [OIII] 5007Å filter and adjacent broad continuum filters. The number detected was consistent with approximately 40% of all the stars in the Fornax cluster being intergalactic, arising from tidal encounters between cluster galaxies. Planetary nebulae emit up to 10% of their total luminosity in the 5007Å line, i.e. up to 500 L(sun) in a line of 30 km s^{-1} width. Since PNe at these distances will be spatially unresolved, the NGST's high-resolution optical camera is the ideal instrument with which to detect them against cluster backgrounds. With 10mas pixels and assuming a PSF peak spread over several pixels, on-line and off-line images will enable PNe to be isolated inside galaxies and within clusters of galaxies, and to be easily distinguished from contaminating sources such as dwarf emission-line galaxies, which will be well-resolved. An [OIII] filter exposure time of 10,000 seconds would enable a typical luminous PNe to be detected with S/N of 10 at a distance of 60 Mpc, i.e. to much larger distances than the Virgo and Fornax clusters. Follow-up spectroscopy at R~1000 could be made of the selected PNe candidates, providing emission-line radial velocities to an accuracy of about 25 km s^{-1}. This would allow cluster mass potentials to be accurately characterized, as well as nebular physical parameters such as temperatures, densities and abundances, thereby enabling a determination of the metallicity distribution of the intergalactic stellar population within clusters. Such a programme would yield unprecedented information on the overall stellar population of galaxy clusters, and provide a powerful probe of cluster mass potentials. Finally, fitting the modelled luminosity-function to the detected PNe populations would allow the determination of distances that are well into the Hubble flow.

1.2.7 Observations of 'Cool' Neutron Stars

The understanding of the surface properties of isolated neutron stars (INS) is severely hampered by their faintness, particularly at optical wavelengths where the most useful diagnostics are available. Very young objects, such as the Crab, PSR0540-69 or PSR0833-45 are characterised by flat synchrotron-like spectra from energetic electron interactions in their magnetosphere (Caraveo, 1998, Adv. In Sp. Res. Vol 21 in press, Mignani, Caraveo & Bignami 1998, A&A, 332, L37). However, for at least three INS, PSR0656+14, Geminga and PSR1055-52, non-thermal magnetospheric emission may have faded sufficiently in the optical waveband to render visible the thermal emission from the neutron star surface (Mignani et al 1998. Adv. In Sp. Res. Vol 21).

Recent measurements of Geminga at the sensitivity limit of the Keck telescope (Geminga has a V magnitude of 25.5) by Martin, Halpern and Schiminovich (1998), ApJ, 494, L211, appear to confirm the suggestions by Bignami et al. (1996), ApJ, 456, L111, that the optical spectrum may include a cyclotron line in emission, from protons in a field of 10^{11}G. Broadband HST and ground-based fluxes have been summarised and modelled in Jacchia et al. (1999, A&A, 347, 494). The spectral sampling of those measurements is 60Å, and represents the limit of what is achievable using 8-10m telescopes from the ground. The flux of Geminga is only 0.5% of the dark sky at Mauna Kea. Observationally, the only route forward to investigate directly the physical characteristics of material under such extreme conditions will be provided by an optical low resolution spectroscopic capability on NGST.

These observations would provide detailed diagnostics for the surface characteristics of isolated neutron stars, and permit direct measurement of the magnetic field. Furthermore, the possible detection of proton cyclotron radiation in Geminga opens up new possibilities to study the astrophysics of matter under extreme conditions.

1.2.8 Optical surveys for Trans Neptunian or Kuiper Belt objects

NGST can play a very important role in furthering our understanding of the Kuiper Belt population of the Solar System (or the Trans-Neptunian Objects) out to large distances from the Sun. Although since formation TNOs have interacted with Neptune, changing their distribution and have undergone collisions etc., their distribution, size, mass and composition constrain Solar System formation models. Some 100 TNOs are known with diameters 50-400 km, orbital semi-major axes 40-50 AU, and orbital inclinations 0-30 degrees. It is unclear why none have been found at > 50 AU (e.g., Jewitt et al., 1998, AJ, 115, 2125).

Imaging in the V, R and I-bands is the most efficient way of detecting TNOs. HST observations to magnitude limits of ~29, provide tantalising clues to the space density of KBOs. For NGST, in the ecliptic plane the prediction is for ~800 TNOs per 5x5 arcmin field-of-view at m_{AB} = 30.5 (R-band). The wide-field visible camera is capable of detect a 5km TNO at 50 AU, or a 100km TNO at 200 AU. In the deepest images a 1-2km TNO could be detected at 50 AU. Combining optical (reflected) and IR (emitted) flux would constrain the albedo, which when combined with the distance gives the size. The distribution of TNOs around the ecliptic would give the resonance structure while the variation with latitude gives the inclination distribution. The 3-dimensional structure of the Kuiper Belt so derived could then be compared with those of the dust disk systems seen around nearby stars and with dynamical simulations of the evolution of Solar-System disk.

Spectral information about TNOs obtained via multi-colour imaging would also be important. Multi-colour photometry/spectroscopy constrain surface properties, formation history and relation to comets/planets. Jewitt & Luu (1998, AJ, 115, 1667) note that the V-J colour is the best suited to distinguishing between different spectral types among the TNO population (Figure 1.2.8-a). IR colours alone provide a poor discriminant.

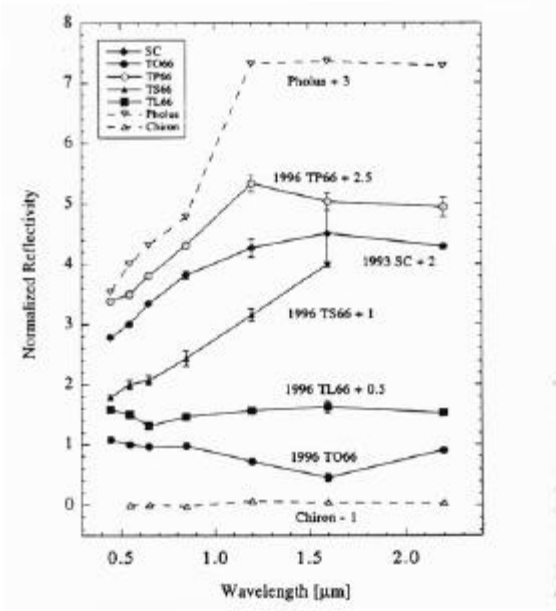


Figure 1.2.8-a: Normalized reflectivity versus wavelength for five TNOs (Jewitt & Luu, 1998)

1.2.9 Comparison of an NGST visible instrument with ground-based and HST

The spectacular results obtained from faint-object optical imaging and spectroscopy by HST has clearly demonstrated the advantages of space-borne relative to ground-based optical telescopes, even in the era of the new 8m-class ground-based facilities. However, by 2010 all of HST's optical instruments (ACS, STIS and WFC3) will be 7-10 years old. Hence, access to state-of-the-art optical capability in space cannot be assured via HST. Provision of an optical camera on NGST using efficient detectors will allow for the HST legacy to be continued and greatly extended for relatively low cost.

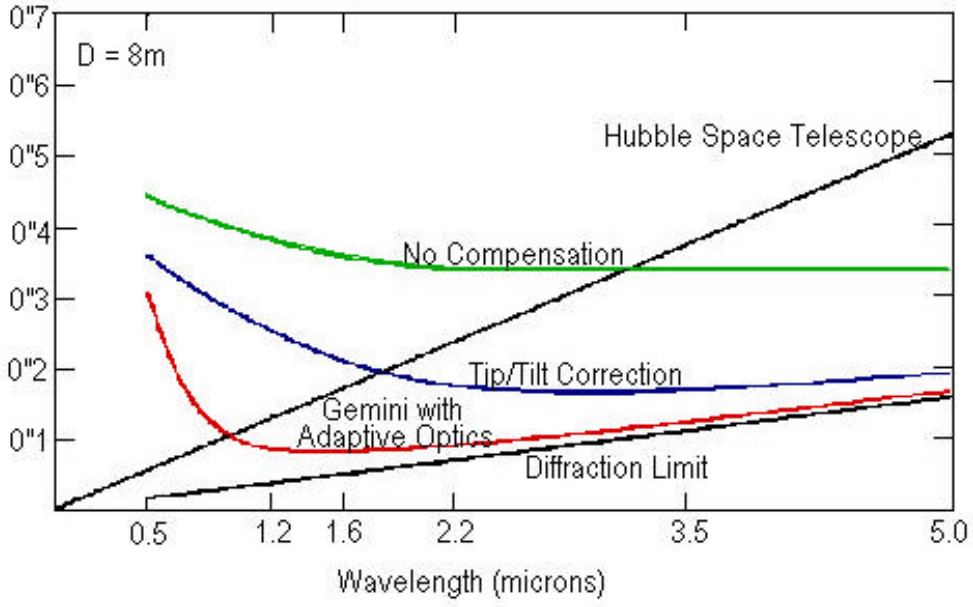


Figure 1.2.9-a: The state-of-the-art image quality available from a ground-based 8m telescope (Gemini) using either tip-tilt or AO correction.
The image quality available from HST and from a diffraction limited 8m telescope are also shown. The potential of NGST to provide superior image quality in the visible is clearly shown

The provision of adaptive optics (AO) capability on ground-based, 8m telescopes will provide some overlap with current HST optical imaging performance for 'bright' objects (approximately V<28). However, these AO systems are designed primarily for the IR, and will provide only a small fully-corrected field of view (~10 arcsecs) at 0.5 microns with FWHM ~0.1-0.2 arcsecs from the V to I bands (i.e., worse than available from NGST even if NGST is diffraction limited at 2 microns). Over wide-fields (several arcmins), the FWHM at 0.5 microns for ground-based 8m telescopes on good sites (e.g. Hawaii and Chile) will more typically be 0.4-0.7 arcsecs, an order of magnitude larger than NGST will provide.

Even on these best sites, the sky background is higher for a ground-based telescope - by a factor 4 at 0.5 microns and >10 at 0.8 microns relative to HST. If NGST is placed at L2, it will have a similar sky background to HST (i.e. dominated by zodiacal light - about 23 magnitudes/square-arcsec at 0.5 microns) and should have a point-spread function with a superior strehl ratio compared to ground-based telescopes even with AO. Crucially, it will also have a substantially larger field of view over which diffraction-limited imaging is available. The exact performance improvement factors for NGST relative to ground-based telescopes or HST depend on the assumed sky background levels, detector noise characteristics, image quality (telescope tracking, strehl ratio etc.) and detector field of view. The capabilities of the proposed visible instrument are discussed in detail in subsequent sections. Here we provide a brief comparison to illustrate the potential gains:

The proposed visible wide-field imager using an 8kx8k pixel array with spatial scale around 0.04 arcsec/pixel would provide a field-of-view of 5x5 arcmins, giving a performance improvement factor (including the field of view factor) for point-sources of several thousand relative to ground-based 8m telescopes. The NGST camera would have twice the sky area and a slightly better spatial resolution than the forthcoming HST ACS-WFC, coupled with the much larger effective area, improvement in detector efficiency and camera design.

The proposed high-resolution imager, even if diffraction limited at 2 microns, would offer further advantages over HST and ground-based facilities. A 8kx8k pixel array with 0.01 arcsec pixels would provide a field of view of 80x80 arcsecs, giving a performance improvement factor of a thousand or more relative to ground-based 8m telescopes. Such a camera would provide ten times the sky area and twice the spatial resolution of the HST ACS-HRC, again with higher efficiency and telescope area. Within the small AO optical field of view of ground-based 8m telescopes, the NGST performance advantage falls, but the field of view advantage and stable PSF would still be a substantial factor.

The proposed visible integral-field spectrograph, although providing a modest field of view, would also provide a gain over other facilities, while being relatively simple in design, mass, power and operational requirements. The visible IFS would have smaller pixels than available from ground-based AO coupled with a stable PSF, and provides a more sensitivity spectroscopy mode over a wider field of view than available with HST/STIS.

1.3 INSTRUMENT REQUIREMENTS FROM SPACE

Having outlined in the previous section various programmes for which observations at visible wavelengths are desirable, or essential, we can now define the instrumental requirements.

For efficient cosmological surveys to detect high redshift galaxies a wide field of view is necessary. We have noted the need for visible band observations in order to distinguish the numerous medium redshift galaxies from those at high redshift, which will be the focus of so much of the NGST programme. To provide complementarity with survey projects carried out using the near infrared camera, the visible camera should have a similar field of view and pixel scale. One might of course satisfy these requirements by provision of a single wide field instrument. However, a crucial point relevant to this option is the issue of red-leak in the combined visible filter and detector response. Whereas in terms of the infrared background, a specification of red leak less than a few 10ths of a percent will be acceptable, this is not the critical issue. In many cases the energy distributions of high redshift galaxies will be very red, but we will nevertheless require observations at visible wavelengths in order to pre-select candidates, and to study the evolution of their stellar populations. It is therefore the red leak of flux from the object itself that will dominate the uncertainty in the measured optical colours. This effect will result in systematic errors in these colours, which will seriously compromise our ability to make full scientific use of the information obtained at visible wavelengths. Obviously, the use of an instrument with a detector response which has an intrinsic cut-off in the near infrared would solve this problem.

The wide field visible camera should be provided with a set of broad band filters for use in the deepest surveys which require the maximum sensitivity. In addition to this we identify the need for intermediate band filters which may be used to better define the redshifts of galaxies for which the redshift discriminant lies in the visible range. An efficient means to achieve this is by use of a comb filter, which can provide simultaneous imaging in several intermediate bands. This multiplex advantage is gained at the expense of sensitivity. For shallower wide field surveys at higher spectral resolution than can be provided by filters, we propose to use a prism with a resolution of about 100 around the mid-wavelength of its coverage.

For programmes involving follow-up observations of selected targets the most attractive option is to employ some form of field censoring in order to redistribute the use of detector pixels from spatial to spectral coverage. Micro Mirror Arrays (MMAs) are the obvious choice for multi-object spectroscopy (MOS) over a wide field of view. We fully appreciate the dramatic potential of MMAs for background suppressed multi-object spectroscopy over wide fields of view. However, in the context of our study of a separate visible wide field camera, we decided not to duplicate the very considerable effort already devoted in this area in relation to the near infrared instruments (e.g. MacKenty et al, NGST MOS study team). A visible camera must conform to strict cost and complexity constraints, and we felt that a design based on MMAs would not be compatible with these. However, our proposed design of the wide-field camera may in future form the basis for a modification incorporating use of an MMA.

One of the most compelling reasons for inclusion of a visible camera is to fully exploit the best image quality provided by the NGST. Although we will not know the characteristics of the PSF until detailed mirror designs are available, we can make some reasonable assumptions at this stage. Even if the optics are at best only diffraction limited at 2 microns, the PSF at optical wavelengths will still have Strehl ratios of from 0.3-0.5. We therefore propose a high resolution visible camera with a pixel scale that will Nyquist sample the core PSF of an image at around 0.7 micron, implying 10mas per pixel. For an assumed 8k x 8k detector array this provides a field of view of 80 x 80 arcsec. Large fields would of course be preferred, but drive up the size and cost of the camera. The high resolution camera would have broad band filters, and would be uniquely suited to programmes such as stellar population studies, Cepheid projects, micro-lensing, the investigation of small scale sub-structures in distant galaxies, and revealing the presence of young active galactic nuclei in forming galaxies.

Finally, we consider the need for a medium resolution spectroscopic facility at visible wavelengths. This would ideally be provided using MMAs, but for the reasons given above we did not encompass this option in our study. Instead we looked at a design for an integral field spectrograph, giving a resolution of around 1000. Using a microlens array we can fully sample a field 10 x 10 arcsec. with pixels of 0.1 arcsecs. This would be used in a wide range of projects, for example to determine the physical conditions of the ionized gas, as well as giving kinematic information on the stars and gas with a velocity resolution of better than 100 km s^{-1}. To cover most of the visible band, including the H-alpha and the Calcium triplet at low redshift, requires several grisms.

Such an instrument would be a powerful facility, even when compared to AO integral field spectrographs on ground-based 8m telescopes. For example, using the published sensitivity estimates for the Gemini/GMOS instruments, we can obtain S/N=10, R~1000 spectra of faint, small (~0.2 arcsec) galaxies about 5 times faster with the proposed NGST, IFS, albeit over a smaller wavelength range per exposure. The NGST IFS also has twice the field of view and pixels half the diameter of GMOS.

2. INSTRUMENT ENGINEERING

2.1 DESIGN CONCEPT

2.1.1 Instrument requirements

Three different types of visible wavelength instruments are desired from a scientific standpoint:
- a Wide Field Imager,
- a High Resolution Imager,
- an Integral Field Spectrograph.

The main objective of the Wide Field Imager is to image as wide a field of view as possible, which is not possible from ground-based observatories. The sampling is desired to be small but is matched to the possible size of the focal plane array. Spectroscopic capabilities are required for survey programmes, and in particular to detect candidate primordial galaxies.

The High Resolution Imager is designed to exploit the best spatial resolution available at visible wavelengths. The aim is to provide the finest sampling of the visible PSF consistent with the Shannon principle. This instrument will be close to the diffraction limit of the PSF core in the visible. It requires that the line of sight jitter during an observation period be compatible with this precision.

The Integral Field Spectrograph is an alternative to the Wide Field Spectrograph if field censoring is not available. Instead of selection of objects, it provides a systematic spectral analysis of all objects over a small field.

The instrument performance requirements are summarised in Table 2.1.1-a.

Instrument	Performance requirements
Wide Field Imager	field : 5 arcmin square, sampling << 100 mas, objective 37 mas spectroscopic capability with moderate spectral resolution and with multiband, 8 band pass filters + 1 prism + 1 prism with multiband filter.
High Resolution Imager	sampling : 10 mas field : 30 x 30 arcsec² to 1 x 1 arcmin² 6 band pass filters no spectroscopic capability
Integral Field Spectrograph	Field: 10 x 10 arcsec² field sampling: TIGER integral field spectrograph sampling: 100 mas spectral resolution : R = 100 to 1200

Table 2.1.1-a: Performance requirement of the desired four instruments

2.1.2 Optical concept

The optical designs are based on a f/24 telescope (yardstick design) without field curvature. The telescope exit pupil, used as entrance pupil for the instruments, is located at a distance of 2 m from the telescope focal plane towards the telescope, so the entrance beam has diverging chief rays. The impact of telescope field curvature has been analysed. No coupled analysis or optimisation of the instrument with the telescope has been performed; only the best image quality for each standalone instrument was aimed at. The focal plane pixel size is 15 µm (see section 2.1.3.2.1).

2.1.2.1 Wide Field Imager

The Wide Field Imager (WFI) instrument is dedicated to imaging with a field of 5 by 5 arcmin². A set of eight selectable filters has to be accommodated (in a filter wheel), and some low resolution slitless spectrography is needed (in the filter wheel).

The design drivers for this instrument have been:
- to minimise the envelope and mass,
- to provide a reduced diameter beam to accommodate a filter and disperser wheel,
- to provide an image quality of less than 2 pixels in the 0.5 to 1.0 µm spectral band.

The object size at the entrance of the instrument is 279×279 mm², corresponding to a field of 5 x 5 arcmin² in telescope object space. The telescope field curvature is not taken into account in the instrument but its influence has been looked at: the concavity of the telescope image surface is oriented towards the telescope which means that the convex secondary mirror contribution is predominant; a all mirror re-imager with a global positive power (power made mainly by concave mirrors).will naturally compensate this curvature.

An $8k \times 8K$ detector array with 15 µm pixels is considered to get the wider possible field (5 x 5 arcmin²) with good sampling (37 mas). The detector array size is 123 x 123 mm² and the Wide Field Imager magnification is 1 x 0.44 with an image aperture of f/10.6.

An all mirror optical concept (see Figure 2.1.2-a) has been selected and preferred to a dioptric, centred system and a catadioptric system. The major feature of an all-reflective design is obviously its independence from wavelength, allowing an extension of the spectral domain below 500 nm. Another important point is the capability to get a homothetic behaviour with temperature - therefore image quality is invariant - if using the same material for structure and mirrors.

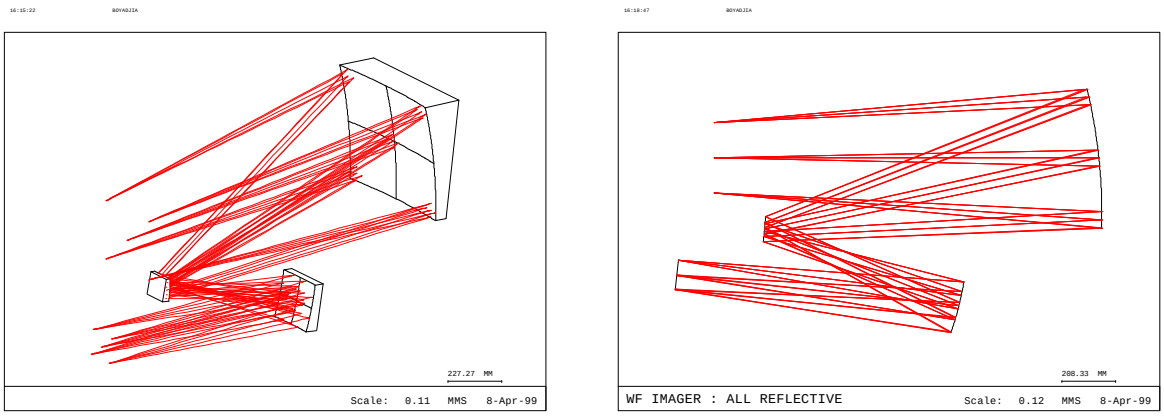


Figure 2.1.2-a: All-mirrors Wide Field Imager optical concept

The proposed design uses a three-mirror configuration. The M1 (first mirror) is an anamorphic conic, whereas the two others are de-centred ant tilted conics. The size of the M2 (secondary mirror) and M3 (tertiary mirror) has been limited to a maximum 20-cm side square aperture (10 cm for M2 and 20 cm

for M3). The primary mirror, whose size (54×54 cm²) is fixed by the OTA focal plane distance of 1.5 m, needs to be asymmetrical in order to correct the pupil aberrations, which would otherwise increase dramatically the filters diameter.

The pupil diameter is 76 mm and the associated field angles are small (10° total angle). The pupil is located in a significantly non-parallel beam (Diverging beam, with a numerical aperture 0.045): in fact this does not imply any drawback, since the filter plate defocus and small aberrations can be taken into account during the system final optimisation, with the only constraint to use the same thickness for all the filters.

The overall optical dimensions, without mechanical parts, are 1.7×0.9×0.5 m^3.

Optical performances in Imaging mode

This design has a good imaging performance: the geometric spot diameter containing 90 % of the total energy is lower than 20 µm (< 1.3 pixel) over the whole field (see Figure 2.1.2-b) . The distortion is of asymmetrical type, and in the range of ±3% in one direction. There is an anamorphose of ±3%.

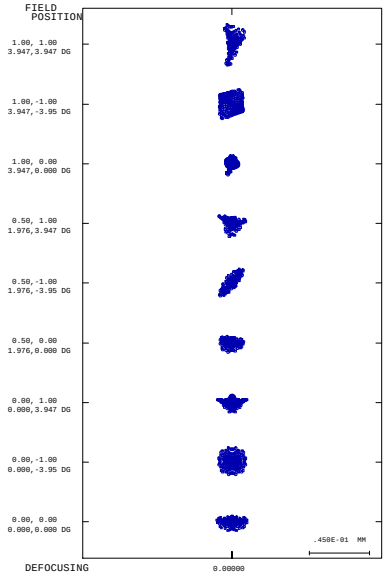


Figure 2.1.2-b: WFI spot diagram in imaging mode (3 pixels scale bar)

Optical performances in low-resolution spectroscopic mode

The WFI has to provide a slitless spectrograph mode with low resolution (spectral resolution R = $\Delta\lambda/\lambda$ about 100) on full field. Inserting a dispersing element in the filter wheel gets this mode. Since the disperser is a removable element, it must have a small deviation. Classical candidates for this purpose are a double-prism with different glasses or a grism. Since grisms are not suitable for very low dispersion (very large groove spacing, diffraction orders non-separability, bad efficiency at spectrum edges…) a prism solution has been selected, despite the strong non-linearity due to glass dispersion law. In this study, a simple BK7 - SF6 double prism has been used, without trying to select the best glass combination to improve the dispersion linearity (furthermore, this would require more knowledge on glasses partial dispersion at low temperatures). Special care has been taken in the image quality analysis, whose optimisation yields in several cases to slightly curved prisms (spherical surfaces instead of flat ones). The very low resolution spectrometer is such that the spectrum extent at focal plane is 3.4 mm. Due to glasses dispersion law, the spectral resolution R = $\lambda/\Delta\lambda$ varies significantly over the spectral domain.

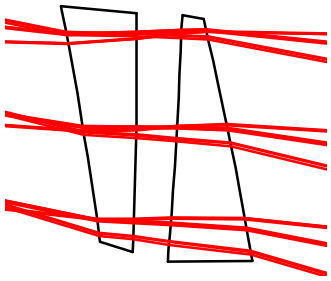


Figure 2.1.2-c: WFI : prisms for spectrograph mode

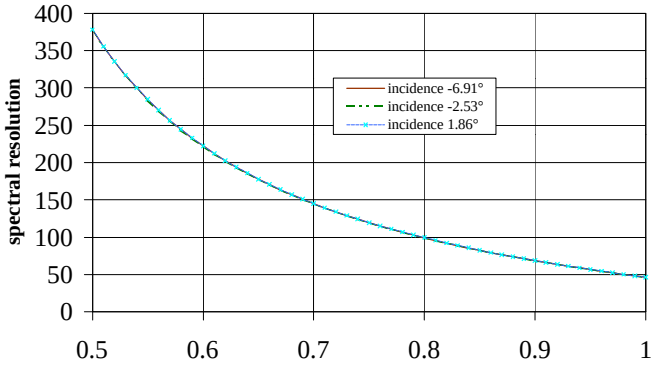


Figure 2.1.2-d: Spectral resolution of the Wide Field Imager in spectrograph mode

Prisms need to have slightly curved surfaces. The small field angle allows small prism angles and a good homogeneity of the disperser characteristics over the field of view. The prism angles are 11° and 15°, with a thickness of 40 mm at centre, and a height of 86 mm. The prism faces are slightly curved (R ≈ 1.5 m),

The spectrum size variation over the field of view is very low: ± 0.6 %. The spot diameter (maximum monochromatic spot size at 90 % energy for 3 different wavelengths in 500-1000 nm range) is 35 µm (i.e. 0.09 arcsec - better than the 0.3 arcsec need).

Tolerance analysis

A sensivity analysis on each mirror position has been performed for the WFI, which appears to be the most demanding instrument. The image quality degradation criterion is an increase of 3 µm of the spot diameter for each perturbation.

Element	Axial position without refocus	Axial position with refocus	Transverse position	Tilt
M1	± 200 µm	± 2 mm	± 1 mm	± 1 mrad
M2	± 50 µm	± 500 µm	± 80 µm	± 200 µrad
M3	± 20 µm	± 200 µm	± 60 µm	± 100 µrad
Focal plane	± 35 µm	NA	-	± 500 µrad

Table 2.1.2-a: Tolerance requirements

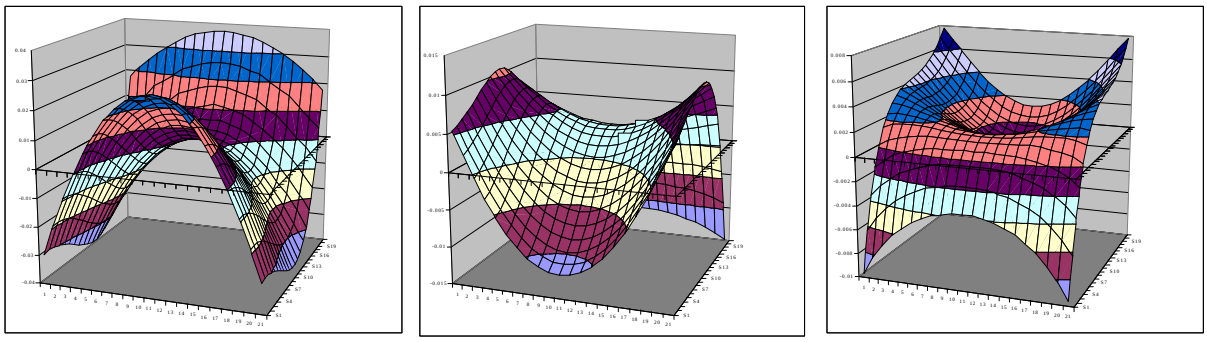
It can be pointed out that :
- for small defocus compensations, either M2 or focal plane can be moved ; the corresponding mechanisms are quite similar (nearly the same displacements are necessary),
- with homogeneous material from telescope image to final image, no image quality variation occurs while cooling down,
- for in-orbit fluctuations, the "without refocus" tolerances have to be applied.

Manufacturability of the mirror surfaces

Two criteria are used to verify the feasibility of the surfaces :

- the rms sag with respect to the best sphere which characterises the amount of material to be removed in order to get the right surface profile.

- the so-called "DY" parameter, distance on the surface for which the wavefront error in reflective mode is λ/4 (at λ = 500 nm). It is directly related to the slope with respect to the reference sphere. Values of the order of 10 µm are difficult to achieve if diffraction limited performance is needed.



Mirror 1 *Mirror 2* *Mirror 3*

Figure 2.1.2-e: Surface plots of the three mirrors with respect to the best rms sphere

The values for these two criteria are given in Table 2.1.2-b and are quite reasonable proving the manufacturing feasibility.

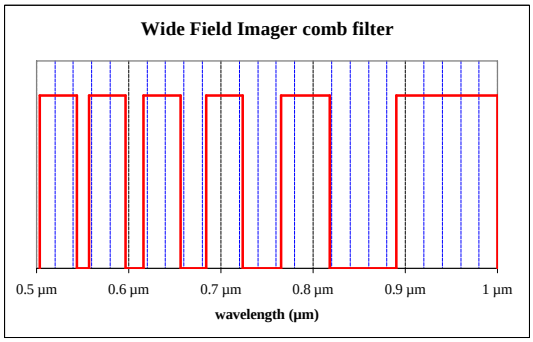
Mirror	Surface type	RMS sag / sphere	Minimum DX	Minimum DY
Primary	Anamorphic conic	18 µm	160 µm	120 µm
Secondary	Off-axis conic	6 µm	100 µm	50 µm
Tertiary	Off-axis conic	3 µm	440 µm	190 µm

Table 2.1.2-b: Criteria for WFI mirror manufacturing

The comb filter mode

The comb filter mode of the Wide Field Imager is similar to the spectrograph mode for spectral dispersion but with addition of a multi band filter. The sub bands of the comb filter are defined in such a way that the sub bands and the inter band gap spread in the focal plane over at least 0.5 arcsec (equivalent field in object space). The need of a spreading on at least 0.5 arcsec stems from the typical primordial galaxy size of couple of tenths of arcsec. This spreading avoid any significant overlapping of the galaxy image. The deviation due to the prism varies between ± 4.1 arcsec. Because of the non linear behaviour of the prism, a sub band of 110 nm is required to get a spreading on 0.5 arcsec at the long wavelength edge (1µm) while a sub band of 10 nm would be enough at the short wavelength edge (0.5µm). Narrow sub band would be a penalty for the signal to noise ratio and a width of 40 nm has been chosen as a minimum. Conditions of spreading and minimum width lead to the spectral definition of comb filter given in Figure 2.1.2-f.

With addition of the signal on the sub band, this mode become equivalent to an imaging mode but with the multiplex advantage of images in several bands in one observation.



Band	Band definition nm	Band width nm	Field spread arcsec	Pixels spread
1	504 – 544	40	1.84	50
2	557 – 597	40	1.26	34
3	616 – 656	40	0.87	24
4	684 – 724	40	0.59	16
5	766 – 818	53	0.50	14
6	890 – 1000	110	0.50	14

Figure 2.1.2-f: Definition of the comb filter of the WFI in multiband spectrograph mode

2.1.2.2 High Resolution Imager

The High-Resolution Imager (HRI) is an imaging instrument, which performs an oversampling of the telescope image pattern with a sampling of 10 milli arcsec. There is no spectroscopy option considered for this instrument. A 4k x 4k focal plane array (smaller than WFI instrument, to limit the mass/envelope of the HRI) with 15 µm pixels has been considered. With a 10 milli arcsec sampling: the magnification from OTA image is therefore 1 x 1.61 and the field of view is 41×41 arcsec².

A separate instrument has been preferred to solutions with High Resolution mode inside the Wide Field Imager using either switchable lens or switchable mirror. The choice of an independent instrument allows simultaneous operation of HRI and WFI, and suppresses the failure drawbacks related to switching paths. The image quality of the instrument can be furthermore well optimised.

The selected optical design is a 3 mirror system derived from the Offner relay with only spherical surfaces (see Figure 2.1.2-g). Spherical mirror manufacturing is very easy. The overall length is about 50 cm. The intermediate pupil is not accessible, but the filter diameter is very small, less than 20 mm. The optical characteristics are summarised in Figure 2.1.2-h.

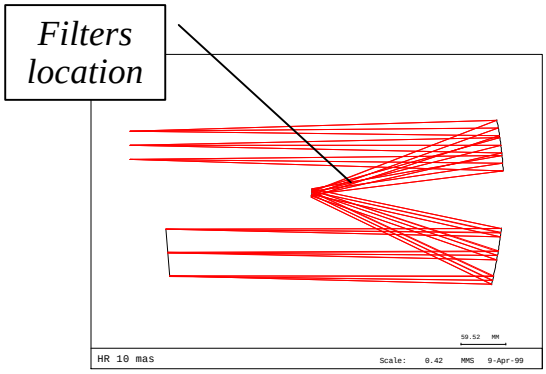


Figure 2.1.2-g: HRI optical design

Sampling	10 mas
Magnification	1.61×
Field of view	41 arcsec
Optical design	3 spherical mirrors
Dimensions (Optical)	0.50×0.22×0.08 m^3
Filters (Shifted from pupil)	∅ 16 mm
Image quality	23 µm (90 % energy)

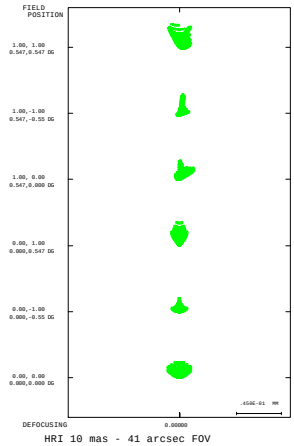


Figure 2.1.2-h: HRI optical design synthesis and spot diagrams

With such a field of view, the telescope field curvature effects are negligible (1 µm increase of spot diameters). The geometric spot diameter at 90 % energy is 24 µm (1.6 pixel).

A possible improvement of the High Resolution Imager is the use of an 8k x 8k array with a field extension to 82 x 82 arcsec². The 3 mirror concept can offer such a field of view but aspherisation of optical surfaces is needed to cope with the field. Optical performance is excellent with spot diagrams better than 1 pixel.

2.1.2.3 Integral Field Spectrograph

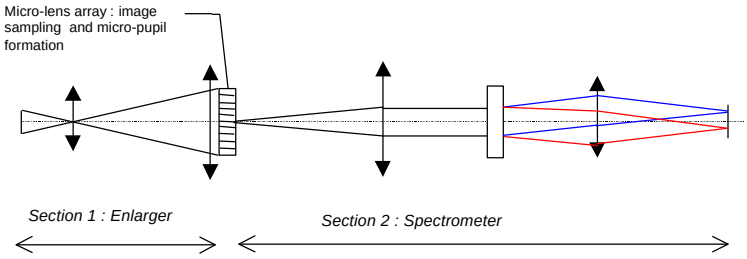


Figure 2.1.2-i: Principle of the Integral Field Spectrograph

The principle of an Integral Field Spectrograph (IFS) is to realise a sampling of the field and to generate a spectrum of each field sample. The sampling is here performed by a micro-lens array.

The Integral Field Spectrograph includes three sections (Figure 2.1.2-i). The telescope field is first enlarged in order to match the microlens array dimensions. The microlens array performs then the field sampling and focuses the incoming light on micropupils. The micropupils are then imaged and dispersed through a classical spectrograph with collimator, dispersing element and focusing optics (camera). The disperser is slightly rotated in order to avoid the overlapping of the spectra. Each section has been optimised separately.

The requirements for the Integral Field Spectrograph are a field of 10 x 10 arcsec^2, a sampling of about 100 mas and a spectral resolution $R=\Delta\lambda/\lambda$ from 100 to 1200.

A all dioptric solution derived from the TIGER design has been selected (see Figure 2.1.2-j) because the mirror solutions, which have been investigated, have all led to very large instruments. The instrument requires a compensation of the secondary spectrum, so anomalous dispersion glasses are needed. The overall optics envelope is a cylinder of 8 cm diameter and 1 m long. The advantage of this design is its small envelope, but its performance fully depends on glass behaviour at low temperature.

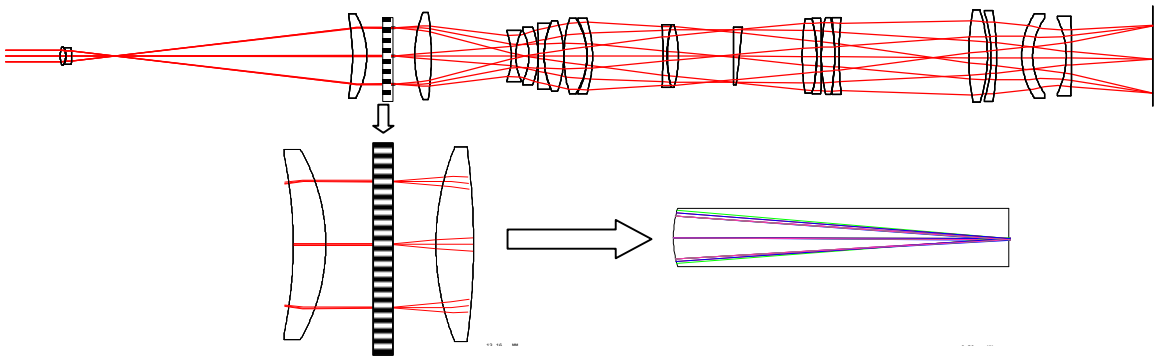


Figure 2.1.2-j: Integral Field Spectrograph optical design, with details of microlens array and microlens pupil formation

Enlarger section

The Field magnifier optics adapts the focal length of the telescope in order that the 0.5 mm microlens pitch matches with the field sampling at 0.1 arcsec. This requires an overall focal length of 1031 m and an enlarger magnification factor of 1 x 5.4. With a telescope diameter of 8 m, the beam speed is f/129 at the microlens array entrance.. In addition, the optical beam needs to be telecentric to provide a regular mesh of micropupils. The enlarger is made with a doublet and a field lens to ensure the telecentricity. The image quality is compatible with the microlens size (spot diameter far smaller than the micro-lens dimension).

Field censor section

The sampling of field is performed by an array of microlenses of 50 mm x 50 mm size with a sampling of about 0.5 mm x 0.5 mm. Three different microlens shape can be thought at :
- a circular microlens with a square sampling which results in a loss of flux due to the 79% fill factor,
- a square microlens requiring a low spectrograph f-number for the to cope with the diagonal field,
- a hexagonal microlens which is considered as the best compromise between efficiency and f-number and which is, thus, selected as baseline.

With an hexagonal microlens array, the horizontal or column sampling with 100 hexagons of 0.5 mm pitch leads to an hexagon side of 289 µm. The interline distance or vertical sampling pitch is 433 µm leading to 115 lines. So the 10 arcsec square field of view is sampled with a 0.1 arcsec pitch horizontally and 0.087 arcsec vertically.

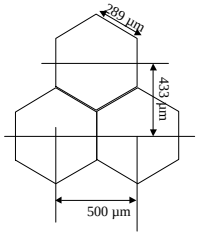


Figure 2.1.2-k:
Microlens hexagonal geometry

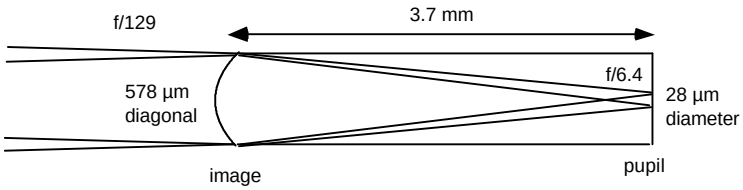


Figure 2.1.2-l:
Optical parameters of the microlens array

The output face of the microlens array is a pupil plane. The f-number of microlenses results from a compromise. A small f-number leads to a reduced pupil diameter with a high power on the microlens and a fast and more difficult spectrograph section. On the other hand, a small f-number produces small micropupils and less confusion in the focal plane. A f-number of f/6.4 has been selected as reasonable compromise and the microlens focal length is thus 3.7 mm. The microlens pupil diameter results from the microlens focal length and beam f-number at microlens entrance and is equal to 28.7 µm.

Micro-pupil dimensions have been checked by footprint calculations, for 3 different field positions, with an optical model including the field expander and a micro-lens. The micropupils have an elliptical shape. The total footprint dimensions (geometric 100 % energy) are given in Table 2.1.2-c. The increase of the spot diameter with respect to the paraxial size is due to the axial chromatism of the microlenses despite use of low dispersion glass (silica) and to the distortion related to aperture variations which gives the elliptical shape. A more sophisticated field lens (aspherical lens or doublet) would reduce these effects.

Wavelength/Radial FOV	0 mm	25 mm	35.4 mm
1000 nm	32.2 µm	33.4×32.2 µm	34.6×33.0 µm
640 nm	31.8 µm	34.6×31.6 µm	37.8×32.8 µm
500 nm	36.0 µm	38.9×35.2 µm	41.8×36.4 µm

Table 2.1.2-c: Total footprint dimensions within the field

Figure 2.1.2-m: Micro-pupil footprints, for edge diagonal field and 3 wavelengths

Spectrograph section

The spectrograph section reimages the micro-pupils, considered now as the "objects" on the detector. The pupil of the spectrograph section is determined by the shape of the microlenses and has an hexagonal form (design is made to cope with hexagon diagonal which defines the f/6.4 speed). The spectrograph section has a magnification of 1 x 1.2 which results in a pupil image diameter of 34 µm i.e. about 2 pixels.

The spectrograph design is adjusted to get disperser size of 40 mm. The collimator has a focal length of 250 mm and an aperture of f/6.25. The field in the pupil space is 16.6°. The micro-lens system provides telecentric beams. The image quality of the collimator (geometric polychromatic spot diameter at 90 % energy) is equivalent to 1.9 pixel (at detector plane). The camera has a focal length of 300 mm and f/7.5 aperture. There is no telecentricity requirement. The camera achieves an optical quality of 0.95 pixel spot diameter at 90 % in polychromatic better than the collimator performance.

The image quality of the complete spectrograph section has been checked without disperser to assess its intrinsic quality. The polychromatic spot diameters are less than 2.2 pixels at 90 % energy on the whole spectral band, with a noticeable influence of the secondary axial chromatism, despite the use of special crown. This chromatism could be reduced if needed by a more suitable choice of glasses.

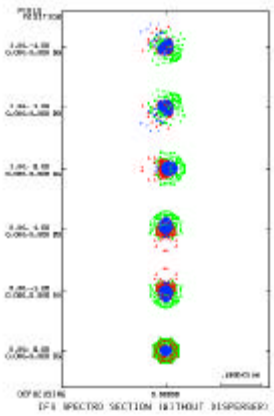


Figure 2.1.2-n: Spot-diagram of IFS spectrograph section, without disperser

Various types of dispersers can be inserted in the intermediate pupil, between collimator and camera. The required high spectral resolution R is about 1200 and is performed with 5 high spectral resolution grisms in BK7 covering the 629-910 nm spectral range. The grisms are blazed at the central wavelength and at first order and comply the zero deviation condition. The required low spectral resolution R of about 100 is performed by a zero deviation low spectral resolution double-prism covering the whole spectral range.

The definitions of the grisms for the high spectral resolution modes are given in Table 2.1.2-d. All the grisms are in BK7 and are blazed at the central wavelength and without deviation at this wavelength. The spectral resolution is defined with a spectral element spread on 2 pixels.

BK7 grism	Spectral domain	Groove number	Blaze and prism angle	Spectral resolution	Blaze efficiency	diameter 90% energy
1	629-684 nm	180gr/mm	13.28°	1188-1367	>99%	15.9 µm
2	678-736 nm	170gr/mm	13.55°	1205-1387	>99%	16.0 µm
3	729-791 nm	160gr/mm	13.75°	1215-1402	>99%	15.9 µm
4	784-850 nm	150gr/mm	13.89°	1222-1411	>99%	14.5 µm
5	840-910 nm	140gr/mm	13.91°	1222-1411	>99%	17.1 µm

Table 2.1.2-d: Integral Field Spectrograph in high spectral resolution mode definition of grisms and spectrograph performance

The spectrograph focus is optimised for the sub-bands 1-4. The grism of sub-band 5 requires a very small curvature correction on the non-diffractive surface (less than 1 µm sag) to correct from focus error. Table 2.1.2-d gives the worst case diameters within the field and the wavelength range for each sub-band ; the worst case diameter is equivalent to 1.1 pixel.

Two possibilities have been looked out for the disperser of the low spectral resolution mode, a grism and a prism. The aim is to cover the complete spectral range from 0.5 to 1 µm in one shot.

A grism (transmission grating with prism to compensate the deviation) with 20 grooves /mm, blazed at 0.75 µm at first order provides a dispersion of 20.67 µrad/nm at the blaze wavelength. The required blaze angle, which is also the prism angle, is 1.679°. The spectral resolution is in the range 110-207 and the theoretical blaze efficiency goes down to 34%.

The alternative is a direct vision double prism LaSFN30/SF6 with prism angle of 28° and 27.978° respectively. The condition of zero deviation is got at 0.609 µm with a 0° incidence angle. The wavelength of 0.609 µm for zero deviation has been selected to balance the spectra. The spectral resolution is in the range 44-366.

The major drawbacks of the grism is the problem of order overlapping because the band is equal to one octave and the theoretical blaze efficiency which goes down to 34%.The prism drawbacks are the large prism angle of 28° leading to a thick piece of glass (thickness < 35 mm for 50 mm diameter) and its non linear angular dispersion 32°. The prism is nevertheless preferred to grism and thus selected.

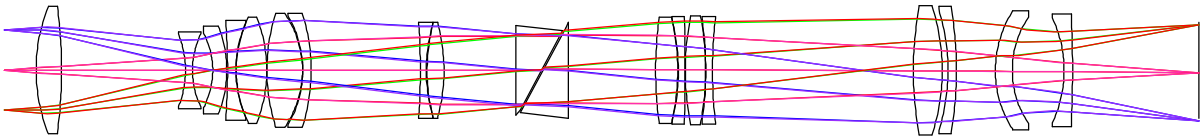


Figure 2.1.2-o: IFS spectrograph section, with double-prism low resolution disperser

The image quality in the low-resolution mode with the double-prism is lower than in the high resolution mode due to the axial chromatism residue of the spectrograph, and to the small beam shifts due to prisms. The spot diameter is 3.3 pixels at 90 % energy, and 2.4 pixels at 80 % energy. This image quality could be improved by further optimisation. The spectrum extent variations are ± 4% within the field.

Accommodation of the spectra in the focal plane

The distance in the focal plane between the pupil images is 600 µm or 40 pixels. The spectrum width is about 2 to 3 pixels. A rotation of the spectrum performed by disperser rotation is needed for an optimum packaging of the spectra. The principle of spectrum rotation is shown in Figure 2.1.2-p.

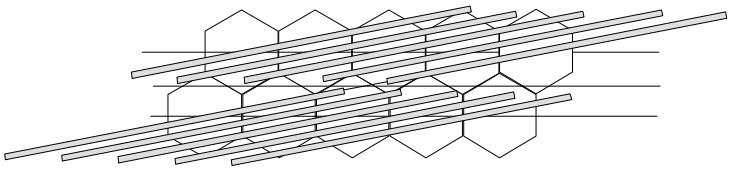


Figure 2.1.2-p: Rotation of spectra for focal plane optimum packaging

Figure 2.1.2-q shows the inter spectrum distance and the spectrum length as a function of the spectrum rotation. With a small angle, long spectra are allowed but the distance between spectra is reduced. A reasonable compromise is needed with a distance between 6 and 7 pixels between spectra. The selected baseline is a rotation of 8.95° with a distance of 6.4 pixels between adjacent spectra. The related spectrum length is 222 pixels which corresponds to 111 spectral elements because the image spot is spread on 2 pixels. The repartition of spectra in focal plane is illustrated in Figure 2.1.2-r.

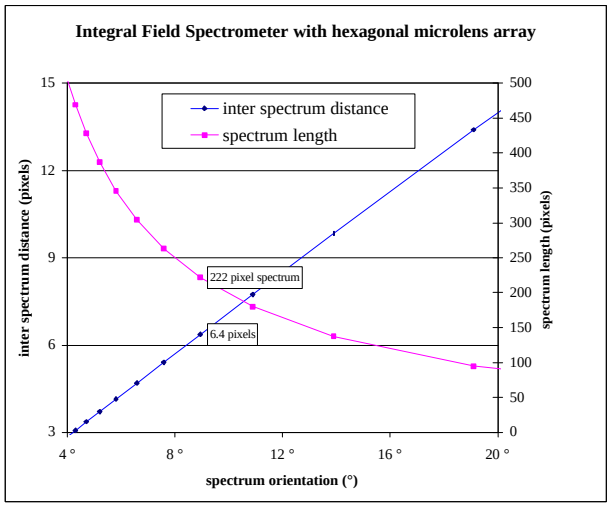


Figure 2.1.2-q: Interspectrum distance and spectrum length as a function of spectrum rotation

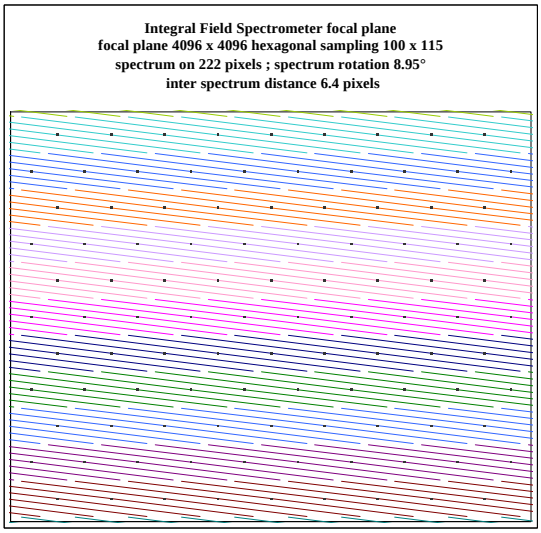


Figure 2.1.2-r: Spectrum arrangement in focal plane for a 8.95° spectrum rotation

Addition of a PYTHEAS type mode

The addition of a Fabry Perot within the enlarger of the Integral Field Spectrograph would provide a mode with very high spectral resolution of several thousands as in the ground based PYTHEAS spectrograph and will allow the astrophysical applications foreseen in science chapter.

Change of enlarger

A set of enlargers on a wheel mechanism will allow us to use different field/sampling couple around the nominal couple 10 arcsec/0.1 arcsec. Note, nevertheless, that a field increase results in an increase in proportion of the micropupil diameter which reduces the margin between spectrums and also the spectral resolution. Maximum possible extension of the field is about 15 arcsec.

In case of a movable microlens array, it could be possible also to have an imaging mode in a relatively large field with a high angular resolution comparable to the High Resolution Imager.

2.1.3 Focal plane

2.1.3.1 Detector trade offs

The main electro-optics parameters required for the NGST visible instrument (see Table 2.1.3-a) can only be reached by high performance photon detectors.

Parameters	Requirements
Quantum efficiency	80% from 0.5 to 0.6 µm / > 70% from 0.6 to 0.8 µm > 60% from 0.8 to 0.9 µm / > 50% from 0.9 to 1 µm
Read-out noise	< 3 e-rms. for a single read-out (for spectrograph)
Dark current	<0.0003 e-/s (mainly for spectroscopy)
Dynamic	100 000 e- total / 50 000 e- linear

Table 2.1.3-a: Main required performances for the NGST visible instrument focal planes

2.1.3.1.1 CCD performances and constraints

Indeed, CCDs offer excellent performances, such as excellent quantum efficiency (up to 95% peak) over a large spectral range a low read out noise (few electrons up to 1Mpixels/s) and low dark current. CCDs are used in nearly all the visible astronomical instruments on ground or in space. It is a mature technology, very large monolithic devices (up to 9k x 9k pixels) can be manufactured and CCDs can feature small pixel size.

However, it is also known that for non scientific applications, CCDs have today to compete with the so-called Active Pixel Sensors (APS) which offer several advantages, such as random access to the pixels and far less power dissipation.

As the ISIM is mainly dedicated to infrared instruments, its temperature is in the 30 to 35 K range to minimise the infrared arrays dark current. Buried channel CCDs (BCCDs) are not able to be operated at so low a temperature due to the effect of carrier freeze out. For this reason, the charge transfer efficiency (CTE) for temperature typically lower than 100K is in general not acceptable for BCCDs. On the other hand, surface channel CCDs, used for a long time as cooled IR array read-out devices, cannot be used for low light level applications, as their CTE is bad (even if less degraded at low temperature). CTE improvement requires therefore to heat up the CCD in order to reach a temperature around 100 K. As the CCD dissipates non negligible amount of power when dynamically operated, this power could be used to heat up the focal plane but requires an “inverse” cryostat to insulate the CCD from the rest of ISIM.

The elementary CCD sub-module which has been considered is equal to 2k x 4k (which is the today workhorse for modern astronomical optical instrumentation) with a single video output stage per CCD. As there is a need for having the read-out time negligible when compared to the elementary integration time (around 1000 s), a pixel rate of 125kpixels/s was selected, which allows to read out with low read-out noise the full frame in 60 s. The duty cycle is equal to 94% which was considered as acceptable. As the static power dissipation of the output amplifier can be stopped during integration (30 mW per output stage when switched on), and as the energy required for the dynamic read out of a 2k x 4k array is around 300mJ, the 8k x 8k focal plane, featuring 8 output amplifiers, would produce a power dissipation as high as 280mW during the read-out peak, which corresponds to 17mW averaged during the elementary integration period. As the two other focal planes of the instrument together correspond to 50% of the 8k x 8k, the total averaged power dissipation of the focal plane is around 25 mW, without considering any preamplifier in the cold part of the instrument. This is not far from the maximum tolerated power dissipation in the cold part for the visible instrument (50mW). In addition, even if non simultaneous read out of the elementary CCDs would allow to smooth the focal plane dissipation versus time, there will be peaks for power dissipation.

The selection of CCDs is therefore questionable, as:

- the power consumption is high, and marginal when compared to the preliminary NGST visible instrument dissipation budget in the cold focal plane
- the temperature of the CCD would change with time
- even if well baffled, there would be leaks of straylight which could bias the measurements for the IR channels (it would be almost impossible to manufacture a window with nearly perfect optical transmission in the 0.5 to 1 µm and nearly perfect blocking outside of this band)

2.1.3.1.2 CCD versus CMOS

CMOS read-out integrated circuits (ROIC) are today used for all the imaging applications, which need to be operated at low temperature, in particular for infrared array. They allow very good performances to be achieved in terms of read-out noise and dynamic. The first possible solution to get a visible CMOS detection array is to use monolithic CMOS image sensors, also called Active Pixel Sensors (APS). However, although very attractive in terms of cost and available capabilities, these devices should not be suitable for the NGST application. Indeed, standard monolithic APS offers good electro-

optics performances but presents also as a major drawback a relatively poor quantum efficiency when compared to back-side illuminated CCDs.

The poor quantum efficiency stems from three main reasons:

- the non 100% fill factor: in monolithic APS because a part of the pixel area is devoted to the active part of the pixel. For non smart imagers, this corresponds typically to 3 transistors for photodiode pixels and 5 transistors for photoMOS pixels. Even if this active area is partly sensitive to light, it reduces the global efficiency of the photodetector In the near future compatible with NGST planning, this limitation should disappear, at least for detectors which do not require complex processing in the pixel (which is the case for the visible instrument) because the use of deep submicron CMOS technology and the technology improvement will allow a strong decrease of the active function area. As an example, and for a photoMOS fill factor around 60%, the pixel pitch has to be of the order of 30, 20 and 10 µm respectively for 0.7, 0.5 and 0.25 µm design features.

- the front illumination: monolithic APS cannot be back-side illuminated and/or thinned without specific processes (not offered today by back-end CMOS foundries, and which would be contradictory with the general monolithic APS philosophy). This means that for photoMOS, the poly gate will absorb part of the incoming flux, particularly for the short wavelength. Using photodiode structures allows to increase this quantum efficiency, but values are still lower than what can be achieved using scientific CCDs.

- the use of thinner active layer: in order to improve the density of CMOS devices (microprocessors, memories…), the market asks for deeper and deeper feature size. Following Moore's law, 0.18 µm is expected as minimum lithographic feature size for 2001, and 0.1 µm for 2007. For feature size beyond 0.25 µm, SOI substrates will be used, with advantages and drawbacks when compared to the bulk substrate case. The obvious advantage is the increase of the fill factor. One of the main drawback will be the reduction of the photo-charge collection depth, especially for long wavelength photons as silicon becomes very transparent up to 0.7 µm. On the contrary, dedicated CCD technologies have always customised the fabrication process to properly position the junctions and depletion depths for optimal spectral sensitivities and minimum cross-talk.

While standard CMOS processes (down to 0.5 µm feature size) are today used for APS design and manufacturing without any process change, it is expected that some modifications to the fabrication process will be requested in order to enable CMOS to perform good quality imaging at the 0.25 µm technology generation and beyond. For high performance imaging devices in the visible, it is foreseen that the silicon hybrid selected for NGST and presented below is the good way for the next decade.

The hybrid approach

The hybrid principle is the one used for cooled infrared arrays manufacturing. It should be recalled that the first attempt for multiplexed IR array manufacturing at the end of the 1970s was to build monolithic devices using the same material for detection array and for read-out circuit. For instance, InSb CID and Ge or MCT CCD were produced by few companies and more recently InGaAs FET arrays were manufactured. Due to low metallurgic detection material quality, performances of such devices were poor and the trend was then to use a first layer of exotic material detectors hybridised on the second layer constituted by the read-out circuit in silicon (CCD were first used, then replaced by CMOS read-out circuit due to the reasons explained above). The hybridisation is mainly made by matting the two layers using indium bumps (see Figure 2.1.3-a and Figure 2.1.3-b), but other techniques were also proposed (for instance the loophole technique, uses by GEC Marconi IR (UK)). The obvious advantage of the hybrid concept for photon detection is that it allows to separately optimise the detection layer and the read-out circuit. Coupled with CMOS read-out, this approach will offer excellent electro-optics performances, and will provide all the advantages of APS versus CCD.

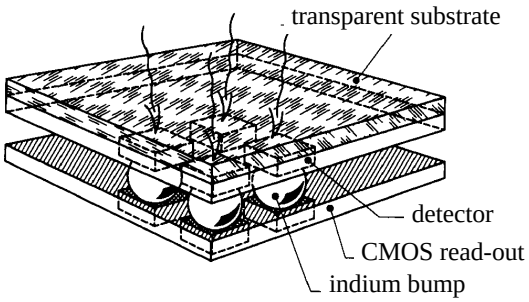


Figure 2.1.3-a:
Principle of hybrid 2-D arrays

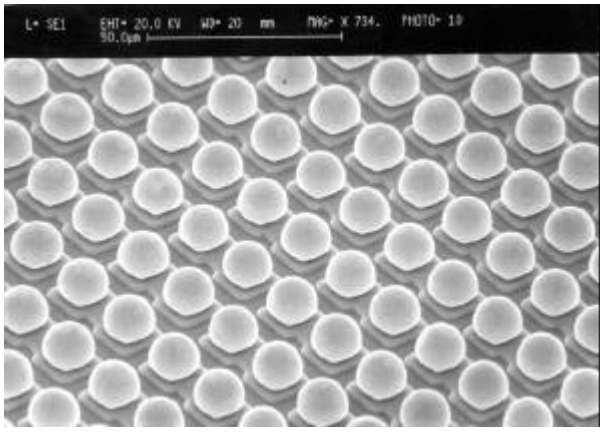


Figure 2.1.3-b:
Example of indium bump array (20 µm pitch) used for IR CMOS hybrid (from LETI/LIR)

The main advantages of the hybrid approach in the framework of the NGST visible instrument are:

- the read-out flexibility: as for a RAM, the pixels of the CMOS imagers can be randomly addressed, on the contrary to a CCD which can only be sequentially read-out (see Figure 2.1.3-c and Figure 2.1.3-d). This could be of interest for NGST as it allows specific features such as fast windowing read-out or local exposure control. As the read-out can be non destructive, multiple read-out can be used, which is of great interest for NGST, as it allows to reduce the read-out noise by averaging non correlated sources of noise, to remove cosmic hits and to improve useful scene dynamic. These advantages are already extensively used by astronomical infrared cameras using IR hybrid, on ground or in space.

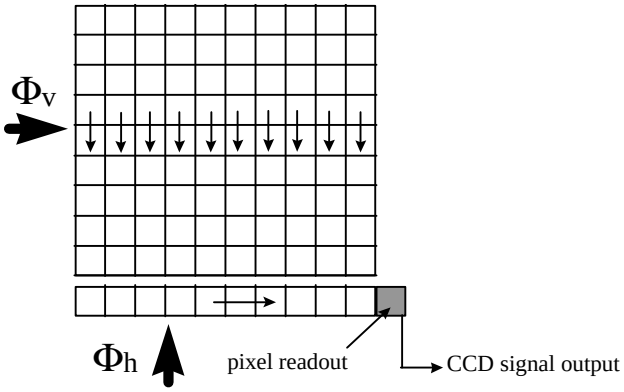


Figure 2.1.3-c:
sequential mode of CCD read out

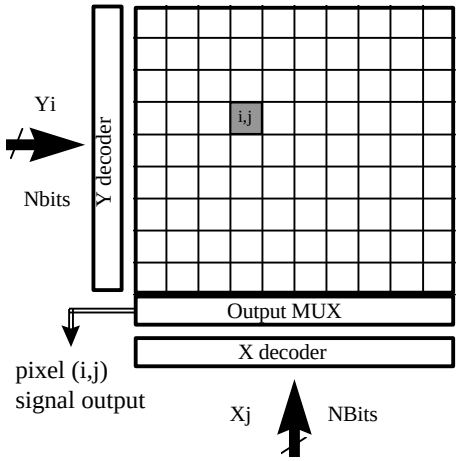


Figure 2.1.3-d:
random address mode of CMOS read out

- the very low read-out noise: noise can be reduced by multiple sampling for slow read-out rate. Another way can be used, which may allow CMOS imager to be designed as nearly photon counting imagers. Indeed, the read-out noise in electrons depending on the noise floor of the read-out circuit and the conversion factor, which mainly depends on the capacitance used for the charge to voltage conversion (this capacitance being located in each pixel for CMOS read-out and at the output stage level for CCD). State-of-the-art CCDs feature conversion factor around 6 to 10 µV/e-. On the other hand, existing CMOS read-out can feature figures as high as 20 µV/e- for photodiode pixels (typical stray capacitance would be around 10fF for a 15-20 µm silicon photodiode pixel pitch) and 100 to 150µV/e- for photoMOS pixel (capacitance of floating diode can be as low as 1fF). Typical noise at

the CMOS read-out output is around 50 µV rms for cryogenic operation and low pixel rate (i.e. few 100s pixels/s). Therefore, read-out noise of few electron rms can be easily achieved (see Figure 2.1.3-e). However, a trade-off has to be made with respect to the acceptable full dynamic (in electrons), as this parameter decrease with increasing conversion factor. It should also be noticed that CMOS imagers would benefit from low noise thanks to a semi-parallel read-out which would relax the noise bandwidth to be taken into account for the in-pixel source follower noise budget.

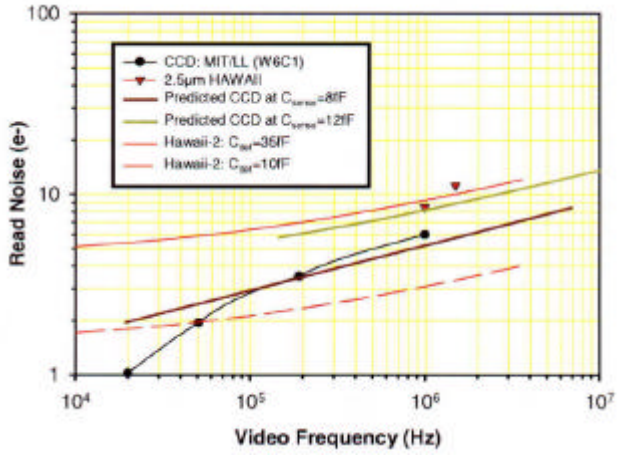


Figure 2.1.3-e: Comparison of read-out noise for various photodiode array types hybridised with CMOS ROIC

- the snapshot mode without shutter: high performances 2-D CCDs are full frame (FF) or frame transfer (FT) types. FF cannot be used without shutter in most of the applications as pixel wells will continue to integrate during CCD read-out. The situation is more favourable for FT CCDs as charges are transferred very fast from image zone to memory zone in order to be read-out. However, FT CCDs can feature non negligible smearing for contrasted scenes as the transfer from image to memory is not zero. In addition, the use of a memory zone has to be avoided for applications requiring mosaics focal planes, such as for the NGST. The use of interline CCDs can solve this problem, the price to pay being a large degradation of the quantum efficiency, non acceptable for NGST. CMOS hybrid would allow to be totally snap-shot without any electro-optics degradation, as it will be possible to store the analogue information within the light-masked pixel prior reading it out. In addition, the duty cycle will be equal to 100%, as integration and read-out can be operated simultaneously.

- the low power dissipation: CCD are "capacitive" device, due to the use of long and numerous gates for charge transfer. This means that for large device, the on chip power dissipation will be large, which leads to very demanding thermal requirements. In addition, the operating electronics required by a CCD is also power consuming (use of specific biases (up to 30V), sequencer and clock drivers, video chain(s), pre-amplifiers…), even if this point is less critical for NGST. CMOS hybrid intrinsically features low power dissipation, as its uses CMOS process and as charge to voltage conversion appears in the pixel. This is why for same format and operating conditions, CMOS hybrids dissipate less than 4% of CCD with corresponding format (also switched on). For NGST WFI focal plane, 8kx8k CMOS hybrids, when switched on, would dissipate less than 10 mW to be compared to 280mW for 8kx8k CCDs. Low power dissipation of the CMOS ROIC will permit to keep the read-out permanently switched on, which would allow multiple read-out operations, while CCDs require duty cycle less than 4% to reach equivalent average dissipation.

- the radiation tolerance: In addition to the cosmic glitches transient effects, CCD mainly suffers of two radiation degradation types. First, total dose create ionisation in the oxide, which would mainly produce an increase of surface dark current generation and flat band voltage shifts. These effects are now well known and can be well managed for NGST. The displacement damage created by high

energy particles (mainly protons) is a more serious problem for low level applications as shown by HST results. Indeed, the traps created by the high energy particles will strongly decrease the charge transfer efficiency at unacceptable levels for certain applications due to the impact on signal amplitude (photometry) and maximum energy location (astrometry). As they are using thinner oxide layers, the total dose effects are less dramatic for most of the CMOS sensors. In addition, the absence of charge transfer over macroscopic distance will solve the problem linked to transfer inefficiency. In addition, it should be noticed that CMOS ROIC will benefit of all the improvements of design rules researches which are today under way in order to make CMOS devices more radiation tolerant.

- the intrinsic anti-blooming: charge wells in a CCD show blooming effects unless specific features have been implemented in the device. Vertical anti-blooming (AB) is not compatible with back-side illumination. Dynamic AB is limited in terms of efficiency and possible operations. Lateral AB generally implies a decrease of fill factor and full well capacitance and can be non compatible with inversion mode. Due to their design flexibility, it is easily possible to design a CMOS ROIC pixel in order to avoid voltage saturation. This is currently demonstrated without any impact on functionality and performances.

- the easier operation and the lower cost electronics: even if the NGST CMOS read-out circuit cannot take all the benefit of the possibility to integrate on chip electronics functions (which is one of the major advantage for APS) as the on chip power dissipation minimisation is one of the highest priority for NGST, CMOS ROIC will required less sophisticated electronics than CCD (easier clocking, no bias voltage higher than 5V, no clamp stage in the video chains if CDS function is implemented on chip…).

2.1.3.1.3 Choice of the detection material

Questioning the use of silicon as the detector material of choice for a visible instrument may appear unusual. However, for NGST, the possible choice of an hybrid approach and the need for detection up to 1 µm for a detector operated at 30-35 K changes the situation. Indeed, silicon has two properties, which have to be taken into account for the NGST. First, the silicon energy bandgap is around 1.17 eV at 30K, which means that photons with wavelength shorter than 1.057 µm will create a detectable photoelectron. Second, the optical transparency of silicon dramatically increases with wavelength, and also increases for decreasing temperature, as shown by Table 2.1.3-b. This explains why for nearly all the CCDs, quantum efficiency starts to fall far below 1.1 µm limit. Arguments apply both to CCD and to silicon hybrid, the only difference is that the situation will be a little bit more favourable for a CCD at 100K when compared to a silicon hybrid at 30K. As no available data about silicon optical properties at 30K was found, 77K figures have been used. This point needs further assessment.

Lambda	T=300K	T=120K	T=77K
400 nm	0.12 µm	0.20 µm	0.28 µm
600 nm	1.65 µm	2.88 µm	3.88 µm
800 nm	8.1 µm	15.9 µm	22.5 µm
900 nm	22.9 µm	52 µm	80 µm
1000 nm	107 µm	398 µm	858 µm

Table 2.1.3-b: Silicon transparency for various wavelengths and T° (from D.E. Groom et al, 1999)

In order to improve the 0.7 to 1 µm quantum efficiency when using silicon, the only solution is to increase the thickness of the active layer in order to absorb the near infrared photons. This approach, well known from CCD manufacturers, has been called the "deep" CCD. It has another advantage, as it strongly improves the spectral response flatness. Indeed, a "fringing" pattern is created by multiple reflections in thin CCD layers when illuminated by nearly monochromatic light (which is the case for spectroscopic applications). A thick silicon layer will strongly reduce this pattern. However, this approach also features drawbacks. The first one is the dark current increase, as more volume of silicon

is involved in the detection process. However, based on the very low operating temperature, this is not deemed to be a problem for NGST application. For instance, the last measurements performed by the Lawrence Berkeley National Laboratory (LBNL) team feature 0.001 e-/s/pixel at 140 K, 80 V bias, 10000 Ohms.cm resistivity and 300 µm depth. The second one is the cosmetic quality of the detector. In order to increase the depth of the depletion region, and as high voltage cannot be used, a higher resistivity substrate has to be used (around 1000 Ohms.cm instead of 10 to 100 Ohms.cm for standard CCDs). The crystalline quality of such substrate is not as good as the standard ones, and more material defects would create more cosmetic defects. The third drawback is the sensitivity to high energy particles, as a deeper active layer will stop more solar or cosmic rays. Finally, a deep detection layer can imply cross-talk at short wavelength as corresponding photons will be converted in electrons far from the depletion zone, and can migrate in neighbouring pixels. The solution, recently demonstrated by a team from LBNL is to use a very high silicon substrate resistivity -10000 Ohms.cm- coupled with a medium voltage bias in order to fully deplete the substrate (see Figure 2.1.3-g and Figure 2.1.3-g).

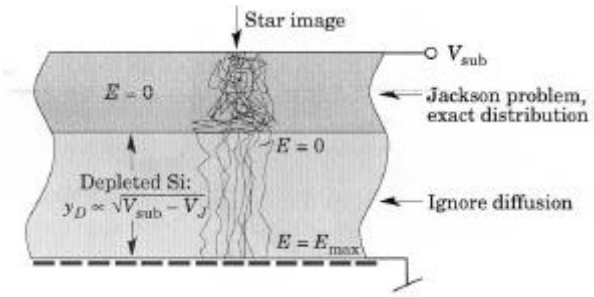


Figure 2.1.3-f:
Electrical field is not sufficient for full depletion, and cross-talk will appear.

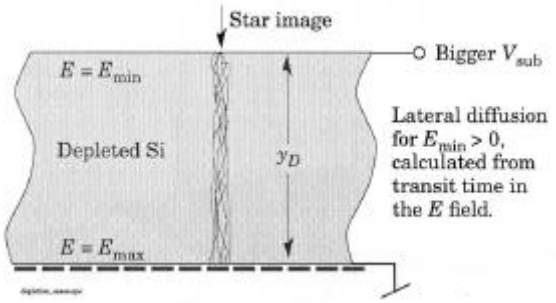


Figure 2.1.3-g:
Full-depletion is achieved, reducing pixel to pixel cross-talk (from D.E. Groom et al, 1999)

Silicon is not the only material able to detect the 0.5 to 1 µm range. Even if InP cuts too short for this application (cut-off wavelength near 0.9 µm), bicolour detectors composed of InP and InGaAs material layers can be foreseen. The main advantages of such detectors could be the lower dark current when compared to other "non silicon" material. However, considering that MCT and InSb material, which are the today candidates for the 1 to 5 µm spectral range detectors, would meet the dark current requirement for the NIR range at 30K, it is interesting to get an idea of the possible properties of these two IR materials in the 0.5 to 1 µm as they are used to build today's largest IR detectors, with extensive new developments funded for the NIR instruments of the NGST.

Raytheon IR Centre of Excellence (RIRCE) has already developed InSb for visible detection thanks to thinning the substrate, used for instance for the HYDICE hyperspectral imager. MCT optimised for the visible is currently developed by Rockwell Science Centre (RSC), thanks to selected etching to remove the CdZnTe cap.

However, the use of MCT and InSb materials for the focal planes of the NGST visible instrument also features drawbacks. For instance, the dark current of these materials will be higher at 30 K than the one of silicon (even for deep silicon). This is particularly true for InSb for which the cut-off wavelength cannot be tuned to shorter wavelength. Thanks to its ternary alloy properties, MCT dark current can be optimised by changing the alloy compound. Cosmetics will also be better for silicon (even for deep silicon). At 5µm wavelength cut-off, cosmetics should be better for InSb than for MCT over large arrays, but once again, MCT can be tuned to shorter wavelength cut-off, which will strongly improve its metallurgic quality. Read-out noise can also be better for silicon, as conversion factor can be higher (smaller stray capacitance for photodiode). InSb would feature the highest stray capacitance. In addition, silicon approach is compatible with the use of photoMOS (see Figure 2.1.3-h and Figure 2.1.3-i) which would feature even lower read out noise than photodiode. This is due to the fact that

conversion factor for the photoMOS is not defined by the detection diode stray capacitance but by the conversion floating diode, which capacitance can be tiny for deep submicron processes (around 1fF), offering conversion factor as high as 150µV/e-. In addition, photoMOS would allow real correlated double sampling operation (CDS) which is not possible for photodiode. On the other hand, feasibility of multiple read-out using photoMOS needs further assessment. Table 2.1.3-c summarises the trade-off between Si, InSb and MCT.

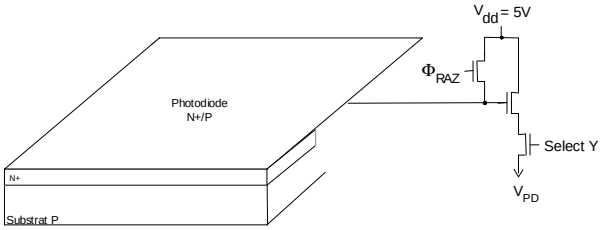


Figure 2.1.3-h:
Photodiode intra-pixel architecture

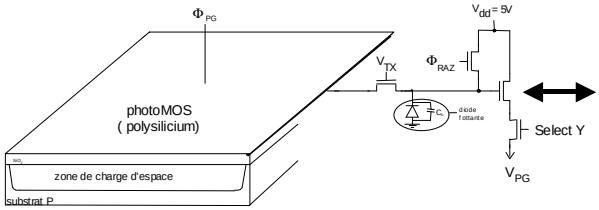


Figure 2.1.3-i:
PhotoMOS intra-pixel architecture

Criteria	Si	InSb	MCT	Comments (*)
Quantum efficiency	++*	+	+	excellent for deep substrate
Dark current	++	-	+*	for MCT tuned for short wavelength
Read out noise	++	+*	+	lower conversion factor and noise model problem for "high energy" photons
Glitches sensitivity	High*	Low	Low	due to deep substrate: need trade-off
Cosmetics	++*	+	+	even if worse than standard substrate
PhotoMOS possibility	Yes	No	No	
Thermal mismatch	No	Yes*	Yes*	well mastered for format up to 1k x 1k
Ambient T° operation	Yes*	No	No	system requirement to be assessed
Maturity	+*	++	++	no foreseen difficulty for Si
Total cost (*)	Medium	High	High	development plus production

Table 2.1.3-c: Detection material trade-off for NGST visible (0.5 to 1 µm) instruments

The major drawback of the silicon option is its sensitivity to cosmic rays for deep silicon. As the deep silicon is used in order to improve the quantum efficiency in the 0.8 to 1 µm range, a trade-off shall be made in order to define the optimum acceptable thickness. It should be recalled that the use of CMOS read-out would also help to restore the useful information after high energy particle hits. As nearly all the other criteria are better for silicon, and even if further assessments are needed (for instance about residual images effects and cosmic rays time constant), this material is today proposed as baseline for the NGST visible instrument.

2.1.3.1.4 CCD versus silicon hybrid

As this material has been selected, a global trade-off (including material and read-out circuit) between the silicon CCD and the silicon CMOS (i.e. silicon detection layer hybridised on top a CMOS read-out circuit) needs to be performed. Table 2.1.3-d summarises the trade-off.

Despite less technological maturity and some unknowns about silicon optical properties at 30K, and based on its numerous advantages over the CCD, silicon hybrid is selected as the baseline detector array for the NGST visible instrument, mainly because it can be operated at 30K and would feature low power dissipation.

Criteria	CCD	Silicon hybrid	Comments (*)
Operating temperature	above 100 K*	30 K is OK	Low CTE below 100 K
Thermal dissipation	marginal	compliant	for all focal planes
Minimum pixel size	9 µm	15 µm	for high performances
Quantum efficiency	high for deep silicon	high for deep silicon	lower at 30K than at 100K
Dark current	OK*	OK	higher at 100K
Noise/detector dynamic	3 e- rms /3 10^5 e-	<3 e- rms/10^5 e^{-*}	open trade-off
Multiple read out during integration	no	yes *	method/power dissipation to be assessed
Dynamic improvement	no	yes *	using multiple read out
Analogue binning	yes	no	not a driver for NGST
Windowing	no	yes	not a driver for NGST
Snap shot mode	no *	yes	CCD need a shutter
Cosmic ray removal	no	yes *	using multiple read-out
Sensitivity to radiation	high *	low	mainly CTE degradation
Glow effects	no	possible*	amplifiers are switched on for multiple read-out
Associated electronics	complex	simple	
Thermal accommodation	cryostat plus baffle	simple	
Technology status	mature	in development *	no problem expected
Offset (FPN)	no	yes *	easy to calibrate and stable
Linearity	Excellent	Acceptable *	easy to calibrate and stable
Blooming	need special features	implicit	
Cosmetics	Excellent	Good *	excellent for photoMOS

Table 2.1.3-d: Trade off between CCD and silicon hybrid

2.1.3.1.5 Potential manufacturers

Monolithic APS detectors for standard applications can be designed by teams with a good experience in full-custom circuits design. The silicon hybrid does however feature specificities with regard to monolithic APS. The main one is the need for hybridisation. Even if there are teams working on hybridisation method alternatives, the most standard one for high performances is the indium bump method used for cooled IR hybrid focal planes. Therefore, potential manufacturers for silicon hybrid arrays can be the same as the ones building the cooled IR array, replacing the IR exotic material (MCT, InGaAs, InSb, QWIP, Si:x…) by the silicon photodetector layer. In addition, the use of silicon hybrid for low temperature operations would require that the manufacturer have experiences with CMOS process characteristics at low temperature. Once again, this is the case for cooled IR array manufacturers.

In the US, there are at least two IR manufacturers -RSC and RIRCE-, which have announced their interests and developments in this field. RSC have already manufactured demonstrators and preliminary data sheets are now available (see Figure 2.1.3-j). In Europe, MMS have technical exchanges with LETI/LIR, which is involved in several new technological developments for cooled and uncooled IR arrays. Figure 2.1.3-b shows a field of indium bump manufactured by LETI/LIR with pitch equal to 20 µm, and Figure 2.1.3-k shows a wafer of read-out circuit prior to collective hybridisation, which has been developed by LETI/LIR. Using such method, 100% interconnection efficiency can be guaranteed, with a proven reliability of the bonds due to true welding and a self alignment better than +/1µm. Concerning the silicon detection layer (photoMOS or photodiode), no specific problem is expected about its realisation, and several companies or institutes can be foreseen in Europe or in the United States.

Whatever the selected manufacturer for these large focal planes, electro-optics selection of the best devices will require a very accurate and operational test set-up facility. Indeed, representative conditions are required and a lot of electro-optical parameters need to be measured. In addition, tests using real astronomical targets would be an excellent way to determine the real capabilities of these new focal planes.

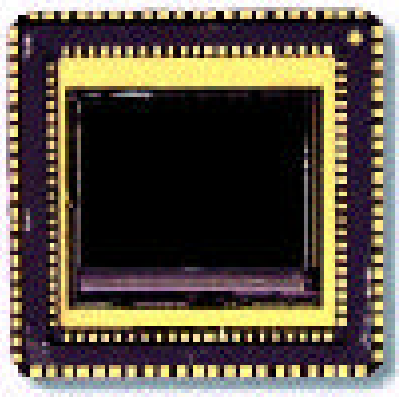


Figure 2.1.3-j:TMC 6600 Hybrid Visible Silicon Imager from RSC

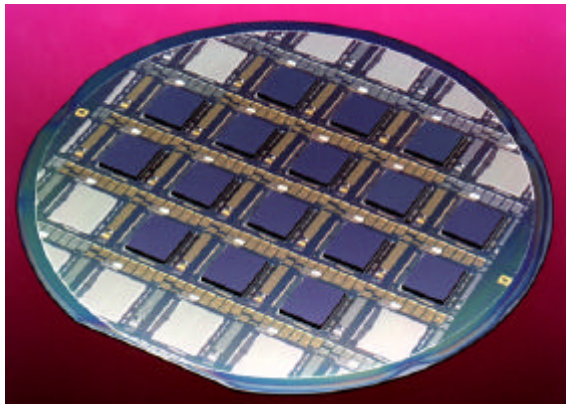


Figure 2.1.3-k: IR infrared read-out CMOS circuit on a wafer (from LETI/LIR)

2.1.3.2 Detector design

2.1.3.2.1 Choice of pixel size

The choice of the best pixel size results from several constraints. Table 2.1.3-e shows that some aspects would require small pixels while others would require larger pixels. From this trade-off, a pixel size of 15 µm was considered as the best compromise.

Field	item	optimised for	comments
System	Mechanical & thermal	Small pixel size	- easier accommodation for smaller focal plane - no impact on power dissipation
	Opto-mechanical	Medium pixel size	- manufacturing more difficult for pixels <10µm - accommodation more difficult for pixels >20µm
	Electrical	Neutral	No impact on FPA lead count
Technology	Yield	?	- easier intra-pixel connections for large pixels - less sub arrays for small pixels
Electro-optics	Dark current	Small pixel size	Smaller volume for thermal generation
	Read-out noise	Small pixel size	Only for photodiode (higher conversion factor)
	Cosmic ray	Small pixel size	Lower sensitivity as lower peripheral surface
	Uniformity	Large pixel size	Better intra-pixel uniformity for large pixels
	Cross-talk	Large pixel size	Related to diffusion length
	Saturation	Large pixel size	Only for photodiode (smaller conversion factor)

Table 2.1.3-e: Trade-off about optimised pixel size

2.1.3.2.2 Focal plane arrangement

As the 8k x 8k for WFI is the largest desirable FPA, the following paragraph will describe the proposed approach for this array, but this approach is also applicable to the 4k x 4k focal planes of the two other instruments. Based on pixel pitch equal to 15 µm, the 8k x 8k FPA represents a 120 x 120 mm^2 surface. The choice for the arrangement of the focal plane would mainly depend on the maximum size, which can be achieved for the sub-arrays used in order to manufacture the focal plane assemblies.

As it is not expected to achieve sufficient technological yield to build a full-monolithic detector, and as hybridisation operation would also not be possible for such a large FPA, mechanical butting (see Figure 2.1.3-l) is proposed in order to cover the full focal plane.

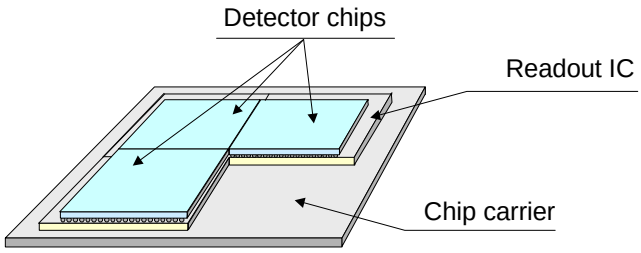


Figure 2.1.3-l: Mechanical butting of sub-arrays is proposed for NGST focal plane assembly

Due to limitation of the field of wafer steppers, which are used for deep sub-micron processes, the maximum die size can be around 20 x 20 mm^2. With this limitation, 36 sub-arrays are required in order to build the entire 8k x 8k FPA. As can be seen on Figure 2.1.3-o, peripheral sub-arrays will be 2 or 3-side butted. However, the central ones will be 4-side butted, which is a problem as there is no available room in order to make the connection with the outside. Indeed, there is a need of room for bond pads is order to allow this connection. With this interconnection approach, there will be non negligible dead zone at butting gap (between 1 to 2 mm), which is bad in terms of FPA fill factor, even if dithering mode can be used for imaging observation. An alternative can be the use of intra-connection (or Z connection) today in development, in order to allow connections at the back of the sub-arrays. In this case, the dead zone can be strongly reduced (see Figure 2.1.3-m and Figure 2.1.3-n). Even if breadboards have been demonstrated by LETI, maturity of this technology needs to be assessed with respect to NGST planning.

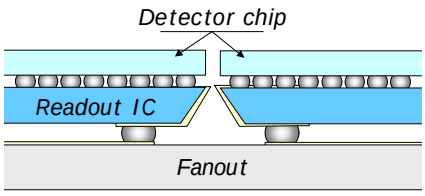


Figure 2.1.3-m:
Intra-connection principle

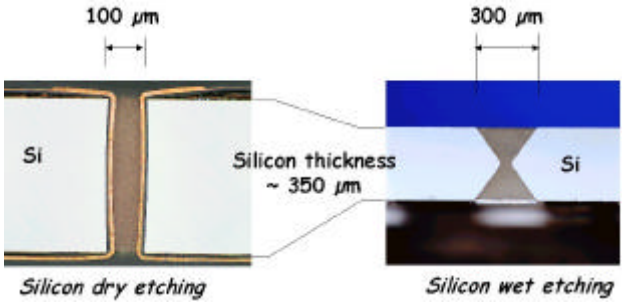


Figure 2.1.3-n:
Intra-connection state of the art at LETI/LIR

Another way can be to use non deep sub-micron processes for the read-out production, say 1 µm, which allow to use wafer-scale photolithography with no size limitation. In this case, circuits of reasonable size in order to guarantee a good yield can be manufactured. Figure 2.1.3-p is based on 2k x 2k arrays, with 20 µm pixel pitch. The pixel pitch would be larger because the more conservative CMOS process would need more space in the pixel in order to integrate all the necessary functions.

Finally, the last approach can be the use of optical stitching in order to make larger sub-arrays by moving the wafer stepper with excellent accuracy (0.05 µm is reported by RSC) in order to build a large monolithic sub-array using deep sub-micron processes (which are expected to feature an excellent yield). In order to keep one small side due to hybridisation constraint it is proposed to build 20 x 60 mm^2 sub-modules, which avoids central sub-modules, and simplifies their inter-connections. Figure 2.1.3-q represents the 8k x 8k focal plane, made of 12 sub-modules of 4000 x 1330 pixels. Three of these sub-modules are proposed for the 4k x 4k FPAs of the other visible instruments. This is the today proposed baseline, but Figure 2.1.3-r configuration is kept as back-up in case of too short

development time with respect to NGST schedule. Indeed, 2k x 2k read-out circuit blocks should already be designed and manufactured for the IR instruments

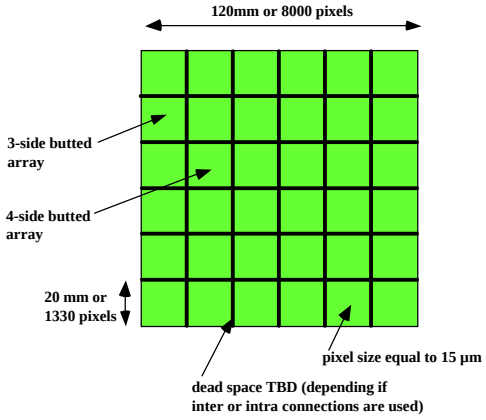


Figure 2.1.3-o:
Expected topology with deep sub-micron process and 20 x 20 mm^2 sub-array size.

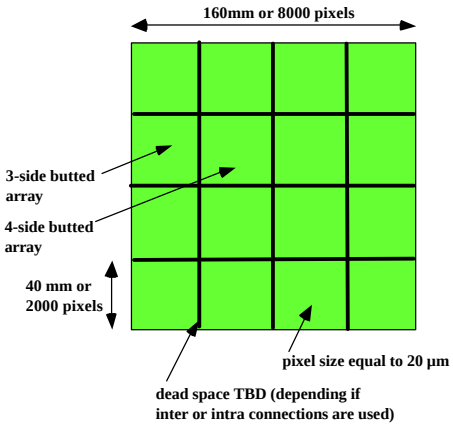


Figure 2.1.3-p:
Possible use of 2k x 2k sub-arrays using conservative processes

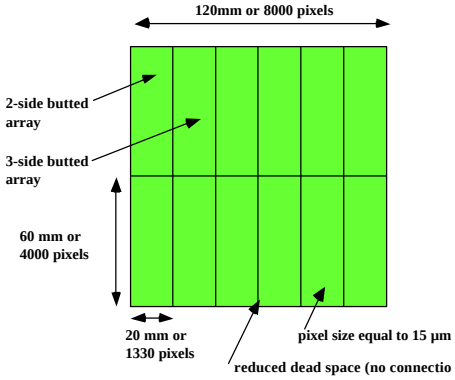


Figure 2.1.3-q: 8k x 8k FPA baseline topology optical stitching and no central sub-modules

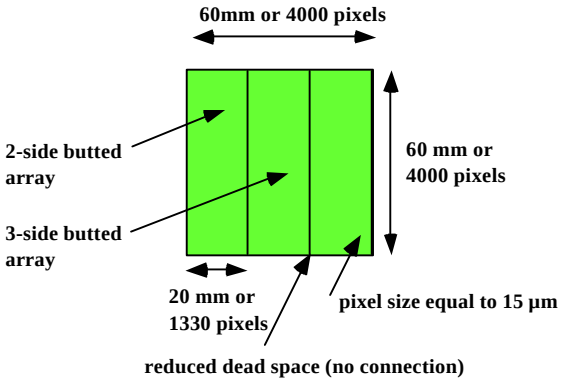


Figure 2.1.3-r: 4k x 4k FPA baseline topology optical stitching and no central sub-modules

The sub-module packages have to be mechanically robust and have to guarantee a good flatness, in order to be compliant with required focal plane planarity. As butting is mandatory, this package should also protect the edges of the detectors and should minimise the distance between edges. A connector would be necessary in order to provide the electrical interfaces. The packages should also feature adequate thermal properties. It is assumed that the focal plane will not feature any particular geometry and/or will not require any adjustment and therefore, the sub-arrays will be mounted with all the detection lines and columns parallel one to each other's. More complex assemblies would be possible, for instance a curved FPA in order to avoid optical field flatteners, but with risk and cost impacts. The reparability of defective sub-arrays should be addressed, as it impacts the complexity of the assembly. The reliability / redundancy should also be discussed during the study follow-on, as it would have a large impact about leads count. Finally, the cleanliness policy should be defined in order to select the integration of a window in front of the focal plane and/or to decide whether or not the FPA will be in vacuum or in neutral nitrogen. Qualification of the assemblies will requires special tests, using qualification models, such as vacuum baking, thermal cycling, thermal chocks and mechanical vibrations and chocks.

2.1.4 Instrument overall configuration

2.1.4.1 Wide Field Imager

The Wide Field Imager overall configuration is shown in Figure 2.1.4-a.

2.1.4.2 High Resolution Imager

The High Resolution Imager overall configuration is shown in Figure 2.1.4-b.

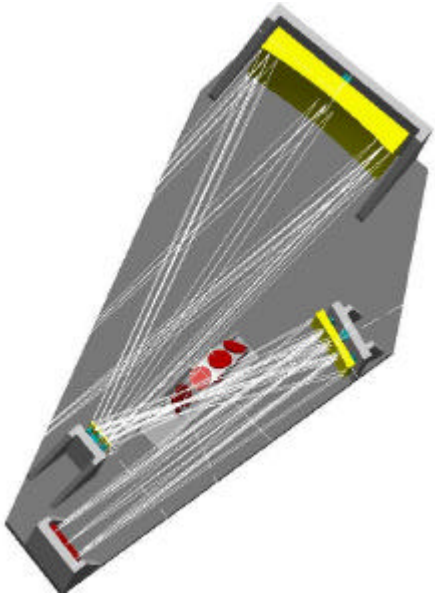


Figure 2.1.4-a: Overall configuration of the Wide Field Imager

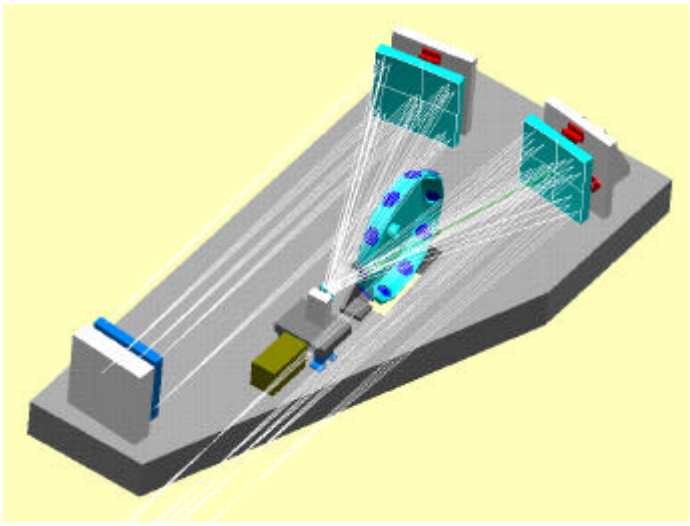


Figure 2.1.4-b: Overall configuration of the High Resolution Imager

2.1.4.3 Integral Field spectrograph

The Integral Field Spectrograph overall configuration is shown in Figure 2.1.4-c.

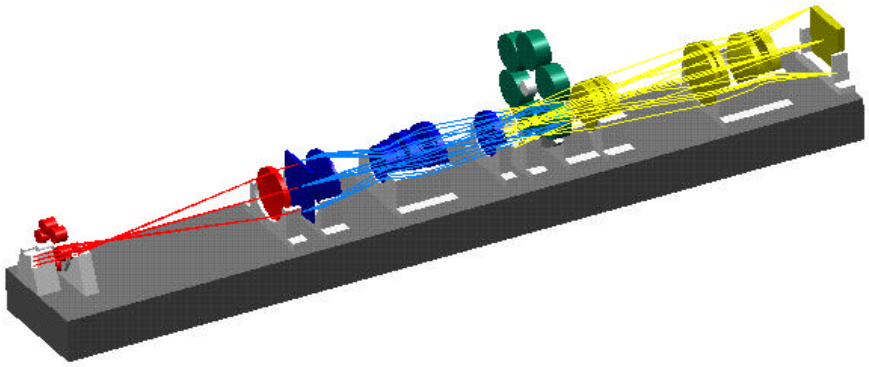


Figure 2.1.4-c: Overall configuration of the Integral Field Spectrograph

2.1.5 Structure design and mechanical analyses

The exact positioning of the individual optical components during operation is an essential condition for a satisfactory optical function of the investigated instruments. The structures of the instruments have to ensure, that the components are fixed with sufficient stability under the expected environmental influences. The influences of temperature and gravity are regarded as most critical for the structure dimension stability. Temperature changes generate undesired displacements due to the thermal expansion of the structure material. The thermal deformations are particularly critical, if the geometrical temperature distribution is uneven. The uniformity of the temperature distribution is strongly influenced by the thermal conductivity of the structure material. The influence of gravity during instrument integration and adjustment on ground can lead to misalignments in the orbit. Density and stiffness of the structure material influence the gravity deformations. The same material properties determine the mechanical structure eigenfrequencies, which are essential for the structure stress levels during launch.

A suitable structure material should have a low coefficient of thermal deformation, high thermal conductivity, low density and high stiffness. Among a number of candidate materials C-SiC and SiC were selected as most promising structure materials for the planned application. The materials combine low thermal deformation with high thermal conductivity, low density and high stiffness. A disadvantage of C-SiC and SiC are their brittleness. Therefore connection parts of INVAR material were foreseen in the load introduction areas. The thermal deformations of INVAR and C-SiC/SiC match quite well in the applicable temperature range and consequently the structure deformations during cooling down will be even to a far extent.

A minimum eigenfrequency of 60 Hz was desired for all instruments. Static accelerations of 30g in x-, y- and z-direction were used for investigation of mechanical stresses. The used material properties of C-SiC are listed in Table 2.1.5-a.

Young's modulus	230000 MPa
Poisson's ratio	0.25
density	2700 kg/m^3
allowable stress	100 MPa

Table 2.1.5-a: Properties of C-SiC material

2.1.5.1 Wide Field Imager

A lightweightened C-SiC plate was selected as the primary structure for the WFI instrument (see Figure 2.1.5-a). The optical components are mounted to the baseplate via C-SiC brackets. The optical C-SiC mirrors are fixed to the brackets by 3 point isostatic mounting using flexible INVAR mounts.

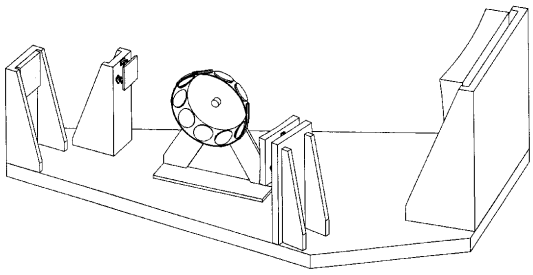


Figure 2.1.5-a: Design of the WFI instrument

A finite element model was established as basis for a preliminary analytical investigations. The baseplate was idealised by QUAD4 plate elements. The properties of the stiffening ribs were simulated by smeared plate properties. The mounting brackets and the optical components were considered as rigid solid elements. The actual mass distribution was represented by adapted material

densities. The model was isostatically supported at 3 points. With the described model an eigenfrequency analysis was performed. The lowest eigenfrequency is 61 Hz.

Acceleration loads of 30g in x-, y- and z-direction were calculated as static load cases. The according maximum VON MISES stresses within the baseplate are summarised in Table 2.1.5-b which shows, that the allowable stress of 100 Mpa is only slightly exceeded in the case of 30g acceleration in z-direction. A reduction of the stress level can be reached without mass increase by proper tuning of skin thickness, total plate thickness, rib thickness and rib distance.

location	30g in x-direction	30g in y-direction	30g in z-direction
skin	55	40	121
ribs	51	37	117

Table 2.1.5-b: WFI structure maximum VON MISES stresses in Mpa

The following aspects in the mechanical field of the Wide Field Imager are rated as critical areas:
- the refocus mechanisms (the experience for operating at a temperature of 30K is very limited)
- the large structure components out of C-SiC (the manufacturing of lightweighted parts with such large dimensions is not proven)
- the mechanical load introduction into C-SiC parts
- the material is rather fracture sensitive
- thermal stresses caused by load introduction elements can be critical
- the knowledge about the stability of connections is insufficient

2.1.5.2 High Resolution Imager

A lightweighted C-SiC plate was selected as the primary structure for the HRI instrument (see Figure 2.1.5-b). The optical components are mounted to the baseplate via C-SiC brackets. The optical C-SiC mirrors are fixed to the brackets by 3 point isostatic mounting using flexible INVAR mounts.

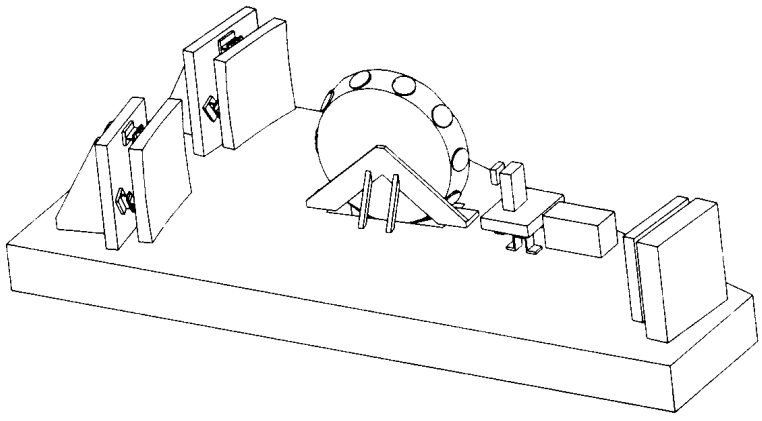


Figure 2.1.5-b: Design of HRI instrument

A finite element model was established as basis for a preliminary analytical investigations. The baseplate was idealised by QUAD4 plate elements. The properties of the stiffening ribs were simulated by smeared plate properties. The mounting brackets and the optical components were considered as rigid solid elements. The actual mass distribution was represented by adapted material densities. The model was isostatically supported at 3 points.

With the described model an eigenfrequency analysis was performed. The lowest eigenfrequency is 271 Hz. Due to the small instrument size the eigenfrequencies are high.

Acceleration loads of 30g in x-, y- and z-direction were calculated as static load cases. The according maximum VON MISES stresses within the baseplate are summarised in Table 2.1.5-c which shows, that due to the small instrument size the applied stress levels are clearly lower than the allowable stress of 100 Mpa.

location	30g in x-direction	30g in y-direction	30g in z-direction
skin	27	22	22
ribs	25	18	24

Table 2.1.5-c: HRI structure maximum VON MISES stresses in MPa

The following aspects in the mechanical field of the High Resolution Imager are rated as critical areas:
- the refocus mechanisms (the experience for operating at a temperature of 30K is very limited)
- the mechanical load introduction into C-SiC parts

2.1.5.3 Integral Field spectrograph

A lightweighted C-SiC plate was selected as the primary structure for the IFS instrument (see Figure 2.1.5-c). The optical components are mounted to the baseplate via C-SiC brackets.

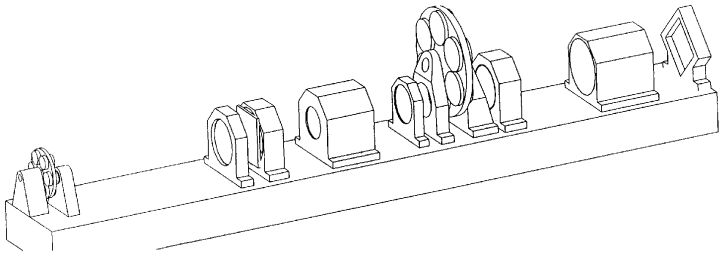


Figure 2.1.5-c : Design of the IFS instrument

The following aspects in the mechanical field are rated as critical areas:
- the mechanical load introduction into C-SiC parts
- the suspension of the lenses
- the thermal lens deformations do not match to the primary structure material
- the available load introduction area is strongly limited, because large areas are needed for the optical function
- severe requirements are expected for the allowable lens deformations
- the adjustment of individual lenses within a lens package
- bad accessibility due to compact arrangement
- complication due to requirements for lens suspension

2.1.6 Thermal engineering

The major thermal problem for the visible instruments of NGST is the particular environment of the ISIM with an operating temperature of 30 K needed for the infrared detector performances. The visible instrument has to operate at this cryogenic temperature and in addition has to minimise its power dissipation and heat leaks in order to cope with the current radiator size.

The cooling is a passive cooling of the complete ISIM and the contribution of the visible instrument to ISIM heat load budget is limited to instrument internal dissipation (detectors and mechanisms) and heat leaks through the harness. Heat exchanges between the visible instrument and the rest of the ISIM either by radiation or by conduction are only internal exchanges for the ISIM standpoint and do not contribute to the ISIM heat load.

The total power dissipation (detectors and mechanisms) at 30 K, presented in the power budget of section 2.1.10.2 is 5.5 mW for WFI and 2.1 mW for HRI and IFS.

Assessment of heat leaks through harness can be done with 440 (for WFI), 150 (for HRI) and 120 (for IFS) stainless steel wires of 100 µm diameter with 3 m long between 30 K and 200K. Assuming a mean thermal conductivity of 10W/m/K for stainless steal (see Figure 2.1.6-a), the heat leaks are respectively of 2 mW, 0.7 mW and 0.5 mW. A thermalisation of the harness at 100 K after a distance of 2 m would reduce this heat leak by a factor 4.

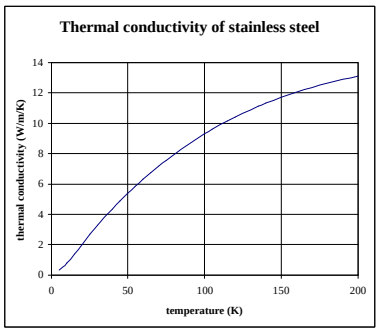


Figure 2.1.6-a: Thermal conductivity of stainless steel at cryogenic temperatures

Some thermal protections will be likely necessary to minimise the gradient inside ISIM. Some baffles will be likely necessary too to minimise the straylight towards the long wavelength detectors.

2.1.7 Mechanism design

2.1.7.1 Filter wheel mechanism (for WFI and HRI)

The general principle of the filter wheel mechanism of the HRI is shown in Figure 2.1.7-a. The filter wheel mechanism of WFI is based on the same concept but with larger dimensions.

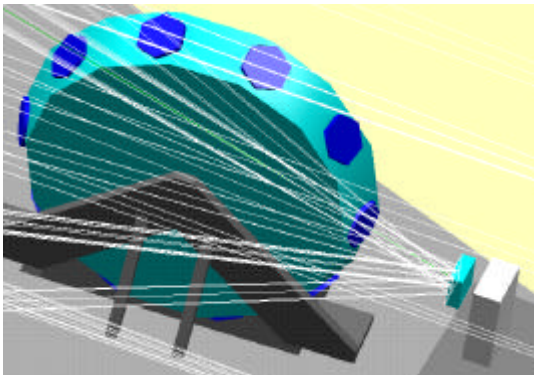


Figure 2.1.7-a: The filter wheel mechanism

The proposed design concept for the filter wheel mechanisms is derived from the ISOPHOT filter wheel (see Figure 2.1.7-b). Inside the wheel flange a step motor is positioned. For every filter change one step is performed, a half step for acceleration and a half step for deceleration. The positioning is secured via a dent by a spring induced centring between ball bearings. Position sensors are included. Since the wheel is perfectly balanced, no launch lock is needed. The actual position after launch is detected by sensors. The ISOPHOT filter wheel has a diameter of 110 mm and is operating at 4K. The dissipation of the mechanism during filter change is 4 mW per step. In non operating condition the dissipation is zero. For larger filter wheels the dissipation per step will increase approximately quadratic with the wheel diameter.

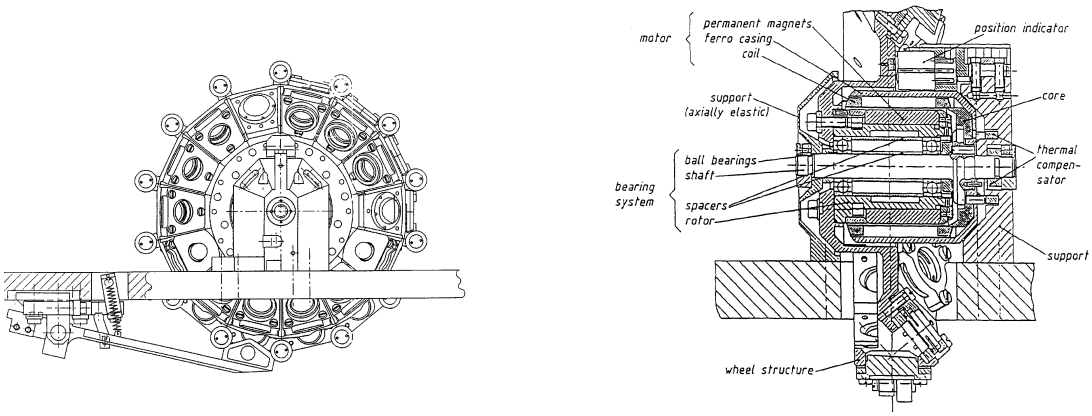


Figure 2.1.7-b: ISOPHOT filter wheel

2.1.7.2 Refocusing mechanism (for WFI and HRI)

As refocusing actuator for the different instruments a sequential working linear inchworm type is foreseen, which is under development at Burleigh Instruments for NGST and is already working as prototype. The essential actuator data are a resolution better than 20 nm, a stroke larger than 6 mm, a holding force about 4.5 N, a heat dissipation of 5 mW during calibration and of 0.05 mW during operation. The size of the prototype is about 50 x 50 x 100 mm.

Due to the limited holding force of this actuator type hold down and release mechanisms have to be foreseen for the refocusing mirrors. The principle of such a mechanism is shown in Figure 2.1.7-c. During launch the mirror support frame is fixed by a latch with conical stud. After launch the release is actuated via toggle by a gear motor connected to an eccentric pulley. After release the mirror is completely free. The mechanism works without pyrotechnics and has no dissipation in power-off condition. The gear motor is redundant and combines low mass with high torque.

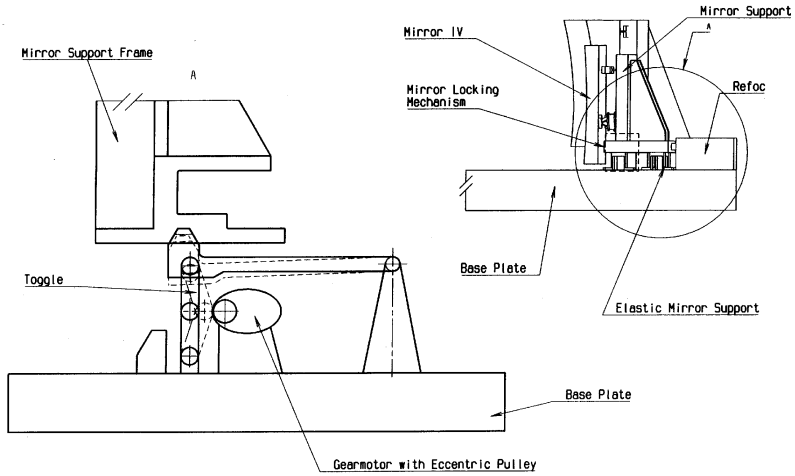


Figure 2.1.7-c: Hold down and release mechanism for refocusing mirrors

2.1.8 Electronics architecture

2.1.8.1 Read out strategy

Before defining the electronics architecture, it would be necessary to define which strategy will be selected for the read-out of the arrays. There are mainly four read-out strategies that can be foreseen: fast sampling, correlated double sampling, Fowler sampling and continuous sampling up the ramp.

Fast sampling consists in the read-out of a single sample referenced to electrical ground. This method is well adapted to high speed applications that are not demanding in terms of radiometric performances. Therefore, this method is no more considered for NGST.

Correlated double sampling (CDS) is the most used read-out technique, for both CCDs and APS. CDS consists in sampling both reference level and useful signal in order to remove the kTC thermal noise of the output floating diode. In addition, when performed with minimum time gap between the reference sampling and the useful signal sampling, CDS will allow to reject low frequency noises and drift (such as 1/f noise and thermal drifts). For a CCD, the reference level corresponds to the output node floating diode reset. Interest of multiple sampling is therefore limited for CCD as it requires an increase of the electrical bandwidth for the video chain. Therefore, what can be gained by multiple sampling will be lost due to the noise bandwidth increase.

For a CMOS read-out, the situation is not the same, as each pixel has its own floating capacitance. In addition, the situation is different for photoMOS and photodiode. For a photoMOS (see Figure 2.1.3-i) and after gate reset, charges are integrated under the MOS gate before reading. At the end of the integration, and just before injection of the charges from the gate to the floating diode, the diode is reset and this reset level is stored using CDS stage (on or off-chip). Then, transfer occurs, and the useful signal is memorised. Figure 2.1.8-b represents this true CDS operation, as the samples are close one from each other in the time domain, and as the CDS is applied on two samples from the same frame.

Things are different for a photodiode (which is the standard situation for hybrid FPA used for IR detection). Indeed, for a photodiode, the integration and the charge to voltage conversion are provided by the same stray capacitance of the photodiode. Therefore, if the reset level of the corresponding pixel is sampled, the useful signal sampling will be made at the end of the integration time, which means that the low frequency noise and drifts will be poorly rejected (see Figure 2.1.8-a). If the reset level of the next pixel is sampled, then the delay between the two samples will be short, but the kTC thermal noise will not be removed and their noises will not be correlated.

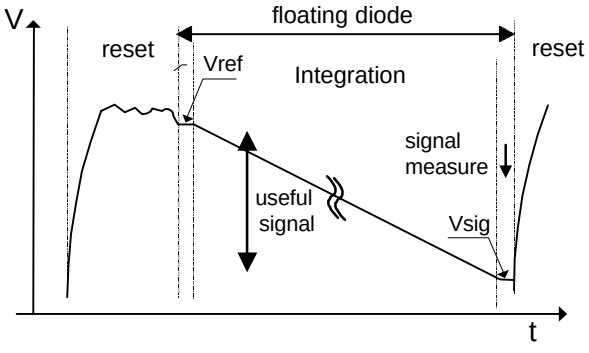


Figure 2.1.8-a:
Reference and signal sampling operation for a photodiode

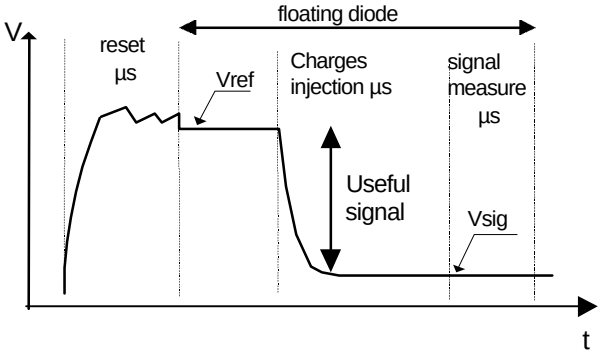


Figure 2.1.8-b:
Reference and signal sampling operation for a photoMOS

Fowler sampling method main goal is to minimise the electronics read-out noise. First, the line (or the full array) is reset. Then a series of N/2 samples are fast acquired at the beginning of the observation (N typically between 16 and 64). The same N/2 samples are acquired at the end of the integration time (see Figure 2.1.8-c for a photodiode). Up to the ramp sampling offers several advantages over Fowler

sampling. First, the line (or the full array) is reset. Then, N+1 samples are acquired at regular intervals during the observation, including one at the very beginning (see Figure 2.1.8-d for a photodiode).

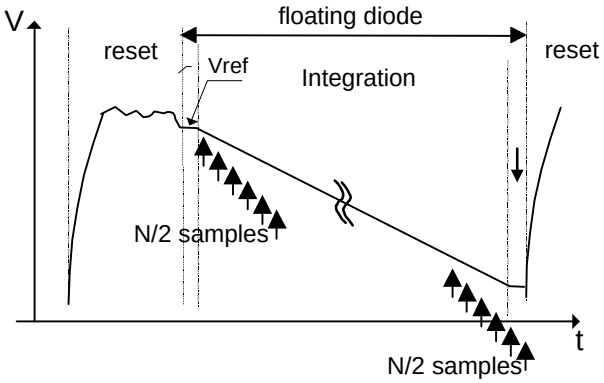


Figure 2.1.8-c:
Fowler sampling method

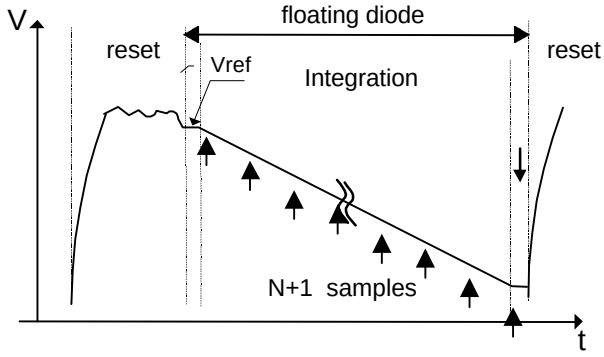


Figure 2.1.8-d:
Up to the ramp sampling method

In order to choose between the different methods, the trade-off has to be made at system level, applying the good weight for the various criteria. Table 2.1.8-a presents a preliminary trade-off. As the SNR at very low flux (i.e. not in photon noise regime) should be one of the major criteria for NGST (at least for the spectroscopic instruments), multiple sampling methods are preferred rather than the CDS method. As cosmic hit rejection is an other very important issue, particularly for thick silicon layer, sampling up to the ramp method is today selected as baseline.

Criteria	Correlated Double Sampling	Fowler Sampling	Up to the ramp sampling	Notes (*)
Read out noise reduction	Medium	Good *	Excellent	For white read out noise larger electrical bandwidth
Photon noise regime	Excellent	Good *	Good *	Noise from samples is correlated
Low Frequency noise regime	Excellent	Medium *	Medium *	LF noise could be reduced by "double reset"
Dynamic enhancement	No	No	Yes *	Requirement to be assessed
Cosmic rays rejection	No	No	Yes *	Very important issue for NGST
Processing implementation	Easy	Simple	Complex *	Limitation to be assessed
Compatibility with photodiode	No	Yes *	Yes *	with today ROIC architecture
Compatibility with photoMOS	Yes	TBC *	TBC *	Should be OK
Thermal dissipation and glow effects	Optimum	Not optimum *	Not optimum *	Amplifier only on during signal read-out
Observation efficiency	OK	Not optimum	OK	For Fowler samples, large number of read-out at the end of the integration time
Synthesis & conclusion	Real CDS not feasible with photodiodes	Noise performances depending of duty cycle	Today baseline better for time varying phenomena	

Table 2.1.8-a: Preliminary trade-off between the possible read-out strategies

2.1.8.2 Electronics architecture

The electronics architecture will be a challenging development for two main reasons. First, there is a need for very low noise in order to keep an excellent SNR for faint targets, even when a long distance (nearly 3 m) would separate the focal plane unit (FPU) from the medium temperature electronics (>200K). Second, the FPU has to dissipate as low power as possible, which means that additional electronics in the FPU has to be minimised. An important choice for mass/power budget is the decision about the redundancy philosophy. To secure the instrument operations, and in a conservative way, a full cold redundancy has been selected, (all boards duplicated and redundant items switched off in nominal mode). No redundancy of pre-amplifiers, if required, is planned (even cold one) in order to avoid signal degradation.

The number of preamplifiers (on chip or off chip) has been minimised to reduce thermal dissipation, which leads to a relatively high pixel rate per FPA output with nevertheless an acceptable performance (even if cosmic ray rejection would need further investigations). Based on a minimum of 18 video chains, one 8k x 8k and two 4k x 4k arrays to be read out, it would be possible to read-out 35 samples at 200 kpixels/s output rate for each pixel during the 1000 s of elementary integration time.

On the other hand, as the exact topology of the focal plane and the exact number of outputs per elementary sub-array are not frozen, the number of video chains taken into account for the power and mass budgets is higher in order to get a conservative limit. Table 2.1.8-b summarises the assumptions for both performances and mass/power budgets for the 3 focal planes. Video signal multiplexers are not considered further, as efficient multiplexing requires cryogenic devices. This would increase the complexity of the sequencer clocking, and would degrade the instrument performance.

FPA configurations	Total number of outputs	Impact on performances	Impact on budget	Choice
18 sub-arrays with one output per sub-array	18	+	++++	baseline for performances
18 sub-arrays with three outputs per sub-array	54	+++	++	-
24 sub-arrays with one output per sub-array	24	++	+++	-
24 sub-arrays with four outputs per sub-array	96	++++	+	Baseline for budget

Table 2.1.8-b: Selection of the baseline configurations for both performances and mass/power budget assessments (for the three focal planes)

Operating electronics

The operating electronics include the functions that are needed in order to operate the detectors and the mechanisms. The corresponding boards are located in the medium temperature area of the satellite. Thanks to their CMOS ROIC, the visible focal planes are easy to operate when compared to CCDs. Indeed, the sub-arrays will only require few 0 to 5 volts clocks (say 3 to 6) and few 0 to 5 volts DC bias (say 10-12). In addition, each sub-array will feature one temperature probe and one on-chip reference in order to reject temperature fluctuation or spikes detection. Clocking diagram for ROIC read-out is expected to be simple, with full frame and line reset for the initialisation of the integration time. As the visible FPAs are not required for the FSM control (more stars would be observed using the NIR cameras), no windowing capabilities are planned today for the visible FPAs.

The first proposed type of board in a control board, which includes a satellite interface (1553), a FPGA sequencer, the clocks for the Focal Plane Unit (FPU), the command for the stepper motors and the operation and control of the temperature probes and other sensors/housekeeping's (position sensors, voltage and current monitoring…). The second type of board is a power supplies board, which provides the FPU, supplies, the electronics supplies and the command for the inchworms. Passive

filtering of the DC bias signals will be implemented in the cold part of the instrument, close from the sub-arrays in order to reject EMI extra-noise. It is expected that in flight timing diagram and DC bias reprogramming would be necessary, mainly in order to reduce the residual images and the cosmic ray effects, using bias and timing diagram/cycles optimisation.

Video electronics

The video electronics has to drive the useful analogue signal coming out of the sub-arrays along nearly 3 meters and has to convert this analogue signal into a digital one. Even if the properties of the harness are not known at this time, and in addition to expected excellent linearity and temporal stability/reproducibility, it is expected that the matching between the output of the sub-arrays and the harness would need to avoid:

- too long rise and fall time constants in order to allow stabilisation of the useful signal before digital conversion,
- high resistance along the 3 meters harness in order to avoid electro-magnetic coupling which would introduce extra noise,
- loss of useful signal amplitude by non optimised coupling.

Therefore, there is a need for a buffer located between the sub-array internal data bus and the 3 meters length of the cable in order to reduce the resistance of the output stages of the sub-arrays, assuming that the harness would also need optimisation in view of reducing its capacitance. The use of previous or on going space programs dealing with cryogenic links between IR detectors and deported electronics (ISO, SIRTF, Meteosat…) would be a good starting point for the study and development of the NGST harness. Cryogenic harness developed for ISO by GORE (using stainless steel wires) should be usable for NGST and is considered as the baseline.

The buffer can be on-chip or off-chip, and on-chip solution would be preferred as its power consumption is expected to be lower than an off-chip one. Today existing on chip buffers for infrared arrays offer 100 ohms for the output resistance, with capability for 300pF driving. If confirmed for NGST, and assuming a resistance matching at the other extremity of the harness, this would minimise any extra power dissipation in the cold part of the instrument. If higher output resistance is featured by the selected sub-arrays, an external preamplifier needs to be added on each video line. As the power dissipation for each off-chip pre-amplifier is expected to be around 1mW, this could be acceptable for the 18 video chains configuration (see Table 2.1.8-b) but is non acceptable for the 96 video chains configuration. The main drawback for on-chip preamplifier is the potential risk of glow effect. An other possible drawback could be the non optimisation of its noise spectrum when compared to a discrete pre-amplifier, to be build around MOSFET transistor type well known for their good noise behaviour at cryogenic temperature. However, even with a conservative 30nV/Hz½ noise density and based on an electrical bandwidth equal to twice the pixel rate –400 kHz- the electronics noise is around one e- thanks to the baseline conversion factor of 16 µV/e-.

As the signal provided by the video output of the sub-arrays are internally sampled (which is not the case for CCDs), the electrical bandwidth can be reduced to approximately twice the pixel rate (depending of the rise and fall time constants). In the medium temperature zone, the video chain would mainly be composed of a differential amplifier and the ADC. 16 bits ADCs are required in order to minimise the quantization noise, which would be around 0.4 electrons for 100 000 electrons dynamic. ADC could not be located in the cold part of the instrument, which would be the best way to avoid signal degradation for its transmission along the 3 meters of the harness thanks to digital transmission. Indeed, such a device is expected to dissipate too much power (at least few tens of mW for future generation ADC, based on a conversion rate equal to 200 kpixels/s), and cryogenic 16 bit ADC developments should be in a too early stage with respect to NGST schedule. Space qualified 16 bits operated at 200 kpixels/s should be available as other existing space programs (COROT, IASI…) and future programs (GAIA…) would require the same type of device.

Manufacturing of the video electronics would be possible using hybrid electronics (see Figure 2.1.8-e). This would allow to mount 32 video hybrids per board, which means that 3 boards will be sufficient for the most conservative approach, i.e. 96 chains (see Table 2.1.8-b). Video electronics should also contain its own internal electrical stimulus in order to check calibration changes and normal operations (on ground and in flight).

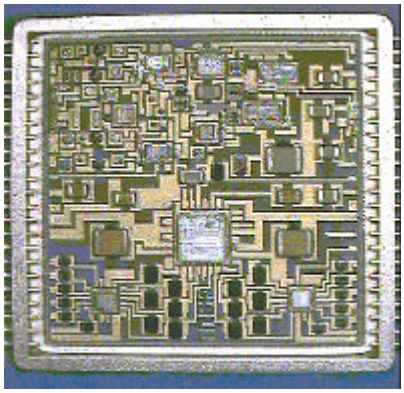


Figure 2.1.8-e: Example of video hybrid for space applications, developed by MMS

Figure 2.1.8-f is a synoptic of the electronics required for the NGST visible instrument. It is considered that the scientific processing of the digital data provided by the ADC is not part of this electronics.

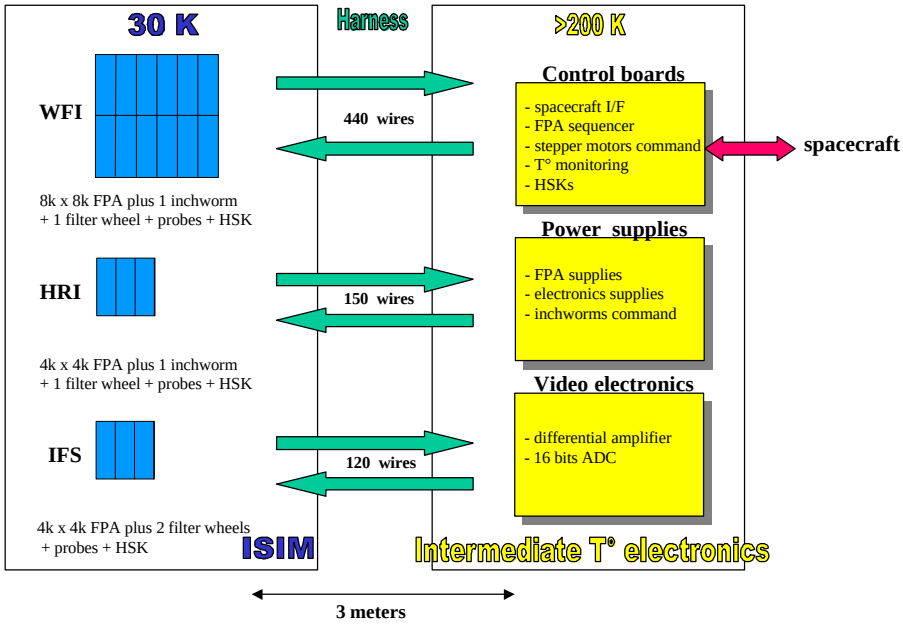


Figure 2.1.8-f: Synoptic of NGST visible instrument camera electronics

2.1.8.3 Electronics mass and power budget

The mass and power budgets have been estimated for the most conservative case (see Table 2.1.8-b) assuming independent visible instruments. Full cold redundancy for non cryogenic electronics is considered for this budget.

2.1.8.3.1 Power budget

Heat load at cryogenic temperature

The heat load at cryogenic temperature is due to the dissipation of the focal plane and the mechanisms and to the leaks along the harness.

From preliminary estimates and inputs from RIRCE, low power source follower per detector (SFD) would need 2nJ/pixel/sample. The focal plane dissipation is thus evaluated with a complete read out at each sampling period of 28 seconds. No preamplifier has been taken into account, as the FPA baseline is based on the use of the on-chip buffer. If required in the future, the power dissipation of such MOSFET pre-amplifier is expected to be around 1 mW, which shows that the 18 video outputs configuration may be manageable, but not the 96 video outputs configuration.

The number of leads required for the harness is around 440 for the WFI, 150 for the HRI and 120 for the IFS (based on 22 wires per sub-modules, 4 wires per filter wheel, 42 wires for an inchworm, plus contingencies for temperature probes, position sensors…). As the stainless steel wires would feature 100 µm diameter along the harness (3 m long), the conductive leaks should be around 2 mW for WFI, 0.7 mW for HRI and 0.5 mW for the IFS, based on an average thermal conductivity equal to 10W/m/K between 30 and 200K.

The inchworms are expected to dissipate around 5 mW during calibration and 0.05 mW during operation, and the stepper motor may take benefit of the use of high temperature super conducting materials. Figures measured for ISO are taken as reference, i.e. 4mW during filter wheel step and no power dissipation when non operated. For each detection module (WFI, HRI and IFS), the average power dissipation for the mechanisms will highly depend on the frequency of mechanisms activation which is expected to be very low (at least no change during 1000 s). Therefore, it is expected that for each instrument, the average power dissipation would be less than 1 mW, with peaks around few mW.

Dissipation of the other parts (temperature probes…) is considered as negligible. Table 2.1.8-c summarises the expected heat load at cryogenic temperature for the most conservative case, i.e. the use of 64, 16 and 16 output amplifiers respectively for WFI, HRI and IFS (see Table 2.1.8-b). 50% margin has also been added as this budget is very preliminary. It can be seen that the total power dissipation in the cold area is 19 mW compatible with the allocated 50mW.

Power dissipation at 30 K (mW)	Wide Field Imager	High Resolution Imager	Integral Field Spectrograph
mechanisms	1	1	1
focal plane	4.5	1.1	1.1
harness	2	0.7	0.5
total with margin	11	4	4

Table 2.1.8-c: Expected heat load at cryogenic (conservative case)

Intermediate temperature electronics dissipation

The intermediate electronics (located at temperature > 200K) will be composed of the control board, the power supply board and the video electronics boards. Table 2.1.8-d summarises the expected power dissipation (yield of the DC/DC supply is taken equal to 75%) for the most conservative case, i.e. the use of 96 video chains which leads to 94 W. The most favourable case would lead to 53 W instead of 94 W.

Power dissipation at 200 K (W)	Wide Field Imager	High Resolution Imager	Integral Field Spectrograph
control board	8	8	8
supply board	16	7.5	7.5
video board	26	6.5	6.5
total with margin	50	22	22

Table 2.1.8-d: Expected power dissipation at intermediate temperature (conservative case)

2.1.8.3.2 Mass budget

Table 2.1.8-e summarises the estimation of mass budget for the detection parts of the 3 instruments, for the most conservative case. The total mass of the 3 instruments focal plane and electronics is 55 kg.

Mass budget (kg)	Wide Field Imager	High Resolution Imager	Integral Field Spectrograph
FPA plus bracket	2.5	0.6	0.6
harness	5	3	3
electronics	16	12	12
total with margin	23.5	15.6	15.6

Table 2.1.8-e: Expected mass budget for the focal plane plus electronics (conservative case)

2.1.9 Instrument performance

2.1.9.1 Assumptions for the radiometric performances

The system of magnitudes used in the performance evaluation is the AB magnitude system defined by AB magnitude = zero point – 2.5 $\log_{10}$ (Irradiance per unit frequency band) where the zero point is 48.58 for irradiance per unit frequency band in erg/cm²/s/Hz. With this system, the irradiance (or flux) per unit frequency band of a point source of AB magnitude zero is 3698 Jy (Jansky) and remains constant throughout the spectrum. Extended sources are defined in AB magnitude per arcsec².

The zodiacal light is modelled by a blackbody at colour temperature 5600 K (diffusion of solar radiance) with a mV magnitude of mV=23/arcsec². Sensitivity to zodiacal light level from mV=25/arcsec² to mV=21/arcsec² has been looked at. The self emission of the optics is definitely negligible with respect to the zodiacal light in the visible range.

In the radiometric performance assessment, the telescope is assumed to have a pupil diameter of 8 m without any obstruction. It consists of 5 mirrors with 98% reflection for a total transmission of 90%.

The telescope is assumed in diffraction limit at 2 µm. The diameter of the first dark ring of the diffraction pattern or PSF (Point Spread Function) is 126 mas (milli arcsec). The energy encircled inside a diameter of 126 mas is about 80%. The PSF is assumed to be identical at visible wavelength because of the telescope aberrations. The optical quality of the visible instruments is quite diffraction limited in visible and thus does not affect the overall quality, which is dominated by the telescope performance. It is assumed that the Line Of Sight (LOS) jitter is well inside the PSF and that the long term PSF (convolution of PSF with LOS instabilities during observation period) is not degraded with respect to the instantaneous PSF. In the simulation, the PSF is modelled by a 2 dimensional gaussian function with a standard deviation σ of 35 mas in order to get 80% of energy inside 126 mas diameter.

The imagers oversample the image pattern. A small amount of the PSF is collected by a single pixel. To improve the signal to noise, the signals from adjacent pixels are co-added. The size of the window for summation is selected to have the best possible signal to noise ratio. The pixels of the Wide Field Imager and of the High resolution Imager are 37 mas and 10 mas respectively and the optimum corresponds to about 3 and 11 pixels. The Integral Field Spectrograph microlenses have a diameter of 100 mas close to the optimum. The proportion of energy collected within 100 mas diameter is about 64%.

A maximum integration time of 1000 s is defined in order to limit the low frequency noise. A reset of detector is performed at the beginning of each integration time. The observation time considered in the performance evaluation is 10^5 s. During the observation time, 100 periods of integration times are cumulated. The cumulation provides a gain of 10 on the signal to noise ratio.

The following assumptions are made for detector performances:

- read out circuit: one output stage per elementary module (8192x1365) ; therefore, the 8192x8192 focal plane array of WFI has 12 output stages and the 4096x4096 of HRI and IFS 3 output stages; an up-the-ramp sampling is performed during the integration time of 1000 s with 35 read-out leading to a pixel frequency of 200 kHz.

- noise and dynamic: the capacitance is 10 fF leading to a conversion factor of 16 µV/e^-.; the voltage noise is assumed to be 50 µV rms equivalent to 3.125 e^- rms per read out. The total dynamic is 1.6 V leading to 100 000 e^- at saturation while the linear dynamic is 0.8 V leading to 50 000 e- for the linear domain.

- dark current: in a very conservative way, the dark current density at 300 K is assumed to be 1000 pA/cm². An extrapolation of dark current as a function of temperature is made using the rule " half the dark current each 7 K temperature decrease " down to T = 120 K. It is assumed, again in a conservative way that the dark current does not decrease from 120 to 30 K. This leads to a dark current density of 0.35 fA/cm² and then to a dark current of 0.0003 e^-/s per pixel.

- quantum efficiency: it is assumed to decrease from 80% at 0.5 µm to 50 % at 1 µm.

- the so called up-the-ramp technique is used during the integration time of 1000s. It consists in making a reset at the start of the integration period then n regular sampling during the integration time. The major interest of this technique is that it provides an efficient way to remove cosmic rays parasitic signal. It provides a reduction of the read out noise but a slight increase of the photon noise according to the formulas $\sigma^2_{readout\,total} = \sigma^2_{readout}\,\frac{12n}{n^2-1}$ and $\sigma^2_{photonic\,total} = \sigma^2_{photonic}\,\frac{6}{5}\frac{n^2+1}{n^2-1}$

- the background charge amount $N_{background}$ is assumed perfectly removed from adjacent pixels for example with the exception of its own photon noise.

2.1.9.2 Wide Field Imager performance

The Wide Field Imager optics consists of 3 mirrors with 98% reflection and 1 filter with 80 % transmission. The overall optics transmission including the telescope is 68%. The pixel size of the Wide Field Imager is 37 mas in the object space. The optimum integration window for detection is 3x3 pixels, which provide about 78% of the PSF energy when the point spread function is centred on the window. In spectroscopic mode, the spectral element spreads on 2 pixels. In that case, the window used for performance assessment is a 2x3 pixels which provides about 62% of the PSF energy when the point spread function is centred on the window.

Performances in imaging modes

Table 2.1.9-a and Table 2.1.9-b provide the Wide Field Imager sensitivity (faintest point source detectable in 10^5s with a signal to noise ratio of 10) and dynamic range (saturation levels) with standard filters for point source and extended source respectively in terms of AB magnitude of point source and AB magnitude / arcsec² of extended source respectively.

The saturation for the last sample corresponds to a saturation for the complete integration time of 1000 s. It is recalled that the up the ramp sampling provides an extension of the dynamic range because a bright star which saturates for the complete integration time can nevertheless be analysed using the data from the first samples. With the proposed design with 35 read outs during the integration, the first sample occurs after 28 s and the brightest star which can be analysed is just at saturation after 28s (saturation for first sample). The dynamic range is thus increased by 3.9 magnitude.

filter	Wavelength range	Spectral R (λ/Δλ)	AB magnitude for SNR=10	AB magnitude at saturation	
				last sample	first sample
V	0.5 – 0.6 µm	5.5	31.6	21.6	17.8
R	0.59 – 0.81 µm	3.2	31.7	22.2	18.4
I	0.78 – 1 µm	4.0	31.4	21.8	17.9

Table 2.1.9-a: Wide Field Imager sensitivity and dynamic range in imaging mode with standard filters: AB magnitude of point source at detection limit and at saturation

filter	Wavelength range	Spectral R (λ/Δλ)	AB mag/arcsec² for SNR=10	AB mag/arcsec² at saturation	
				last sample	first sample
V	0.5 – 0.6 µm	5.5	25.8 / arcsec²	16.4 / arcsec²	12.6 / arcsec²
R	0.59 – 0.81 µm	3.2	26.0 / arcsec²	17.0 / arcsec²	13.1 / arcsec²
I	0.78 – 1 µm	4.0	25.7 / arcsec²	16.6 / arcsec²	12.7 / arcsec²

Table 2.1.9-b: Wide Field Imager sensitivity and dynamic range in imaging mode with standard filters: AB magnitude/arcsec² of extended source at detection limit and at saturation

For the Wide Field Imager in imaging mode with the R filter at detection limit with a signal to noise ratio of 10 (point source of AB magnitude 31.7), more than 98% of the signal comes from the sky background and more than 99% of the noise budget comes from the photon noise on the sky background. The photon noise from the dark current and the read out noise are negligible.

Figure 2.1.9-a provides the impact of the read out noise and the zodiacal light on the Wide Field Imager sensitivity with the R filter. The performances are clearly dominated by the zodiacal light photon noise.

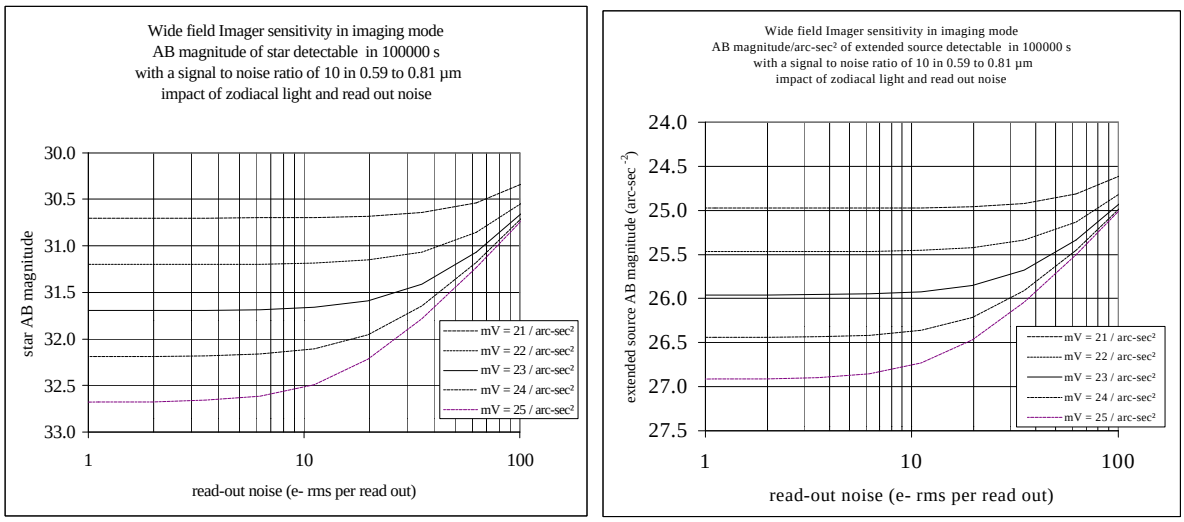


Figure 2.1.9-a: Impact of the read out noise and the zodiacal light on the Wide Field Imager sensitivity with the R filter

Impact of telescope and instrument optical quality

The baseline telescope with instrument optical quality corresponds to diffraction limit at 2 µm. The PSF energy collected within 126 mas diameter is 80 %. The optimum window selected is a 3 x 3 pixel window in the WFI radiometric performance, which collects 78% of energy.

With an upgraded telescope with instrument in diffraction limit at shorter wavelength, the size of the window can be reduced and the sensitivity improved. For example, if diffraction limit is at 1.3 µm wavelength, a 2x2 pixel window will collect the same energy and the signal to noise ratio will be

improve by a factor 3/2 or in other words, the limiting magnitude of point source just detectable is increased by 0.44 magnitude.

The sensitivities (SNR=10 in 10^5 s) of WFI in R band for point sources as a function of telescope optical quality are given in Table 2.1.9-c.

Diffraction limit	2 µm	1.3 µm	0.7 µm
Window	3 x 3	2 x 2	1
Limiting AB magnitude	31.7	32.14	32.9

Table 2.1.9-c: Impact of telescope and instrument optical quality on WFI limiting magnitude in imaging mode

Performances in Spectroscopic modes

The spectrograph mode of the Wide Field Instrument is got by inserting a direct vision prism. Figure 2.1.9-b provides the spectral resolution and the spectral width of the Wide Field Imager in spectrograph mode. The spectral resolution $R=\lambda/\Delta\lambda$, where $\Delta\lambda$ is the spectral interval on 2 pixels, varies from 379 to 47 as the wavelength evolves from 0.5 to 1 µm while the spectral dispersion varies from 1.3 to 21 nm per spectral element of 2 pixels.

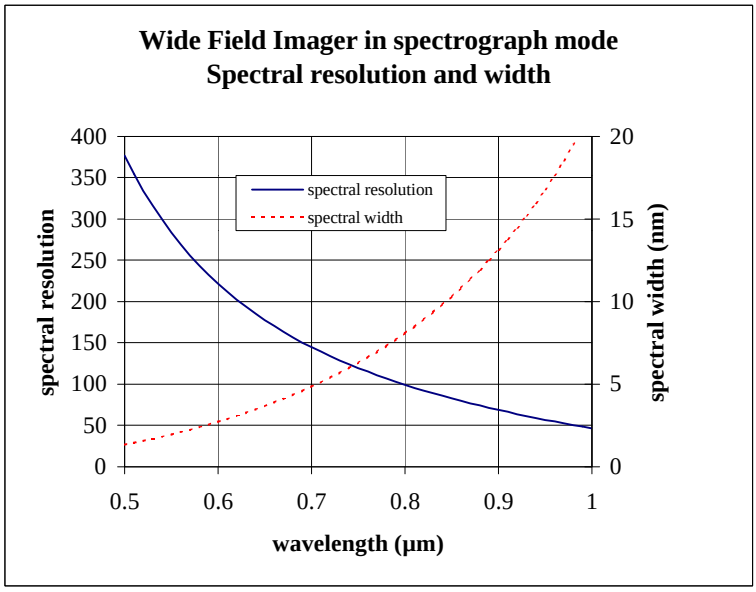


Figure 2.1.9-b: Spectral resolution and spectral width (on 2 pixels) of the Wide Field Imager in spectrograph mode

Figure 2.1.9-c provides the Wide Field Imager sensitivity for a point source (AB magnitude of point source detectable in 10^5 s with a signal to noise ratio of 10) and the dynamic range (saturation levels) as a function of wavelength in the spectrograph mode. The AB magnitude of the faintest point source detectable with a signal to noise ratio of 10 is in the range m_{AB} = 26.4 to m_{AB} = 28.2 according to the central wavelength. The AB magnitude of the brightest point source just at saturation in spectrograph mode is in the range 17 to 19 for the last sample and in the range 13 to 15 for the first sample according to the central wavelength. As in imaging mode the signal and noise budgets are dominated respectively by the zodiacal light and its photon noise.

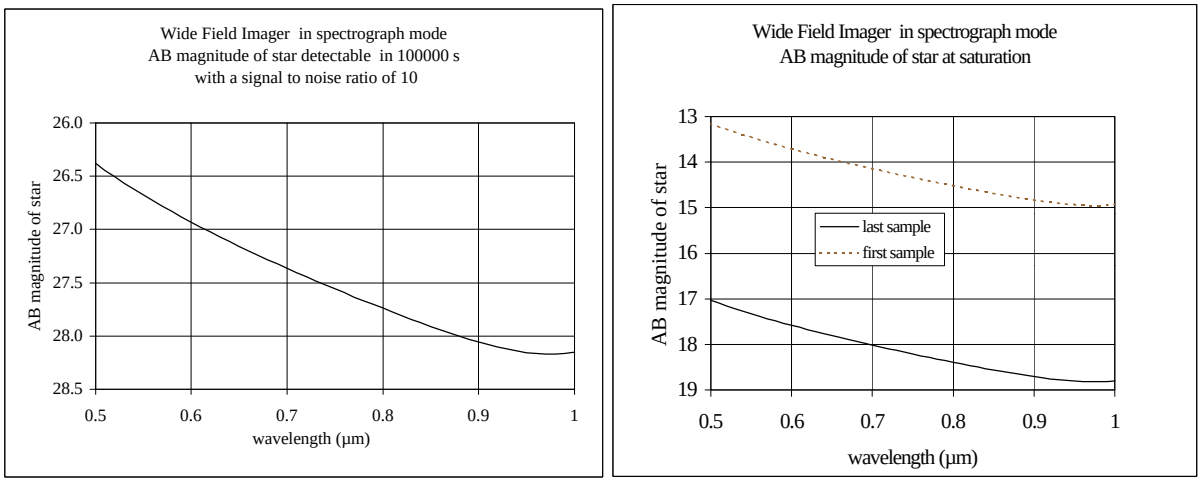


Figure 2.1.9-c: Sensitivity (AB magnitude of point source detectable in 10^5 s with a signal to noise ratio of 10) and dynamic range (saturation levels) of the Wide Field Imager for point source in spectrograph mode

Performances in multiband spectroscopic mode

This mode uses the same disperser as the spectroscopic mode but with a multiband filter instead of a wide band between 0.5 µm and 1 µm. The multiband filter has 6 bumps and covers a total wavelength range of 323 nm instead of 500 nm. The background is thus reduced in proportion and is 62% of the background with the wide band filter. The photon noise on this background (which is the dominant noise) is thus 78% of the photon noise with the wide band filter . The related gain in sensitivity is 0.26 magnitude. Another gain in sensitivity is due to the cumulation of pixels within the sub-band. For band 1 for example, the spectrum of the sub band spreads on 50 pixels. The signal can thus be cumulated on a window of 50x3 pixels instead of 2x3 pixels in spectroscopic mode. The gain in signal to noise ratio is 5 equivalent to 1.75 magnitude. The total gain is 2 magnitude in this band.

The performances in multiband mode can be equivalently deduced for performances in imaging mode assuming 6 different filters equivalent to the bumps of the multi band filter. The first loss with respect to imaging mode performances is due to the spectrum spread. For band 1 the signal is spread on 50x3 pixels instead of 3x3 pixels thus a loss in sensitivity of 1.53 magnitude. The second loss is due to the background. For band 1 the background with single filter is 13.6% of the background with multiple filter and the loss of performances is $\sqrt{0.136}$ equivalent to 1.08 magnitude. The total loss of performances for band 1 is 2.6 magnitude (that is the price to paid to the multiplex advantage). This loss reduces to 1.6 magnitude for band 6.

Table 2.1.9-d provides the sensitivity to point source in spectroscopic mode at the band central wavelength and in imaging mode in equivalent band and the derived sensitivity in multiband mode after cumulation on the spectrum. The sensitivity to point source in multiband mode is in the range of 28.5 to 29.4 AB magnitudes. The reduced sensitivity of 1.6 magnitude in spectroscopic mode near 0.5 µm with respect to that near 1 µm is improved by the cumulation on a large spectrum and the variation of sensitivity is reduced down to 0.9 magnitude in multiband mode close to the difference of sensitivity of 0.5 magnitude in imaging mode.

Sub band	Central wavelength (nm)	Band width (nm)	Pixels spread	Sensitivity to point source : AB magnitude		
				spectroscopic mode	imaging mode	multiband mode
1	524	40	50	26.53	31.13	28.5
2	577	40	34	26.81	30.96	28.6
3	636	40	24	27.10	30.83	28.7
4	704	40	16	27.38	30.73	28.8
5	792	53	14	27.71	30.72	29.0
6	945	110	14	28.14	30.96	29.4

Table 2.1.9-d: Sensitivity of the Wide Field Imager in multiband spectrograph mode for point source: AB magnitude of point source detectable in 10^5 s with a signal to noise ratio of 10

2.1.9.3 High Resolution Imager performances

The High Resolution Imager optics consists of 3 mirrors with 98% reflection and 1 filter with 80 % transmission. The overall optics transmission including the telescope is 68%. The pixel size of the High Resolution Imager is 10 mas in the object space. The optimum window for detection is 11x11 pixels, which provide about 80% of the PSF energy.

Table 2.1.9-e and Table 2.1.9-f provide the High Resolution Imager sensitivity and dynamic range with standard filters for point source and extended source respectively in terms of AB magnitude of point source and AB magnitude / arcsec² of extended source at detection limit and at saturation. The performances are very similar to those of the Wide Field Imager with the assumed PSF (see discussion on PSF impact below). This is easily explained as the two instruments have the same transmission, the same detector and the energy collected in the 11x11 pixel window of the High Resolution Imager is equal to the energy collected in the 3x3 pixel window of the Wide Field Imager.

filter	Wavelength range	Spectral R ($\lambda/\Delta\lambda$)	AB magnitude for SNR=10	AB magnitude at saturation	
				last sample	first sample
V	0.5 – 0.6 µm	5.5	31.5	19.5	15.6
R	0.59 – 0.81 µm	3.2	31.7	20.0	16.1
I	0.78 – 1 µm	4.0	31.4	19.6	15.7

Table 2.1.9-e: High Resolution Imager sensitivity and dynamic range with standard filters: AB magnitude of point source at detection limit and at saturation

filter	Wavelength range	Spectral R ($\lambda/\Delta\lambda$)	AB mag/arcsec² for SNR=10	AB mag/arcsec² at saturation	
				last sample	first sample
V	0.5 – 0.6 µm	5.5	27.1 / arcsec²	16.5 / arcsec²	12.7 / arcsec²
R	0.59 – 0.81 µm	3.2	27.3 / arcsec²	17.0 / arcsec²	13.2 / arcsec²
I	0.78 – 1 µm	4.0	27.0 / arcsec²	16.6 / arcsec²	12.8 / arcsec²

Table 2.1.9-f: High Resolution Imager sensitivity and dynamic range with standard filters: AB magnitude/arcsec² of extended source at detection limit and at saturation

As for the Wide Field Imager, the signal and noise budgets of the High resolution Imager are dominated respectively by the zodiacal light and its photon noise, but the read out noise becomes more important and have to remain below about 5 electrons rms per read out.

Impact of telescope and instrument optical quality

The baseline telescope with instrument optical quality corresponds to diffraction limit at 2 µm. The PSF energy collected within 126 mas diameter is 80 %. The optimum window selected is a 11 x 11 pixel window in the HRI radiometric performance, which collects 78% of energy.

With an upgraded telescope with instrument in diffraction limit at shorter wavelength, the size of the window can be reduced and the sensitivity improved. For example, if diffraction limit is at 1.1 µm wavelength, a 6x6 pixel window will collect the same energy and the signal to noise ratio will be improve by a factor 11/6 or in other words, the limiting magnitude of point source just detectable is increased by 0.66 magnitude. The sensitivities (SNR=10 in 10^5 s) of HRI in R band for point sources as a function of telescope optical quality (assuming Gaussian PSF) are given in Table 2.1.9-g.

Diffraction limit	2 µm	1.1 µm	0.55 µm
Window	11 x 11	6 x 6	3 x 3
Limiting magnitude	31.7	32.36	33.11

Table 2.1.9-g: Impact of telescope and instrument optical quality (Gaussian PSF) on HRI limiting magnitude in imaging mode

The HRI is designed to Nyquist sample the "core of the PSF" in the optical. Its performance is directly driven by the width of the PSF and the related window. The PSF shape is more complex than the Gaussian used for above simulations. Accurate sensitivity estimates are not possible for the HRI because these depend on how crowded the field is i.e. the wings of the PSF profile will overlap and provide an extra source of the background. If we refer to J. Burge paper "Simulated optical and ultraviolet performance for the University of Arizona NGST Mirror System Demonstrator", we can hope that the PSF at 0.5 µm, even if the optics are diffraction limited at 1.5 µm, will present a strong central peak and that a substantial percentage of the total flux will be in the core of the PSF at visible wavelengths. The FWHM of this central peak is typically 15 mas and a 3x3 window will collect completely the energy within the peak. A magnitude limit close to 33 can thus be expected for HRI.

2.1.9.4 Performances of the Integral Field Spectrograph

The Integral Field Spectrograph optics consists of 20 lenses (including microlens) with 98% transmission each, 1 filter with 90 % transmission and 1 prism for low resolution mode or 1 grism for high resolution mode both assumed with 70% transmission. The overall optics transmission including the telescope is 45%. The energy from a point source, which is collected by a microlens, is 63.8%.

Performance in low spectral resolution mode

The low spectral resolution is got by inserting a direct vision prism. Figure 2.1.9-d provides the spectral resolution and the spectral width (nm/pixel) of the Integral Field Spectrograph in low spectral resolution mode. The spectral resolution $R=\lambda/\Delta\lambda$, where $\Delta\lambda$ is the spectral interval on 2 pixels, varies from 354 to 45 as the wavelength evolves from 0.5 to 1 µm while the spectral dispersion varies from 0.71 to 11 nm per pixel.

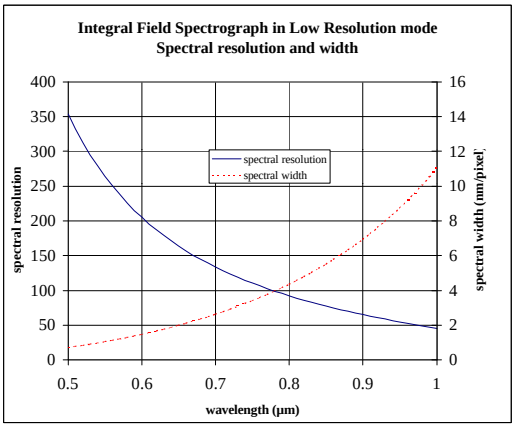


Figure 2.1.9-d: Spectral resolution and width of the Integral Field Spectrograph in low spectral resolution mode.

Figure 2.1.9-e provides the Integral Field Spectrograph sensitivity (AB magnitude of point source detectable in 10^5 s with a signal to noise ratio of 10) and dynamic range (saturation levels) in low spectral resolution mode for a point source as a function of wavelength. The AB magnitude of the faintest point source detectable by the Integral Field Spectrograph in low spectral resolution mode with a signal to noise ratio of 10 are in the range of m_{AB} = 28.5 to 29.5 according to the wavelength. The AB magnitude of the brightest point source just at saturation in the Integral Field Spectrograph in low spectral resolution mode are in the range of m_{AB} = 16.5 to 18.2 (last sample) and of m_{AB} = 12.5 to 14.5 (first sample) respectively according to the wavelength. The signal and noise are dominated by the sky background.

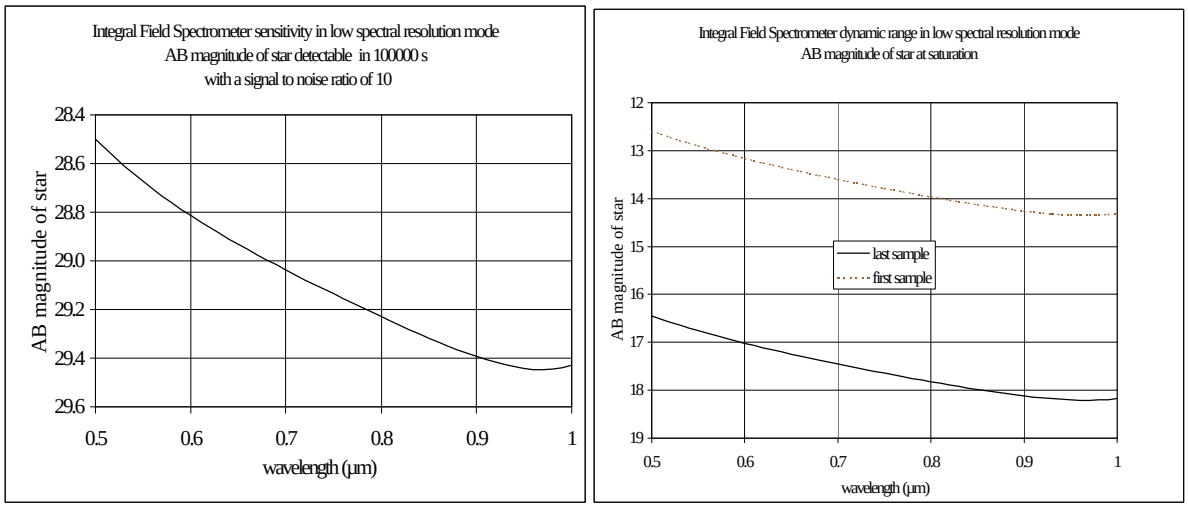


Figure 2.1.9-e: Sensitivity (AB magnitude of point source detectable in 10^5 s with a signal to noise ratio of 10) and dynamic range (saturation levels) of the Integral Field Spectrograph for point source in low spectral resolution mode

Performance in High spectral resolution mode

The high spectral resolution is got by inserting one of the 5 grisms of the disperser wheel. The spectral resolution is about 1200 with any of these grisms.

Figure 2.1.9-f provides the Integral Field Spectrograph sensitivity (AB magnitude of point source detectable in 10^5 s with a signal to noise ratio of 10) and the dynamic range (saturation levels) in high spectral resolution mode for a point source as a function of wavelength. The AB magnitude of the faintest point source detectable by the Integral Field Spectrograph in high spectral resolution mode with a signal to noise ratio of 10 are in the range of m_{AB} = 26.8 to 27.4 according to the wavelength. The AB magnitude of the brightest point source just at saturation in the Integral Field Spectrograph in high spectral resolution mode are in the range of m_{AB} = 14.6 to 15.2 (last sample) and of m_{AB} = 10.7 to 11.3 (first sample) according to the wavelength. The signal is dominated by the zodiacal light but the noise budget is dominated by the read out noise.

Impact of telescope and instrument optical quality

The baseline telescope with instrument optical quality corresponds to diffraction limit at 2 µm. The PSF energy collected within 126 mas diameter is 80 %. The energy collected by a microlens is 64%. With an upgraded telescope with instrument in diffraction limit at 0.5 µm, 80 % of the energy would be collected within 31 mas diameter. The energy collected by a microlens would be higher but even with a 100% collected energy, the improvement in limiting magnitude would be less than 0.5.

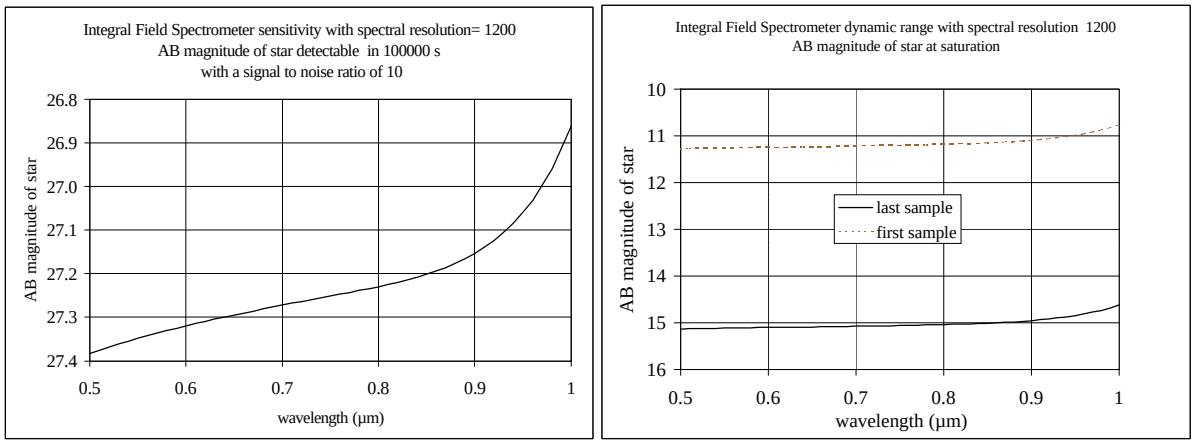


Figure 2.1.9-f: Sensitivity (AB magnitude of point source detectable in 10^5 s with a signal to noise ratio of 10) and dynamic range (saturation levels) of the Integral Field Spectrograph for point source in high spectral resolution mode

2.1.10 Instrument budgets

2.1.10.1 Mass budget

2.1.10.1.1 Wide Field Imager

The mass budget of the Wide Field Imager is provided in Table 2.1.10-a. It does not include the ISIM mounts. Tilting the base plate in relation to the optical configuration could significantly reduce the total mass. In this way on the detector side the offsets of the optical components can be reduced and the brackets can be designed smaller and lighter.

Items	Mass (kg)
Primary mirror with fixations and support bracket	14.5
Secondary mirror with Refocusing mechanism	1.4
Tertiary mirror with fixations and support bracket	1.8
Filter wheel with mount	6.2
Detector package and support bracket	2.5
Thermal hardware and baffles	2
Optical bench	38
Harness	5
Video and mechanism drive electronics	16
Total	87.5

Table 2.1.10-a: Wide Field Imager mass budget

2.1.10.1.2 High Resolution Imager

The mass budget of the High Resolution Imager is provided in Table 2.1.10-b. It does not include the ISIM mounts.

Items	Mass (kg)
Primary mirror with fixations and support bracket	0.5
Secondary mirror with Refocusing mechanism	0.45
Tertiary mirror with fixations and support bracket	0.52
Filter wheel with mount	1.1
Detector package and support bracket	0.6
Thermal hardware and baffles	0.3
Optical bench	3.6
Harness	3
Video and mechanism drive electronics	12
Total	22

Table 2.1.10-b: High Resolution Imager mass budget

2.1.10.1.3 Integral Field spectrograph

The mass budget of the Integral Field Spectrograph is provided in Table 2.1.10-c. It does not include the ISIM mounts.

Items	Mass (kg)
Lenses with barrels and support brackets	2
Grism wheel with mount	0.7
Enlarger mechanism with mount	0.4
Detector package and support bracket	0.6
Thermal hardware and baffles	0.3
Optical bench	4
Harness	3
Video and mechanism drive electronics	12
Total	23

Table 2.1.10-c: Integral Field Spectrograph mass budget

2.1.10.2 Power budget

2.1.10.2.1 Wide Field Imager

The dissipation of WFI mechanism at 30 K is about 1 mW in average with a peak value of few tens of mW. The 8k x 8k WFI focal plane dissipation is about 4.5 mW. The heat leak through harness is 2 mW. The total heat load of WFI at 30 K is 7.5 mW and 12 mW with 50% margin with the assumption that no preamplifier is required. The dissipation of preamplifier is about 1 mW. With the present design of 12 outputs, the extra dissipation with amplifiers would be of 12 mW and the total dissipation of WFI would be 27 mW compatible with the 50 mW requirement. The power dissipation budget of WFI is summarised in Table 2.1.10-d. The dissipation and electrical consumption of WFI warm electronics is 50 W.

WFI dissipation at 30 K	Mechanism: 1 mW 8kx8k focal plane : 4.5 mW harness : 2 mw total : 12 mW with 50% margin option preamplifier : + 12 mW
Consumption/dissipation of WFI warm electronics	50 W

Table 2.1.10-d: Power dissipation budget for the Wide Field Imager

2.1.10.2.2 High Resolution Imager

The dissipation of HRI mechanism at 30 K is about 1 mW in average with a peak value of few tens of mW. The 4k x 4k HRI focal plane dissipation is about 1.1 mW. The heat leak through harness is 0.7 mW. The total heat load of WFI at 30 K is 2.8 mW and 4 mW with 50% margin with the assumption that no preamplifier is required. The dissipation of preamplifier is about 1 mW. With the present design of 3 outputs, the extra dissipation with amplifiers would be of 3 mW and the total dissipation of HRI would be 7.5 mW compatible with the 50 mW requirement. The power dissipation budget of HRI is summarised in Table 2.1.10-e. The dissipation and consumption of HRI warm electronics is 22 W.

HRI dissipation at 30 K	Mechanism: 1 mW 8kx8k focal plane : 1.1 mW harness : 0.7 mw total : 4 mW with 50% margin option preamplifier : + 3 mW
Consumption/dissipation of HRI warm electronics	22 W

Table 2.1.10-e: Power dissipation budget for the High Resolution Imager

2.1.10.2.3 Integral Field spectrograph

The dissipation of IFS mechanisms at 30 K is about 1 mW in average with a peak value of few tens of mW. The 4k x 4k IFS focal plane dissipation is about 1.1 mW. The heat leak through harness is 0.5 mW. The total heat load of WFI at 30 K is 2.6 mW and 4 mW with 50% margin with the assumption that no preamplifier is required. The dissipation of preamplifier is about 1 mW. With the present design of 3 outputs, the extra dissipation with amplifiers would be of 3 mW and the total dissipation of IFS would be 7 mW compatible with the 50 mW requirement. The power dissipation budget of IFS is summarised in Table 2.1.10-f. The dissipation and consumption of IFS warm electronics is 22 W.

IFS dissipation at 30 K	Mechanism: 1 mW 8kx8k focal plane : 1.1 mW harness : 0.5 mW total : 4 mW with 50% margin option preamplifier : + 3 mW
Consumption/dissipation of IFS warm electronics	22 W

Table 2.1.10-f: Power dissipation budget for the Integral Field Spectrograph

2.2 TECHNOLOGY READINESS

2.2.1 Hardware description

2.2.1.1 Wide Field Imager

The Wide Field Imager (WFI) is a big instrument of envelop 1.7 m x 0.9 m x 0.5 m. The total mass including the electronics is about 85 kg. The optical concept is a 3 mirror optical relay with a primary mirror of 54 cm square. The mirrors are in SiC. Fixations of mirrors are either a central fixation in SiC (small mirrors) or a more conventional 3 Invar blade fixation (primary mirror). The structure is a flat optical bench in SiC or C/SiC with a set of brackets (also in SiC or C/SiC) for mirror and mechanism supports. The focal plane consists of an 8k x 8k array of Si diodes hybridised on Si CMOS read out circuits. The instrument includes 2 mechanisms: a filter wheel and a refocusing mechanisms. The instrument electronics is removed from the cryogenic package at a distance of about 3m. Harness from detectors to removed electronics has to be thermalised at intermediate temperature. Table 2.2.1-a provides a description of the Wide Field Imager hardware.

Sub system	description
Optics	3 mirrors in SiC M1 : square 540 mm x 540 mm M2 : square 100 mm x 100 mm M3 : square 200 mm x 20m mm 10 elements Φ 80 mm: 8 filters + 1 prism + 1 multi band prism
Detector package	1 8 k x 8 k hybrid array of Si diodes hybridised on CMOS read out circuit ; dimension 120 mm x 120 mm Proximity electronics at 30 K / preamplifiers
Mechanisms	1 filter wheel mechanism with 10 positions filter Φ 90 mm ; wheel Φ 400 mm 1 refocusing mechanism on M2
Structure	1 SiC or C/SiC optical bench 1.5 m long Invar support brackets and isostatic fixations for mirrors, mechanisms and detector
Thermal and stray light protection	Thermal blankets and insulators A set of thermistors Baffles for straylight Thermal straps for harness
Remote Electronics at 200 K	Detector Drive Electronics Analogue to Digital Processing Electronics Thermal Control Electronics Mechanism Drive Electronics Interface with Spacecraft
Harness	From ISIM to Remote Electronics (3m) From Remote Electronics to Spacecraft Electronics at 300 K (6m)

Table 2.2.1-a: Hardware description of the Wide Field Imager (Flight Model)

2.2.1.2 High Resolution Imager

The High Resolution Imager (HRI) is very similar to the Wide Field Imager but with a smaller envelop of 0.5 m x 0.22 m x 0.08 m. The optical concept is a 3 mirror derived from Offner relay. The focal plane is 4k x 4 k instead of 8k x 8k. It includes as the WFI a filter wheel mechanism and a refocusing mechanism. Electronics is removed out of the cryogenic vessel 3 m apart. Table 2.2.1-b provides a description of the High-Resolution Imager hardware.

Sub system	description
Optics	3 mirrors in SiC M1 : square 70 mm x 70 mm M2 : square 15 mm x 15 mm M3 : square 80 mm x 80 mm 1 set of 6 filters Φ 15 mm
Detector package	4 k x 4 k hybrid array of Si diodes hybridised on CMOS read out circuit ; dimension 60 mm x 60 mm Proximity electronics at 30 K / preamplifiers
Mechanisms	1 filter wheel mechanism with 6 positions filter Φ 15 mm ; wheel Φ 100 mm 1 refocusing mechanism on M2
Structure	1 SiC or C/SiC optical bench 0.5 m long Invar support brackets and isostatic fixations for mirrors, mechanisms and detector
Thermal and stray light protection	Thermal blankets and insulators A set of thermistors Baffles for straylight Thermal straps for harness
Remote Electronics at 200 K	Detector Drive Electronics Analogue to Digital Processing Electronics Thermal Control Electronics Mechanism Drive Electronics Interface with Spacecraft
Harness	From ISIM to Remote Electronics (3m) From Remote Electronics to Spacecraft Electronics at 300 K (6m)

Table 2.2.1-b: Hardware description of the High Resolution Imager (Flight Model)

2.2.1.3 Integral Field Spectrograph

The Integral Field Spectrograph (IFS) is a spectrograph where the field is sampled by an array of microlenses. It includes an enlarger to adapt the focal plane scale factor to the microlens size. After sampling by the microlens array the object points are dispersed within a spectrograph made of a collimator, a set of dispersers and a camera. .It is an all-dioptric optical system. The structure is a SiC or C/SiC optical bench with brackets to support lens barrels and mechanisms. It includes a disperser wheel mechanism and an enlarger mechanism to change the resolution and field of the spectrograph. Dispersers are prisms and grisms according to the required dispersion. Taking into account the large f-number, a refocusing mechanism is not foreseen. The focal plane is 4k x 4 k Si diodes array hybridised on CMOS read out circuit as for the HRI. Table 2.2.1-c provides a description of the Integral Field Spectrograph hardware.

Sub system	description
Optics	1 (or few) dioptric enlarger (2 lenses Φ 12 mm) 1 fixed filter (0.5-1 µm band) Φ 3 mm 1 field lens Φ 75 mm 1 100x115 hexagonal microlens array (50 mm x 50 mm x 4 mm) 1 dioptric collimator (9 lenses Φ 80 mm) 1 set of dispersers : 1 prism + 5 grisms Φ 45 mm 1 dioptric camera (8 lenses Φ 80 mm)
Detector package	1 4 k x 4 k hybrid array of Si diodes hybridised on CMOS read out circuit ; dimension 60 mm x 60 mm Proximity electronics at 30 K / preamplifiers
Mechanisms	1 disperser wheel mechanism with 6 positions filter Φ 45 mm ; wheel Φ 200 mm 1 enlarger mechanism with 4 Φ 15 mm positions
Structure	1 SiC or C/SiC optical bench 1 m long barrels for lenses Invar support brackets and isostatic fixations for lens barrels , mechanisms and detector
Thermal and stray light protection	Thermal blankets and insulators A set of thermistors Baffles for straylight Thermal straps for harness
Remote Electronics at 200 K	Detector Drive Electronics Analogue to Digital Processing Electronics Thermal Control Electronics Mechanism Drive Electronics Interface with Spacecraft
Harness	From ISIM to Remote Electronics (3m) From Remote Electronics to Spacecraft Electronics at 300 K (6m)

Table 2.2.1-c: Hardware description of the Integral Field Spectrograph (Flight Model)

2.2.2 Technical difficulties and critical technological areas

The major difficulties come from the cryogenic operating conditions. The instruments have to operate at 30 K with a minimum power dissipation (<< 50 mW).

The potential critical points requiring more investigations and in advance technological developments for some of them are :

- the focal plane with Si diodes hybridised on Si CMOS read out circuits. This technology is feasible on paper but has been never realised except some preliminary development at Rockwell. CEA/LETI/LIR in Grenoble (France) can be candidate for this manufacturing . The major difficulty is the size of the required array (8192x8192 for WFI, 4096x4096 for HRI and IFS) with requires both stitching and mechanical butting. Cryogenic CMOS (< 40 K) with small pixels (15 µm) is also a point of feasibility to be demonstrated. Other critical area is the detector internal preamplifier for which the output resistance, noise spectrum and glow effect have to be investigated. All the electro optical performances of the detector have to be measured on a mock up.

- the cryogenic (30 K) mechanisms with very low thermal dissipation; the choice of materials (in particular use of superconductor materials for coils) , the inchworm behaviour and the tribology at 30 K have to be analysed with mock up during phase A.

- for Wide Field Imager, the large SiC or C/SiC structure can be critical and some technological works has to be performed in phase A in particular to determine cryogenic characteristics of materials.

- the development of the cryogenic harness and related interconnection which needs to have both low thermal conductance and low electrical capacitance; depending on outcomes of technological works on harness and detector, the technological program could be modified with development of cryogenic amplifier with very low noise and very low power dissipation.

- for the Integral Field Spectrograph, the optical characteristics of glasses at 30 K and the behaviour under irradiation have to be measured during phase A.

Other critical points need to be investigated further without requiring technological predevelopment:

- the method for testing the off axis aspherical mirrors,

- the mirror alignment and the stability of this alignment during the cooling down; the choice of SiC mirrors with C/SiC structure with quite similar coefficient of thermal expansion could avoid a cryogenic alignment,

- the behaviour of the filter coating at 30 K; the spectral shift due to cooling has to be accounted for,

- the fixations of the lenses in their barrels have to support cooling down and require particular design efforts,

- the qualifications of the microlens array and of grisms (mechanically ruled without replica) have to be investigated,

- the development of a 16 bit video hybrid part with very low noise.

Subsystem	Critical area	Comment
Optical system	Testing of mirrors	Method for mirrors figure accuracy to be investigated
	Mirror alignment and stability	SiC mirrors with C/SiC structure can avoid cryogenic alignment
	Filter coating at 30 K	Standard in IR, in visible ? spectral shift to be accounted
	Behaviour of IFS glasses at 30 K and under irradiation	Characterisation program recommended during phases A/B
	Fixation of IFS lenses in barrel	Adequate design to avoid strength during cooling down
	IFS grisms	Mechanically ruled without replica
Structure	WFI large C/SiC or SiC structure	Technological program recommended in phase A/B
	C/SiC and SiC behaviour at 30 K	
Mechanisms	Coils with superconductor material at 30 K & related power dissipation	Technological program recommended in phase A/B
	Inchworm operation at 30 K	
	Tribology behaviour at 30 K	
Focal plane	Detector internal preamplifier	Output resistance, noise spectrum and glow effect
	Electro optical performance at 30 K	Technological program recommended in phase A/B
	Focal plane assembly (stitching and butting)	
Harness	Development of the cryogenic harness with low thermal conductivity and low electrical capacitance	Technological program recommended in phase A/B
	Development of cryogenic preamplifier with very low noise and very low power dissipation	Depends on outcomes of technological programs on detectors and harness
Electronics	Development of a 16 bit video hybrid part with very low noise	

Table 2.2.2-a: Potential critical areas

2.3 DEVELOPMENT PLAN

2.3.1 Development schedule

The constraints on the development schedule are:
- a phase A start in September 2000
- a phase B start in March 2002 (ESA budget constraint)
- a delivery of instrument flight model in January 2006 (NASA request).

Taking into account these constraints , it is proposed to reduced to the maximum extent the phase B (12 months) in order to have 33 months for the phase C/D, and to prepare the C/D by advance technological developments during phases A and B for critical technologies described in section 2.3.3.

The development schedule is presented in Figure 2.3.1-a. Note that Hi Rel components will have to be purchased in advance.

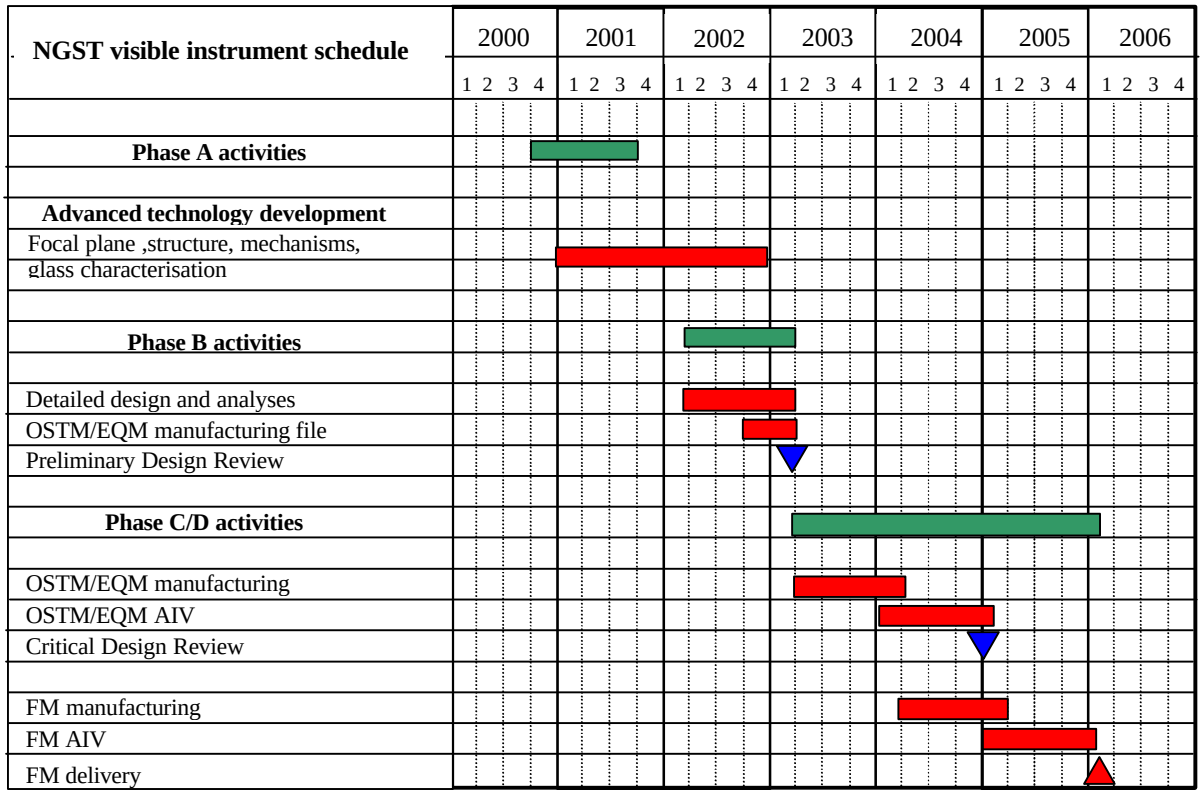


Figure 2.3.1-a: NGST visible instrument development schedule

2.3.2 Development approach

The required development schedule (NASA request of instrument flight model in January 2006) with a C/D phase in 33 months is very unusual for a scientific instrument. The credibility of such a development schedule is questionable.

Two prerequisites are mandatory to have a chance to develop the instruments in time:

- the instrument interfaces must be frozen very early,
- all the critical technologies must be validated at end of phase B.

The approach is therefore to develop and validate all the critical technologies during the phases A and B. The major risk is that one or several technologies are not fully available at start of phase C/D which would imply some delay in the schedule.

2.3.3 In advance technology activities

Taking into the very short schedule of phase C/D and the critical technologies addressed in section 2.2.2, it is necessary, in order to secure the development schedule, to undertake technology activities in advance for the potential critical areas. These areas are defined in Table 2.3.3-a.

The following technological activities are recommended in parallel with phase A and B.
- Si hybrid focal plane,
- Cryogenic mechanisms,
- SiC or C/SiC structure (WFI),
- Electrical characterisation of cryogenic harness (and optionally development of preamplifier),
- Characterisation of glass at 30 K for IFS.

Subsystem	Potential critical area
Focal plane	Detector internal preamplifier Electro optical performances at 30 K Focal plane assembly (stitching and butting)
Harness	Low thermal conductivity & low electrical capacitance cryogenic harness Cryogenic preamplifier with very low noise and very low power dissipation
Mechanisms	Use of superconductor material at 30 K and related power dissipation Inchworm and tribology at 30 K
Structure	WFI large C/SiC or SiC structure and material behaviour at 30 K
Optics	Behaviour of IFS glasses and coatings at 30 K and under irradiation

Table 2.3.3-a: Potential critical areas requiring pre developments during A/B phases

2.3.4 Philosophy of models and tests

Taking into account the very short development schedule, it is proposed to have two engineering/qualification models developed in parallel then the flight model.

The development is based on the following models:
- an Optical Structural Thermal Model (OSTM) representative of the structure, thermal hardware, mechanisms and optical concept. This model will be used to qualify the optical and structural concepts and the mechanisms,
- an Electrical Qualification Model (EQM) which will be representative of the focal plane and the electronics chain integrated on a jig and will perform the qualification of the detection / electronics chain,
- a Flight Model (FM).

The test philosophy is defined in Table 2.3.4-a.

Test	OSTM	EQM	FM
Functional tests	Mechanisms at ambient temperature and at 30 K during thermal vacuum tests	Electrical performances of electronics	Instrument functional tests
Optical tests	Wavefront error tests at ambient temperature and at 30 K during vacuum tests	no	Wavefront error tests at ambient temperature
Thermal vacuum tests	Thermal qualification of hardware at 30 K	Thermal qualification of hardware at 30 K	Thermal acceptance of hardware at 30 K
Mechanical test	vibrations	vibrations	no
EMC	no	yes	yes
Performances tests	Thermo elastic behaviour during thermal vacuum tests	Test of detector at 30 K and electronics with optical stimuli during thermal vacuum tests	Overall performances at 30K during thermal vacuum tests

Table 2.3.4-a: Test matrix

2.3.5 Development risks

Major risks are directly related to the very short schedule. The OSTM and EQM tests are completed at the end of the FM manufacturing. Any problem found during OSTM or EQM tests requiring modification of hardware will delay the beginning of FM integration and consequently likely the delivery of this model.

Except the risks related to the development of focal plane , cryogenic mechanisms and problem of removed electronics, which are common to the 3 instruments, some specific risks can be identified for each instrument:

- Wide Field Imager:

  development of the large SiC or C/SiC structure and cryogenic behaviour,

  alignment of optics at cold temperature,

- High Resolution Imager:

  alignment of optics at cold temperature,

- Integral Field Spectrograph :

  glass optical behaviour at cryogenic temperature,

  evolution of design (Fabry Perot could be an improvement desired by scientist),

  provision for a refocusing mechanism.

2.4 CONCLUSIONS

Three different concepts of visible instruments have been defined, a Wide Field Imager, a High Resolution Imager and an Integral Field Spectrograph. The Wide Field Imager is a relatively large (more than 1 m^3) and heavy (80 kg) instrument while the two other ones are very small instruments with less than 8 kg inside ISIM. Both imagers are all mirror system while the spectrograph is a pure dioptric system. All the instruments use a hybrid Silicon focal plane that operates at cryogenic temperatures with a very low thermal dissipation.

The very short schedule requires pre development of the most critical technologies during A and B phases, namely, the focal plane, the harness, the mechanisms, the SiC mirrors and the C/SiC structure. A characterisation of the visible glasses at cryogenic temperature is also required for the spectrograph.