

13.2 GMTNIRS – The High Resolution Near-IR Spectrograph for the Giant Magellan Telescope

13.2.1 Introduction - GMTNIRS

This document puts forth a conceptual design for a powerful, high-resolution near-infrared spectrograph for the Giant Magellan Telescope (GMT). This instrument is referred to as the Giant Magellan Telescope Near-Infrared Spectrograph (GMTNIRS). Our design utilizes silicon immersion gratings as the primary dispersing elements. This results in a compact instrument with high resolution and broad wavelength coverage.

In this document we provide a basic configuration and a workable optical and mechanical point design. It includes an examination of the packaging issues and the critical optical and detector technologies. As part of the optical design, we produce an estimate of system sensitivity as a function of wavelength and resolving power. A section devoted to the silicon immersion grating technology assesses their current status and discusses grating production risks and schedule issues. We discuss some ancillary requirements and plans for installation and maintenance. In some cases, additional supporting materials are being provided to the GMT project office.

13.2.1.1 GMTNIRS personnel

This conceptual design is authored by a team led by Daniel Jaffe, at the University of Texas at Austin. The optical design was produced under contract by David Warren, who also co-authored Section 13.2.2 of this document. The prototype silicon immersion gratings shown in Section 13.2.4 were fabricated by Doug Mar and Jasmina Marsh. Pedro Segura provided support for the engineering design.

13.2.1.2 Science Case

A high-resolution near-infrared spectrograph operating at 1-5 μm plays an important role in two of the main science priorities called out in Chapter 3: unlocking the secrets of planet and star formation and understanding the evolution of stellar populations and chemical elements. We reiterate some of the points made in Sections 3.1, 3.2, and 4.2 here.

13.2.1.2.1 Unlocking the Secrets of Planet and Star Formation

As outlined in section 3.1.3, we can study the chemistry and physics of protostellar disks and the dense interstellar medium (ISM) using near-IR absorption and emission lines. The wavelength range from 2 to 5 μm gives us access to vibrational transitions of many of the most abundant molecules in the protostellar environment: H_2 , H_2O , CO , OCS , as well as a large variety of molecules with C-H bonds. The ISM studies will help us to understand the conditions in molecular clouds before and after the onset of star formation. The inner parts of the disks can be studied by observing molecular emission

lines and the cooler outer disks and circumstellar envelopes are approachable via absorption line measurements. Resolved line profiles of the emission lines allow us to map the radial distribution of abundance and excitation for molecular species in protoplanetary disks.

Chapter 3.1.6 discusses the initial mass function. Studies of the initial mass function will only be as good as our understanding of the properties of the very young objects from which the IMF is constructed. We can characterize the photospheres of protostars by measuring T_{eff} , $\log(g)$, radial velocity, $v\sin(i)$, the magnetic field strength, and elemental abundances. Young stellar objects are not only highly obscured at visible wavelengths, but also have spectra contaminated by non-photospheric continuum emission from the ultraviolet to the submillimeter. Analysis of high-resolution infrared spectra can overcome these obstacles by penetrating the overlying dust and providing information about relative line strength and about line shape for many lines. The scaling of Zeeman splitting with wavelength means that Zeeman-sensitive lines in the near-infrared provide a better estimate of the magnetic field strength than optical lines. Finally, the infrared provides access to many photospheric absorption lines of hardy molecules (notably OH and CO) as well as lines of neutral metals, thereby permitting abundance and isotopic abundance analysis for cool stars.

We can take high resolution spectra of young, planetary mass objects, something that cannot be done at visible wavelengths. A 2000 K brown dwarf has an I-K ~ 5.5 and a V-K ~ 10 (Chabrier et al. 2000). The theoretical models tell us that the proposed spectrograph on the GMT should be able to take $S/N=50$, $R=25,000$ spectra in an hour of $5 M_{\text{Jupiter}}$ objects if they are at the distance of Ophiuchus. These results could test evolutionary models for the lowest mass free-floating objects and allow us to learn about the nature of objects similar to young gas giants in environments where their light does not need to be teased out of the radiation from a much more luminous parent object.

13.2.1.2.2 Understanding the Evolution of Stellar Populations and Chemical Elements

Section 3.2 deals with stellar populations. In 3.2.2, the science case includes a discussion of optical and infrared spectroscopy as applied to population and abundance studies. The advances in capability of the instrument proposed here will change the relative importance of optical and near-IR studies of stellar populations. The use of an immersion grating will allow GMTNIRS to realize a sensitivity gain proportional to the area of the telescope, rather than the diameter. More important, however, will be the gains in wavelength coverage. GMTNIRS will provide an order of magnitude improvement in instantaneous coverage at $R=50,000$.

We can perform high-resolution spectroscopy of red giants in open and globular clusters, the Galactic bulge and center, as well as the more luminous giants and supergiants in Local Group galaxies. The spectra, notably in the H-band, will provide the metallicity for chemical evolution studies.

13.2.1.3 Existing State of the Art

The proposed GMTNIRS science program will not take place in a vacuum. It is therefore worth understanding the competitive landscape during the period leading up to GMT first light. There are two instruments on 8 meter class telescopes that cover the same wavelength range at comparable resolving power. These are PHOENIX (Hinkle et al. 2000), currently at Gemini South as a quasi-visitor instrument, and CRIRES (Wiedemann et al. 2000) at the VLT. Both of these are single-order instruments with maximum resolving powers in the range of 70,000 for slit sizes of 0.15-0.2". CRIRES has the larger instantaneous wavelength coverage but this only amounts to $\lambda/70$. Gemini considered construction of a cross-dispersed spectrograph for the near-IR (HRNIRS) with $R \sim 10^5$, but it now appears that this instrument will not be funded. To the best of our knowledge, there are no other high resolution near-IR instruments in the pipeline for 8-10 meter class telescopes. GMTNIRS will therefore realize very substantial gains over the instruments now coming on line. The use of an immersion grating and a larger array will permit us to have a slit width comparable to that of the 8 meter telescopes but with better throughput at a given real resolving power. We will therefore improve sensitivity faster than as the area of the telescope. At the same time, the instantaneous wavelength coverage of GMTNIRS will be about 20 times that of CRIRES.

13.2.1.4. Baseline Instrument Specifications

Based on the list of science projects in Section 13.2.1.2, we have developed the following list of instrument requirements:

1. Operating Wavelength Range: 1.1 - 5.5 μm .

There are two reasonable choices for the lower limit of this range: the wavelength at which the quantum efficiencies of the short-wavelength HgCdTe detectors begin to become competitive with Si charge coupled detectors (0.7-0.8 μm), or the wavelength where the quantum efficiency of silicon detectors drops precipitously. We have chosen the latter boundary both to reduce instrumental complexity and to permit us to use silicon immersion gratings as the main dispersing elements. Large Si immersion gratings will not operate efficiently below 1.1 μm , due to absorption. The atmospheric cutoff at the upper end of the M-band sets the high end of the operating range.

2. Resolving Power: 2×10^4 - 2×10^5 .

This specification may depend on goals at particular wavelengths. Many of the interstellar applications of this instrument require the highest possible resolving power while observations of stars may not suffer much once R is greater than 3×10^4 . The goal of the design is to provide sufficient optical quality and pixel sampling to permit observations at $R \sim 10^5$ while allowing for work at lower resolving power at higher throughput and no loss of sensitivity due to detector readout noise.

3. Coverage: All of H, K, L, or M in a single exposure.

Broad simultaneous wavelength coverage at high spectral resolving power will be a new capability in the near-IR. Given the rich spectra of protostars, late-type stars, and brown dwarfs, the large coverage will permit new kinds of quantitative spectroscopy.

4. Noise Limits: source photon noise (1.1 - 2.4 μm), background (2.5 - 5 μm).
At 1.1-2.4 μm , the instrument should be limited by photon shot noise from the source at the shorter wavelengths and higher resolving powers and background limited at longer wavelengths and/or lower resolving powers. As far as possible into the K-band, the instrument should be source noise-limited at S/N=100 with the best possible throughput. At 2.5-5 μm , the system should be background-limited with the lowest possible background. In the L and M bands, background becomes an issue even at high spectral resolution. The instrument should insure the highest possible sensitivity by minimizing the effects of this background.

13.2.2 Optical Design

The instrument is a cross-dispersed echelle spectrograph. In the initial design, we assume availability of 4096^2 focal plane arrays, a reasonable extrapolation from current technology, given the timescale to GMT first light. We make use of silicon immersion gratings, a developing technology pioneered by our group. Section 13.7.4 describes the current and likely future state of the art of silicon immersion gratings. Based on reasonable extrapolation of current capabilities, the optical design is constructed around a silicon R3 (71.565°) immersion echelle that accepts a beam with an effective diameter of 83 mm. This results in a compact instrument design that fits within the provided space on the GMT instrument platform.

Alternatives to this design that do not use silicon immersion echelles are possible, but require collimated beam diameters that are larger (~285 mm) in order to achieve the same resolving power for a given slit width. The associated optical, optomechanical and mechanical systems will be correspondingly larger, more complex, and more expensive.

13.2.2.1 Goals

The primary goal of the instrument design is to provide GMTNIRS users with as much discovery space as possible. We expect to make gains over existing facilities through the increased aperture of the GMT, through expanded instantaneous wavelength coverage, and through improved instrument throughput. Many spectrographs for 8-10 meter class telescopes have had their designs driven to lower throughput by user demands for versatility. We have consciously chosen to avoid this. Given the different factors driving the design, it is clear from the outset that it is not possible to produce an instrument that is optimal across the entire proposed wavelength range. As a starting point for this study, we therefore have separate modules for the short and long wavelength bands.

13.2.2.2 Short Wavelength (1.1-2.4 μm) Module (SWM)

At the shorter end of the 1.1-2.4 μm range, where Strehls tend to be lower, or at lower resolving powers where the slit will be fairly large, a module designed for native seeing should have significantly greater throughput than that of an AO spectrograph. A native seeing instrument will also be able to take advantage of atmospheric conditions that are not optimal for AO. The instrument will use the largest practical grating size (300 mm) leading to an R3 immersion echelle with an 83 mm collimated beam. Given the refractive index of Si (~ 3.4 throughout the near-IR), this grating will have a diffraction limited resolving power, $R_{\text{diff}} = 1.72 \times 10^6 / \lambda$ where the wavelength λ is given in microns. The diffraction-limited spatial resolution of the telescope is $\theta_{\text{diff}} = 8.4 \times \lambda$ milliarcsec. These parameters lead to a resolving power of 10^5 for a 0.15" slit and 2.5×10^4 for a 0.6" slit. These slit width-resolution products are comparable to those available on high resolution IR spectrographs currently available on 4 and 8 m telescopes – CSHELL (Greene et al. 1993), PHOENIX (Hinkle et al. 1998), CRIRES (Kaüfl et al. 2004), meaning that comparable instruments will gain signal as the square of the aperture.

One goal of the design is for the optical performance to be capable of providing slit-limited resolution at $R = 50,000$ and as close as possible to slit-limited performance with an $R = 100,000$ slit. To keep pixel sampling from being a limiting factor, we set the plate scale of the camera to put 2.25 pixels across the narrowest (0.15") slit. At the same time, we would like to produce continuous spectra of each of the infrared atmospheric windows (K: 2.0-2.5 μm ; H: 1.4-1.8 μm ; J: 1.15-1.35 μm) in a single exposure. The J window is shortened at the short wavelength end by absorption in the silicon immersion grating. To permit sky subtraction for point sources while always integrating on the target, we set a goal of a 5 arcsec slit length with a minimum of 6 arcsec between spectral orders. Finally, we tried to optimize the design for maximum throughput.

13.2.2.2.1 First Order Calculations

In the optical design, we assume that the 4096×4096 HgCdTe array has a pixel pitch of 18 μm , although field-of-view limitations by the camera optics may not allow such a large array to be fully exploited. The requirement of 2.25 pixels per 0.15-arcsec slit sets the range of incident angles seen by a single pixel in the focal plane, or the instantaneous field-of-view (IFOV), at 1/15 (0.0667) arcsec. The effective focal length (EFL) at the array is then given by:

$$\frac{\text{pixelpitch}}{EFL} = 0.0667 \text{ arc sec} = 0.3232 \mu\text{rad} \Rightarrow EFL = 55,691 \text{ mm}$$

The bounding aperture diameter of the GMT primary is 24.5 m, giving an f /ratio at the focal plane of 2.27. Since we are basing the design of the system on an R3 immersion echelle able to accommodate an 83 mm beam, the EFL of the spectrograph camera will be approximately $(83 \text{ mm}) \times 2.27 = 188.7 \text{ mm}$. There will be anamorphic magnification in the spectrograph due to the echelle and cross-disperser being used off the Littrow condition. The cross-disperser typically produces the larger anamorphic contribution.

The external dimension of the 4096^2 array is $4096 \times 18 \mu\text{m} = 73.728 \text{ mm}$ along each side. At a distance equal to the camera focal length, the array subtends a field of view of

$$\theta_{cam} = \pm \tan^{-1} \left[\frac{36.864}{188.7} \right] = \pm 11.05^\circ$$

The maximum potential full field of view of the camera is therefore 22.1×22.1 square degrees, or 31.3 degrees across the diagonal. When combined with the relatively fast focal ratio and significant beam diameter, field of view becomes the most formidable challenge to the spectrograph camera design.

13.2.2.2.2 Slit Dimensions

The EFL of the GMT at the f/8 Gregorian focus is 196,000 mm. At the GMT Gregorian focus, a 5 arcsec slit is 4.75 mm long and a 0.15 arcsec slit width corresponds to $142.5 \mu\text{m}$. For f/8 and K-band wavelengths, the diffraction point spread function ($2.44 f\lambda$) is approximately one-third of the slit width.

13.2.2.2.3 Collimator

A major fixed constraint to the system is the maximum likely physical size available for the silicon immersion grating. To make the most effective use of a given-sized grating, the collimator should form a pupil image on the echelle surface. To produce an 83 mm collimated beam at f/8, the collimator EFL should be $(83 \text{ mm}) \times 8 = 664 \text{ mm}$. For a 4.75 mm slit length and collimator EFL of 664 mm, the collimator full field of view is only 24.6 arcmin. A paraboloid of this focal length used about 80 mm off-axis works very well. The foreoptics that produce an initial pupil where we can place a cold stop for K-band work have not been included in this conceptual design but should be similarly undemanding.

13.2.2.2.4 Echelle Orders and Cross-Dispersion

13.2.2.2.4.1 Camera Field of View Trade

The desire to cover either of the H or K atmospheric windows in a single exposure drives the design of the SWM spectrograph camera. A 2048^2 array does not provide adequate spectral and spatial sampling along a long enough slit when one desires to cover an entire near-IR atmospheric window at $R = 10^5$. However, making full use of a 4096^2 array requires a camera field of view of 22 degrees square. We have carried out design studies of all-reflective, three mirror anastigmat (TMA) cameras and find that fields of approximately 14×14 to 14×20 degrees are the upper limit for acceptable image quality at the required f/# ratios. These fields are also the extreme upper limit for refractive camera designs due to size limitations on available lens materials.

Empirically, it appears that about 250 “square degrees” of camera coverage are required to meet the goals of the most stressing bands (the J-band is actually the most stressing and K the least). Because of anamorphic magnification effects, it turns out to be most effective to make the Y field as large as the optics will tolerate (see optical design report in the appendices). This study fixes the Y (immersion echelle) field at 18 degrees and allows the X-field to grow to meet the design requirements, such as minimum order separation. This approach can result in X (cross-disperser) fields up to 16 degrees for the J-band. This trade space should be more carefully explored during a phase A design study.

We note that an 18 degree field of view corresponds to a pixel count of 3321 in the immersion echelle direction. The necessary pixel count in the cross-dispersion direction will be given for each band in the discussion for that band’s cross-disperser.

13.2.2.2.4.2 K-band

The resolving power of the immersion echelle at a given slit width is set by the maximum size of available high-purity silicon boules. Our initial design calculations were carried out with a spreadsheet that inverts the grating equation in a refractive medium, solving for the wavelength that falls at a specified angle δ' away from the Littrow angle α . The spreadsheet uses a dispersion model for the refractive index of silicon to correct for the variation in refractive index across the band, and λ is solved iteratively for $\pm\delta'$. As discussed above, δ' is fixed at 9° , representing the maximum camera field of view.

Since photolithography masks can be manufactured with any possible line spacing, the groove density of the immersion echelle is a free parameter that can be used to optimize the design. With an echelle groove width of $25.1\ \mu\text{m}$, placing $2.50\ \mu\text{m}$ in echelle order 66 gives continuous spectral coverage on the array while covering the K-band down to $2.05\ \mu\text{m}$ in 15 orders, as shown in Table 13.2-1. There are $3321/2.25 = 1476$ resolved elements across each order. At the center of the K-band, the width of order 72 is $\Delta\lambda = 33.935/1476 = 0.023\ \text{nm}$, and the resolving power is consequently $2275/0.023 = 98,000$ and varies by about 1000 across the K-band. Note that the groove width is equivalent to a width of $86\ \mu\text{m}$ in vacuo, considerably coarser than any commercially available ruled gratings.

order	$\lambda -$ (nm)	λ center (nm)	$\lambda +$ (nm)	order overlap	nm / order
66	2462.241	2482.064	2499.358		37.117
67	2426.004	2445.522	2462.552	0.311	36.548
68	2390.840	2410.063	2426.836	0.831	35.995
69	2356.704	2375.639	2392.162	1.321	35.458
70	2323.549	2342.206	2358.486	1.782	34.936
71	2291.337	2309.722	2325.766	2.216	34.429
72	2260.026	2278.147	2293.961	2.625	33.935
73	2229.580	2247.445	2263.036	3.010	33.455
74	2199.965	2217.579	2232.952	3.372	32.988
75	2171.146	2188.517	2203.678	3.714	32.533
76	2143.092	2160.226	2175.181	4.036	32.089
77	2115.774	2132.677	2147.431	4.339	31.657
78	2089.163	2105.841	2120.399	4.625	31.236
79	2063.233	2079.691	2094.058	4.895	30.825
80	2037.958	2054.201	2068.382	5.149	30.424

Table 13.2-1. K-band orders and coverages.

A 354 lines/mm cross-disperser used in-plane (no γ tilt) at an incidence angle $\alpha = 10.96^\circ$ disperses the echelle orders on the focal plane. The separation of the highest orders is 6.0 arcsec while the lowest orders are separated by almost 10 arcsec. There are 2189 pixels in the cross-dispersion direction.

The β angle for the center of the K-band ($\lambda = 2.275 \mu\text{m}$) is 37.96° , giving an α - β of 27° to allow the cross-dispersed beam to clear the echelle on its way to the camera. To maintain a fixed camera location, α - β is maintained for all the SWM cross-dispersers. The behavior of the cross-disperser gratings at these operating angles should be modeled with a rigorous vector code to assess efficiency and polarization effects.

13.2.2.2.4.3 H-band

In order to cover the H-band while maintaining the minimum order spacing of 6 arcsec, it is necessary to use a different immersion echelle than that used in the K-band. This change has consequences both for the mechanical design and for the immersion grating production program, but in both cases, the substantial benefits of the change outweigh the costs. In our H-band baseline design, we use a groove width of $18.2 \mu\text{m}$. The 17 echelle orders then cover the H-band and the resolving power at the band center of $1.62 \mu\text{m}$ in order 74 is 102,000.

order	$\lambda -$ (nm)	λ center (nm)	$\lambda +$ (nm)	order overlap	nm / order
67	1770.455	1784.419	1796.616		26.161
68	1745.132	1758.879	1770.887	0.432	25.755
69	1720.554	1734.089	1745.915	0.783	25.361
70	1696.689	1710.019	1721.666	1.112	24.977
71	1673.506	1686.637	1698.110	1.422	24.604
72	1650.978	1663.915	1675.219	1.713	24.241
73	1629.078	1641.825	1652.965	1.987	23.888
74	1607.779	1620.342	1631.323	2.245	23.544
75	1587.059	1599.443	1610.267	2.488	23.208
76	1566.893	1579.103	1589.775	2.716	22.882
77	1547.262	1559.300	1569.825	2.931	22.563
78	1528.143	1540.016	1550.396	3.134	22.252
79	1509.518	1521.229	1531.468	3.324	21.949
80	1491.369	1502.921	1513.022	3.504	21.654
81	1473.676	1485.074	1495.041	3.673	21.365
82	1456.425	1467.671	1477.508	3.831	21.083
83	1439.598	1450.698	1460.406	3.981	20.808

Table 13.2-2. H-band orders and coverages.

The groove density of the H-band cross disperser is 536 lines/mm. The incidence (α) angle remains nearly the same as for the K-band: 13.01°. The separation of the highest orders is 6.0 arcsec while the highest orders are separated by almost 11 arcsec. There are 2664 pixels in the cross-dispersion direction.

13.2.2.2.4.4 J-band

The 18.2 μm groove width echelle used for the H-band can be used to cover the J-band from 1.35 μm down to 1.15 μm . At the shortest wavelengths, internal absorption in the silicon immersion echelle may limit the system efficiency. Table 13.2-3 shows the 17 orders required. The resolving power in order 99 at 1.25 μm is 108,000.

order	$\lambda -$ (nm)	λ center (nm)	$\lambda +$ (nm)	order overlap	nm / order
90	1332.520	1342.676	1351.565		19.044
91	1318.602	1328.634	1337.416	4.895	18.814
92	1304.994	1314.906	1323.582	4.980	18.588
93	1291.688	1301.481	1310.055	5.060	18.367
94	1278.673	1288.350	1296.823	5.135	18.150
95	1265.940	1275.503	1283.877	5.204	17.937
96	1253.480	1262.932	1271.209	5.270	17.729
97	1241.285	1250.628	1258.810	5.330	17.525
98	1229.347	1238.583	1246.672	5.387	17.325
99	1217.658	1226.788	1234.787	5.440	17.129
100	1206.211	1215.238	1223.147	5.489	16.936
101	1194.998	1203.924	1211.745	5.534	16.747
102	1184.012	1192.839	1200.574	5.576	16.562
103	1173.247	1181.976	1189.627	5.615	16.380
104	1162.697	1171.330	1178.898	5.650	16.201
105	1152.355	1160.894	1168.380	5.683	16.025
106	1142.215	1150.662	1158.068	5.713	15.853

Table 13.2-3. J-band orders and coverages (18.2 μm groove width).

While the J-band can use the same immersion grating as the H-band, a different cross-disperser is required to spread the orders efficiently. A 881 lines/mm cross-disperser used at $\alpha = 20.99^\circ$ results in a minimum order spacing of 6 arcsec at order 106 and gives an order spacing of 10.5 arcsec between orders 90 and 91. The X-field of the camera is about $\pm 7.96^\circ$ and the array needs 2930 pixels in the cross-dispersion direction.

13.2.2.2.4.5 Summary, SWM dispersion

Two echelles and three cross-dispersers meet the design goal of single-exposure coverage in the K, H, and J-bands with a separation of at least 6 arcsec between orders. The field of view of the spectrograph camera is up to 18×16 degrees with a focal length of 188.7 mm. A focal plane of 3331×2930 18 μm pixels is required, the cross-dispersion dimension being dictated by the J-band.

13.2.2.2.5 SWM Baseline Configuration

The choice of a baseline configuration for the short wavelength module follows from an exploration of the camera design, the most difficult part of the optical train. Figure 13.2-1 shows the end-to-end baseline optical layout for this module, covering bands K, H, and J. The $f/8$ beam directly from the GMT Gregorian focus or following the insertable cold pupil stop optics (1:1 at $f/8$ as well) is first collimated by a paraboloid of 664 mm focal length used 80 mm off-axis. After being dispersed and cross-dispersed, the beams from

each wavelength are re-imaged onto the focal plane array (FPA) by an all-reflective three-mirror anastigmat (TMA) with an effective focal length of 188.7 mm. The TMA functions equally well for all bands. In order to meet the spectral coverage and slit-length requirements, the echelle and cross-disperser must be switched between K-Band and H-Band. The H-Band echelle also covers J-Band, but requires a third cross-disperser.

Figure 13.2-2 shows how the K-band orders fall on the FPA. Orders 66 and 67 (see Table 13.2-1) represent the longest wavelengths. Orders 72 and 73 are near the center of the band, and orders 79 and 80 are at the short wavelength end. The short lines in each order represent the length of the 5 arcsec slit. The minimum order separation of 6 arcsec occurs between the top (short wavelength) ends of orders 79 and 80. Figure 13.2-3 shows geometric spot diagrams for the representative positions on the orders of Figure 13.2-2. The square in each position is the width of the 0.15 arcsec slit: 2.25 pixels, or 40.5 μm . The optical performance is even better than it looks. There are 316 rays traced in each wavelength, so the spots are unresolved in their cores and do not give a true picture of the compactness of the central spike. We have calculated the fraction of rays landing on a single 18 μm pixel. At the center of each order, all rays land within the pixel. At the corners of the spectrum, the fraction of included rays is typically 95-99% and, in the worst case, only drops to 85% (see Table 11 of Appendix 1 for more detail.)

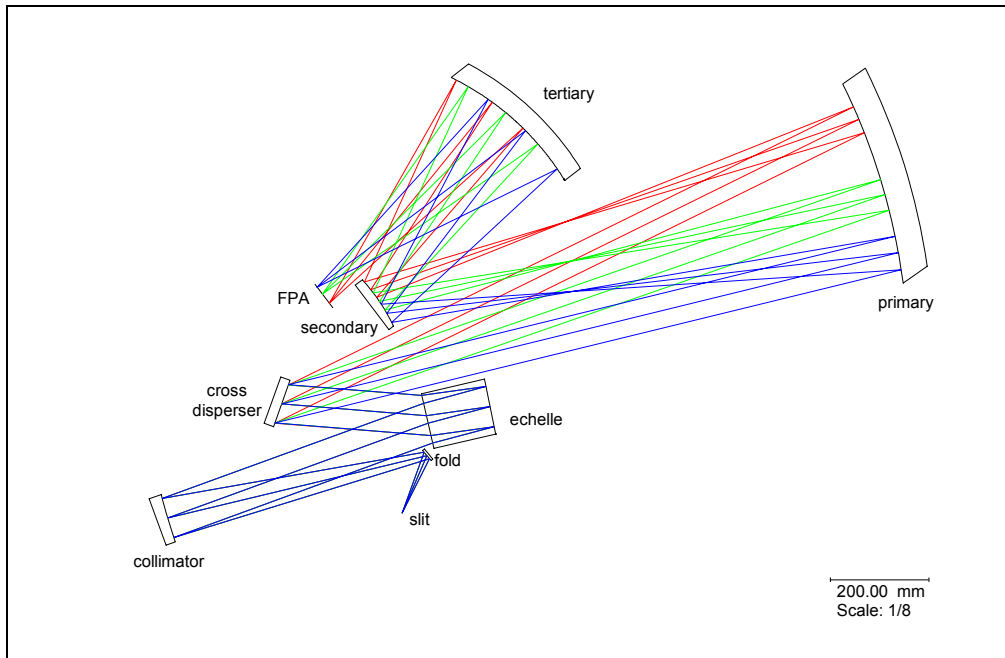


Figure 13.2-1. End-to-end optical model of the short wavelength module.

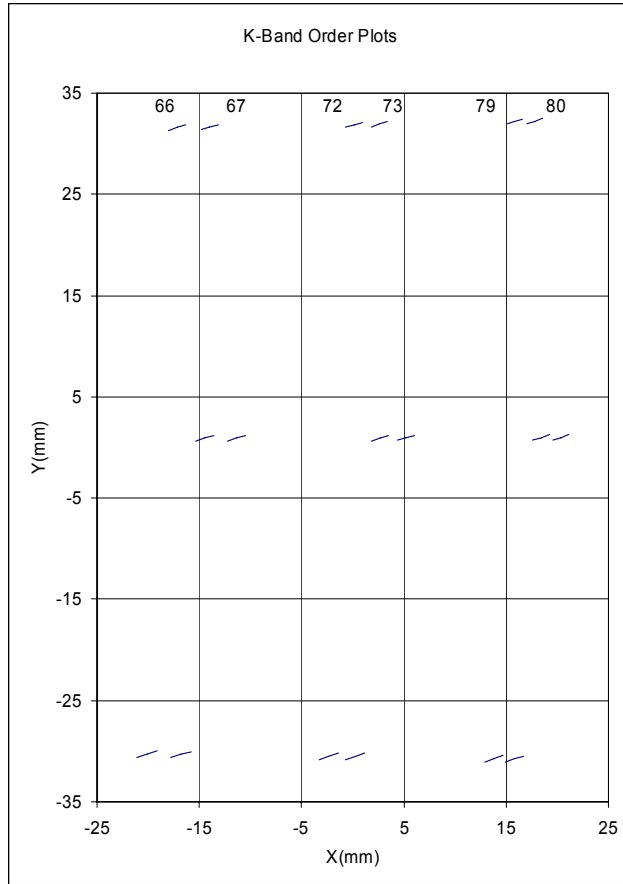


Figure 13.2-2. Layout of K-band orders on the FPA. The stripes represent 5 arcsec long slits at the beginning, middle, and end of representative orders.

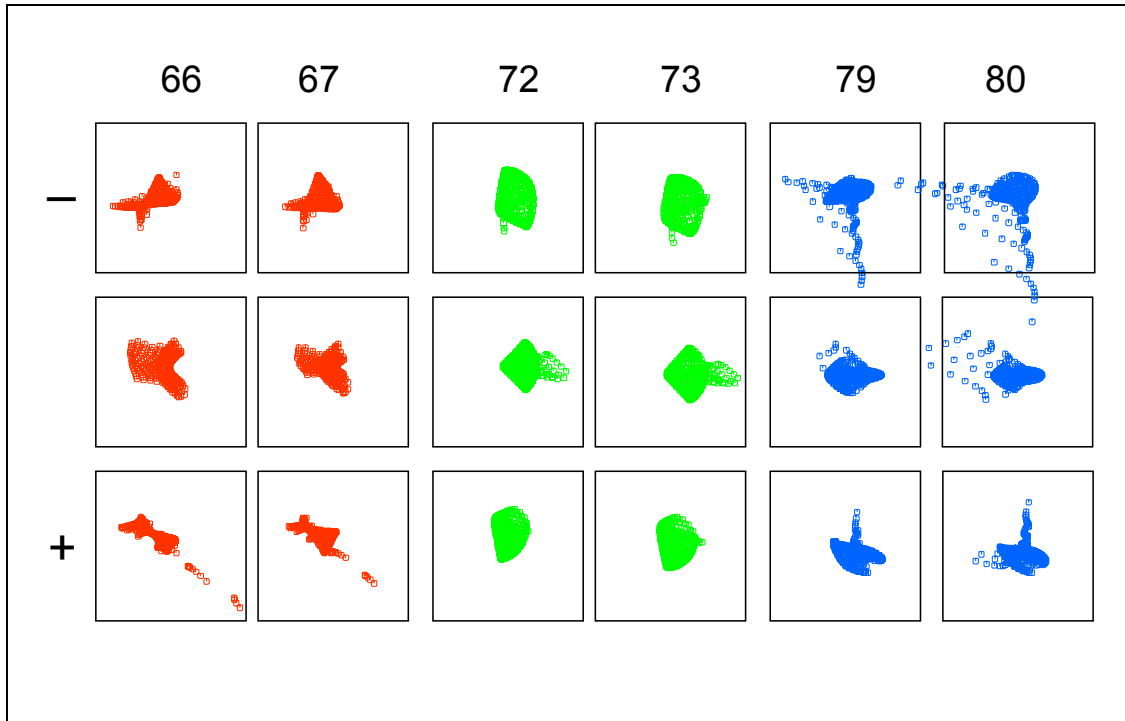


Figure 13.2-3. Geometric spot diagrams for K-band orders. Each square has a width of 40.5 μm , the width of an R=100,000 slit.

13.2.2.2.5.1 Echelle Geometry

In order to clear the beam entering the echelle from the collimator and the dispersed beam heading from the echelle to the cross-disperser, the echelle is tipped by 7 degrees with respect to the beam from the collimator. The direction of this tip is perpendicular to the plane of the echelle dispersion. In immersion, at the grating itself, this cant is effectively smaller by a factor of 3.4.

To avoid a ghost reflection from the front surface of the echelle, the front and rear faces of the echelle are wedged with respect to each other by a small angle. Most of this reflection from the front surface is eliminated by a good antireflection coating (see Section 13.2.4.4.4.1).

13.2.2.2.5.2 Three Mirror Anastigmat, baseline

Table 13.2-4 lists the parameters for the three mirrors comprising the TMA for the SWM. To get a handle on fabrication and alignment sensitivities, each mirror of the baseline system was perturbed manually in each of its rigid body degrees of freedom. Except as noted, the only compensation permitted was a tilt and refocus of the focal plane. For mirror radii of curvature, an axial respace was also allowed, based on the assumption that the TMA will be re-optimized before assembly based on the knowledge of the as-built radii.

The minimum perturbations assumed that the rigid body positions of fiducial references on the mirrors could be measured to an accuracy of 0.010 inch (0.25 mm) relative to each other. This is well within the ability of industrial coordinate measuring systems (e.g. Romer Arm). The level of positional accuracy implies an angular control of about 625 μ rad, or 0.036 degrees for the primary and tertiary mirrors (~400 mm).

	Primary	Secondary	Tertiary
Radius of Curvature (mm)	1201.558 ccv	409.909 cvx	569.439 ccv
Conic Coefficient	-0.262316	+5.763239	+0.060340
Conic Type	ellipsoid	oblate spheroid	oblate spheroid
General Aspheric coefficients (mm)	n/a	n/a	$-9.015e^{-17} r^6$
Lateral Displacement X (mm)	343.114	6.924	22.896
β Tilt (XZ-plane) (deg)	-12.865	-1.868	-2.041
Parent diameter (mm)	1200	160	600
Used section size (X x Y, mm)	326 x 600	110 x 80	340 x 440
Radial offset of section center X (mm)	371.5	20.5	95.5
Parent max departure from best-fitting sphere (mm)	0.644	0.098	0.126

Table 13.2-4. TMA mirror parameters for the short wavelength module.

A single merit function has difficulty capturing the effect of tolerances on total system performance. The ensemble of images is necessarily a compromise, and some image positions actually improve at the expense of others as a result of alignment perturbations. For now, a simple mean and standard deviation of the ensemble 80 percent geometric encircled energies has been adopted (see Table 13 in Appendix 1). Large growth in either the mean or standard deviation is a signal of potential sensitivity, either from an overall degradation or a major deterioration in a few images. If the minimum perturbation did not produce a significant change, the magnitude was increased until its effect was observable in the metrics. The changes in the metrics appear to be fairly linear with the magnitude of the perturbation.

For the primary mirror, spacing to the secondary is the most sensitive parameter. Radius of curvature error can be up to 9 mm (0.7%), provided it can be measured and re-spaced.

The secondary radius of curvature needs to be held to better than 0.5% because the secondary needs to interact with both other mirrors and a respace that helps one side probably hurts the other. Because the secondary is smaller than the primary there is a small lever arm to control angular position through linear position measurements. A tighter positional tolerance (e.g. 0.125 mm) should be adequate.

The tertiary radius also needs to be known to better than 0.5%. Again, knowledge is more important than absolute control of the radius. Tilt of the tertiary is also sensitive.

The worst case is a composite of perturbations to all three mirrors, compensated only by respacing the mirrors and refocusing the FPA. Performance is worse, but not catastrophic, for this situation. For a more rigorous prediction, this case should be repeated in a Monte Carlo fashion for a realistic distribution of errors in knowledge and compensator adjustments. In practice, the as-built mirror parameters will be re-optimized before alignment begins. The alignment itself can then be guided by interferometric testing provided the initial adjustments allow the mirror to be placed within the capture range of the interferometric test and subsequent adjustments are repeatable and orthogonal to each other. In order for the alignment to converge, the mounting structure also needs to be predictable (non-hysteretic) between room temperature and operating temperature, so that adjustments made at room temperature translate reliably into locations at operating temperature.

This initial look indicates that the fabrication and alignment tolerances, while requiring care, are not impossibly tight.

The fabrication of the three TMA mirrors is potentially challenging, but well within the state-of-the-art of modern computer-controlled optical surfacing (CCOS). The secondary and tertiary are small enough to be fabricated from their parent surfaces if this proves to be advantageous. The primary, because of its larger parent, would likely be generated directly as an off-axis section.

High precision cryogenic mirrors are typically fabricated by diamond-turning electroless Ni on 6061-Al, with typical rms roughness of 50-70 Angstroms for plano and spherical surfaces and ~100 Angstroms for off-axis aspheres. This roughness is detrimental for short-wavelength applications (visible, ultraviolet) because diffractive effects from the grooves left by the diamond-machining process have been unacceptably large. Subsequent polishing to remove these grooves can harm the surface figure. A further complication of using Ni on Al is the different coefficients of thermal expansion (CTEs) between the two materials, leading to bimetallic stresses that can distort the figure when the mirror is cooled.

New methods have been developed by several optics companies to deal with this problem. One method employs a thin (3-5 μm) layer of high-purity Al (Alumiplate™) rather than nickel. This matches the thermal expansion coefficients of the mirror surface material to the 6061 Al substrate. Surface rms roughness numbers of 20-40 Angstroms for diamond-turned Alumiplate™ have been reported by II-VI, Inc. (AlumiPlate, Inc. 1998). Some care with substrate preparation is necessary to ensure that the Al overlayer adheres properly.

Another attractive possibility for fabricating the TMA mirrors is the Corning NetOptix LEC process in which the mirror surface is diamond-machined directly into the Al 6061-T6 surface without producing grooves that have residual scattering and/or diffractive effects. Surface rms errors of 20-50 Angstroms have been reported by Corning. They claim that the resulting plot of power spectral density appears smooth and diffractive

effects are eliminated (Corning NetOptix 2004, see Appendix). Such mirrors would perform satisfactorily into the UV region and be more than adequate for the IR needs in GMTNIRS.

13.2.2.2.5.3 TMA alternative configurations

In the course of exploring the baseline TMA design for the short wavelength module, two derivative forms showed some merit and could be developed in the future. The first is a variation of the TMA which is more compact, at the expense of more aspheric departure on the surfaces. This design is shown in Figure 13.2-4.

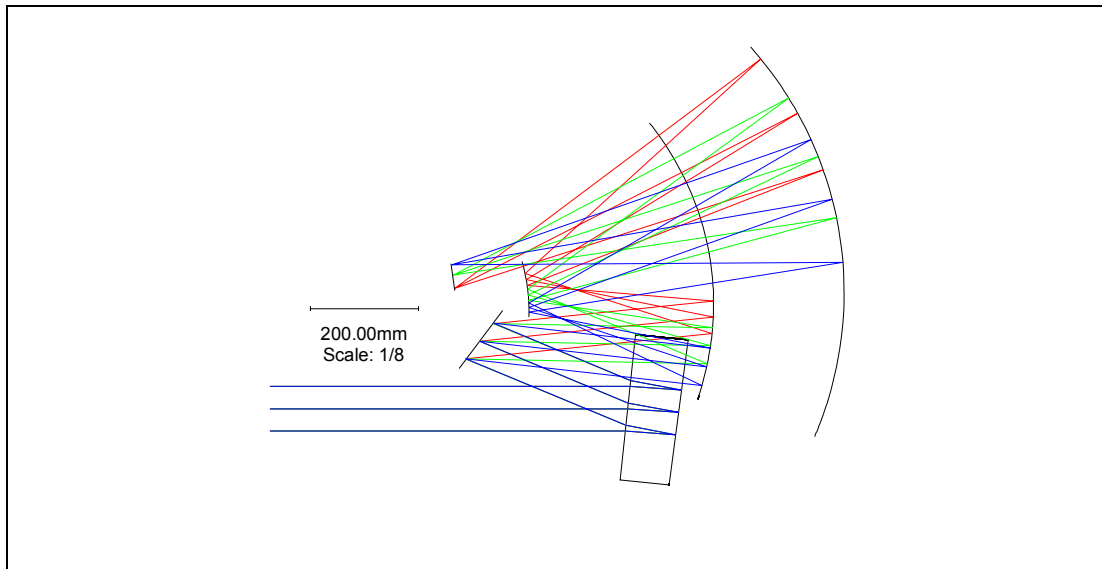


Figure 13.2-4. Compact form of an alternative TMA (parent surfaces shown).

The second alternative reflects a different design philosophy with significant advantages and disadvantages relative to the baseline. Figure 13.2-5 shows how a TMA is used in one dimension only to reimage the collimated beams dispersed by the echelle. An image of the slit is formed with all the echelle orders superimposed. A refractive camera is then used in double pass with the cross-disperser near Littrow to separate the echelle orders. The elements of the echelle 1-D TMA are smaller and less eccentric, and therefore considerably easier to fabricate and align. Because the TMA is working over a smaller field, the image quality at the slit image is better than for the 2-D baseline TMA in Section 13.2.2.2.5.2. The 1-D TMA can also be designed to reform a pupil near the cross-disperser (i.e. a “white pupil” design), so that the size of the cross-disperser is reduced and the sizes of the lens elements become more manageable.

Offsetting these advantages are the greater transmission losses at the 12+ surfaces of the lens elements. Even with good anti-reflection coatings, overall losses could be as high as 24 percent. However, the cross-disperser is operating more nearly at Littrow and consequently should have better diffraction efficiency than the cross-dispersers of the

baseline with their relatively large α - β angles. The example shown uses two silicon elements and one germanium element, and so is not suitable for H and J bands. Achromatizing and anti-reflection coating the refractive collimator/camera across K, H, and J bands is a significant challenge, and separate cameras could be required to optimize both image quality and transmission. However, the smaller element dimensions permit a wider choice of lens materials. A second reflective camera is also possible, but would be considerably larger.

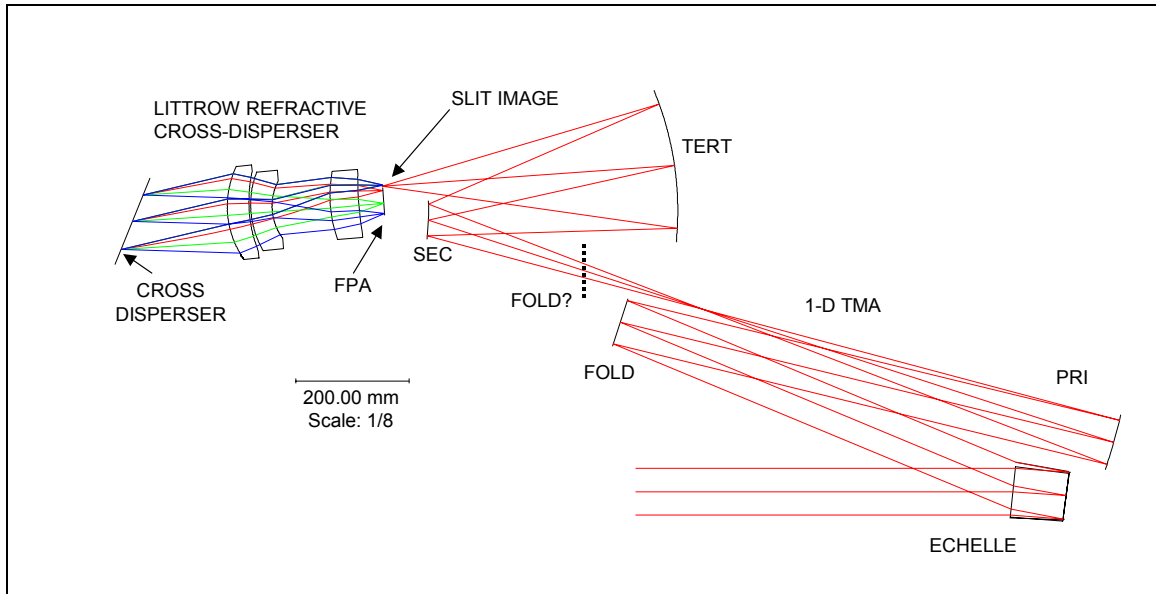


Figure 13.2-5. One-dimensional TMA with refractive cross-disperser section.

13.2.2.3 Long Wavelength (2-5 μm) Module (LWM)

At the longer wavelengths where background radiation can seriously compromise sensitivity even at high spectral resolution, it is desirable to use adaptive optics to reduce the warm solid angle seen by the system. In our strawman design, we therefore assume that the long wavelength module (LWM) is used with adaptive optics (AO). Our design assumes a slit width of $2\lambda/D$ to capture most of the power in the diffraction spike. In this conceptual design, the LWM requires Si immersion gratings that can accommodate a 45 mm diameter collimated beam. Since these echelles are considerably smaller than the maximum feasible size, the collimated beam diameter and the blaze angle of the immersion grating can be optimized in subsequent versions of the design. With a single camera designed to provide 3 pixels per slit at $\lambda = 3.5 \mu\text{m}$, we would have a slit width of 60 milliarcsec and a resolving power of 133,000 at that wavelength.

At $\lambda = 4.6 \mu\text{m}$, we would have an 80 milliarcsec slit sampled by 4 pixels and a resolving power of 100,000. As with the short wavelength module, it is a requirement that the long wavelength module cover entire atmospheric windows (L-band: 2.90-4.15 μm ; M-band: 4.58-5.40 μm) in a single exposure. The small size of the diffraction-limited beam

permits observations of the source plus sky simultaneously using very modest slit lengths. In this design, we set a minimum order separation of 0.6 arcsec as a requirement. For this module, we assume availability of 2048×2048 focal plane arrays with $18 \mu\text{m}$ pixels. These arrays are already being developed for JWST. In several places in the discussion, we point out where a larger array might improve the instrument performance.

13.2.2.3.1 First Order Calculations

The requirement of 3 pixels per 0.06-arcsec slit sets the pixel IFOV at 0.02 arcsec. The effective focal length (EFL) at the array is then given by:

$$\frac{\text{pixelpitch}}{EFL} = 0.02 \text{ arc sec} = 0.097 \mu\text{rad} \Rightarrow EFL = 185,638 \text{ mm}$$

The bounding aperture diameter of the GMT is 24.5 m, giving an f /ratio at the focal plane of $f/7.58$. The mid-wave echelle is baselined as R3 or less with a beam diameter of 45 mm. Therefore the EFL of the spectrograph camera will be approximately $(45 \text{ mm}) \times 7.58 = 341 \text{ mm}$.

The external dimension of the 2048^2 array is $2048 \times 18 \mu\text{m} = 36.864 \text{ mm}$ along each side. At a distance equal to the camera focal length, the array subtends a field of view of

$$\theta_{cam} = \pm \tan^{-1} \left[\frac{18.432}{341} \right] = \pm 3.09^\circ$$

The full field of view of the camera is therefore 6.19×6.19 square degrees, or 8.75 degrees across the diagonal. These camera specifications are considerably easier than those for the short-wave module.

13.2.2.3.2 Slit Dimensions

The EFL of the GMT at the $f/8$ Gregorian focus is 196,000 mm. At the GMT Gregorian focus, a 0.6 arcsec slit is 0.57 mm long. The L-band 0.06 arcsec slit width corresponds to $57 \mu\text{m}$, and the 0.08 arcsec slit for the M-band corresponds to $76 \mu\text{m}$. For $f/8$, the respective diffraction point spread functions (PSFs) at 3.5 and $4.6 \mu\text{m}$ are $68 \mu\text{m}$ and $90 \mu\text{m}$, meaning that there will be partial coherence effects from the slit that should be modeled with a rigorous (i.e. non-FFT) physical optics code.

13.2.2.3.3 Collimator

To produce a 45 mm collimated beam at $f/8$, the collimator EFL should be $(45 \text{ mm}) \times 8 = 360 \text{ mm}$. For a 0.57 mm slit length and a collimator EFL of 360 mm, the collimator full field of view is only 5.4 arcmin. A paraboloid of this focal length used about 50 mm off-axis works very well. The collimator will form a pupil image approximately one focal length from the paraboloid. This pupil should fall on the echelle surface to minimize the size of the echelle. The collimator can be folded as necessary for packaging.

13.2.2.3.4 Echelle Orders and Cross-Dispersion

13.2.2.3.4.1 L-band

The L-band design cannot quite meet the goal of covering 2.90-4.25 μm in a single exposure within the limits set by the 2048×2048 array. The Y-field of view is therefore increased from 6.19 to 8.0 degrees, corresponding to an increase in the Y focal plane size to 2650 pixels. For purposes of calculating resolving power, there are $2650/3 = 883$ resolved elements across the array.

An R2.68 echelle with a groove width of 86 μm covers the L-band in 58 orders with no gaps. Table 9 of Appendix 1 gives complete details of the order lengths and separations. The resolving power at $\lambda = 3.5 \mu\text{m}$ in the central order (158) is then $3500/(26.366/883) = 117,000$. Even better resolving power could be achieved with a slightly larger array format and field of view if it would benefit the science utility.

A 106 lines/mm L-band cross-disperser at $\alpha = 2.54^\circ$ and $\alpha - \beta = 17^\circ$ gives a minimum separation between orders of 0.6 arcsec and ranges up to 1.27 arcsec for the lowest orders. The X-field of the camera is $\pm 4.1^\circ$ and the X-dimension of the array is 2700 pixels.

13.2.2.3.4.2 M-band

As with the short-wavelength module, the LWM optical solutions perform considerably better if we can exchange both the immersion echelle and the cross-disperser for different atmospheric windows. The M-band design meets the goal of covering 4.58-5.40 μm in a single exposure with a 2048×2048 array. An R3 echelle with groove widths of 160 μm covers the band in 36 orders (see Table 10 of Appendix 1). With the 2048×2048 array and 4 pixels per slit width, there are 512 resolved spectral elements, so the resolving power at 4.6 μm in order 226 is $4600/(24.039/512) = 98,000$.

A 113 lines/mm cross-disperser operating at $\alpha = 8.06^\circ$ and $\alpha - \beta = 17^\circ$ has good optical performance and results in a minimum order separation of 0.85 arcsec.

13.2.2.3.5 LWM Baseline Configuration

Figure 13.2-6 shows the end-to-end baseline optical layout for the long wave module covering the L and M bands. The f/8 beam from the GMT Gregorian focus is first collimated by a paraboloid of 360 mm focal length that is used 70 mm off-axis. This collimator also creates a pupil image for use as a cold stop which is then relayed to the echelle by another pair of paraboloids. For packaging, the collimator offsets can be in the direction perpendicular to the plane of the figure. The slit should be perpendicular to the offset direction. After being dispersed and cross-dispersed, the beams from each wavelength are reimaged onto the FPA by an all-reflective three-mirror anastigmat (TMA) with an effective focal length of 341 mm. The TMA functions equally well for both bands. The echelle must be switched between L-band and M-band. The L and M-Band cross dispersers are similar enough (106 and 113 lines/mm) that a common cross disperser might be possible.

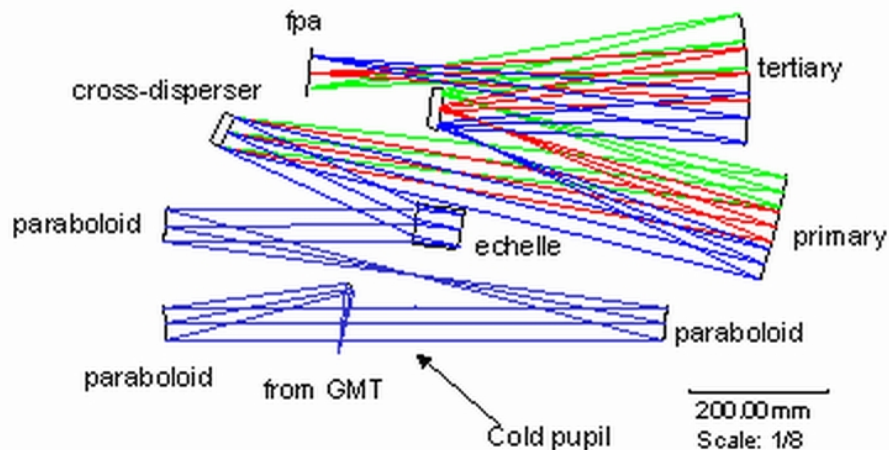


Figure 13.2-6. End-to-end optical model of the long wavelength module.

Figure 13.2-7 shows how the L-band orders fall on the focal plane array. Figure 13.2-8 shows geometric spot diagrams for the representative positions on the orders of Figure 13.2-7. The square in each position is one pixel, or $18 \mu\text{m}$. The diffraction central maximum at the shortest L wavelength is still over $50 \mu\text{m}$, so the images are diffraction limited. The wavefront quality produced by the overall optical system is excellent, with rms optical path differences < 0.012 waves over the whole spectrum (Table 14 of Appendix 1).

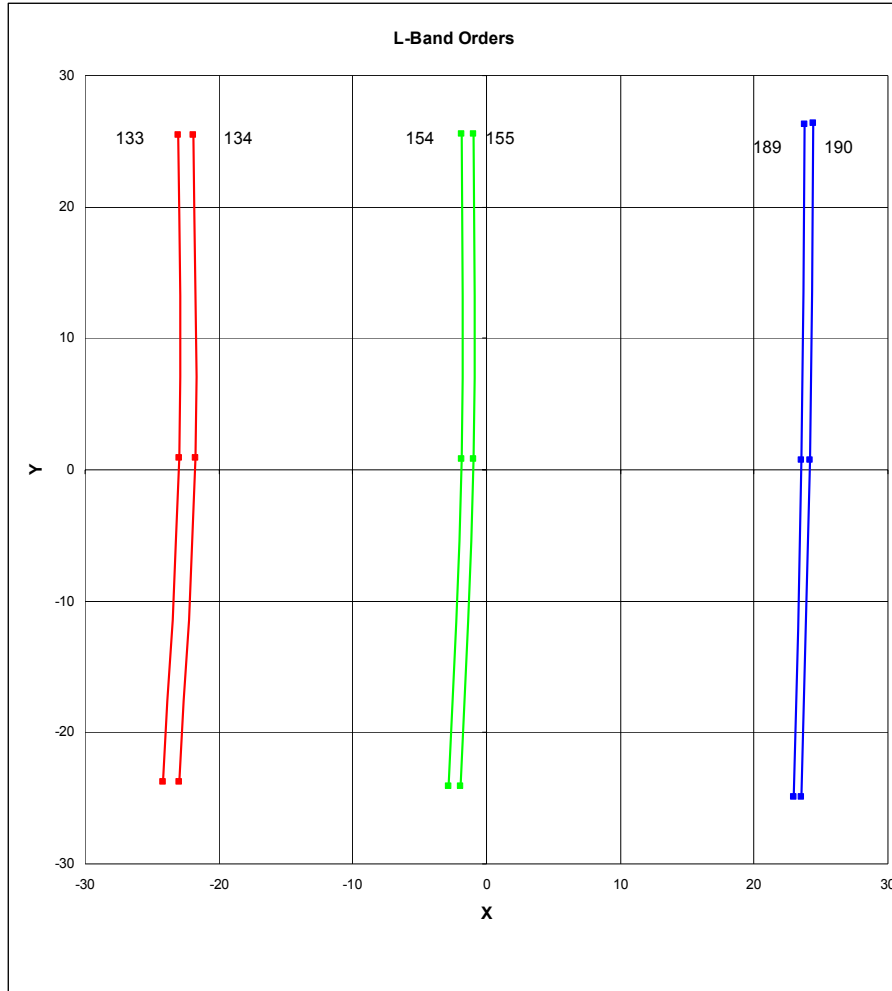


Figure 13.2-7. L-band orders. Orders 133 and 134 (see Table 13.2-5) represent the longest wavelengths. Orders 154 and 155 are near the center of the band, and orders 189 and 190 are at the short wavelength end. The slit length is not shown.

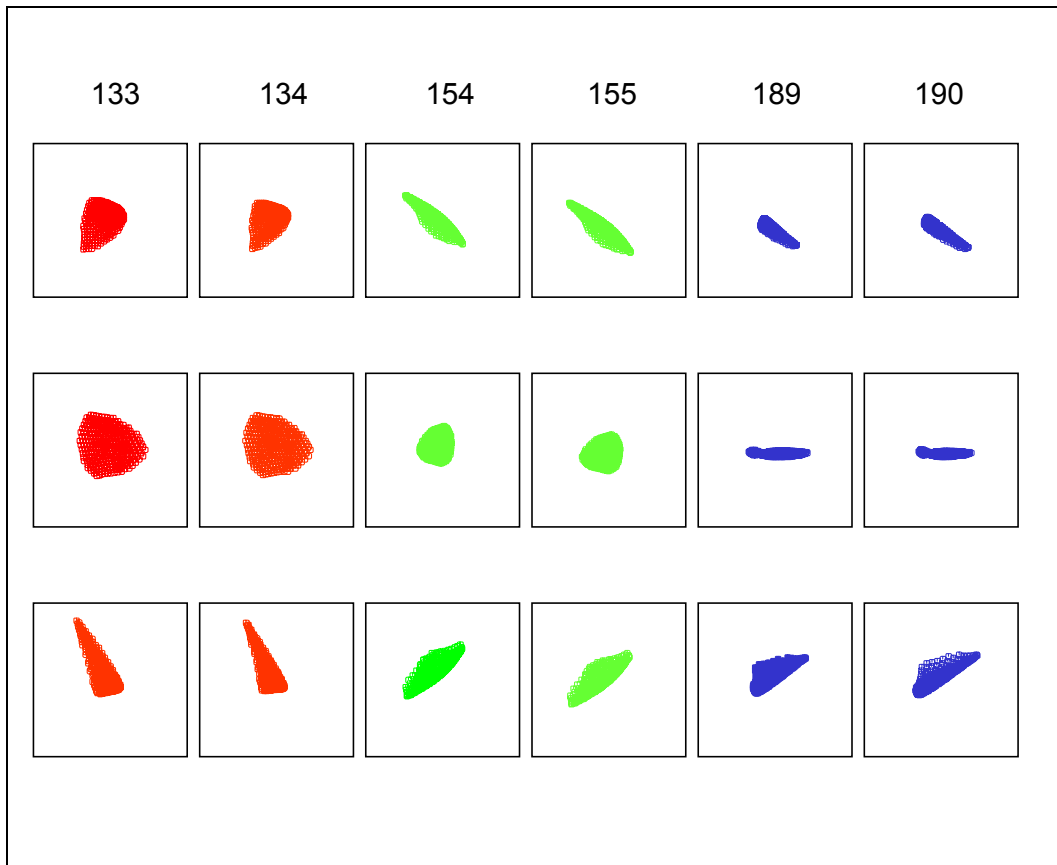


Figure 13.2-8. Geometric spot diagrams for L-band orders (each square is $18\ \mu\text{m} \times 18\ \mu\text{m}$).

Table 13.2-6 lists the parameters for the three mirrors comprising the TMA for the LWM. Although it follows the same principles as the short-wave module's TMA, this system is considerably easier in all respects.

order	Wavelength (nm)	P-V OPD (λ)	RMS OPD (λ)
133	4150	0.031	0.008
133+	4165	0.043	0.007
133-	4133	0.061	0.011
134	4119	0.032	0.008
134+	4134	0.040	0.007
134-	4103	0.057	0.009
154	3587	0.038	0.007
154+	3600	0.050	0.010
154-	3573	0.051	0.008
155	3564	0.040	0.008
155+	3577	0.053	0.011
155-	3550	0.053	0.009
189	2927	0.037	0.008
189+	2938	0.043	0.008
189-	2916	0.061	0.010
190	2912	0.034	0.007
190+	2923	0.050	0.009
190-	2901	0.068	0.012

Table 13.2-5. Imaging metrics for the L-band end-to-end system.

	Primary	Secondary	Tertiary
Radius of Curvature (mm)	820.764 ccv	350.782 cvx	556.662 ccv
Conic Coefficient	-0.678092	0	+0
Conic Type	ellipsoid	sphere	sphere
Lateral Displacement X (mm)	113.457	-2.952	32.560
β Tilt (XZ-plane) (deg)	-2.981	0.244	-3.442
Parent diameter (mm)	450	85	280
Used section size (X \times Y, mm)	170 \times 220	70 \times 60	206 \times 124
Radial offset of section center X (mm)	130	6	44
Parent max departure from best-fitting sphere (mm)	0.105	n/a	n/a

Table 13.2-6. Long wave module TMA mirror parameters

13.2.2.4 Baffling

Optical baffling in the SWM and LWM does not appear to present major problems, although a rigorous raytrace analysis has yet to be performed.

In our conceptual design, the dewar radiation shields (see Section 13.2.5) should form effective cold-cavities that isolate the optics from thermal background radiation. At the

optical input, cold pupil stops help control stray light and thermal background (in the SWM this stop is inserted for K-band operation). Further downstream, baffling structures consisting of aluminum shields and enclosures will isolate the optical train from scattered light and from any thermal background from warm (>80 K) cryostat structures such as supports and wiring.

In both modules, potentially critical areas for baffling are at the cross-dispersers and immersion echelles, where scattered and aliased light may be generated, and at the spectrometer and guide camera detector arrays, where they may contaminate signal. The physical space for baffle clearance is somewhat tight near the echelles (particularly where the beams from the cross-dispersers pass their respective echelles on their way to the TMAs), and also near the secondary mirror of the TMA. The baffling problem is made simpler because these optics either do not move, or if they do move, they translate without rotating. The alternative refractive cross-disperser configuration in Section 13.2.2.2.5.3 should be easy to baffle also.

The housings for the spectrometer and guide camera FPAs will enclose as much as possible of the solid angle surrounding the detectors. These enclosures should provide good optical baffling as well as stable thermal environments around ~35 K for the detector arrays. In the case of the long wavelength module, the baffling from the FPA housing should shield a significant fraction of the detector's potential view of the dewar interior. It may also be possible to install anti-reflection coatings on the FPAs that have some out-of-band rejection qualities.

13.2.2.5 Sensitivity Estimates

There are three relevant sensitivity regimes for high resolution spectroscopy in the near-infrared. At low S/N, at higher resolving power, and at the shorter wavelengths, read noise dominates. When performing high S/N spectroscopy at short wavelengths, the photon shot noise of the source dominates. At longer wavelengths and lower resolving power, background noise becomes most important.

The number of background photons n_{bg} incident on a single pixel is given by

$$n_{bg} = (B_v/h\nu) \Delta\nu A\Omega m \varepsilon,$$

where B_v is the blackbody function at the ambient temperature, $\Delta\nu$ refers to the bandwidth transmitted by the slit, $A\Omega$ is the diffraction limited area-solid angle product and is equal to c^2/v^2 , m is the number of diffraction limited modes incident on a pixel, and the combined emissivity ε of the atmosphere, telescope, and warm instrument surfaces is listed as "Net Emissivity" in the final row of Table 13.2-7. We therefore have

$$n_{bg} = 2\nu (e^{h\nu/kT} - 1)^{-1} (\varepsilon m/R),$$

where R is the slit-limited resolving power. For the short wavelength module, each pixel has a width of 0.0667 arcsec while, for the long wavelength module, the pixel width is 0.02 arcsec. With these scales, we have

$$\begin{aligned} m &= 52.3/\lambda^2 && \text{(SWM);} \\ m &= 4.7/\lambda^2 && \text{(LWM).} \end{aligned}$$

Table 13.2-7 gives the instrument throughput and emissivity at different wavelengths and resolving powers. The boxes list the estimated optical efficiency of the various elements of the system. The net throughput is the product of these efficiencies times the estimated atmospheric transmission (at zenith), telescope throughput, and slit throughput from Table 13.2-9. The net emissivity takes into account the atmosphere and the warm optical elements in the beam. It assumes that a cold stop has masked the unfilled portions of the telescope aperture.

	1.2 - 2.4 μm			3.6 μm	4.8 μm
Resolving Power	25,000	50,000	100,000	133,000	100,000
Spectrograph Optics	0.7	0.7	0.7	0.7	0.7
Filter	0.85	0.85	0.85	0.85	0.85
Grating ¹	0.7	0.7	0.7	0.7	0.7
Cross-disperser	0.8	0.8	0.8	0.8	0.8
Detector QE	0.8	0.8	0.8	0.8	0.8
Instrument Throughput	0.27	0.27	0.27	0.27	0.27
Net Throughput, η	0.11	0.09	0.07	0.09	0.08
Net Emissivity ² , ϵ	0.18	0.18	0.18	0.42	0.52

Table 13.2-7. Instrument throughput and efficiency. ¹The grating efficiency is given on the blaze. In the source noise limit, it may be possible to have a fairly flat response if the detector sees past the ends of the orders. Otherwise, the efficiency will be ~50% of this value at the ends of the orders. ²Includes 2% for the dewar window.

The number of noise electrons in a given exposure time t, taking into account all noise sources, is then given by

$$(n_{\text{noise}})^2 = n_{\text{read}}^2 n_{\text{pix}} + n_{\text{source}} \eta t + n_{\text{bg}} n_{\text{pix}} \eta t,$$

where n_{pix} is the number of pixels included in the average (usually the number of pixels corresponding to the slit width times the full width to half maximum of the image), η is the optical throughput of the atmosphere, telescope, and instrument times the QE of the detector (listed as “Net Throughput” in Table 13.2-7) and n_{source} , the number of source photons contributing to the signal, is given by

$$n_{\text{source}} = N_{\text{zm}}(v) (v/R) 10^{-0.4 m(v)} \eta t$$

where $N_{zm}(\nu)$ is the number of photons per unit frequency incident on the telescope from a zero magnitude star and $m(\nu)$ is the source magnitude at the same frequency.

Table 13.2-8 gives our assumptions about the behavior of the atmosphere, the telescope, and the AO system. The telescope emissivity assumes that the system includes a cold pupil stop. The low backgrounds at J and H mean that such a stop is not necessary at those wavelengths. Table 13.2-7 gives our best estimate of the instrument throughput. The last two rows of this table give estimates of the photon-to-electron efficiency from the top of the atmosphere and of the total emissivity of the system plus telescope and atmosphere. This efficiency does not allow for truncation losses at the cold pupil stop. The number given for the grating is at the blaze wavelength. In the background and read noise limits, performance will be poorer off the blaze. For many orders, we will be able to recover signal in adjacent orders for wavelengths off the blaze and improve effective throughput when working in the source noise limit.

	1.2 - 2.4 μm			3.6 μm	4.8 μm
Resolving Power	25,000	50,000	100,000	133,000	100,000
Atmosphere	0.9	0.9	0.9	0.88	0.78
Telescope	0.92	0.92	0.92	0.92	0.92
AO System	1.0	1.0	1.0	0.8	0.8
Slit (effective loss)	0.5	0.4	0.3 ^{note 1}	0.5	0.5
Throughput	0.41	0.33	0.25	0.32	0.29
Emissivity	0.18	0.18	0.18	0.4	0.5

Table 13.2-8. Sky to Slit Throughput and Emissivity. The emissivity and transmission of the atmosphere are calculated at the zenith. ¹The R=100,000 estimate assumes better seeing conditions than the average conditions assumed for R=25,000.

Table 13.2-9 gives the limiting magnitudes for observations of point sources at S/N = 10 and S/N = 100. These numbers are calculated assuming two half-hour integrations. In the background and source noise limits, the exposures can be broken into smaller units with no loss of sensitivity. For J, H, and K, we calculate the limiting magnitudes for R = 50,000. For L and M, where AO is used, we calculate the limiting magnitudes for R = 133,000 and 100,000, respectively. In all cases, the S/N is for a single resolution element. Performance will be somewhat poorer off the blaze and at higher airmass.

band	λ μm	R	Limiting Magnitude	
			S/N=10	S/N=100
J	1.22	50,000	19.6	16.0
H	1.65	50,000	19.1	15.5
K	2.16	50,000	18.1	14.1
L	3.55	133,000	14.5	11.9
M	4.77	100,000	12.1	9.6

Table 13.2-9. Limiting magnitudes for point sources. The table assumes two 30 minute integrations with the source on the slit in both exposures.

Some science programs may wish to observe extended sources of line emission rather than continuum point sources. Table 13.2-10 therefore lists the 10σ limiting surface brightness for a one-hour observation consisting of 30 min on source and 30 min off source. The beam size used in the calculation is listed in the table.

band	λ μm	R	beam size arcsec	Limiting Surface Brightness $10^{-6} \text{ erg cm}^{-2} \text{ sec}^{-1} \text{ ster}^{-1}$ S/N = 10, (30 min on, 30 min off)
J	1.22	50,000	0.3×0.6	0.4
H	1.65	50,000	0.3×0.6	0.3
K	2.16	50,000	0.3×0.6	0.25
L	3.55	133,000	0.06×0.06	14
M	4.77	100,000	0.08×0.08	30

Table 13.2-10. Limiting surface brightness for extended sources. The table assumes two 30 minute integrations with the source on the slit in both exposures.

13.2.3 Mechanical Design

The short and long wavelength modules take their beams from different pickoff positions and have different cryogenic requirements for the detectors and optics. As a result, it makes sense to house them in independent cryogenic containers.

13.2.3.1 Short Wavelength Module

In this section, we build an illustrative version of the short-wavelength module, working from the inside to the outside. This module consists of six sections: the tertiary mirror assembly, the calibration system, the pupil stop assembly, the guide camera, the grating/cross-disperser assembly, and the TMA/detector assembly. We have produced mechanical designs for the last two sections which are the largest and most demanding. The short wavelength module has its own, deployable tertiary mirror. This is a flat inserted to the telescope optical axis, 0.7 m above the platform. Outside the cryostat, the instrument has a calibration box. In normal operation, this box is bypassed and contributes no optical losses. The calibration box contains a flat-field lamp, emission line lamps, and an absorption cell that can be used either in front of a continuum source or inserted directly into the telescope beam. When observing at K-band, it helps the instrumental sensitivity to have a cold pupil stop. At J and H, the thermal background is not high enough for the lower photon noise to offset the losses entailed by the additional optical elements (for sensitivity calculations, see Section 13.2.2.5 above). The cold pupil assembly, located immediately inside the dewar, consists of an insertable system with a pair of fold mirrors, a collimator, the stop, and a reimager. The assembly can be deployed by the same mechanism that switches between the K-band and J/H-band dispersive configurations. Figure 13.2-9 shows the optical ray trace (right) and a solid model of the optical elements (left). In this image, the J/H-band immersion grating and the H-band cross-disperser are illuminated and the K-band immersion grating and cross-disperser lie below them.

The spectrometer requires five cold actuators, two of which will be ganged together: the filter wheel, the slit wheel, the cold-stop assembly, a changer for the J/H to K-band configurations (ganged with the cold-stop), and a selector between the J and H cross-dispersers. This count is actually quite modest for a cross-dispersed spectrograph and should be easy to deal with, given the space and access constraints. We have had good experience with our 5-25 μm cross-dispersed echelle (Lacy et al. 2002) using external stepper motors, ferrofluidic feedthroughs at the vacuum wall of the dewar, and G10-to-stainless driveshafts internally. Most of the actuators move elements between two fixed positions, so encoding will not be an issue. The filter wheel position is not particularly sensitive. For the slit wheel, the only movable element with critical registration requirements, optical feedback and confirmation will be available from the slit-viewing guide camera.

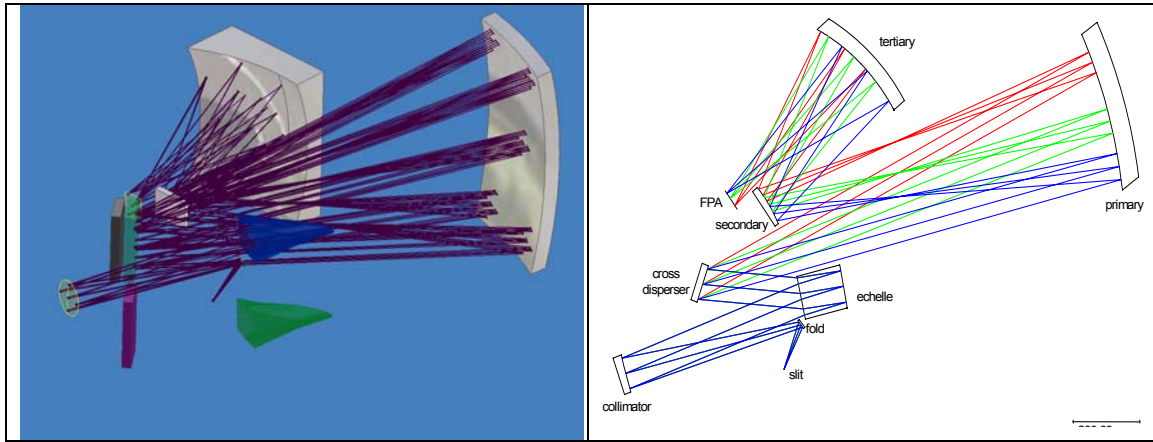


Figure 13.2-9. (left) Solid surfaces for the H-band configuration of the short wavelength module. The immersion grating is in blue and the H-band cross-disperser is in light green. The dark green J-band cross-disperser is mounted on the same shaft as the H-band cross-disperser. The three mirrors of the TMA are shown in grey. The green immersion grating and the purple cross disperser are part of the K-band configuration. The slit wheel is omitted for clarity. (right) Optical layout corresponding to the left panel, with labels for the various components (see Figure 13.2-1).

Figure 13.2-10 shows the grating/cross-disperser assembly and the TMA/detector assembly mounted on their cryogenic optical bench. The reflective slit mask sends the parts of the source that are not coupled through the slit, as well as the surrounding field, to an IR imaging system that serves as a focal plane guider. The three vertical rails are guides for the grating/cross disperser assembly. This assembly travels vertically between two three-point end stops when the system switches from J/H to K-band operation. The mounts for the TMA elements are purely representative at this point since it is unclear whether the camera will be furnished as a set of individual elements or as a pre-aligned assembly.

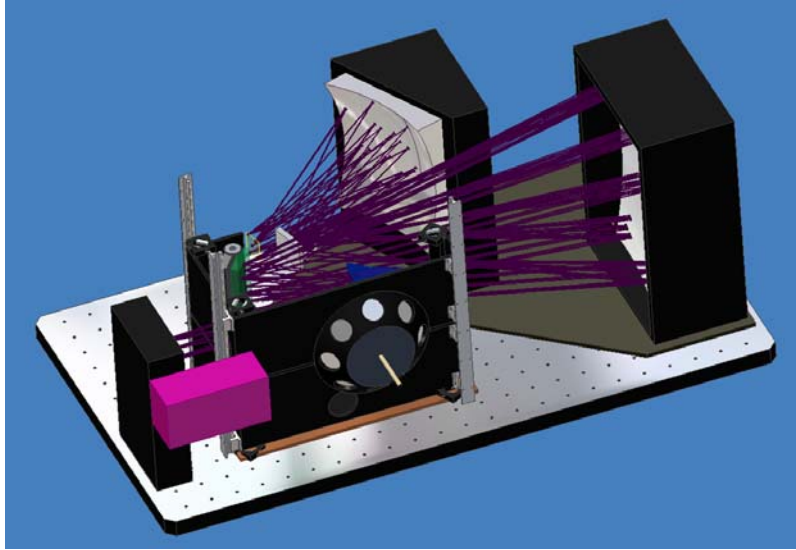


Figure 13.2-10. Full view of the grating/cross disperser and TMA/detector assemblies. The slit wheel is in the foreground. The purple box is a representative slit-viewing IR guide camera.

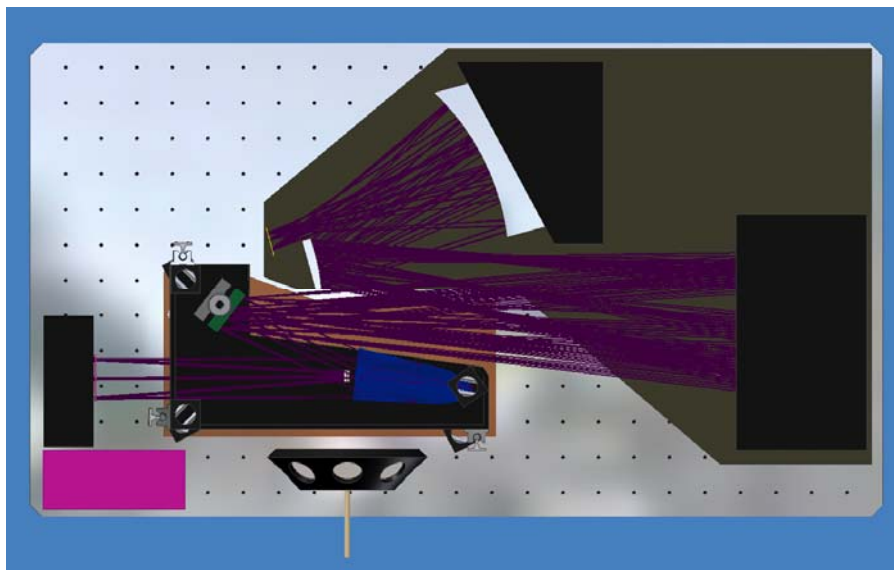


Figure 13.2-11. Top view of the grating/cross-disperser and TMA/detector assembly. The light enters the system from the lower left and is folded onto the collimator at the extreme left. The three TMA mirrors are to the right and the focal plane array is at the center of the optical bench.

Figure 13.2-11 shows a top view of the grating/cross-disperser assembly, and the TMA/detector assembly. This view shows the three locating pads for the grating assembly near each of the vertical rails. The J and H cross-dispersers are mounted back to back on the same shaft and switch between two fixed positions. The silicon immersion grating rides with the cross-dispersers on a platform to which it is attached by a fixed, stress-free mount. The focal plane array will be thermally isolated from the 77 K optical bench and enclosed in a small housing. Both this housing and the chip carrier will be cooled to ~ 35 K via a heat strap tied to the second stage of the cryocooler.

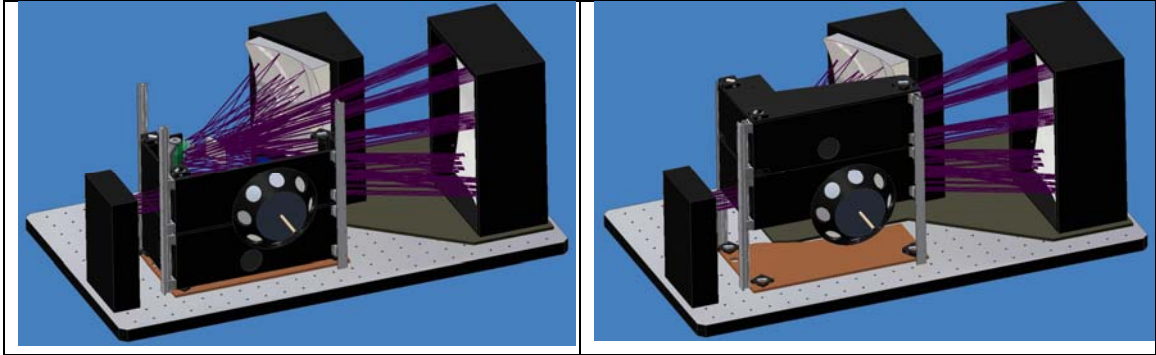


Figure 13.2-12. (left) GMTNIRS grating/cross disperser assembly in the J/H-band configuration. (right) The instrument in its K-band configuration.

Figure 13.2-12 illustrates the motion of the grating/cross disperser assembly when switching between J/H-band and K-band operation. The immersion gratings and cross-dispersers are packaged into optical assemblies that are optically aligned and move as a combined unit. In the K-band assembly, the cross-dispersion grating is fixed. In the J/H assembly, the cross-disperser assembly is driven between two hard stops to select the appropriate cross-disperser. Either assembly can be switched into the optical path, as illustrated in Fig. 13.2-12. There are lands for the K and J/H-band locating pins at the top and bottom (respectively) of the vertical drive.

In Figure 13.2-13, we show a cutaway view of the cryostat. The exterior dimensions are $2.25 \text{ m} \times 1.45 \text{ m} \times 0.95 \text{ m}$. The 77 K optical bench is suspended from a framework that also serves as a skeleton for the radiation shield. The space between the outer wall and this shield will be filled with multiple layers of aluminized Mylar insulation. The triangular fiberglass supports extending from the room temperature floor of the cryostat to the top of the framework on each of the four sides will be made up either of individual flats or of rectangular box-beams. This design will need to be subjected to a rigorous finite element analysis to determine its sensitivity to flexure and vibration during the Phase A design study.

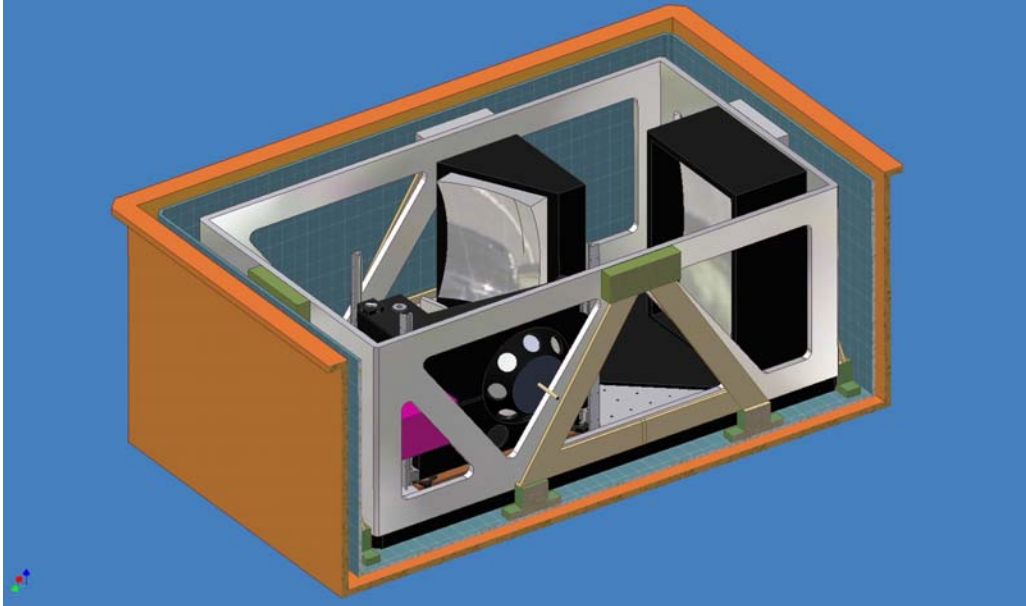


Figure 13.2-13: Cutaway of the GMTNIRS short wavelength module cryostat (orange box). The beam enters the dewar from the missing side. Mylar superinsulation and the non-structural parts of the heat shields have been omitted for simplicity. The fiberglass triangles, of which one of four can be seen, connect the top of the 77K shield to the bottom of the dewar. A heat strap will connect the focal plane assembly and a small housing to the second stage of the cryocooler.

In Fig. 13.2-14, we show the cryostat and some representative racks in the nominal GMT instrument envelope. These racks will house electronics that support operation of the FPA and guide camera, the mechanical actuators, and cryogenic and optical diagnostics.

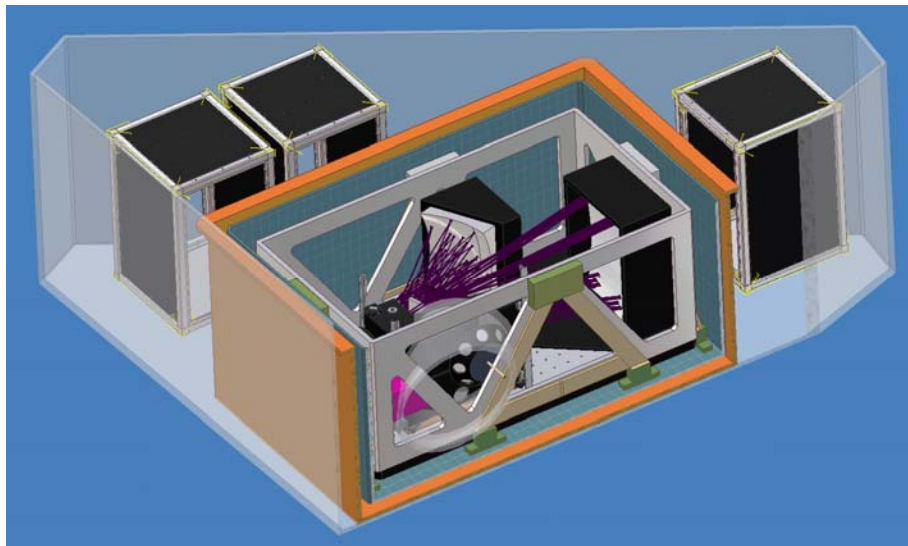
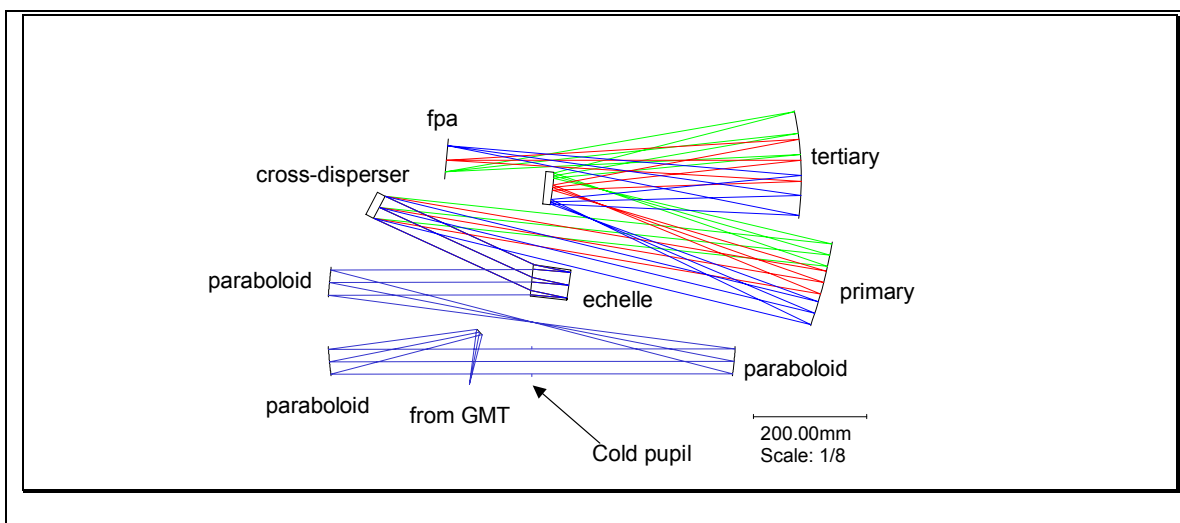


Figure 13.2-14. Cutaway of the GMTNIRS dewar (orange box) and auxiliary electronics (black boxes) within the nominal instrument envelope (transparent wedge).

13.2.3.2 Long Wavelength Module

The smaller collimated beamsize and smaller angular divergences in the long wavelength module lead to a much more compact design for the long wavelength module. The exterior of the cryostat measures 1.35 m × 0.9 m × 0.55 m. The top panel of Figure 13.2-15 repeats the optical layout for this module (Figure 13.2-6), while the bottom panel shows a solid-model realization of the optical components. Only the L-band immersion grating and cross disperser are shown. The M-band gratings will be above or below the L-band components, with an elevator-changer, similar to the one used in the short wavelength module. As with the short wavelength module, the design does not include the calibration system or a fully realized design for the infrared slit viewing camera. The pupil reimaging optics are present in both figures. The output focus of the AO module is at the bottom-center of the figure. A fold mirror brings the light to the pupil reimaging mirror on the left. The pupil stop is at the bottom-center of both drawings, followed by a reimaging mirror on the right. Beyond that, the optical and mechanical geometry is similar to that of the short wavelength module. Figure 13.2-16 shows the long wavelength module optical bench. The black assembly in the foreground supports the reimaging optics and the pupil stop. The blue box at the center is the elevator assembly containing the immersion echelle/cross-disperser pairs for the L and M bands. The green box at the rear is the housing for the focal plane array.

Since the AO feed for 3-5 μm lies above the position for the 1-2.5 μm feed (1.2 m vs. 0.7 m), it makes sense to position the long wavelength module on top of the short wavelength module. Figure 13.2-17 shows the two modules mounted together within the instrument envelope. The relative heights of the beams are close to but not exactly consistent with those of the feeds since these were not known when the mechanical design was being carried out. The design is sufficiently robust to allow adjustments at a later stage. The figure gives a good indication of the relative size of the two modules.



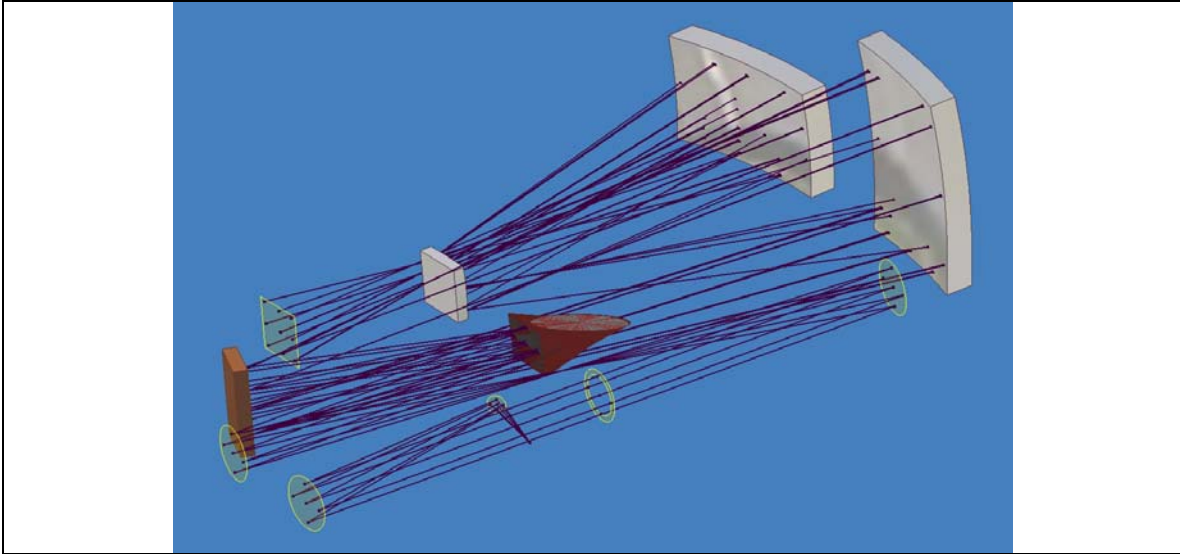


Figure 13.2-15: (top) Labeled optical layout for the long wavelength module. (bottom) Solid surfaces for the pupil imager, grating/cross disperser, and TMA/detector assemblies in the long wavelength module.

Figure 13.2-18 shows two views of the complete GMTNIRS in its envelope on the instrument platform. The figure does not include the warm calibration systems for the two modules or a possible deployable tertiary mirror to feed the short wavelength module.

13.2.3.3 Weight

A crude upper limit weight for the GMTNIRS dewar assembly, including both modules and all internal and dewar-mounted assemblies (but excluding external electronics racks), is ~2000 kg. The mass of the cold portions is about 25% of this figure, with the majority of the cold mass in the short wavelength module.

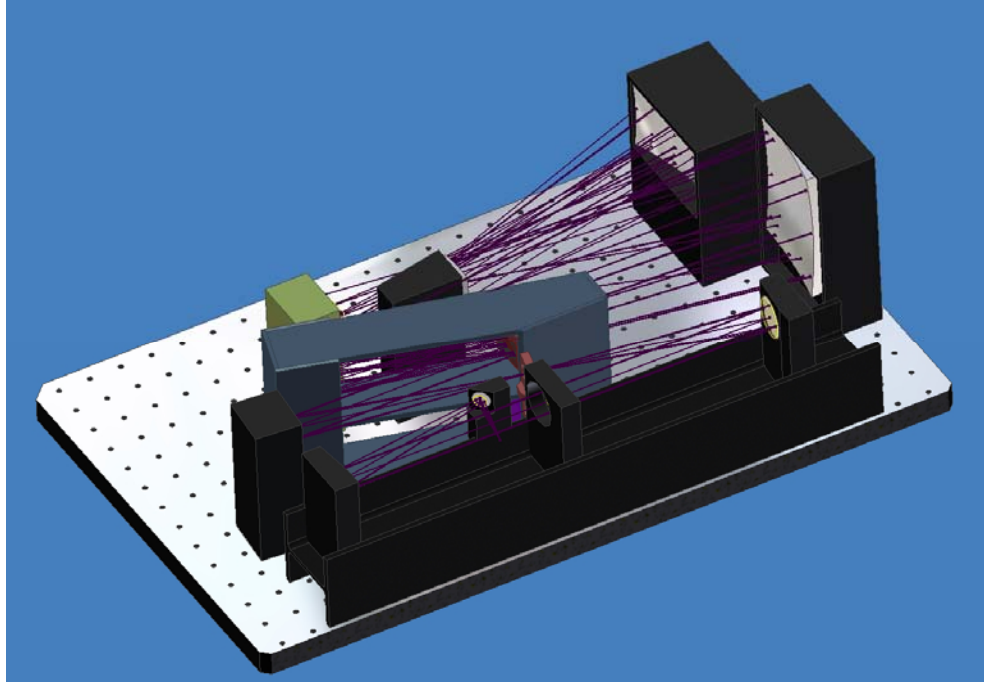


Figure 13.2-16: Long wavelength module optical bench.

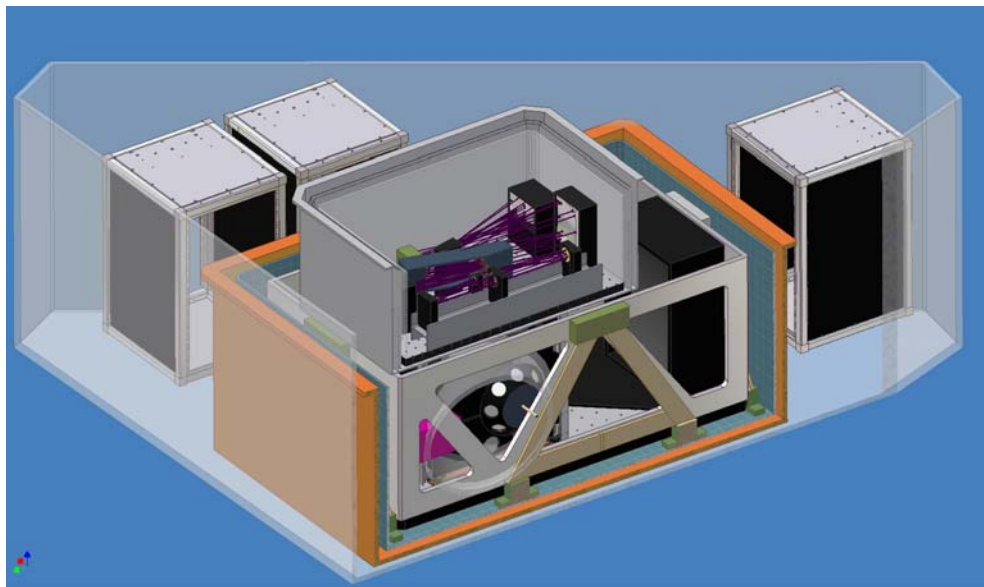


Figure 13.2-17: Long wavelength module in its housing (gray box), mounted on top of the short wavelength module. The transparent envelope indicates the nominal instrument envelope.

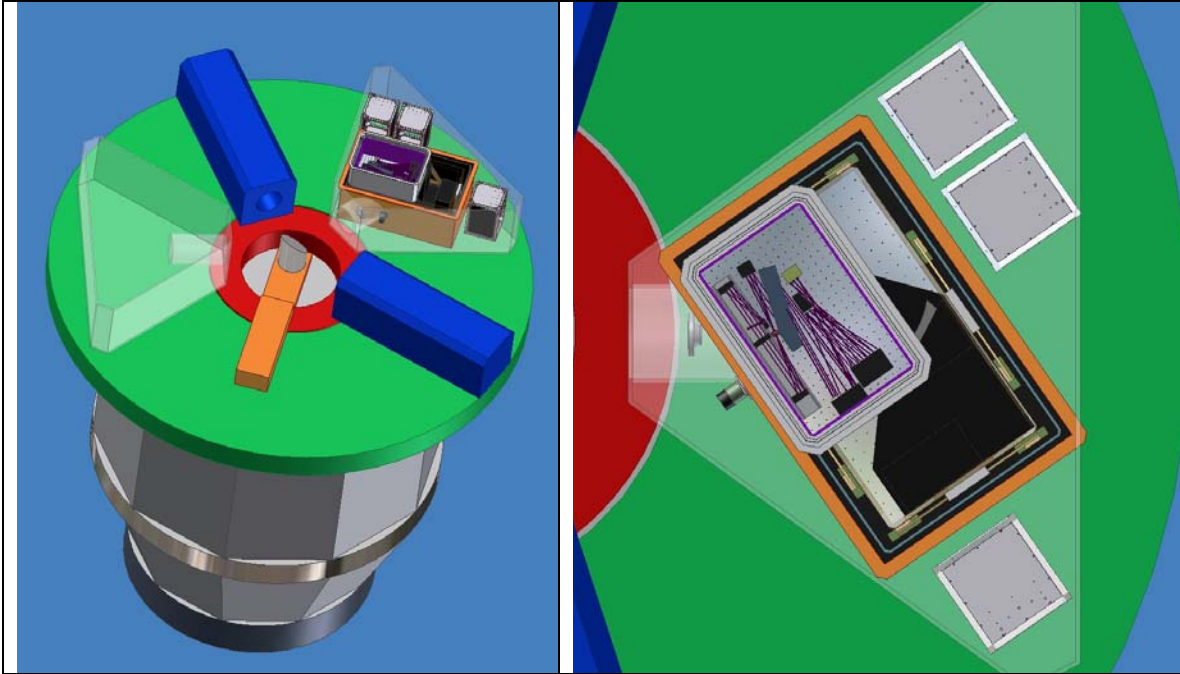


Figure 13.2-18. GMTNIRS in the instrument mounting envelope (transparent wedge). The short and long wavelength modules are shown co-mounted within the envelope. Grey boxes outside the dewars indicate boxes housing support electronics. The calibration systems are not shown.

13.2.4 Silicon Immersion Gratings

Silicon immersion gratings are a critical enabling technology for the development of a compact, high-resolution near-infrared spectrograph such as GMTNIRS. This section describes their performance status and the issues with scaling up to provide the required optics for GMTNIRS.

13.2.4.1 Introduction

Silicon immersion gratings, in which the infrared light is incident on a diffraction grating on the inside of a prism-shaped piece of crystalline silicon (see Fig 13.2-19) offer significant advantages for high resolution infrared spectroscopy. Because the refractive index of silicon (3.44) results in a larger phase difference along the grating, a spectrograph using a silicon immersion grating will have an order of magnitude smaller collimated beam area than the beam area needed with a conventional front-surface grating for a given slit width-resolving power product. In addition, silicon micromachining techniques permit us to make very coarse gratings that enable cross-dispersed spectrograph designs with continuous rather than fragmented wavelength coverage.

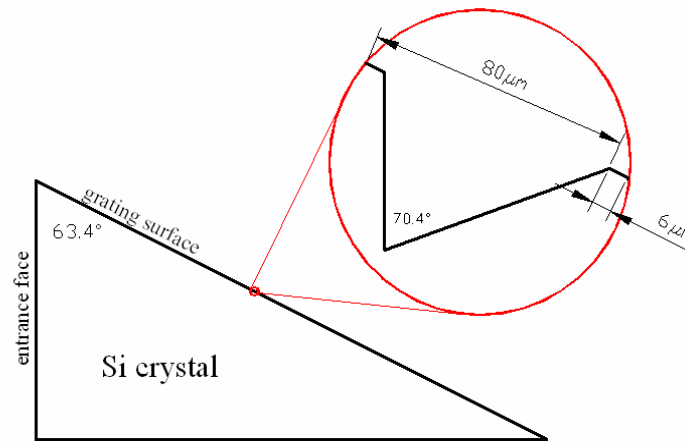


Figure 13.2-19. Schematic diagram of a silicon immersion grating. The light enters and exits through the planar surface at the left and is diffracted by the vertical facets on the downward sloping surface. In this example, the blaze is 63.4° and the groove spacing is $80 \mu\text{m}$.

The University of Texas IR group led by Dan Jaffe has spent the past 15 years developing techniques to produce silicon immersion gratings (Moore et al. 1992, Graf et al. 1994). Our approach has been to fabricate, test, model, and understand the gratings step by step and to publish detailed results (Jaffe et al. 1998, Ershov et al. 2001, Keller et al. 2000, 2002, Marsh et al. 2002). With support from the David and Lucile Packard Foundation, NASA, and UT, we have produced several generations of silicon micromachined

diffraction gratings leading up to the current generation of scientific-grade immersion gratings blazed for R2. One immersion device in this generation is complete.

The use of immersion gratings as dispersers permits designs for near-infrared spectrographs with high spectral resolution but relatively small size and low weight. The optical and mechanical layouts described in Sections 13.2.2 and 13.2.3 fit within the physical envelope and weight limits set by the GMT telescope design. Without the use of immersion grating technology, it is unlikely that a near-IR spectrograph of comparable spectral performance could be designed to fit within the allowable space.

13.2.4.2 Grating requirements for GMTNIRS

As specified by the optical design in Section 13.2.2, the collimated beam diameters for the short wavelength ($\lambda = 1.1\text{-}2.5 \mu\text{m}$) and long wavelength ($\lambda = 2.9\text{-}5.4 \mu\text{m}$) modules measure 83 mm and 45 mm respectively. For each of these modules we will need to fabricate two monolithic silicon immersion echelles with sufficiently large gratings to cover the elliptical projections of the beams. The largest echelles are for the short wavelength module (SWM), requiring illuminated elliptical grating areas with minor axis = 83 mm and major axis = 249 mm (for R3, grating angle = 71.565°). This size requirement represents a scale-up of approximately a factor of 3 to 4 in each linear dimension from the best grating (R2) currently in hand (see Figure 13.2-20). To provide for mounting and alignment, the dimensions of the entrance face will be slightly larger, approximately $90 \text{ mm} \times 90 \text{ mm}$. The echelle size requirements for the long wavelength module (LWM) are approximately half the size for the SWM and less demanding.

Relevant parameters for the grating surface are given in Table 13.2-11.

band	R	grating angle	grating area (mm \times mm)	grating period (μm)	groove error (μm)
J and H	3.00	71.565°	83×249	18.2	0.07
K	3.00	71.565°	83×249	25.1	0.07
L	2.68	69.538°	45×135	86	0.1
M	3.00	71.565°	45×135	160	0.1

Table 13.2-11. Parameters for the immersion gratings required for GMTNIRS. The grating period is the length of the grating divided by the number of lines. The groove error specification is a maximum deviation (from zero) between any pair of lines, and is analogous to a peak-to-valley specification for surface figure. The rms groove position errors will usually be smaller than this parameter.

The groove spacing errors specify maximum errors between corresponding edges of *any* pair of lines in the illuminated area. To maintain coherence over the entire illuminated area of the grating the entrance face must be cut at a precise angle (to within approximately 0.4° in both tip and tilt directions) and polished to high flatness. The flatness requirement is set by the shortest internal wavelength of interest, i.e., $(1.15 \mu\text{m})/3.4 = 0.34 \mu\text{m}$. For planar surfaces on large silicon parts, an optical polishing surface figure of $632.8 \text{ nm}/20 \text{ rms}$ is routine, and $632.8 \text{ nm}/50$ or even better is not too difficult, provided the substrate is greater than several cm thick.

Three of the immersion echelles are R3 (grating angle = 71.565°) prisms and one is R2.68 (grating angle = 69.538°). The entrance faces are cut at small ($\sim 1.5^\circ$) angles with respect to the grating surface to minimize spurious reflections. This requires the formation of compound angles but is straightforward. The bottom surface is canted and polished to help direct any internal reflections away from the optical beams. Our preliminary analysis suggests that a modest angle ($3\text{-}4^\circ$) should be sufficient. For a robust design, this issue should be explored further using a raytrace analysis.

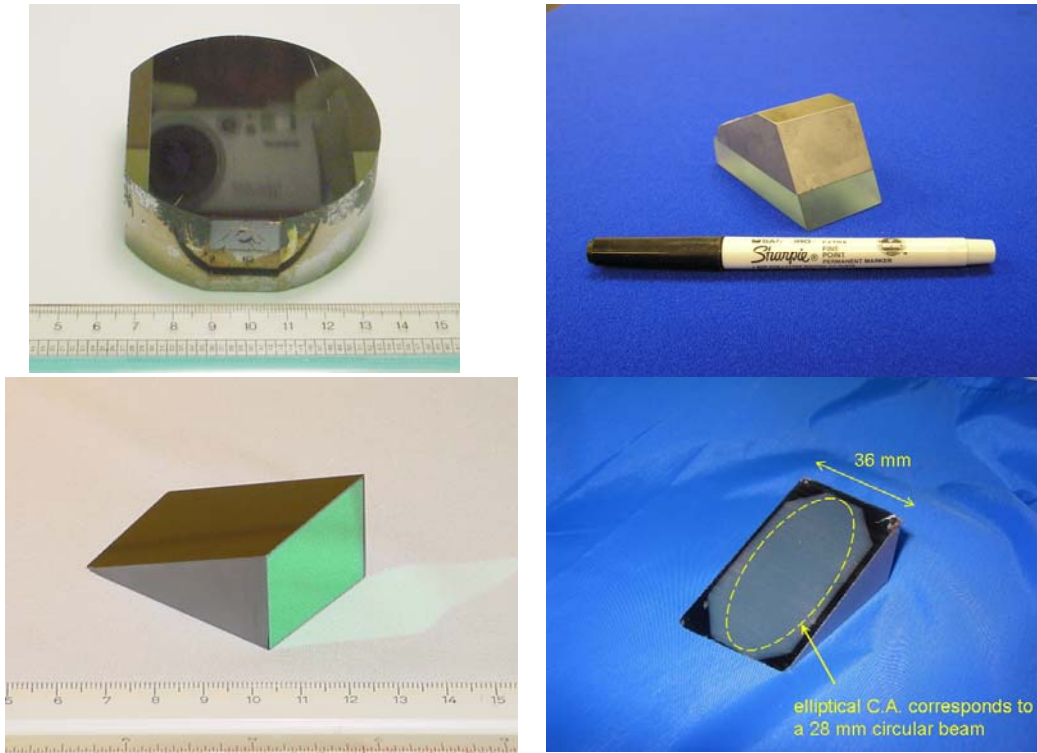


Figure 13.2-20. Images of a silicon R2 immersion grating during fabrication steps: (upper left) after the grating has been etched into the top surface of a puck-shaped substrate. A reflected image of the camera is visible in Littrow; (upper right) after rough cutting but before polishing of the entrance face; (lower left) after final shaping and application of a $1.0\text{-}5.0\ \mu\text{m}$ broadband antireflection coating to the polished entrance/exit face; (lower right) completed, after metallization by sputtering aluminum onto the grating facets.

To avoid reflection losses at the entrance/exit faces of the immersion echelles, broadband antireflection (BBAR) coatings will be applied to these surfaces. In GMTNIRS, each immersion grating will operate only over a single atmospheric window (except in the case of J and H), so the individual BBAR coatings can be optimized for fairly narrow wavelength ranges. To increase reflectivity, the groove facets are metallized by sputtering an aluminum film onto them. This metallization layer is shown in the final panel of Fig. 13.2-20. The BBAR and metal coatings are discussed further in Section 13.2.4.4.4.

13.2.4.3 Overview of fabrication steps

The process of making a silicon immersion grating has three basic steps: preparation of the substrate, fabrication of the blazed grating on the silicon surface, and preparing the final piece. Details of the processing steps will be available in the PhD dissertation of Jasmina Marsh (Marsh 2006).

13.2.4.3.1 Substrate preparation

Our process of fabricating micromachined diffraction gratings in silicon starts from a cylindrical boule or ingot of monocrystalline high-resistivity, float-zone monocrystalline silicon. The boule is oriented using X-ray crystallography to establish the crystal orientation to within $\sim 0.3^\circ$. Precision flats are ground into the side of the boule to indicate the orientation in later processing steps. The boule is rotated to the desired blaze angle and sliced into disks and wafers of the proper thickness. Disks become substrates for actual immersion grating devices, while wafers are useful for process verification and prototyping. The cut surfaces are then etched to relieve saw damage, ground, and polished using chemical-mechanical planarization (CMP) to high optical flatness (for thick disks the surface figure can approach 632.8 nm/50 or better). All parts are then coated with a thin (~ 60 nm) passivation layer of silicon nitride. On the face where the grating is to be fabricated, the thickness of this layer is held to better than 5% across the aperture. All of these steps are carried out by commercial vendors.

13.2.4.3.2 Grating fabrication

The actual fabrication of the grating involves semiconductor lithography processes. A uniform layer of photosensitive resist is applied to the substrate surface. A quartz mask contains the desired line pattern as a set of precisely-spaced chrome lines of uniform width. The mask is placed in contact with the photoresist such that the lines are oriented perpendicular to the wafer flat. The photoresist is then exposed to ultraviolet light through the mask. The substrate and photoresist are immersed in a liquid developer to pattern the photoresist. This layer then serves as a mask in a dry (plasma) etch process that transfers the pattern to the silicon nitride passivation layer. After stripping the photoresist, the silicon nitride serves as a patterning mask for a KOH-based wet etch process that creates the final Si groove surfaces. The passivation layer is removed, resulting in a grating of the precise period and blaze angle etched into the surface of the substrate (see first panel in Fig. 13.2-20).

Although the typical feature dimensions are very large (10-200 μm), the large bulk of the substrates and the requirement of uniform groove spacing across the entire surface require special attention to each step in the lithography. For example, when spin-coating thick substrates with elliptical cross-sections, one must suppress aerodynamical effects that can harm the uniformity of the photoresist layers.

13.2.4.3.3 Prism shaping and coatings

After the grating is fabricated on the surface of the disk substrate, the disk is cut to form a prism such as the one shown in Fig. 13.2-20. The entrance face is polished and coated for anti-reflection, and the grooves are metallized to complete the device. These steps are performed by commercial vendors.

13.2.4.4 Current state of the art

The current state-of-the-art in immersion grating fabrication can be summarized by describing a part we have recently completed for a new near-IR spectrograph for the NASA Infrared Telescope Facility. This part is shown in Fig. 13.2-20.

13.2.4.4.1 Size and orientation.

The physical grating area measures approximately 70 mm long \times 36 mm wide. The grating wedge angle is $\gamma = 63.4^\circ \pm 0.04^\circ$ (R2). The orientation of the entrance face relative to the grating face is specified in both tip and tilt directions to within 0.04° to manage undesirable reflections. The entrance face measures 36 mm wide \times 26 mm high and accommodates a beam diameter up to 24 mm in the dispersion direction.

13.2.4.4.2 Optical quality

The overall optical quality of the device in Fig. 13.2-20 is very good. This has been verified from external measurements taken using visible laser light (632.8 nm and 543.5 nm, corresponding to application wavelengths of 2.15 μm and 1.84 μm) and in immersion using near-IR light at 1.523 μm

From an interferogram taken in reflection using red HeNe light, we have the surface error map shown in the left panel of Fig. 13.2-21. The surface figure of the grating over a 25 mm aperture is 0.04 waves rms at 632.8 nm. We have achieved this quality or better on silicon substrates of comparable size, with different grating angles and groove constants (6.16° , 25 μm and 32.6° , 87 μm). On these other devices we have measured surface figures nearing 10^{-2} waves over 25 mm apertures (devices and data not shown). By scaling and inverting the error map, we obtain the point spread function (PSF) shown in the center ($\lambda = 1.1 \mu\text{m}$) and right ($\lambda = 3.0 \mu\text{m}$) panels of Fig. 13.2-21. Some low level satellites are seen in the PSF corresponding to $\lambda = 1.1 \mu\text{m}$, the shortest wavelength for GMTNIRS.

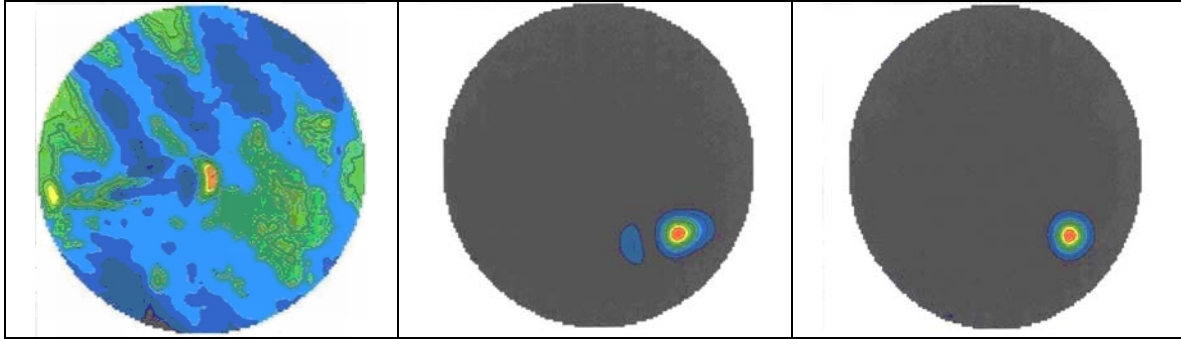


Figure 13.2-21. (left) wavefront error map of the grating in Fig. 13.2-20, for a 25 mm diameter collimated beam at 632.8 nm. Each color level in the surface map represents 0.022 waves; (center) and (right): point spread functions (PSFs) for 1.1 μm (center) and 3.0 μm (right), derived from the wavefront error map. In the PSFs, each color gradation represents approximately 11% of the peak value. This data was taken before metallization of the groove surfaces.

Data taken in immersion are consistent with the reflection data above. In Fig. 13.2-22 we show the two-dimensional spectral PSF taken in immersion using a 10 mm diameter beam at $\lambda = 1.523 \mu\text{m}$, for the same device. The color map has been chosen to better show low-lying features. Two off-axis satellites are visible at 30 degrees from the horizontal axis, symmetrically located on each side of the central peak. The peak amplitude of these satellites is about 9% of the central peak and the integrated strength of both together is 12% at 1.523 μm . The external wavefront measurements are consistent with this data (note the low level corrugation in the left panel of Fig. 13.2-21) and predict that these features diminish over the H band and vanish at longer wavelengths.

By collapsing the data of the central peak along the cross-dispersion direction, we obtain a one-dimensional spectral PSF. This is shown in Fig. 13.2-23, where we have used only those data from within 5 pixels on either side of the central peak, a truncation process mimicking the intended application in a spectrograph. The PSF from the echelle obtained in this way is virtually indistinguishable from that obtained from reflection from a reference mirror, demonstrating diffraction-limited immersion performance over a 10 mm diameter beam. Including the convolution with camera pixels, the demonstrated resolving power in immersion Fig. 13.2-23 is approximately 67,000; this is consistent with the data obtained from the front-surface measurements, after accounting for the difference in beam diameters.

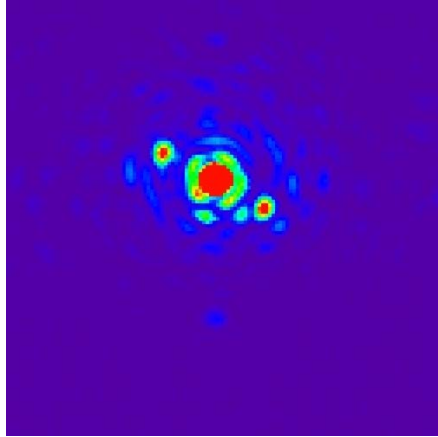


Figure 13.2-22. Two-dimensional spectral PSF of the brightest order, taken in immersion using a 10 mm diameter beam at 1.523 μm . The dispersion and cross-dispersion directions in the figure are horizontal and vertical, respectively. The color map is saturated to better show low-lying features. Two off-axis peaks are visible at approximately 30 degrees from the horizontal axis.

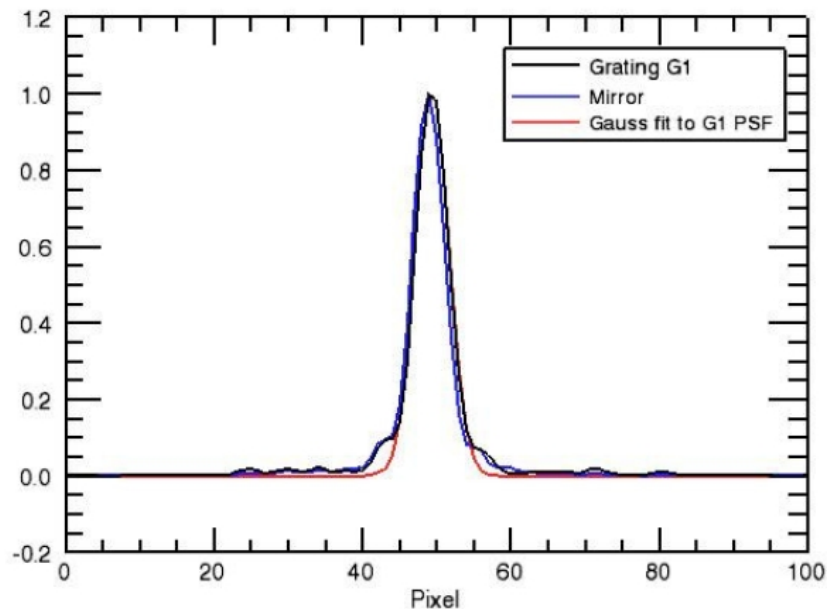


Figure 13.2-23. One-dimensional spectral PSF measured in immersion for the grating (black) and for reflection from a reference mirror (blue), obtained by summing the data within 5 pixels of the central peak. The data are obtained using a 10 mm diameter beam at 1.523 μm . The cutoff partially removes contributions from the satellite peaks in Fig. 13.2-22. The immersion PSF of the grating is indistinguishable from the reflection PSF of a reference mirror. The demonstrated resolving power as determined from the FWHM of the gaussian fit (red) is 6.7×10^4 .

We believe that the presence of the satellites in Fig. 13.2-22 is due to residual lithography error and are currently taking steps to eliminate these errors in our fabrication steps. Such satellites are not present in other immersion devices that we have made. We expect that the optical quality of subsequent immersion devices will be improved. This and

other issues associated with development work on silicon grating fabrication are discussed below in Section 13.2.4.5.

13.2.4.4.3 Spectrum in immersion

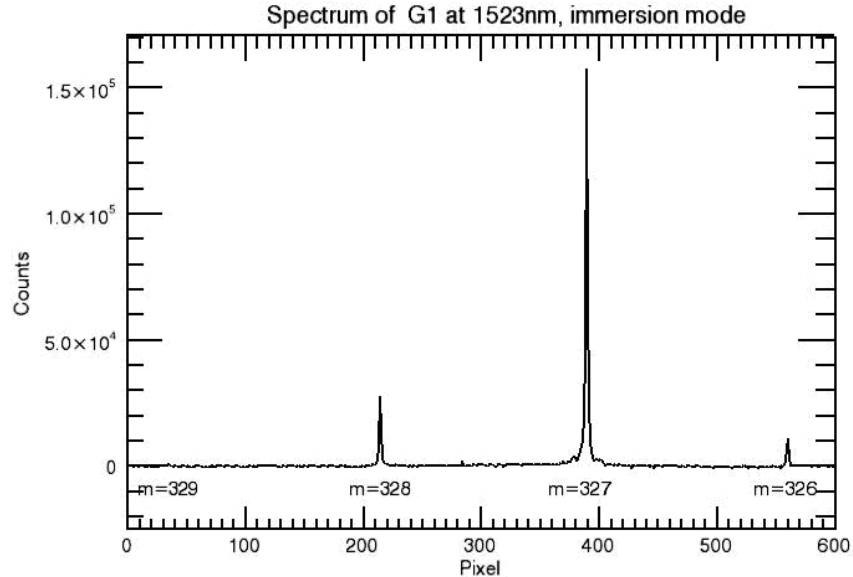


Figure 13.2-24. Spectrum of the immersion grating echelle shown in Fig. 13.2-20, taken in immersion using 1523 nm light over a 10 mm diameter beam. This wavelength corresponds to 443 nm inside the silicon. No ghosts are visible in the spectrum. The numbers immediately below the curve designate the diffraction orders. These data were taken after applying the AR coating to the entrance/exit face and metallization of the grating surface. Summing the power in the four brightest orders yields a throughput of 82% relative to the incident beam.

The immersion spectrum of this grating is shown in Fig. 13.2-24. By summing the total power in the brightest orders, the measured throughput in immersion at 1.523 μm for the 22 mm aperture is $75\% \pm 5\%$, relative to the incident beam. This should be equal to the diffraction efficiency on the blaze. This figure is consistent with an estimate (82%) that is derived from the interferometric data in Fig. 13.2-21 when one accounts for losses from two passes through the BBAR coating and the reflection at the Si-metal interface.

13.2.4.4.4 Performance of coatings

Coatings are applied to the silicon immersion grating to enhance the infrared throughput. For GMTNIRS, these coatings must have good mechanical and cryogenic properties.

13.2.4.4.1 Broadband anti-reflection (BBAR) coating

A commercial 1.00-5.0 μm BBAR coating with $\sim 2\%$ average reflectance and less than 5% peak reflectance has been applied to the entrance/exit face of the part shown in Fig. 13.2-20. The coating appears green in the lower left panel. The theoretical performance of this coating is shown in Fig. 13.2-25. It is designed to have 3% average reflection from 1.00 to 5.0 μm , with modest short wavelength absorption (approximately 20% absorption at the 77 K silicon cutoff at 1.07 μm , see (MacFarlane et al. 1958)). The measured reflectance of the coating averages about $\sim 2\%$ from 1 to 5.0 μm and the ripples closely match the design curve in Fig. 13.2-25. Using silicon witness pieces we have confirmed that this coating is durable by subjecting it to multiple rapid cooling cycles between room temperature and 77 K. Because the individual immersion gratings for GMTNIRS will operate over smaller wavelength ranges than that shown in Fig. 13.2-25, it should be possible to fine-tune the coatings to maximize the transmission at a particular wavelength region.

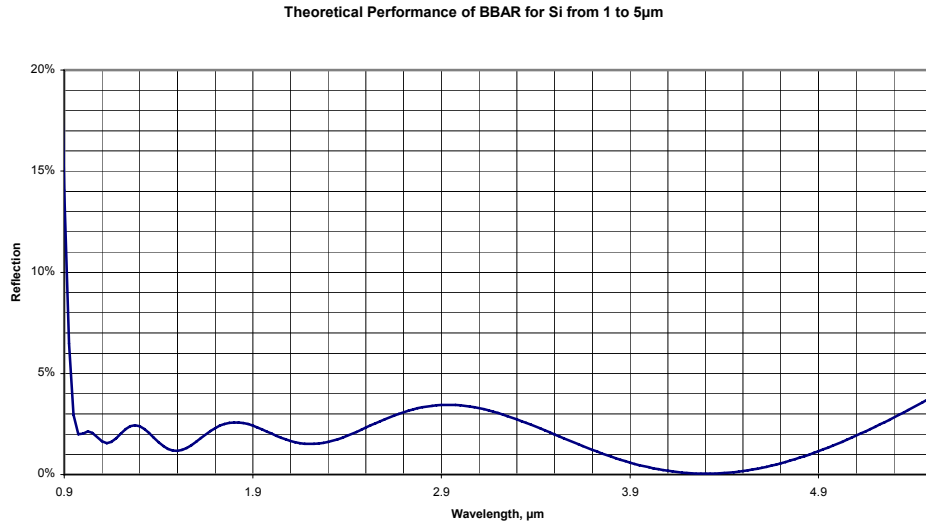


Figure 13.2-25. Theoretical single-pass reflection performance of the commercial broadband antireflection coating applied to the entrance/exit face of the immersion grating in Fig. 13.2-20.

13.2.4.4.2 Metallic overlayer

We have also deposited a metallic layer on the facets of the device in Fig. 13.2-20. This layer consists of sputtered aluminum and is visible in the lower right panel of the figure. We have verified using SEM micrographs of Si grating witnesses (see Fig. 13.2-26) that these aluminum layers are free from voids and adhere well, even under very rough thermal treatment (multiple rapid thermal cyclings to 77 K by dunking into liquid nitrogen in air). The high electrical conductivity of Al results in a highly reflective surface ($>90\%$) over a wide wavelength range (~ 200 nm to 10 μm). Because the

reflecting surface is internal, neither overcoating nor cleaning should be required. If necessary, the metal can be stripped from the silicon and a new layer can be deposited.

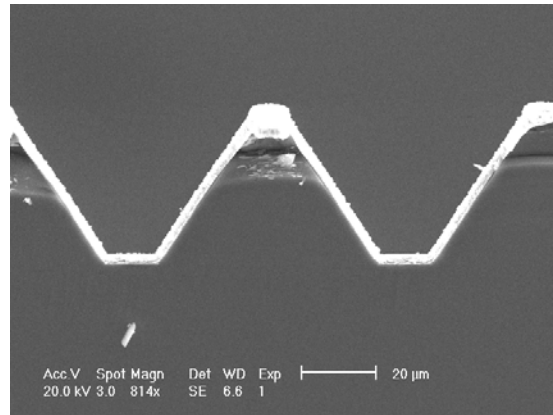


Figure 13.2-26. SEM micrograph of sputtered aluminum on a silicon witness piece into which a grating has been partially etched. The silicon is at the bottom of the figure. The aluminum layer appears bright in the micrograph. It is contiguous and completely covers the grating facets.

13.2.4.5 Scale-up issues (Technical Risks)

In going from the completed device described in Section 13.2.4.4 to the immersion echelles required by GMTNIRS, a size scale-up is required in each linear dimension. For the SWM and LWM echelles, these enlargement factors are approximately 3.5 and 2.0, respectively. Some technical issues and risks are associated with this scale-up. In approximate order from greatest to least risk, they are discussed below.

13.2.4.5.1 Material availability and quality

High purity silicon ingot is commercially pulled from molten silicon (Czochralski process) in diameters of 50, 75, 100, 150, 200, and 300 mm. The orientation of the seed crystal (usually {100} or {111}) is specified to within approximately 2 degrees.

Float-zone (FZ) material involves refining a silicon rod by subjecting a section of it to a radiofrequency field. This field locally melts the silicon and segregates impurities such as oxygen in the melted zone, which can be swept slowly through the rod and guided to one end, thereby concentrating the impurities in a section that can later be discarded. Because there is no contamination from a melt crucible, the float-zone process results in the material with the highest purity and the lowest infrared absorption (Kaiser, Keck, and Lange 1956). Diameters of 150 mm and 200 mm are now industry standard sizes for most commercial silicon (but not FZ). At present, FZ material is grown at 200 mm diameter only as a test and development product (i.e. it is not yet available without special circumstances). It is likely that, in the years required for the development of GMTNIRS, the FZ silicon industry will be able to supply 200 mm diameter material routinely, and 300 mm diameter FZ on a contract basis. To allow for the expected yield

of our process, three or more boules will be necessary for process development, prototyping, and final science-grade parts. Because of the extremely large cost of silicon fabrication facilities, it seems unlikely that silicon ingot diameters will grow beyond 300 mm, although this could change. These issues will impact lead time and price. The immersion echelles required for the short wavelength module of GMTNIRS will be fabricated from 300 mm diameter material. If 300 mm diameter FZ does not become available within a few years, it may be possible to work with smaller ingot diameters if alternative lithographic methods are developed.

Another material consideration is the absorption loss due to the length within the silicon through which the near-infrared light passes. High purity silicon transmits well in the infrared between 1.1 and $\sim 8 \mu\text{m}$ (Becker and Fan 1949). The short wavelength cutoff occurs at a wavelength of approximately 1.1 to 1.2 μm (Dash and Newman 1955, MacFarlane et al. 1958), depends weakly on temperature, and is attributed to lattice vibrations.

The bandwidth for both modules of GMTNIRS is 1.14 to 5.40 μm . Over this band the Si absorption is expected to be largest at the short wavelength cutoff. Using a 8 mm thick double-sided polished FZ Si disk of high purity (resistivity $> 1000 \text{ ohm-cm}$), the measured absorption coefficient at 77 K is $\alpha = 0.0133 \text{ cm}^{-1}$ at $\lambda = 1.10 \mu\text{m}$, in good agreement with the data from MacFarlane et al. (1958) and the dependence found by Svantesson and Nilsson (1979) on "high-purity" Si: $\alpha(T) = T/(172.3 \text{ K})^{4.25} [\text{cm}^{-1}]$ at $\lambda = 1.064 \mu\text{m}$.

For the R3 echelles in the short-wavelength module, the roundtrip distance at the beam center is $(90 \text{ mm}) \times (3/2) \times 2 = 270 \text{ mm}$. The corresponding absorption loss in these echelles is expected to be $1 - \exp(-0.0133 \text{ cm}^{-1} \times 27 \text{ cm})$, or $\sim 30\%$ at $\lambda = 1.10 \mu\text{m}$. At $\lambda = 1.15 \mu\text{m}$, the absorption coefficient has decreased slightly to $\alpha = 0.0085 \text{ cm}^{-1}$, and the absorption loss is correspondingly smaller ($\sim 20\%$). For λ between 1.2 μm and 2.5 μm the absorption is negligible.

For the long wavelength module, the round trip distance is half that of the short-wavelength module. We do not have reliable absorption data at 77 K for FZ Si in the L and M bands, but can extrapolate modest-resistivity Si data (Runyan 1965) to estimate the absorption coefficient for high purity FZ material to be $\alpha \sim 5 \times 10^{-4} \text{ cm}^{-1}$ at 77 K for $\lambda = 2\text{-}5 \mu\text{m}$. Absorption in the long wavelength echelles is therefore expected to be about 1% or less.

In addition to the aforementioned absorption loss, there is a loss of coherence due to the differential absorption across the tapered beam. This will lower the peak of the PSF and lead to a very slight broadening in the direction of variation, as shown in Fig. 13.2-27. These effects are small and do not significantly degrade the capabilities of the instrument, even with the large optical path lengths through the echelles.

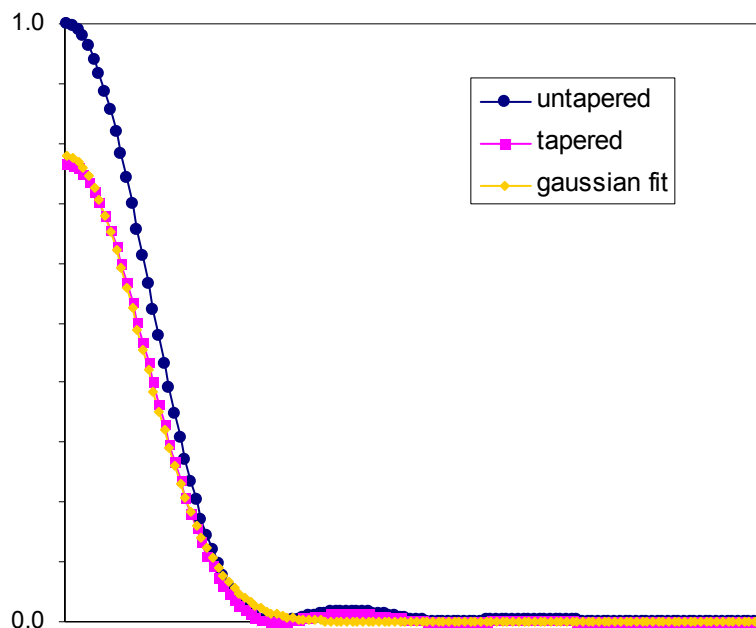


Figure 13.2-27. Point spread function for transmission of an 83 mm diameter collimated beam through an R3 silicon immersion grating ($\gamma = 71.565^\circ$), showing effects of absorption and tapering due to differential absorption in the silicon across the beam. The untapered (blue) and tapered (magenta) curves are respectively calculated for zero absorption ($\alpha = 0 \text{ cm}^{-1}$) and for $\alpha = 0.0085 \text{ cm}^{-1}$, corresponding to the absorption of $\lambda = 1.15 \text{ }\mu\text{m}$ light in high-purity Si at $T = 77 \text{ K}$. The peak intensity has dropped by 24% at the central peak and the width of the best gaussian fit (yellow) to the tapered profile has increased by 0.2% over the width of the best fit (not shown) to the untapered profile.

13.2.4.5.2 Processing Risks

Historically, the commercial equipment and capabilities for the silicon electronics industry has moved towards use of the largest diameter material available. We expect this size-scale up trend to continue, so that equipment and vendor capabilities helpful to fabrication of the immersion gratings for GMTNIRS will develop and become available, even if they are not industry standard today.

13.2.4.5.2.1 Orientation and grating surface

Precise orientation of the boule is accomplished using X-ray diffraction. The size of the boules may become an issue for the orientation equipment—in the past boules that were too long had to be rough cut first. It is not clear yet what constraints exist upon the size of boule that can be oriented. These constraints will depend on the combined diameter and length of the boule. Larger pieces of silicon are better able to hold their shape during machining of the boule, so that angular tolerances (e.g. the blaze angle) will not be degraded, and may actually improve slightly.

13.2.4.5.2.2 CMP Polishing

The cost of polishing will be a significant fraction of the external vendor cost. The large diameters (above 100 mm diameter) and thicknesses probably demand that each part be polished serially. The larger sizes, however, should affect neither the polishing flatness (it is generally easier to achieve excellent surface figures with larger silicon substrates), nor the uniformity of the silicon nitride passivation layer.

13.2.4.5.2.3 Masks

The overall size and accuracy of the mask may be the limiting factor for large area immersion gratings. Standard chrome-on-glass masks are 9 inches (225 mm) square and the pattern area should be able to cover most of a 200 mm diameter wafer. Masks for 300 mm diameter wafers surely exist, but we have no firm data about their tolerances. The maximum size seems to be set by the largest substrates that can be accommodated by the equipment that is used to produce masks, rather than any fundamental limit.

13.2.4.5.2.4 Lithography

There are no major issues foreseen with scaling up the lithography to large substrates. Custom equipment will need to be designed and fabricated. Disk substrates for the largest echelles in GMTNIRS are 300 mm in diameter and ~85 mm thick, and weigh nearly 14 kg. Spinning such large (and expensive) substrates to 3500 rpm obviously will require careful attention to mounting, balance, and safety in addition to raw torque. A custom UV exposure system will also be required to handle large substrates and masks. Successful lithography may depend on good thermal management of the mask and substrate. To this point we have not had to deal with this issue, so thermal management during exposure is a new engineering problem. Custom handling equipment will also have to be designed and fabricated for wet chemistry and possibly some vendor processing steps such as applying silicon nitride. Certain modifications will need to be performed on commercially purchased equipment. In particular, the plasma etcher will require an enlarged vacuum chamber to accommodate the thick substrates. An additional benefit of the larger chamber is improved directionality in the plasma etch; this will help maintain the accuracy and precision of the relative groove spacing. We have discussed this issue with several plasma etcher manufacturers and they seem receptive to this work.

The main risk here seems to be to the project schedule, because of the time necessary to perform the engineering modifications and properly check out the equipment. Some laboratory space will have to be reserved for the specialized lithography and wet etch processing. This may require some facility renovations.

13.2.4.5.2.5 Prism cutting and polishing

There are no major risk issues here, other than cost. Larger silicon pieces are generally easier to work with, as long as they fit into the polishing and shaping apparatus. It is easier to achieve tight angular and surface figure tolerances on large bulky parts.

13.2.4.5.2.6 Coatings

There don't seem to be any major risk issues here, although it's not clear yet what is the largest area that can be coated uniformly. The uniformity across the aperture may be a problem if deposition relies on thermal processes. The separate SW and LW modules will allow the optical performance of the coatings to be tuned up somewhat from the current coatings.

With respect to the metallization, uniformity issues may arise when attempting to deposit metal films on bulky prism-shaped substrates. A specialized vacuum chamber may need to be built. It is likely that a commercial vendor can be found for this task.

13.2.4.5.3 Issues with schedule and effort

We estimate that the production cycle for the actual immersion echelles is at least 4 years, assuming that all major fabrication issues have been resolved for processing ingots of the specified diameter and that all of the necessary equipment and vendors are in place. The relatively long cycle time reflects the serial nature of the processing steps, and builds in some necessary slack to accommodate potential delays from any step in the chain.

Before reaching this point, some effort will be required to develop and demonstrate a reliable production process. This will permit engineering to develop equipment and procedures for working with larger substrates, from device sizes slightly larger than the current state-of-the-art all the way up to the final production devices. We can take advantage of this development time to also improve upon current technical specifications and to develop new lithographic methods that may result in a substantial performance improvement as well as reducing risk and cost.

A potential risk to maintaining the ability to produce silicon immersion gratings of the required size and quality is that vendor capabilities may change over time. This can occur due to changes in key technical personnel, upgrades to equipment or processing, and attention to other trends in semiconductor technology (e.g. the trend has been towards use of 200 and 300 mm diameter material for most consumer electronics; the situation for float-zone material is less likely to demand the largest diameters). It is also important to realize that current vendor capabilities can become unavailable or develop scheduling and risk issues. It is vital for the project to maintain some continuous level of technical contact with as many of the silicon processing vendors as possible.

13.2.4.6 Device development plan

Here we plot out a rough timeline of the silicon immersion grating effort. Note that the grating effort spans the entire time from now until delivery of GMTNIRS. Year 1 may be assumed to be approximately equivalent to calendar year 2006.

- yr 1: assess resistive Czolchralski material as substitute for float-zone
improve UV photolithography exposure system design
construct improved UV exposure apparatus
develop substrate holders for plasma etcher to improve uniformity
specify and order moderate diameter (100 mm or 150 mm) diameter material
begin vendor processing of moderate diameter material (orientation and cutting
optical polishing, passivation)
- yr 2: assess interference lithography methods, compare with contact photolithography
begin testing with oriented larger diameter material
begin design of custom equipment to handle larger (200-300 mm diameter)
material, including spin table, dry and wet etch substrate handling
equipment.
- yr 3: assess lithographic procedures and specify large diameter material
vendor processing of large diameter material (orientation and cutting, optical
polishing, passivation)
fabricate custom equipment for large diameter material
procure laboratory space and renovate if necessary
order plasma etcher with modifications
- yr 4: equipment checkout
wafer tests on large diameter materials
begin grating prototypes
- yr 5: continue prototyping tests
order final flight ingot material
- yr 6-7: process final immersion gratings and spares
optical evaluation of final parts and spares
- yr 8-9: final gratings finished, installation into GMTNIRS
cold testing within GMTNIRS.

13.2.4.7 Assessment

In-hand parts are already of sufficient quality and size for astronomical spectrographs with beam sizes approaching 25 mm in diameter. There are good prospects for improvement in the optical quality of the gratings by devoting more care to the photolithographic steps in efforts to avoid introducing errors. We have several ideas and can start testing some of them relatively quickly (within ~1 year). Some of this effort may be funded from outside sources, and reduce the cost and risk for immersion grating development.

There are some technical risks involved in scale-up, but these are modest. Some of these rely on the ability of vendors to continue being able to perform certain critical tasks (orientation, polishing). Others rely on the semiconductor industry being able to provide Si material at the appropriate diameter and purity. Committing resources to ongoing device development will reduce risk and cost. We support a long-term strategy in which the design and fabrication of the instrument and the development of silicon echelles proceed in parallel. This maximizes the likelihood of completing GMTNIRS with the highest optical performance.

13.2.5 Cryogenics and Vacuum

The detector and optics of the GMTNIRS instrument operate in vacuum at cryogenic temperatures, with the likely operating temperature equal to ~35 K for the focal plane array (FPA) and associated electronics and somewhat higher temperatures for the optics (80 K and 50 K for the short and long wavelength modules, respectively).

13.2.5.1 Short wavelength module

For the short wavelength module, the relatively low thermal background implies that to be source-limited the detector should operate below ~80 K and the optical system within the dewar can be somewhat warmer, around 120-140 K. For the best noise performance, the IR detector within the instrument should operate at the lowest temperatures consistent with the thermal power dissipation of the FPA and its support electronics. This temperature will be ~35 K. To meet these requirements, we baseline a cryostat for the short wavelength module in which the optics are at a temperature of 80 K and the FPA operates at 35 K. The stability requirement of these temperatures will depend respectively on the thermal sensitivities of the optics, the optomechanical design, and the FPA performance.

In our conceptual design, the vacuum jacket for the short wavelength module is an aluminum box measuring 2.25 m × 1.45 m × 0.95 m. Within this box is an aluminum radiation shield that is actively cooled and blocks radiation from the warm external wall of the cryostat. The shield is suspended by fiberglass G10 triangle structures. Several possible strategies exist for reducing the large 290 K to 80 K radiative heat load (~40

mW/cm²) to a more manageable level. One or more "floating" shields may be placed between the cryostat wall and the 80 K shield. Another strategy that can be used is to cover the shield(s) in multiple layers of aluminized Mylar superinsulation. This stack of intervening intermediate shields is effective in reducing thermal radiation and may be characterized by an "effective thermal conductivity" of less than 10⁻⁵ W/cm-K (Richardson and Smith 1988). Wrapping the shield in ~1 cm of superinsulation reduces the heat leak to 2 mW/cm². Assuming that the emissivities can be kept below 0.1, we expect the heat load on the 80 K shield from radiation and subsequent "conduction" through the superinsulation to be ~28 W or less.

In addition to the radiative load, thermal conduction also occurs along fiberglass supports, wiring assemblies, and motor shaft mechanical feedthroughs. There is also a very small direct contribution from the guide camera. We compute their values using radiation transport theory and list conservative estimates of these values in Table 13.2-12. These estimates are consistent with experience drawn from other cryogenic near-IR spectrographs such as SpeX (Rayner et al. 2003) and PHOENIX (Hinkle et al. 2000).

At the cold stage that is connected to the FPA housing, the heat loads are radiation from the 80 K environment (radiation shield, optical bench and optics), thermal conduction along the supports, wiring, and heat generated by the FPA and associated support electronics. These loads are listed in Table 13.2-13.

Radiation	28 W
Conduction (supports)	10 W
Conduction (wiring)	10 W
Conduction (motor shafts)	4 W
Guide camera	1 W
Entrance window	2 W
TOTAL	55 W at 80 K

Table 13.2-12. Estimated heat loads on the first (80 K) cooling stage in the short wavelength module.

Radiation from shield	0.5 W
Conduction (supports)	3 W
Conduction (wiring)	1 W
Electronics (FPA, ASIC)	0.5 W
TOTAL	~5 W at 35 K

Table 13.2-13. Estimated heat loads on the second (35 K) cooling stage in the short wavelength module.

Since the radiation load on the cold stage is only about 10% of the remaining parasitic loads, fluctuations in the closed cycle cooling to the radiation shield should not have a significant effect on the temperature of the cold stage.

In our conceptual design of the short wavelength module, the total cooling is achieved by connecting the first stage of a cryocooler to the cold shield and to the optical bench within it. This shield operates approximately at 80 K and sinks the external radiation and conductive loads. The interior surfaces of the shield are polished to reduce the emissivity, reducing the heat load on the FPA housing and any cold optical baffling structures that are attached to it. A second stage of a cryocooler is thermally anchored to the FPA housing, and operates at ~ 30 K. This stage maintains the temperature of the FPA at approximately 35 K by sinking the residual 80 K radiation from the cold shield and the heat generated by the FPA itself. Cooling connections to the 80 K stage should be duplicated for additional cooling power during cooldowns (see Section 13.2.5.3), and for redundancy.

This system is simple, robust, and does not require the use of liquid cryogen. However, we consider the benefits of using cryogen in Section 13.2.5.4.

13.2.5.2 Long wavelength module

The long wavelength module requires a colder optics temperature than the short wavelength module. Assuming a detector cutoff of $5.5\ \mu\text{m}$, temperatures of less than ~ 60 K are required to keep the thermal radiation from the cryostat interior to negligible levels. Accordingly, we baseline a cryostat for the long wavelength module in which the detector array is at ~ 35 K and the inner shield and optics are held at 50 K. As with the short wavelength module, the stability requirement of these temperatures will depend respectively on the thermal sensitivities of the optical design and the FPA performance.

In this conceptual design, the cryostat dimensions are $1.35\ \text{m} \times 0.90\ \text{m} \times 0.55\ \text{m}$. Within this box are two concentric shields. The outer shield is a “floating” shield designed to intercept the majority of the thermal radiation from the warm outer walls of the cryostat. This shield should have low emissivity so as to reflect as much of the radiation as possible, but does not have to be completely light-tight and need not be connected to a source of cooling. Within this shield is a radiation shield that is connected to the first stage of a cryocooler that actively cools it and the long wavelength optical bench to 50 K. Between these shields is a ~ 1 cm space that is filled with superinsulation. The suspension for these structures can be a fiberglass triangle system as in the short wavelength module. Assuming that the emissivity of the floating shield can be kept below 0.1, the radiative heat load at the 50 K shield is then estimated to be about 16 W. This load and other thermal loads from conduction through fiberglass supports, wiring assemblies, and motor shaft mechanical feedthroughs, and the guide camera are listed in Table 13.2-14.

Radiation from floating shield	16 W
Conduction (supports)	8 W
Conduction (wiring)	9 W
Conduction (motor shafts)	4 W
Guide camera	1 W
Entrance window	2 W
TOTAL	40 W at 50 K

Table 13.2-14. Estimated heat loads on the first (50 K) cooling stage in the long wavelength module.

At the cold stage that houses the FPA, the heat loads are radiation from the 50 K shield, thermal conduction along the supports and wiring, and heat generated by the FPA and associated support electronics. These loads are listed in Table 13.2-15.

Radiation from 50 K shield	0.3 W
Conduction (supports)	2 W
Conduction (wiring)	1 W
Electronics (FPA, ASIC)	0.5 W
TOTAL	~3.8 W at 35 K

Table 13.2-15. Estimated heat loads on the second (35 K) cooling stage in the long wavelength module.

Since the radiation load on the cold stage is only about 10% of the remaining parasitic loads, fluctuations in the closed cycle cooling to the radiation shield should not have a significant effect on the temperature of the cold stage.

The conceptual design of the long wavelength module is similar to that for the short wavelength module. The first stage (50 K) of a cryocooler is thermally anchored to the inner shield and to the optical bench within it. The interior surfaces of this shield can be polished to reduce its emissivity, reducing the heat load on the FPA housing and any cold optical baffling structures that are attached to it. The second stage of a cryocooler is thermally anchored to the FPA housing, and operates at ~30 K. Cooling connections to the 50 K stage should be duplicated for additional cooling power during cooldowns (see Section 13.2.5.3), and for redundancy. It is also possible to connect cooling power to the outermost (“floating”) shield in order to reduce cooling requirements at the colder stages.

13.2.5.3 Cryocoolers and cooldowns

The cooling requirements to sink the operating heat loads estimated above for the short wavelength module (55 W at 80 K, 5 W at 35 K) and the long wavelength module (40 W at 50 K, 4 W at 35 K) can be provided by closed-cycle cryocooling. Several alternatives exist for potential cryocoolers. The two-stage CTI 1050 cryocooler used in the IRTF spectrograph SpeX is rated at 70 W at 80 K and 8 W at 20 K. A pair of CTI 1050s operating in parallel is used to cool the PHOENIX cryostat to 50 K. An alternative

cryocooler is the Cryomech two-stage PT810, a pulse-tube cryocooler rated for 80 W at 80 K and 14 W at 20 K with low vibration and low servicing requirements.

A larger cryocooler cooling capacity (particularly at the warmer stages) is necessary for faster cooldowns. To estimate this capacity we assume cold masses of 400 kg and 100 kg for the respective cold masses in the short and long wavelength modules. These cold masses are largely composed of 6061-aluminum, for which we estimate the average specific heat from 300 K to 50 K to be 650 J/kg-K. Thus the amounts of heat to be removed are 57 MJ and 16 MJ for the two modules. Because they are separate dewars, they can be cooled simultaneously, but separate cooling power and connections must be provided. Assuming that we wish to cool both dewars in 48 hours, we require ~330 W at 80 K for the short wavelength module and ~95 W at 50 K for the long wavelength module. These numbers are probably optimistic, since they do not account for thermal loads when cold and imperfect thermal coupling to the cryocooler. On the other hand, most cryocoolers have a slight performance boost at warmer temperatures. Cooling faster than 48 hr is undesirable, lest delicate components such as optical coatings become damaged by thermal gradient stresses.

Gifford-McMahon and Stirling cryocoolers with large power ratings (at 60 Hz operation) are commercially available (Cryomech model AL 300, 320 W at 80 K; Cryomech model AL 330, 100 W at 30 K; Qdrive model 2S241K, 200 W at 80 K). For redundancy and thermal stability it is wise to incorporate multiple fittings for coolers to the instrument. This arrangement also has the advantage of allowing added cooling power to be delivered to the instrument for cooldowns, and also to separate locations within the cryostat. For example, it is desirable to cool the radiation shield before the optics, so that any residual volatiles that are present condense on noncritical surfaces within the cryostat.

If it is efficient in terms of space and resources, the Observatory should provide a common cryocooling facility, with common fittings to meet the needs of various instruments with similar cooling needs.

13.2.5.4 Liquid cryogen assist

To help speed cooldowns, and to help stabilize the temperature of the radiation shield assembly, it is possible to install modest liquid nitrogen (LN₂) reservoirs into the GMTNIRS cryostats. This has several potential benefits, including faster cooling of the cryostats for testing without sensitive optics, precooling in the event that cryocooling is not available, and the ability to temporarily hold stable temperatures in certain situations (e.g. cryocooler failure, transport between instrument staging areas and the telescope, etc.). The hold time of a 10 liter LN₂ reservoir is expected to be approximately 24-48 hr. If LN₂ is used as a precooler, then some careful management of the heat flow within the cryostats may be required to prevent stressing internal systems during the initial rapid cooldown phase.

All optical and mechanical systems of the instrument should function normally when the instrument is cooled with LN₂, so that GMTNIRS can be checked out in the lab with cryogen. However, the detector background will be lowest when the cold stages are cooled to their designed operating temperature of ~30 K.

13.2.5.5 Vacuum

The GMTNIRS cryostats must be evacuated in order to be cooled. The evacuation volume of the short and long wavelength modules is estimated to be 2400 liters and 500 liters, respectively. For these volumes and using reasonable plumbing dimensions (~100 mm diameter), the pumpdown times from atmospheric pressure to $<10^{-3}$ mtorr can be reasonable, using commercially available turbomolecular pumps.

For diagnostics and safety, pressure gauges and pressure relief valves should be installed on the vacuum jackets.

13.2.5.6 Assessment

The preliminary thermal analysis presented here indicates that the cryogenic needs of GMTNIRS can be met by a series of closed-cycle cryocoolers. A more rigorous thermal analysis should be performed to establish the equilibrium temperatures of the various locations within the cryostats and to more accurately determine the heat loads and cooldown curves.

13.2.6 Internal electronics and mechanisms

In the baseline optical design of Section 13.2.2, the infrared light is imaged onto a focal plane array (FPA) detector within the cryostat. There is also a separate IR detector array used as a guide camera. Other in-dewar electronics are mentioned briefly at the end of this section.

13.2.6.1 Spectrometer

We anticipate that the science FPA take the form of a monolithic 4096×4096 HgCdTe array. Although no $4K \times 4K$ format detector is available today in 2006, on the time scale of GMT it is virtually certain that such FPAs will be developed and become available. Several near-IR astronomical instruments on long (5-10 year) development times are aiming at the $4K \times 4K$ format, and at least one vendor (Rockwell Scientific) has said that they are committing resources to development of these FPAs and associated support electronics. Prototypes of the Rockwell Scientific "HxRG" array could be available as early as later this year.

The $4K \times 4K$ HxRG FPA format is currently expected to have a pixel pitch of $15 \mu\text{m}$ (detector area is $61.44 \text{ mm} \times 61.44 \text{ mm}$), set mainly by the size of the detector substrate material. Note however, that the optical design in Section 13.2.2 is based on $18 \mu\text{m}$ pixels, following the size of the currently commercially available $2K \times 2K$ HAWAII-2RG arrays. The 2RG device specifications are listed in Table 13.2-16, along with performance of prototype 4096×4096 mosaic devices.

	2-RG, 2048^2	4096^2 Mosaic prototypes
Readout rate	100 kHz to 5 MHz continuously adjustable	
Spectral response	1.0 - 5.3 μm	0.3 - 6.0 μm , band-gap tunable
Operating temp.	$\sim 30 \text{ K}$	$\sim 30 \text{ K}$
Quantum efficiency	$>65\%$	$>80\%$ with AR coating
Well depth	$100000 e^-$	$>100000 e^-$
Dark current (array mean)	$\sim 0.1 e^-/s$ (2.5 μm @ 77 K)	$0.02 e^-/s$ (2.5 μm @ 78 K, 5 μm @ 37 K)
Fill factor	$>98\%$	$>98\%$
Read noise (array mean)	$\sim 15 e^-/\text{pixel}$ (CDS, 100 kHz)	$\sim 15 e^-/\text{pixel}$ (CDS, 100 kHz)
Integrated noise		$6 e^- \text{ rms}$ (16/16, 100 kHz, 40 K, 1000 s)

Table 13.2-16. Specifications of the Rockwell-Scientific HAWAII 2-RG arrays, and preliminary performance data on Rockwell-Scientific's mosaic 4096×4096 prototypes.

Rockwell-Scientific considers the 2RG 2048×2048 FPAs to be "buttable", however the area on the off-the-shelf chip has a minimum gap between the arrays of 2.79 mm on two

adjacent sides and somewhat more on a third side. The gap can be custom-thinned to a minimum gap of ~ 1 mm by special request. Buttability of the 4096×4096 products has not yet been determined.

The HAWAII-2RG focal plane arrays are designed to work with a companion application-specific integrated circuit (ASIC) called "SIDE CAR" that can operate at a range of temperatures (typically 30 K to room temperature, depending on available cooling power). The ASIC simplifies integration and FPA management, and performs the analog-to-digital data conversion step. By being placed within the dewar in close proximity to the FPA, the ASIC helps to maintain high throughput and low noise by keeping signal lines short and capacitances small. This minimizes weight, power consumption and complexity of the FPA subsystem. The ASIC dominates the cooling requirements, requiring more than twenty times the cooling power needed for the FPA. One important feature of this ASIC is the capability to manage subarrays within the FPA; this is useful for optimizing throughput and minimizing power dissipation. It is likely that the 4096×4096 array will be supported by a similar ASIC with low-voltage differential signaling (LVDS, 250-450 mV) outputs for low-power dissipation and high data throughput. The LVDS outputs connect to computers via an interface card. Such hardware helps standardize the format and reduces cost and personnel requirements for software development.

The total power consumption for the spectrometer detector and associated readout electronics can be crudely estimated by scaling up the numbers for existing products. Some economies will be realized in consolidating the arrays into a single monolithic device and in using LVDS outputs to carry the data out of the dewar. If the frame rate remains the same, the power dissipation in each element will be approximately proportional to the device currents. For a pixel readout rate of 1 MHz without LVDS, the power dissipation estimates are ~ 160 mW for the FPA (30 K) and ~ 4 W for the SIDE CAR electronics (30-80K). Removing this much power should not present any great difficulty. At this readout rate the entire array is read out in less than 20 seconds. Because the duty cycle is $\sim 50\%$, the steady state cooling power required by the cold electronics will be approximately 2W.

In summary, current generation near-infrared FPAs possess specifications that meet all of the required spectrometer design specifications except for size format. At least one company is planning to provide large format arrays and support electronics. Some development test products have demonstrated spectral range and noise performance better than those exhibited by the current generation of arrays, and monolithic $4K \times 4K$ products should be available for evaluation within a year or two.

13.2.6.2 Guide camera

Light for the guide camera is diverted at the slit and relayed to the guide camera. The FPA for the camera is expected to be an FPA similar to the primary spectrograph array,

but a size format of only 512×512 is needed. The operating temperatures for the guide camera FPA and support electronics may be the same as in the previous section.

The guide camera FPA and electronics will probably require a separate housing from the spectrometer FPA. Heat straps connected to the cold head can be used to cool both housings. The heat strap for the guide camera can be considerably longer, since the guide camera electronics should consume only a fraction of the power of the spectrograph array, because of the smaller format: $(512 \text{ pixels}/4096 \text{ pixels})^2 = 1/64$. At a clock rate of 1 MHz, the guide camera is read out in 0.26 seconds. The camera duty cycle can be 100%, so that the steady-state power required for the guide electronics is estimated to be 0.05 W.

13.2.6.3 Wiring

The dewar wiring to the FPAs and their associated electronics comprises both low-noise analog (e.g. voltage references) and digital timing/data lines (LVDS). Signal lines, especially low-impedance lines, will be shielded and the shields will be grounded appropriately. Analog and digital grounds will be kept separate. The grounding interface to the telescope is to be determined.

13.2.6.4 Other dewar electronics and mechanisms

A variety of other electro-mechanical systems are necessary for the operation of the spectrograph. Motorized actuators are required to drive slit and filter wheels, and to switch between the various long and short wavelength modules. Mechanical actuators rated for operation in vacuum at cryogenic temperatures are commercially available from several companies (e.g. Phytron, Empire Magnetics, Mission Research Corporation); these motors use vacuum and cryogenic-compatible bearings and/or lubricant. Cryogenic Phytron motors have been successfully used in other cryogenic infrared instruments such as GNIRS (Elias et al. 1998) and NIRI (Hodapp et al. 2003). These motors have to be served by electrical feeds for control and power, so specifying their use requires management of thermal leakage along the wiring. They also generate heat during operation; for low duty cycle use as envisioned for GMTNIRS, heating should not present much problem.

A different arrangement for rotary actuators that we have successfully used in cryogenic spectrographs is to mount conventional stepper motors on the exterior of the dewar, and then couple the rotary motion into the dewar interior via hollow fiberglass shafts (Lacy et al. 2002). Where the shaft pierces the dewar wall the vacuum is preserved by means of rotary feedthroughs (e.g. Ferrotec Ferrofluidic seals). The conductive heat leak into the dewar interior can be made very small with this motor arrangement, especially if the shaft can mechanically detach by moving along the motor axis. For rotary motions not requiring high linearity between motor location and mechanical placement, this arrangement is easy to test and troubleshoot, and helps maintain a low heat leak. An

additional benefit is that the wiring is external to the dewar, greatly reducing any risk of heat from the motors reaching the detector elements. The risk of electrical interference is also reduced as well in this arrangement.

The remaining electronic systems associated with the dewar should be straightforward and contribute little heat leakage relative to the FPA electronics. These are the digital control electronics for the motors, and simple analog data-acquisition/control for positional sensors (e.g. Hall effect and capacitive sensors), thermal sensors (e.g. Si diodes), a vacuum gauge, and resistive heaters (e.g. Thermofoil™ heaters from Minco) laminated in polyimide. These heaters are vacuum-compatible and are useful for controlled warmup of the instrument.

13.2.7 Support Electronics

This section gives an overview of the GMTNIRS electronics external to the cryostat. We estimate the power and thermal cooling requirements of these systems.

In our design concept, the GMTNIRS instrument is powered and managed by a rack of support electronics that manage the FPA and readout systems, motor systems, temperature control, and optical and mechanical diagnostics. The following items are needed:

computer 1	FPA control and data management
computer 2	guide camera, all other electronics, networking
temperature controller	temperature stability, alarms
power supplies	instrument electronics on rack
motor controller	manages motors for mechanical systems
regulated power supply	within-dewar electronics

These electronics are external to the dewar and are installed in a rack enclosure that may be either physically attached to the dewar or mounted separately and connected electrically to the dewar by cabling. The free-standing configuration is attractive for GMT, in which the instrument platform does not move with the telescope pointing. This separates the warm electronics systems from the cold dewar and also permits personnel to easily access the electronics.

13.2.7.1 Electrical Power and Conditioning

The combined electrical power consumption of the above systems is estimated to be approximately 1.2-1.5 kW, with startup and peak current draws (motors, power supply) under 3 kW.

The instrument requires standard AC current (e.g. 120 VAC, 60 Hz). The power supply circuits of critical systems (e.g. computers, temperature controllers, alarms) within the instrument should be conditioned and/or protected to withstand high-voltage spikes and other dangerous electrical conditions. This can be performed by the Observatory, but it is likely that individual electronics components of the spectrograph will be supplied with backup protection circuitry.

Many manufacturers of rack electronics specify environmental maximums of 40°C and 90% humidity for proper operation. The Observatory should strive to ensure that such conditions are met on the instrument platform. However, the electronics within the rack will generate significant quantities of heat that must be removed. This will require refrigerant-based cooling to be delivered to the electronics rack by the Observatory. A conservative estimate is the steady state power consumption: 1.2 kW at 25°C.

13.2.8 Software

The GMTNIRS software package will handle communication with GMT computers and include the following configuration/diagnostic/operating facilities: instrument observing modes, FPA and SIDECAR modes, management of guide camera images, management of the output data from the FPA, management of diagnostic data from thermal and mechanical sensors within the instrument, and issuing appropriate responses to maintain stable dewar temperatures, or alarms/shutdown commands in unusual circumstances. Some type of QuickLook data reduction is also envisioned.

13.2.9 Assembly, Alignment, and Integration

To assist with facilities planning of the GMT, this section describes some of the issues that may be relevant for delivery, installation, and checkout of GMTNIRS at the Observatory facility. The instrument schedule (including design, development, construction, and checkout) will occur in parallel with that of the telescope itself, and barring significant unexpected delays, delivery of the instrument to GMT is expected in 2016.

13.2.9.1 Handling and storage

Although GMTNIRS will be assembled, tested and checked out before delivery to GMT, the size of the cryostats (approximately $2.3\text{ m} \times 1.5\text{ m} \times 1\text{ m}$ and $1.4\text{ m} \times 1\text{ m} \times 0.5\text{ m}$) and bulk (~2000 kg, including external support electronics), demand that GMTNIRS be shipped in several pieces to GMT, whereupon it can be re-assembled on a handling cart that can deliver the instrument to mount at one of the folded ports of the telescope. The reassembly should take place in a low-particulate, instrument-dedicated space. After assembly it should be checked out again using local optical and communication simulators to test proper integration with the telescope interfaces. This will minimize the work that is required to fully check-out and test the instrument each time it is installed on the telescope.

On the telescope GMTNIRS is deployed at the folded port and should be physically installed in about a half workday. The instrument and handling cart can be physically moved into this instrument space by using the central lift.

When GMTNIRS is removed from the telescope port and stored, it does not need to be stored in a low-particulate space, but the environment should have modest humidity and temperature control to prevent degradation of electronics and other systems. The dewar may be stored on the handling cart. The dewar should be stored under modest vacuum or filled with dry inert gas. It need not be pumped continuously.

13.2.9.2 Space and facility requirements

For assembly and testing of the GMTNIRS instrument before mounting on the telescope, the instrument will require a small laboratory space at the telescope site. This space should include some clean (low-particulate) space for cleaning and final installation of critical optical components. For flexibility, this could take the form of a large portable softwall clean room (e.g. $12\text{ ft L} \times 10\text{ ft W} \times 10\text{ ft H}$). Blowers force air downward through HEPA filters mounted across the ceiling of the portable space, thus providing a constant supply of low-particulate air. The air flows downward and outward at vents near the bottom of the room, sweeping particles out and providing a positive pressure against contamination by particles in the ambient room air. Such rooms are relatively

inexpensive, and provide a flexible clean space that can be used for several different instruments. For general purpose use, a reasonable target for particulate levels is ISO class 5 (“class 100”, or 100 particles of size 0.5 μm or larger in 1 cubic foot of air). This level of cleanliness typically requires personnel in the space to be fully gowned with clean room garments. A gowning area should be constructed adjacent to the instrument clean-space.

Plumbing (cryocooler, glycol coolant) and electrical (conditioned AC, 2 kW) power equal to that of the instrument support hardware on the telescope should be supplied to the test space, in addition to separate electrical power (~ 2 kVA) for running pumps, test equipment, computers, etc. These services should be separate from those on the telescope instrument platform.

Several other useful general purpose equipment and facilities listed below:

- Clean room garments and supplies
- De-ionized water supply
- Pressure-regulated gas handling (dry nitrogen, 0.5-50 psi, filtered to <0.2 μm)
- Helium leak checker (10^{-8} torr or better)
- Dry vacuum pumps (<1 torr)
- Turbopumping system
- High speed network connections (Ethernet)
- Movable cranes (capacity 2000 kg)

13.2.9.3 Facility Simulators

The facility should consider providing a mechanical/optical simulator of the folded port to facilitate optical alignment of various instruments in the assembly area before it is installed on the telescope. This would simulate the mechanical interface at the folded port of the telescope and provide light from a set of visible and infrared light sources.

The facility should also consider providing a communications simulator for testing the instrument interfaces with the telescope control (command/configuration/test/emergency) and data acquisition (readout/archive) systems. This maximizes the amount of software configuration and debugging that can occur off the telescope.

13.2.10 References

- Alumiplate, Inc., 1998, <http://www.alumiplate.com/assets/images/newsletter6.pdf>
- Chabrier, G., Baraffe, I., Allard, F., and Hauschildt, P., 2000, *ApJ* 542, 464
- Corning NetOptix, 2004, <http://www.corningnetoptix.com/LEC.pdf>
- Elias, J.H., Vukobratovich, D., Andrew, J.R., Cho, M.K., Cuberly, R.W., Don, K., Gerzoff, Al, Harmer, C.F., Harris, D. Heynssens, J.B., Hicks, J., Kovacs, A., Li, C., Liang, M., Moon, I.K., Pearson, E.T., Plum, G., Roddier, N.A., Tvedt, J., Wolff, R.J., and Wong, W.Y., 1998, *Proc SPIE* 3354, 555.
- Dash, W.C., and Newman, R., 1955, *Phys. Rev.* 99, 1151
- Ershov, O.A., Jaffe, D.T., Marsh, J.P., and Keller, L.D., 2001, *Proc. SPIE*, 4440, 301
- Graf, U. U., Jaffe, D.T., Kim, E.J., Lacy, J.H., Ling, H., Moore, J.T., and Rebeiz, G., 1994, *Appl. Opt.* 33, 96
- Greene, T.P., Tokunaga, A.T., Toomey, D.W., and Carr, J.S., 1993, *Proc. SPIE* 1946, 313
- Hinkle, K.H., Cuberly, R., Gaughan, N., Heynssens, J., Joyce, R., Ridgway, S., Schmitt, P., and Simmons, J.E., 1998, *Proc. SPIE* 3354, 810.
- Hinkle, K.H., Joyce, R.R., Sharp, N., and Valenti, J.A., 2000, *Proc. SPIE* 4008, 720
- Hodapp, K-W., Jensen, J.B., Irwin, E. M., Yamada, H., Chung, R., Fletcher, K., Robertson, L, Hora, J.L., Simons, D.A., Mays, W., Nolan, R, Bec, Matthieu, Merrill, M., and Fowler, A.M., 2003, *PASP* 115, 1388
- Jaffe, D.T., Keller, L.D., and Ershov, O.A., 1998, *Proc. SPIE*, 3354, 201
- Kaiser, W., Keck, P.H., and Lange, C.F., 1956, *Phys. Rev.* 101, 1264
- Kaüfl, H.U., Ballester, P., Biereichel, P., Delabre, B., Donaldson, R., Dorn, R., Fedrigo, E., Finger, G., Fischer, G., Franza, F., Gojak, D., Huster, G., Jung, Y., Lizon, J.L., Mehrgan, L., Meyer, M., Moorwood, A., Pirard, J.F., Paufique, J., Pozna, E., Siebenmorgen, R., Silber, A., Stegmeier, J., and Wegerer, S., 2004, *Proc. SPIE*, 5492, 1218
- Keller, L.D., Jaffe, D.T., Ershov, O.A., Benedict, T., and Graf, U.U., 2000, *Appl. Opt.*, 39, 1094
- Keller, L.D., Jaffe, D.T., Ershov, O.A., and Marsh, J.P., 2002, *SPIE*, 4485, 385

Lacy, J.H., Richter, M.J., Greathouse, T.K., Jaffe, D.T., and Zhu, Q., (2002), PASP 114, 153

MacFarlane, G.G., McLean, T. P., Quarrington, J. E., and Roberts, V., 1958, Phys. Rev. 111, 1245-1254

Marsh, J.P., Ershov, O.A., and Jaffe, D.T., 2003, Proc. SPIE, 4850, 797

Marsh, J.P., 2006, Ph.D. dissertation, The University of Texas at Austin.

Moore, J., Ling, H., Graf, U.U., and Jaffe, D.T., 1992, Microwave and Opt. Tech. Letters, 5, 480

Rayner, J.T., Toomey, D.W., Onaka, P.M., Denault, A.J., Stahlberger, W.E., Vacca, W.D., Cushing, M.C., and Wang, S., 2003, PASP 115, 362

Richardson, R.C., and Smith, E.N., 1988, Experimental Techniques in Condensed Matter Physics at Low Temperatures, Addison-Wesley

Runyan, W.R., 1965, Silicon Semiconductor Technology, McGraw-Hill

Svantesson, K.G., and Nilsson, N.G., 1979, J. Phys. C 12, 3837

Wiedemann, G., Delabre, B., Huster, G., Moorwood, A., and Sokar, B., 2000, Proc. SPIE 4008, 1076

13.2.11 List of appendices

David Warren	Optical design report, (Microsoft Word document)
Alumiplate	http://www.alumiplate.com/assets/images/newsletter6.pdf
Corning NetOptix LEC	http://www.corningnetoptix.com/LEC.pdf