Micromachined silicon grisms for infrared optics

Douglas J. Mar,1,2 Jasmina P. Marsh,2 Casey P. Deen,2 Hao Ling,3 Hosung Choo,4 and Daniel T. Jaffe2,*

1Liquidia Technologies, Inc., P.O. Box 110085, Research Triangle Park, North Carolina 27709, USA
2Department of Astronomy, The University of Texas at Austin, 1 University Station C1400, Austin, Texas 78712-0259
3Department of Electrical and Computer Engineering, The University of Texas at Austin, 1 University Station C0803, Austin, Texas 78712-0240
4School of Electronic and Electrical Engineering, Hongik University, Seoul, South Korea
*Corresponding author: jasna@astro.as.utexas.edu

Received 26 September 2008; revised 6 January 2009; accepted 6 January 2009; posted 8 January 2009 (Doc. ID 102060); published 11 February 2009

We demonstrate the successful fabrication of large format (approximately 50 mm × 50 mm) gratings in monolithic silicon for use as high-efficiency grisms at infrared wavelengths. The substrates for the grisms were thick (8–16 mm) disks of precisely oriented single-crystal silicon (refractive index, \(n \sim 3.42\)). We used microlithography and chemical wet etching techniques to produce the diffraction gratings on one side of these substrates. These techniques permitted the manufacture of coarse grooves (as few as 7 grooves/mm) with precise control of the blaze angle and groove profile and resulted in excellent groove surface quality. Profilometric measurements of the groove structure of the gratings confirm that the physical dimensions of the final devices closely match their design values. Optical performance of these devices exceeds the specifications required for diffraction-limited performance (RMS wave surface error < \(\lambda/20\)) in the near- and mid-infrared (1–40 μm). Peak diffraction efficiencies measured in the reflection range from 70–95% of the theoretical maximum. Tests of our grisms in the near infrared indicate transmission efficiencies of 30–48% uncorrected for Fresnel losses and confirm excellent performance. In infrared wavelength regions where silicon transmits well, the blaze control and high index permit high-resolution, high-order dispersion in a compact space. The first application of these grisms is to provide FORCAST, a mid-infrared camera on NASA’s airborne observatory, with a moderate resolution (\(R = 100–1000\)) spectroscopic capability. © 2009 Optical Society of America

OCIS codes: 050.1950, 300.6340, 350.1260.

1. Introduction

Transmission gratings mounted on or fabricated on wedged substrates combine the dispersive action of a diffraction grating with the varying optical path length across the prism, and are therefore called grisms, or Carpenter prisms [1]. Typically, grisms are inserted into a beam of collimated or nearly collimated visible or infrared light and used to disperse the light as it is transmitted through the device. The primary geometric parameters are the grism wedge angle \(\delta\) and the grating period \(\sigma\); these specify into which angles the various wavelengths and orders are diffracted. The grating equation applied to a grism is

\[
\frac{m \lambda}{\sigma} = n \sin \left[\delta - \sin^{-1}\left(\frac{\sin \alpha}{n}\right)\right] - \sin \beta,
\]

where \(m\) is the order in which the grating is used, \(\lambda\) is the vacuum wavelength, and \(n\) specifies the index of refraction of the grism material. The angle \(\delta\) is the prism wedge angle, and \(\alpha\) and \(\beta\) specify the angles of the incident and transmitted beams with respect to the optical axis.
to the normals at the entrance and grating (exit) faces of the prism (see Fig. 1). A beam that passes through without deflection satisfies $\beta = \delta - \alpha$, and Eq. (1) becomes

$$\frac{m\lambda}{\sigma} = n \sin \left[ \delta - \sin^{-1} \left( \frac{\sin \alpha}{n} \right) \right] - \sin(\delta - \alpha). \quad (2)$$

For grisms that are blazed such that the facets on the exit face are parallel to the flat entrance face (see Fig. 1, right), the blaze wavelength condition occurs when $\beta = \delta$:

$$\frac{m\lambda_{\text{blaze}}}{\sigma} = (n - 1) \sin \delta \quad (3)$$

and light at the blaze wavelength passes through the grism undeviated. For modest angle grisms ($\delta < 40^\circ$) used in low order, reasoning from scalar electromagnetic theory predicts a maximum in the efficiency at wavelengths near $\lambda_{\text{blaze}}$, although for larger $\delta$, a more rigorous treatment may be necessary [2].

The diffraction-limited resolving power for nearly normal incidence ($\alpha \approx 0$) is given by

$$R = \frac{\lambda}{\Delta \lambda} = (n - 1) \frac{D}{\lambda} \tan \delta, \quad (4)$$

where $D$ is the pupil diameter. For a given wavelength $\lambda$ and desired resolving power $R$, the required $D/\lambda$ ratio is inversely proportional to $n - 1$. Thus, the size of the pupil (and that of the overall optical system) can be reduced by selecting grisms made from high-index material rather than low index material. Equivalently, for a given $D/\lambda$ ratio, the resolving power is increased by choosing a material with a high refractive index (see Table 1), or by going to larger grism angles.

It is possible to relax the condition that the blaze facets be parallel to the entrance face, so that the wedge angle $\delta$ and the blaze angle $\theta$ are not necessarily equal (see the left panel of Fig. 1) and the undeviated beam is no longer on the blaze. For etched silicon gratings, this geometry can be used to reduce the projection of the beam upon the unused facets (Fig. 1) to minimize geometric shadowing losses (see Section 4), but only at the cost of an increase in the overall thickness of the device.

The grating side of a grism is a periodic array of diffracting elements, and is usually formed by one of four methods: ruling, replication, diamond-machining, or patterning/etching. For visible wavelengths, ruled gratings in glass or replica gratings in resin that can be mounted on prisms are commercially available [3,4]. At longer wavelengths, however, optical transmission properties can limit the choice of material, as most resins become absorbing beyond about $3 \mu m$ (see Table 1 for some recently used optical and infrared materials). Also, the groove spacing of an infrared grating is typically larger than that for a visible light grating by about a factor of 3 to 10. This coarseness can preclude the selection of ruled grisms, as it is difficult to control the blaze when removing large amounts of substrate material. Diamond-machining techniques can generate both intricate and coarse structures on many substrates (including metals, Si, ZnSe, Ge, and many polymers) with very low surface roughness ($\sim 5 \text{ nm}$) [5,6], but the surface may still possess large-scale machining defects such as cutting arcs and ripple. For

![Fig. 1. Left: Schematic diagram of a grism with wedge angle $\delta$ and blaze angle $\theta$. The blaze angle is measured between the groove facet and the grating surface. The incidence angle $\alpha$ and diffracted angle $\beta$ are measured with respect to the corresponding normals at the surfaces of the grism, and the sign convention is that both angles are positive in the sense drawn. Center: Detail of a silicon grating surface showing groove period $\sigma$. The plane of the figure is the (110) crystal plane. For silicon material, the facets are adjacent (111) crystal planes and the valley angle measures 70.53°. For the situation in which the facets are nonadjacent, the valley angle is 109.47° (not shown here, but see [52]). As shown, for acute valley angles, the projection along the optical axis of the unused facet and the groove top $r$ partially coincide, reducing the geometric transmission loss. Right: For most of the grisms that we have fabricated, the blazed facets are parallel to the entrance surface so that $\theta = \delta$.](image)

### Table 1. Potential Infrared Grism Materials and Properties

<table>
<thead>
<tr>
<th>Reference</th>
<th>Material</th>
<th>Index$^a$</th>
<th>Grism Type$^b$</th>
<th>Bandpass (micrometers)</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carl Zeiss Inc. [3]</td>
<td>resin/BK-7</td>
<td>1.5/1.5</td>
<td>replica, hybrid</td>
<td>0.3–2.5</td>
<td></td>
</tr>
<tr>
<td>CaF$_2$</td>
<td>1.4</td>
<td>ruled</td>
<td></td>
<td>0.15–8</td>
<td></td>
</tr>
<tr>
<td>Rayner [71]</td>
<td>KRS-5</td>
<td>2.4</td>
<td>ruled</td>
<td>0.5–35</td>
<td></td>
</tr>
<tr>
<td>Ebizuka et al. [72]</td>
<td>LiNbO$_3$/ZnS</td>
<td>2.2/2.2</td>
<td>etched, hybrid</td>
<td>0.35–4.6</td>
<td>birefringent</td>
</tr>
<tr>
<td>ZnSe</td>
<td>2.5</td>
<td>ruled</td>
<td></td>
<td>0.6–21</td>
<td>brittle, low efficiency</td>
</tr>
<tr>
<td>Si</td>
<td>3.4</td>
<td>etched</td>
<td>1.2–8, 20–40</td>
<td>1.8–23</td>
<td>monolithic</td>
</tr>
<tr>
<td>Kaüfl et al. [7]</td>
<td>Si/Ge</td>
<td>3.4/4.0</td>
<td>ruled, hybrid</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$^a$The indices of refraction are for comparison purposes only, since the actual index varies with wavelength and temperature.
$^b$Hybrid grisms are formed by fabricating the grating on a thin substrate and then attaching it to a thicker prism substrate (e.g., resin on BK-7).
large-area gratings, issues can arise with cutting tip wear, due to the serial way in which each groove is created, as well as with the intrinsically slow speed of the serial machining process, which, in turn, places demands on the thermal and mechanical stability during machining. An alternative fabrication method using lithography and anisotropic etching is, by contrast, a parallel method that can produce coarser groove spacing with excellent blaze characteristics and surface quality. It is particularly suited for single-crystal materials in which the crystal plane directions are maintained on macroscopic scales. In this work, we focus on near and mid-infrared applications using monocrystalline silicon (see Section 2).

In spectroscopic applications, gratings are often used as compact dispersers that do not appreciably deviate the direction of a collimated beam at the blaze wavelength. Equation (1) can be expanded in terms of \((\sin \alpha)/n\) to illustrate the dependence of the change in \(\beta\) on the magnitude of change in \(\alpha\):

\[
\sin \beta = -\frac{m \lambda}{\sigma} + n \sin \delta - \sin \alpha \cos \delta + O \left( \frac{\sin \alpha}{n} \right)^2.
\]  

(5)

This equation implies that modest misorientations of the grism [represented by a small change in \(\alpha\) in Eq. (5)] lead to only very small changes in the direction of the diffracted beam since the value of \(\beta\) changes by an almost equal amount [7]. For example, for a Si grism with \(n = 3.4\), \(\delta = 6.16^\circ\), and \(\sigma = 87 \mu m\) operating at \(m = 1\) (\(\lambda_{\text{blaze}} = 22.4 \mu m\)), a device tilt of 1° in the dispersion axis leads to a deflection of the central blaze wavelength of slightly less than 0.001° from that of a device with no tilt. It can also be shown from Eq. (1) that \(d\beta/d\lambda\) is approximately constant with small changes in incidence angle \(\Delta \alpha\). Since the transmitted light through a grism is not very sensitive to the angular orientation of the device, gratings may be mounted in filter wheels or similar inexpensive mechanisms that do not have extremely tight tolerances on the angular positioning.

Because gratings can be designed so that the light rays pass through undepleted or nearly so, downstream optics can support both imaging and spectroscopic modes, depending on whether the gratings are in the path of the beam or not. The use of gratings can therefore simplify the design of a multifunction instrument. This advantage has helped gratings find a place in many near-infrared [8–10] and mid-infrared [11–13] astronomical or imaging systems that double as spectrographs. Other potential applications for gratings include dispersion of wavelength-multiplexed light signals into an array of beams, thereby providing simultaneous demultiplexing with a “single grating coupler” element instead of a bank of filters for optical communication in the near-IR [14,15] or confocal microscopy [16,17] at visible wavelengths. Another potential application uses combinations of gratings to compensate higher-order dispersive effects when compressing and stretching light pulses, a technique that makes them potentially useful for time-domain laser pulse applications [18,19].

The work here demonstrates the fabrication of high-quality silicon gratings with coarsely spaced grooves for near and mid-infrared spectroscopy applications (see Table 2). We discuss the choice of silicon as a suitable grism material, report on the techniques and methods used to fabricate the gratings, discuss factors that can limit their performance, and display finished devices that have high efficiency over large (25 mm and up) aperture diameters. As a direct consequence, the gratings exhibited here will provide a mid-infrared camera on an airborne astronomical observatory with moderate resolution spectroscopy capabilities. Large, coarsely-ruled silicon gratings may be combined in cross-dispersed configurations to enable a new capability: moderate to high-resolution spectroscopy in the near-IR using all-transmissive optics. In an earlier paper [20], we discussed the production and evaluation of high-quality silicon gratings as reflection devices used in immersion.

2. Silicon

Silicon is an important and useful optical material both because of its optical and mechanical properties and because process technologies have been developed for semiconductor VLSI electronics and MEMS applications. When antireflection coated to reduce Fresnel losses, high-purity float-zone silicon transmits well between 1.2 and \(\sim 8 \mu m\) and, as shown in Fig. 2, from \(\sim 20\) to 40 \(\mu m\) (and beyond) [21–25]. Between these regions, infrared lattice absorption due to phonons is observed [25–30] and limits usable optical path lengths to a few mm or less. By using high-resistivity \((\rho > 1000 \Omega\cdot cm)\) float-zone silicon, for which the absorption coefficient can be small (e.g., \(\alpha < 10^{-2} \text{ cm}^{-1}\) for \(\lambda\) between 1.2 and 6.5 \(\mu m\)) [23,24], and by insuring that the fabrication process neither introduces impurities that can scatter light, the absorption losses can be kept at acceptable levels. The lattice absorption can be reduced somewhat (particularly at the edges of the useful wavelength range) by lowering the temperature [25,26,29]. Narrow and strong absorption features due to oxygen may also occur near 9 and 19 \(\mu m\), so for infrared

<table>
<thead>
<tr>
<th>Table 2. Summary of Design Parameters for Silicon Grism Devicesa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grating</td>
</tr>
<tr>
<td>G2</td>
</tr>
<tr>
<td>G3</td>
</tr>
<tr>
<td>G4</td>
</tr>
<tr>
<td>G5</td>
</tr>
</tbody>
</table>

aSee Fig. 1 for the definitions of the various dimensions. 

bFor all gratings in this paper, the prism opening angle (\(\delta\)) is equal to the blaze angle of the grooves (\(\phi\)).

cIntended order and blaze wavelength as used in the FORCAST mid-IR camera.
applications from 5 to 40 μm, lower oxygen content reduces these absorptions [31–33]. Processing silicon at elevated temperatures around 800–1000 °C appears to be beneficial in reducing the infrared activity of oxygen defects [32]. In our case, this processing is achieved incidentally during the deposition of the LPCVD nitride layer. Other absorption features can occur to 40 μm [25–28,34,35] and beyond [25–28,36]. The short wavelength cutoff occurs at a wavelength of approximately 1.1 to 1.2 μm [37,38] at the silicon bandgap. At lower temperatures, this cutoff moves slightly towards shorter wavelengths (∼1.07 μm at 77 K) [22]. For silicon diffraction gratings that are fabricated using wet etch processes to create the diffracting surfaces, a low oxygen content also reduces the facet roughness [39,40], although it is not clear that the surface roughness of the facets is the dominant scattering process [41].

The high index of refraction of silicon (n = 3.44 at λ = 2.5 μm at 295 K) [42] permits large dispersing power in a small device, as the resolving power in Eq. (4) can be larger by a factor of 5 than for a grism made from modest index material such as CaF₂. The refractive index decreases by approximately 0.8% as the temperature is lowered from 295 to 80 K [42].

Mechanically, silicon is hard, possesses a high elastic modulus, and can be polished to high optical flatness [43]. When cooled to the cryogenic temperatures required for sensitive infrared optical measurements, it is mechanically stable and has a modest thermal contraction relative to those of metals and other mounting materials [45]. It is possible to apply antireflection coatings to silicon surfaces to enhance the transmission at the silicon–vacuum (or silicon–air) interfaces.

3. Fabrication

The patterning of precisely positioned periodic grooves in silicon can be accomplished by photolithographic methods [44] that permit precise pattern transfer onto a silicon surface that has been polished optically flat. In combination with anisotropic wet etch techniques that preferentially etch along the ⟨100⟩ directions up to a hundred times faster than along the ⟨111⟩ directions [45], lithographic patterning permits the fabrication of precisely positioned and aligned ⟨111⟩ facets in the grating surface [46]. The groove orientation is controlled by the underlying crystal structure. For single-crystal silicon, the orientation of the grooves can be essentially perfect, provided that the lithographic mask used to pattern the structures (see Section 3) can be aligned precisely to the underlying crystal structure. The high etch anisotropy leads to groove profiles that are flat and smooth from the groove top to the valley (Fig. 3).

In a grism, the grooved surface forms one side of a prism shaped dielectric structure. To produce grisms from monocrystalline silicon using lithographic patterning and anisotropic etching, we must therefore start with substrates at least as thick as the bottom of the final prism. Such substrates are considerably thicker than standard semiconductor wafers and require modifications to standard semiconductor processing methods. Over the past decade, several groups have developed methods for fabricating diffraction gratings on silicon substrates [47–56]. It is also necessary for the production of low order grisms to be able to produce gratings with asymmetric groove profiles (see Fig. 3). This section describes the methods for precision production of such asymmetric grooves [54]. We have been successful in producing high-quality gratings on monolithic substrates up to 25 mm thick, thus producing grisms designed for use at 5–8, 20–28, and 28–38 μm (and suitable for use at 1.1–5 μm as well) that are complete (see Fig. 4 and Table 2) except for commercial antireflection coatings. Our fabrication methods are summarized in this section.

Our production starts from blanks of high-purity monocrystalline silicon. Silicon is commercially available as boules of various diameters (e.g. 75, 100, 150, 200 mm) and resistivities. During wet etch processes, crystal defects can produce pits and hillocks [57] that can scatter light in optical applications. Such issues, as well as the preference for low-oxygen content, have led us to select float-zone (FZ) material. Knowledge of the crystal growth axis of commercial silicon boules is accurate to approximately 1°. In lithographic patterning alignment steps, this level of accuracy is insufficient to prevent dislocations from appearing during the chemical micromachining of long grooves. We therefore orient the boule by using x-ray diffractometry to locate the
crystal directions to within 0.05°. A precision (110) flat is then ground on one side of the boule. This flat is perpendicular to the grating surface and to the groove facets (see Fig. 1) and serves two purposes: it provides a stable platform upon which to mount the boule for subsequent cutting, and it serves as an alignment marker in later lithographic steps. The boule is then sliced into blanks of sufficient thickness (typically 10–30 mm) to contain the grisms and to guarantee their rigidity. The blaze angle θ is determined by bias-slicing the boule at the appropriate angle. For example, if the surface exposed by the slice is a (100) plane, a symmetric (θ = 54.7°) grating results (center panel of Fig. 3), whereas rotating the boule around the (210) axis produces gratings with asymmetric groove profiles (top and bottom panels of Fig. 3). The exposed surfaces are then ground, etched to remove saw damage, and the top surface

Fig. 3. Scanning electron micrographs of gratings with different blaze angles δ and groove constants σ: (a) δ = 6.16° and σ = 25 μm, (b) δ = 54.7° and σ = 25 μm, and (c) δ = 63.4° and σ = 80 μm. The grooves in (b) are symmetric with respect to the top surface. In all three panels the sloping faces are very nearly parallel to {111} crystal planes. From surface profilometry, we obtain valley angles of 72.12° with a measurement uncertainty of 0.05°. The small difference between this value and the cos⁻¹(1/3) = 70.53° angle between nearby {111} planes reflects undercutting arising from the finite etch anisotropy [20]. The top and center panels are taken from [52].

Fig. 4. (Color online) Images taken of silicon gratings after wet etching, before (top) and after (middle, bottom) devices have been shaped into wedges. In the top photograph, the camera flash has been dispersed left–right by the grating. The major (horizontal) axis measures 76 mm. For the grisms in the middle and bottom images, the wedge angles δ are clearly visible and the ruled surfaces are towards the viewer. In the middle image, the polished entrance face of the grism on the right (grating area 51 mm×50 mm) is seen reflected in the ruled surface of the grism on the left (grating area 51 mm × 57 mm). In the bottom image, the corresponding area is 37 mm × 32 mm. Only commercially-available antireflection coatings are needed to complete the wedged grisms. From top to bottom, these four gratings are G2, G5 and G4, and G3 as listed in Table 2.
is polished to optical flatness [typical root mean square (rms) surface figures less than 1/50 of a wave at 632.8 nm] using chemical-mechanical planarization (CMP) processes. This procedure results in an extremely flat surface while minimizing mechanical stresses at and near the surface [58]. The blanks are then coated with a thin (60–100 nm) film of low pressure chemical vapor deposition (LPCVD) silicon nitride as a passivation layer.

After the deposition of the passivation layer, we create a series of regularly-spaced lines on the nitride layer using photolithography. Many of the lithography steps are described elsewhere [49,50,52,54] and additional details of our current process are available in [20]. Our lithographic process employs a positive photoresist that is flood-illuminated by g-line (436 nm) and i-line (365 nm) light from a mercury-gallium lamp. To spin-coat the photoresist onto the massive blanks, we employ a custom-built spin table with sufficient torque to spin the combined moment of inertia of the blank and holder up to several thousand rpm in a period of a few seconds. Once the photoresist has been cured by heating the blank to approximately 100 °C for 20 min, a chrome-on-quartz mask consisting of a series of thin chrome stripes is carefully placed in contact with the photoresist layer. The flood illumination through the mask transfers the mask pattern to the photoresist layer. The exposure system is a custom-designed apparatus that can accommodate a wide range of substrate thicknesses (0.5 to 35 mm). During the exposure step, the temperature of the silicon blank and the quartz mask are held to within a few °C across the grating, thereby preventing potential pattern transfer errors arising from different coefficients of thermal expansion of the substrate and mask.

After the photoresist has been exposed, an image of the mask pattern is produced in the photoresist layer by immersing the photoresist-coated blank in a commercial developer solution. Next, the nitride layer is patterned using a dry (plasma) etch. The photoresist layer serves as an etch mask during this step. Thick substrates undergoing plasma etching can experience nonuniform etch rates due to variations in the electric field profile and in the plasma density within a reactive ion etch (RIE) chamber that is normally used to process thin semiconductor wafers. We have modified our plasma etcher to maintain the uniformity of the plasma in contact with the patterned surface. After the dry etch, the photoresist is stripped by immersion in acetone. The groove facets themselves that form the grating are then created by anisotropic etching in an aqueous solution of potassium hydroxide and isopropanol, maintained at 68 °C by immersion in a recirculating water bath. Ultrasonic vibrations assist in detaching bubbles from the etched surface.

Etching in potassium hydroxide creates a blazed grating over the entire patterned area of the silicon surface. A photograph of a processed blank is shown in the top panel of Fig. 4 and scanning electron microscope (SEM) images of micromachined silicon gratings are shown in Figs. 3 and 5. The nitride strips that protected the groove tops (see Fig. 5) during anisotropic etching are removed by immersing the grating in concentrated phosphoric acid at 140 °C or more. Removal of the nitride promotes adhesion of antireflection coatings that are subsequently applied to the grating surface.

To form a complete grism, the blank is cut into the desired prism shape. For the devices discussed in this paper, the entrance faces are formed parallel to the grating facets (δ = θ). These faces are optically polished to high flatness, with final surface rms figures better than ~1/20 of a wave at 632.8 nm. The center and bottom panels of Fig. 4 show the completed devices.

4. Factors Affecting Grating Performance

For applications such as infrared spectrographs, which demand high sensitivity to faint sources, overall efficiency is a primary consideration. As light passes through the grism substrate and is diffracted by the grating, it is subject to losses that can limit the ideal optical performance of the grating: index mismatch loss at the entrance and exit faces, geometric

![Fig. 5. SEM images of a symmetric (θ = 54.7°) grating immediately after etching in potassium hydroxide, viewed normal to the initial substrate surface. The grating period is σ = 142 μm. The thin dark vertical lines are the groove tops and valleys. The detailed view at bottom corresponds to the white box inset in the top panel. One can see the strip of silicon nitride covering the darker groove top and overhanging by approximately 2 μm at each edge of the groove top. The silicon nitride and the silicon hydroxide precipitates are effectively removed by washing the part in hot (140 °C) ortho-phosphoric acid.](image-url)
losses, absorption and scattering within the bulk, scattering at the surface, and various types of groove position errors [59]. Here we describe these potential sources of error and their possible effects on the grating performance.

Index mismatch losses (or Fresnel losses) take place at interfaces where there is a discontinuity in the index of refraction. Because the index of silicon is large \( n = 3.4 \), the substantial reflection loss at the two interfaces limits the transmission of an uncoated silicon device to \( [4n/(n + 1)]^2 = 49\% \). By applying broadband antireflection optical coatings to the entrance and exit faces of the silicon grism, the transmission can be raised to \( \sim 95\% \). Applying uniform multilayer coatings to the corrugated surfaces of a grating can be challenging but it is possible [60].

Geometrical losses occur where portions of a given grism groove are geometrically shadowed by parts of the adjacent grooves. For a normally-incident (\( \alpha = 0 \)) beam from the left (Fig. 1), most of the light passes through the vertical facet and is diffracted according to Eq. (1). However, in silicon grisms produced by chemical etching, the unused sloping facet and the groove top \( \tau \) intercept a fraction of the beam (see Fig. 1, center). Depending on the value of \( \alpha \) and on the details of the groove geometry, this light is diffracted into other directions and thereby lost. In the case of normal incidence and with the blaze parallel to the entrance face (\( \delta = \theta \)), the loss is minimized by keeping the groove top \( \tau \) as short as possible and having grisms with small opening angles \( \delta \). Figure 6 shows the geometric loss in a silicon grism (valley angle of either 70.53° or 109.47° depending on the crystal orientation) [54] as a function of grism angle \( \delta \), for \( \delta = \theta \) and a fixed ratio \( \tau/\alpha = 0.05 \). The choice of 70.53° is preferable since, in that case, the areas covered by the unused facet and the groove top partially coincide and the beam therefore sustains less shadowing loss. The geometric shadowing provides an estimate of the actual diffraction losses, which we can obtain by a full electromagnetic calculation to model the efficiency behavior of grisms in low order.

We determine the geometry of the grating grooves using surface profilometry and using SEM micrographs showing overhanging nitride edges resulting from etch undercut of silicon (Fig. 5). From the profilometry, we measure valley angles of 72.12° ± 0.05°. This value is close to the ideal value of \( \cos^{-1}(1/3) = 70.53° \) formed by the intersection of adjacent \( \{111\} \) families of crystal planes. The difference in angle results from the finite etch rate in the \( \langle 111 \rangle \) crystal direction which leads to a global tilt of the grating facets. The etch rate ratio implied by the angle difference and by the size of the undercut is \( \sim 60 \). If the anisotropy ratio is known in advance, one should account for the tilt at the orientation step in the processing to prevent a deviation from the desired blaze angle.

Scattering and absorption within the bulk silicon can also lower the optical throughput. To minimize these bulk effects, the optical path length through the material should be kept as short as possible. The optical path length difference across the beam inside the grism is the product of the beam diameter and \( n \tan \delta \). Because the grism is a wedged device, differential bulk absorption/scattering occurs across the aperture. The linearly varying optical path length through the wedge results in an intensity that tapers exponentially across the grating and slightly broadens the point spread function in the direction of the taper, thus reducing the contrast in the side lobe pattern. As illustrated in Fig. 7, these effects are

![Fig. 6. Geometric loss as a function of grism angle \( \delta \), for \( \alpha = 0 \) and fixed valley angles of 70.53° (solid curve) and 109.47° (dotted curve) between the facets of the grooves, for \( \tau/\alpha = 0.05 \). The choice of 70.53° is preferable since, in that case, the areas covered by the unused facet and the groove top partially coincide and the beam therefore sustains less shadowing loss. The geometric shadowing provides an estimate of the actual diffraction losses, which we can obtain by a full electromagnetic calculation to model the efficiency behavior of grisms in low order.](Image 6)

![Fig. 7. (Color online) Computed point spread function for transmission of a 25 mm diameter collimated beam through a silicon grism with \( \delta = 32.6° \), showing the effect of absorption and tapering due to differential absorption in the silicon material across the beam. The untapered (line) and tapered (dots) curves are calculated for Si absorption coefficients \( \alpha = 0 \) cm\(^{-1}\) and 0.2 cm\(^{-1}\), respectively. As shown, the main effect of absorption is to attenuate the intensity across the beam: the peak maximum has dropped by 16%. The width of the best Gaussian fit to the tapered profile (dotted line) has increased only slightly (approximately 0.1%) over the width of the best fit (not shown) to the untapered profile (solid line). At the center of the beam, the path length in the Si is 0.8 cm.](Image 7)
negligible at wavelengths where the silicon absorption is acceptably low from the point of view of efficiency.

Leftover silicon nitride and other debris on the surface (see bottom panel of Fig. 5) can contribute to scattering, but for our devices, the immersion in hot phosphoric acid eliminates this problem. Randomly distributed defects in the groove surface, such as point defects and surface roughness, can also scatter the incident light. To assess the surface condition of the gratings on nanometer length scales, we use atomic force microscopy (AFM). Figure 8 shows an AFM scan of a 5 μm × 5 μm portion of a groove facet of a grism with \( \theta = 6.16^\circ \) and \( \sigma = 25 \mu m \) (G2, see Table 2). As shown, the grooves are smooth: the surface roughness is less than 2 nm rms on the scale shown and the groove facet is free from hillocks and etch pit formations [57,61]. Even if the grooves themselves are smooth (Fig. 8) and flat (Fig. 3), the overall grating performance can be degraded by any errors in the groove orientations and locations. Orientation errors are unlikely since the groove facets are aligned with the underlying silicon lattice, which is monocrystalline. Piston-type groove errors due to variations in groove placement may originate from noise sources associated with lithography. Jog defects—abrupt changes in displacement within a single groove—most likely arise from irregularities in the width of the nitride lines that are patterned by the plasma etch. If these jog displacements are large compared with the wavelength, they lead to inter-order power in the blaze. For the devices described here, jog defect densities are below 10 cm\(^{-2}\) and areal defect fractions are under 0.1%.

The effect of groove position errors depends on their coherence. The grism optical path error is

\[
e = (n - 1)\varepsilon_{\text{rms}},
\]

where \( \varepsilon_{\text{rms}} \) measures the rms of the “piston” error distribution of the grooves along the optical path. Assuming that the groove position errors are Gaussian and uncorrelated, the errors degrade the peak efficiency according to [50]

\[
\frac{\eta}{\eta_0} = \exp \left[ -\left( \frac{2\pi}{\lambda} (n - 1)\varepsilon_{\text{rms}} \right)^2 \right],
\]

where \( \eta_0 \) is the maximum possible efficiency. To maintain at least 80% of the incident power in a diffraction-limited spike, Eq. (7) implies that, in silicon \((n = 3.4)\), the rms error tolerance \( \varepsilon_{\text{rms}} \) must be no more than \( \lambda/32 \), where \( \lambda \) is the vacuum wavelength of the incident beam. If the silicon grating is used as a front-surface reflective device in Littrow, the corresponding expression to Eq. (6) for the optical path errors is \( \varepsilon_{\text{opt}} = 2\varepsilon_{\text{rms}} \) and the corresponding 80% criterion is similar: \( \varepsilon_{\text{rms}} < \lambda/27 \). Measured along the grating surface in the dispersion direction, the actual tolerances in the groove positions are equal to \( \varepsilon_{\text{rms}}/\sin \delta \), so tolerances on groove positioning are less stringent for small grating angles. The groove position errors can be a combination of both Gaussian random errors and slow variations over larger spatial scales. These large-scale errors could be introduced by imperfectly flat substrates, errors in the stripe positions in the mask, imperfectly flat masks and/or mask flexure, imperfect contact between the lithography mask and the substrate, non-uniformity during the plasma etch, or variations in the wet etch environment [59].

5. Performance Measurements

To establish the extent to which the various sources of error could affect the optical performance, we evaluate our fabricated silicon grisms by interferometric measurements of the grating surface in reflection at visible wavelengths, as well as by taking spectra in the visible and near-infrared in reflection and transmission. The information gained is fed back into the development to improve the yield and the quality of the finished gratings [60,62]. Here we demonstrate the excellent optical performance of our fabricated grisms.

A. Reflection Measurements

Immediately after a grating has been etched into the surface of a disk, we can estimate its efficiency and partly predict its final performance as a grism by performing optical tests on the grooves in reflection at visible wavelengths \( (\lambda = 543.5 \text{ nm}, 632.8 \text{ nm}) \). In Fig. 9, we show reflection spectra taken using a collimated 25 mm beam from a green He–Ne laser at \( \lambda = 543.5 \text{ nm} \) incident on the grating in Littrow configuration [50]. Since this wavelength is not on the blaze for any of the grisms shown, the incident beam is diffracted into multiple orders. No diffraction

\[
\eta = \frac{\lambda}{2\pi} (n - 1)\varepsilon_{\text{rms}},
\]

where \( \eta_0 \) is the maximum possible efficiency. To maintain at least 80% of the incident power in a diffraction-limited spike, Eq. (7) implies that, in silicon \((n = 3.4)\), the rms error tolerance \( \varepsilon_{\text{rms}} \) must be no more than \( \lambda/32 \), where \( \lambda \) is the vacuum wavelength of the incident beam. If the silicon grating is used as a front-surface reflective device in Littrow, the corresponding expression to Eq. (6) for the optical path errors is \( \varepsilon_{\text{opt}} = 2\varepsilon_{\text{rms}} \) and the corresponding 80% criterion is similar: \( \varepsilon_{\text{rms}} < \lambda/27 \). Measured along the grating surface in the dispersion direction, the actual tolerances in the groove positions are equal to \( \varepsilon_{\text{rms}}/\sin \delta \), so tolerances on groove positioning are less stringent for small grating angles. The groove position errors can be a combination of both Gaussian random errors and slow variations over larger spatial scales. These large-scale errors could be introduced by imperfectly flat substrates, errors in the stripe positions in the mask, imperfectly flat masks and/or mask flexure, imperfect contact between the lithography mask and the substrate, non-uniformity during the plasma etch, or variations in the wet etch environment [59].

5. Performance Measurements

To establish the extent to which the various sources of error could affect the optical performance, we evaluate our fabricated silicon grisms by interferometric measurements of the grating surface in reflection at visible wavelengths, as well as by taking spectra in the visible and near-infrared in reflection and transmission. The information gained is fed back into the development to improve the yield and the quality of the finished gratings [60,62]. Here we demonstrate the excellent optical performance of our fabricated grisms.

A. Reflection Measurements

Immediately after a grating has been etched into the surface of a disk, we can estimate its efficiency and partly predict its final performance as a grism by performing optical tests on the grooves in reflection at visible wavelengths \( (\lambda = 543.5 \text{ nm}, 632.8 \text{ nm}) \). In Fig. 9, we show reflection spectra taken using a collimated 25 mm beam from a green He–Ne laser at \( \lambda = 543.5 \text{ nm} \) incident on the grating in Littrow configuration [50]. Since this wavelength is not on the blaze for any of the grisms shown, the incident beam is diffracted into multiple orders. No diffraction

\[
\eta = \frac{\lambda}{2\pi} (n - 1)\varepsilon_{\text{rms}},
\]

where \( \eta_0 \) is the maximum possible efficiency. To maintain at least 80% of the incident power in a diffraction-limited spike, Eq. (7) implies that, in silicon \((n = 3.4)\), the rms error tolerance \( \varepsilon_{\text{rms}} \) must be no more than \( \lambda/32 \), where \( \lambda \) is the vacuum wavelength of the incident beam. If the silicon grating is used as a front-surface reflective device in Littrow, the corresponding expression to Eq. (6) for the optical path errors is \( \varepsilon_{\text{opt}} = 2\varepsilon_{\text{rms}} \) and the corresponding 80% criterion is similar: \( \varepsilon_{\text{rms}} < \lambda/27 \). Measured along the grating surface in the dispersion direction, the actual tolerances in the groove positions are equal to \( \varepsilon_{\text{rms}}/\sin \delta \), so tolerances on groove positioning are less stringent for small grating angles. The groove position errors can be a combination of both Gaussian random errors and slow variations over larger spatial scales. These large-scale errors could be introduced by imperfectly flat substrates, errors in the stripe positions in the mask, imperfectly flat masks and/or mask flexure, imperfect contact between the lithography mask and the substrate, non-uniformity during the plasma etch, or variations in the wet etch environment [59].

5. Performance Measurements

To establish the extent to which the various sources of error could affect the optical performance, we evaluate our fabricated silicon grisms by interferometric measurements of the grating surface in reflection at visible wavelengths, as well as by taking spectra in the visible and near-infrared in reflection and transmission. The information gained is fed back into the development to improve the yield and the quality of the finished gratings [60,62]. Here we demonstrate the excellent optical performance of our fabricated grisms.

A. Reflection Measurements

Immediately after a grating has been etched into the surface of a disk, we can estimate its efficiency and partly predict its final performance as a grism by performing optical tests on the grooves in reflection at visible wavelengths \( (\lambda = 543.5 \text{ nm}, 632.8 \text{ nm}) \). In Fig. 9, we show reflection spectra taken using a collimated 25 mm beam from a green He–Ne laser at \( \lambda = 543.5 \text{ nm} \) incident on the grating in Littrow configuration [50]. Since this wavelength is not on the blaze for any of the grisms shown, the incident beam is diffracted into multiple orders. No diffraction
Ghosts are visible between the orders. By summing up the intensities in the series of orders, we can estimate the on-blaze efficiencies for reflective gratings, $\eta_R$, compared to another silicon mirror which eliminates Fresnel losses from the measurement results. For low-order transmission gratings, it is not immediately clear that summing up the power in a series of diffracted orders from a monochromatic source at an off-blaze wavelength gives the peak throughput at the blaze wavelength. These results are used primarily, however, to compare the relative quality of different parts made to the same specification and not to determine absolute performance. For the three spectra in Fig. 9, these reflection efficiencies range from 65–78%.

Interferometric reflection measurements provide the (external) surface error plot shown in Fig. 10. The surface plot is obtained using an optical interferometer using collimated red He–Ne light by illuminating grating G2 in Littrow. As shown, the surface deviations are correlated into structures with spatial wavelengths approaching 10 mm or more. However, these deviations are small and the overall surface figure ($\epsilon_{\text{rms}} < 10^{-2}$ waves rms at 632.8 nm) is excellent. We can expect excellent performance of these gratings at their application wavelengths ($\lambda = 1$ to 40 $\mu$m), now that they have been given their final grism form.

B. Transmission Measurements

Once the grisms are given their final shape (that of a prism with grating grooves on one side), we can evaluate them as transmissive devices. Figure 11 shows transmission spectra for grisms G3, G4, and G5. These data were obtained using a 10 mm beam diameter (limited by our test equipment) at $\lambda = 1523$ nm with fast camera optics resulting in a large field of view. Each of the spectra in Fig. 11 consists of a series of orders as in the case of spectra recorded in reflection, because the laser wavelength is again not on the blaze for these grisms. Between orders, no ghosts are visible in Fig. 11, and in high signal-to-noise measurements, we have verified their absence down to a level of $<10^{-3}$ of the brightest order [20].

Fig. 9. Monochromatic reflection spectra of grisms G3 (top), G4 (middle), and G5 (bottom), taken using a 23 mm (G3) or 25 mm (G4, G5) diameter collimated beam with $\lambda = 543.5$ nm. Order numbers are indicated near the bottom of each spike.

Fig. 10. (Color online) Surface error plot of grism G2 ($\sigma = 25 \mu$m and $\delta = 6.16^\circ$) represents the magnitude of the deviation of measured surface from a perfectly flat surface (expressed in waves), as obtained from front-surface interferometric measurements in reflection using $\lambda = 632.8$ nm laser light. Each contour represents approximately 1/150 of a wave. The rms deviation over the indicated 25 mm diameter aperture is approximately $10^{-2}$ waves.

Fig. 11. Transmission spectra for grisms G3, G4, and G5.
We measured throughput efficiency ($\eta_T$) listed in column 2 of Table 3 using a scanning monochromator in which collimated infrared light is incident perpendicular to the uncoated entrance face of a grism. In this bench spectrograph, efficiencies were measured at each wavelength by fixing the incident angle $\alpha = 0$ and scanning through a range of diffracted angles, $\beta$. The transmitted beam was focused onto a Ge detector mounted on a swinging arm, allowing for throughput measurements at a wide range of diffracted angles. We measured throughput in several orders, for a range of $\beta$ corresponding to the free spectral range (FSR) at the position of each order, by normalizing the measured intensity by the intensity of the beam at the same angle $\beta$ with no grism in the beam. The size of the detector is larger than both the spatial extent of the spectrograph PSF and the angle step sizes, so we subsampled the dispersion pattern by changing $\beta$ in steps smaller than the angular size of the detector and then used the scaled spectrograph PSF to match the observed diffraction pattern and calculate the transmission efficiency. The wavelength range of the measurements is bounded by the transmission cutoff of Si at approximately $1.2\,\mu m$ and the response cutoff of the detector at approximately $1.8\,\mu m$. The results are summarized in three plots in Fig. 12. Measurements for G3 were not done because the FSR of the 74th order is comparable to the spectral purity of the monochromator. The beam diameter was $10\, mm$, and the positioning of the incident beam was also very similar (roughly centered on the entrance face).

In the past, we have used a boundary integral approach to compute reflection efficiencies for etched Si gratings [64] and have confirmed the model results with direct measurements [49]. We have adapted this model to predict the transmission efficiencies of G2, G4, and G5 at their operating wavelengths (given in Table 2) in orders 1, 1, and 2, respectively. The results of the mid-infrared transmission models are summarized in Fig. 13. The models consider only transmission through the grating surface (equivalent to a grism with an uncoated grating and a perfectly antireflection coated flat face). The predicted blaze

| Grism | $\eta_T$ | $|\eta_{b\lambda_{\text{blaze}}} / \eta_{b\lambda_{\text{blaze}}}/0.7|/|\eta_{b\lambda_{\text{blaze}}} (\mu m)|$ |
|-------|---------|---------------------------------|
| G2    | 0.72    | $0.40 \pm 0.018$               |
| G3    | 0.78    | $0.37/0.53/4, 1.67$            |
| G4    | 0.73    | $0.37/0.53/14, 1.58$           |
| G5    | 0.65    | $0.31/0.44/42, 1.54$           |

Table 3. Efficiencies for the Four Silicon Grisms

---

Fig. 11. Monochromatic transmission spectra of grisms G3 (top), G4 (center), and G5 (bottom), taken using a 10 mm diameter collimated beam at $\lambda = 1523\, nm$. Order numbers are indicated near the bottom of each spike.
efficiencies are 0.63, 0.63, and 0.59 for G2, G4, and G5, respectively. In high orders, gratings are scalar devices and the maximum possible throughput can be calculated using Babinet’s principle\cite{65}. For gratings with geometries equivalent to G2, G4, and G5, the scalar blaze efficiencies would be 0.57, 0.61, and 0.61, respectively. The maximum possible efficiencies in orders 4, 14, and 42 where the measurements were made (Fig. 12) should lie between the efficiencies calculated for the operating wavelengths and the scalar values which are, in any case, less than 10% apart. The last column in Table 3 lists the measured values for blaze efficiencies divided by 0.7 to account for Fresnel losses at the flat face and produce numbers directly comparable to the theoretical values. These measured values lie within 70–95% of the theoretical efficiencies.

Fig. 12. (Color online) Plot of efficiency as a function of wavelength in three orders for G2, G4, and G5.

Fig. 13. (Color online) Results of efficiency modeling as a function of wavelength in three orders for G2, G4, and G5 at operating wavelengths (see Table 2). We assume no losses from transmission through the entrance face and we model the efficiency behavior of the diffraction grating using our boundary integral approach\cite{64}.
Efficiencies measured in transmission are close to efficiencies measured in reflection when normalized to the correct reflectivity of silicon, indicating that reflection tests in the visible may be used as surrogate measurements to assess the quality of gratings without requiring a transmission measurement in the infrared. This is reasonable since the two measurements have comparable effective wavelengths and place roughly the same demands on the phase accuracy of the grating surface. Based on the measured blaze efficiencies in Table 3, if we assume an antireflection coating with 95% transmission on both the entrance and the exit face, we predict blaze efficiencies of 57–68% for the grisms described in this paper.

C. Spectral Point Spread Function

To estimate the resolving power and test the diffraction-limited performance of these grisms, we illuminate them with a collimated beam of laser light at 1523 nm and then focus the diffracted beams emerging from the grism onto an InGaAs focal plane array. For the details of the measurement setup, the reader should refer to our previous publications [62,50]. The beam diameter is limited to 10 mm by our test equipment. By a suitable choice of the camera focal ratio, we can obtain a point spread function (PSF) from the intensity profile of a single diffraction order (Fig. 14). The figure shows the normalized one-dimensional PSF of grism G3 (θ = 32.6°, σ = 87 μm, and τ = 6.0 μm, see Table 2). As shown, the shape and width of the diffraction spot are virtually identical to those obtained for the collimated beam with no grism present and agree with the theoretical curve for a circular aperture, verifying diffraction-limited performance over the beam aperture. The measured full width of the profile corresponds to a resolving power of 10^4. Before normalizing the beam and grism PSFs, the peak value of the grism PSF for this single order is ~25% of that measured for the unobscured beam.

6. Conclusions

The silicon grisms whose performance is described in Section 5 were designed to equip FORCAST, a cryogenic mid-infrared (5–40 μm) camera operating at liquid helium temperatures (4 K) with a medium resolution spectroscopic capability [66,67] for use on SOFIA, NASA’s airborne observatory [68], now undergoing initial tests. All four of the grisms have been fabricated successfully and demonstrate optical performance at a level that we can expect diffraction-limited performance over the 22 mm collimated beam of the instrument. To finish these devices requires the application of broadband antireflection (BBAR) coatings to the entrance and grating faces of each grism. A suitable BBAR coating has been developed for the 4.9–8.1 μm wavelength range. This coating can be applied to Si grating facets with good uniformity, is mechanically robust, can survive multiple rapid thermal cycles between 300 and 77 K, and raises the single-interface transmission from ~70% to better than 92% over this region [60]. The mid-infrared (17.1–28.1 μm and 28.6–37.4 μm) BBAR coatings are more challenging, as the choice of available coating materials is limited in this region. Initial development work is encouraging.

All four grisms have been installed into FORCAST to assess their performance in a cryogenic environment [69,70]. An emission spectrum of water vapor obtained using G1 and G2 in a cross-dispersed configuration is shown in Fig. 15. These large, coarsely-ruled silicon grisms can be readily combined in such cross-dispersed configurations to provide moderate resolution spectroscopy in the near-IR using all-transmissive optics. For example, the cross-dispersed configuration (G3 × G2) in FORCAST, provides a resolving power of R = 1200 with a coverage from 4.9 to 8.1 μm in a single exposure.

This work is supported by National Aeronautics and Space Administration (NASA) grants NAG-8858, NRA 03-OSS-01 via the Astrobiology Science and Development (ASTID) program, and NNA 05CS83A via the NASA Astrobiology Institute (NAI), and by funding from the Universities Space Research Association (USRA 8500-98-008) and the McDonald Observatory at the University of Texas at Austin. J.P. Marsh was supported in part by a NASA Graduate Research Program Fellowship (NGT5-137). Fabrication and testing of grating devices were made possible by facilities at the University of Texas at Austin, including the Center for Nano and Molecular Science and Technology (CNM), funded in part by the Welch Foundation, and the Microelectronics Research Center (MRC), funded

![Fig. 14. Normalized one-dimensional point spread function (PSF) taken in transmission using a 10 mm diameter beam with λ = 1523 nm. Filled circles show the PSF measured for grism G3. It is the spectrum of the brightest diffraction order shown in the first plot of Fig. 12 but with slower camera optics that helps fully resolve the image. The horizontal axis is the dispersion direction. Crosses show the corresponding data taken for the beam when no grism is present in the measuring setup. Both the data from the grism and the beam nearly coincide with the theoretical curve (solid line) calculated for a 10 mm circular aperture, indicating diffraction-limited performance. The full width of the profile corresponds to R = λ/Δλ = 1.0 × 10^4.](image-url)
in part by the National Science Foundation (NSF) through the National Nanotechnology Infrastructure Network (NNIN). We thank R. Joyce, K. Hinkle, and G. Poczup of the National Optical Astronomy Observatory for interferometer measurements. We thank K. Allers, S. Joshi, and J. Zhou for assistance in process development, and T. Gaubert and W. Frey for help with AFM measurements, J. Adams and G. Hill for the use of their tunable spectrograph, and K. Ennico, A. Tokunaga, and P. Kuzmenko for their helpful suggestions. Finally, we are grateful to P. D. Ownby and J. E. Peters for providing us with silicon transmission data and information regarding the transmission of CZ and FZ substrates.

References
26. J. E. Peters and P. D. Ownby, complete silicon and germanium far infrared spectroscopy data, Missouri University of Science & Technology, Rolla, Mo. 65401, USA, August 2008 (private communication).


