The case for OH suppression at near-infrared wavelengths

S. C. Ellis[1]⋆ and J. Bland-Hawthorn[2]

[1]*Anglo-Australian Observatory, PO Box 296, Epping, NSW 1710, Australia*
[2]*Institute of Astronomy, School of Physics, University of Sydney, NSW 2006, Australia*

ABSTRACT

We calculate the advances in near-infrared (NIR) astronomy made possible through the use of fibre Bragg gratings to selectively remove hydroxyl emission lines from the night sky spectrum. Fibre Bragg gratings should remove OH lines at high resolution ($R = 10\,000$), with high suppression (30 dB) whilst maintaining high throughput ($\approx$90 per cent) between the lines. Devices presently under construction should remove 150 lines in each of the *J* and *H* bands, effectively making the night sky surface brightness $\approx$4 mag fainter. This background reduction is greater than the improvement adapative optics makes over natural seeing; photonic OH suppression is at least as important as adaptive optics for the future of cosmology.

We present a model of the NIR sky spectrum, and show that the interline continuum is very faint ($\approx$80 photons $s^{-1}\,m^{-2}\,arcsec^{-2}\,\mu m^{-1}$ on the ecliptic plane). We show that OH suppression by high dispersion, that is, 'resolving out' the skylines, cannot obtain the required level of sensitivity to reach the interline continuum due to scattering of light. The OH lines must be suppressed prior to dispersion.

We have simulated observations employing fibre Bragg gratings of first light objects, high-redshift galaxies and cool, low-mass stars. The simulations are of complete end-to-end systems from object to detector. The results demonstrate that fibre Bragg grating OH suppression will significantly advance our knowledge in many areas of astrophysics, and in particular will enable rest-frame ultraviolet observations of the Universe at the time of first light and reionization.

Key words: atmospheric effects – instrumentation: miscellaneous – early Universe – infrared: general.

1 INTRODUCTION

The future of cosmology is dependent in part on the ability to achieve deep observations at near-infrared (NIR) wavelengths. As we strive to observe the early Universe, our observations must be performed at increasingly longer wavelengths, as the diagnostic optical spectroscopic features become more redshifted. At very high redshifts, observations in the optical become futile due to the redshifted Lyman limit. Fig. 1 shows the observed wavelengths of several useful spectroscopic lines as a function of redshift, illustrating the requirement for NIR observations.

The need for NIR observations is well documented. Over the last few years there have been several major reviews of the present status of our knowledge of the Universe, and in particular of gaps in our understanding, serving as a starting point for strategic plans for future observations (e.g. the Australian Decadal Plan for Astronomy 2006–2015,[1] the UK PPARC roadmap,[2] the US National Research

Council's 'Astronomy and Astrophysics in the New Millennium' NRC 2001 and 'Connecting Quarks to the Cosmos' Committee On The Physics Of The Universe et al. 2003).

These reviews provide similar lists of the 'big questions' for the future of astronomy, showing a consensus of opinion at an international level. Some of the key issues they identify are: *What is the nature of dark energy and dark matter? How and when do the first stars form? How and when did the Universe reionize? How and when do galaxies assemble, and how do they evolve? What is the relationship between super-massive black-holes and galaxies? How do stars and planetary systems form? How common are planetary systems capable of supporting life?*

The success of answering the questions above is reliant on NIR observations. The first half of the list depends on the ability to observe the high-redshift Universe, whilst the problems concerning planetary systems will also benefit from dust-penetrating NIR observations, especially since planetary sized bodies emit most of their radiation at NIR wavelengths.

Observations of high-redshift galaxies carried out in the NIR offer significant advantages over optical observations: extinction due to local dust is smaller in the NIR; *k*-corrections are smaller, and therefore less critical; adaptive optics is easier to achieve in the NIR.

⋆ E-mail: sce@aao.gov.au
[1] http://www.atnf.csiro.au/nca/DecadalPlan_web.pdf
[2] http://www.so.stfc.ac.uk/roadmap/rmQuestionsHome.aspx
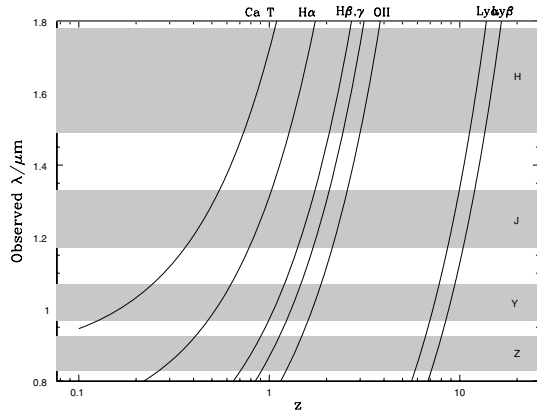
Figure 1. The observed wavelength of several useful spectral lines as a function of redshift. The shaded areas show the wavelength coverage of standard *zJH* filters.

Table 1. Typical sky brightnesses at a good observing site.

| Band | Surface brightness/mag arcsec^{-2} | |
	Natural	No OH emission
U	21.5	–
B	22.4	–
V	21.7	–
R	20.8	21.7
I	19.9	21.7
J	16.0	22.1
H	14.0	22.2

However, there are also significant disadvantages in NIR astronomy, the chief concern being the high background level. Table 1 lists typical surface brightnesses for a good observing site, as used in the ESO Extremely Large Telescope exposure time calculator.[3]

At wavelengths longer than the *H* band, thermal emission from the telescope and surroundings dominate the background emission. Therefore, the detector and camera must be designed to operate at cryogenic temperatures – a significant technical challenge. At shorter wavelengths the thermal background is less problematic but instead there is the supplementary frustration of very bright hydroxyl emission (Meinel 1950; Dufay 1951), resulting from the vibrational decay of an excited OH-radical, formed from the combination of a hydrogen atom with an ozone molecule. Furthermore, the intensity of the OH emission is highly variable (see Section 2.1.2 below), making accurate sky subtraction very difficult (e.g. Davies 2007).

The intensity and variability of the NIR background have traditionally restricted the depth of NIR observations compared to optical observations. However, two attributes of the NIR background spectrum suggest that, in principle, it should be possible to obtain very deep observations in the NIR. First, the OH spectrum comprises nearly all ($\approx$99.9 per cent from our model, see Section 2) of the *J*- and *H*-band background. The interline continuum, due mainly to zodiacal scattered light, is expected to be very faint, for example, the models of Kelsall et al. (1998) give a flux density of only $\approx$77 photons s^{-1} m^{-2} arcsec^{-2} μm^{-1} on the ecliptic plane. Secondly, the OH lines are intrinsically very narrow (Turnbull & Lowe 1983), meaning that large sections of the NIR sky spectrum will be very dark if the OH lines can be efficiently removed, thus enabling deep observations.

[3] http://www.eso.org/observing/etc/bin/ut3/conica/script/eltsimu

There are several methods which can, in principle, exploit the dark interline continuum for deep NIR observations, including narrowband filters which allow observations of a very short bandpass between lines (e.g. DAzLE; Horton et al. 2004), or space-based telescopes [e.g. James Webb Space Telescope (JWST); Gardner et al. 2006]. Here, we discuss ground-based OH suppression, by which we mean the selective removal (by means of low transmission, reflection etc.) of very narrow regions of the spectrum corresponding to the OH lines, whilst maintaining high throughput in the interline regions. There have been several methods proposed to achieve OH suppression, including spectroscopic masking at high dispersion (e.g. Maihara & Iwamuro 2000), rugate filters (Offer & Bland-Hawthorn 1998) and multinotch holographic filters (Blais-Ouellette et al. 2004). We discuss these methods in Section 5.1; each offers potential improvements over natural conditions, but in practice there are problems in achieving *efficient* OH suppression by any of these means. By efficient we mean achieving high suppression of the lines whilst maintaining high interline throughput, for a large number of lines ($\gtrsim$100). In this paper, we make the case for using fibre Bragg gratings (FBGs, see Bland-Hawthorn, Englund & Edvell 2004) to achieve OH suppression (Section 5.2); FBGs do not suffer from any of the drawbacks that have prevented successful application of the methods mentioned above. There is now a real case for believing that highly efficient OH suppression is close to being demonstrated.

Before discussing methods of OH suppression, we first review the background sources contributing to the NIR in Section 2. We then address the scattering properties of dispersive optics (Section 3), the proper understanding of which is essential to achieve efficient OH suppression. The study of scattering shows that the OH lines must be suppressed prior to dispersion; the combination of the extremely bright OH lines and the strong scattering wings, that are an unavoidable consequence of any diffraction system, make attempts to remove or 'resolve out' the OH skylines after dispersion futile. We then make the case for OH suppression, that is, selectively removing the OH lines prior to dispersion (Section 4), including a review of previously proposed methods (Section 5.1), and their limitations. We then show that FBGs offer the prospect of cleanly and efficiently removing OH lines photonically (Section 5.2). We calculate the required level of suppression and spectral resolution to achieve scientifically useful observations. We summarize our case in Section 5.4, and conclude with some remarks on the future in Section 6.

2 THE NEAR-INFRARED BACKGROUND

We now consider the various elements that contribute to the NIR background, focusing on the wavelength range 0.9–2.0 μm, as the region in which OH suppression will be most effective. Following Tokunaga in chapter 7 of Cox (2000), we divide the NIR background into the following components: the OH airglow; the thermal emission from the ground-based telescope; the atmospheric thermal emission; the zodiacal scattered light; zodiacal emission; Galactic dust emission; and the cosmic microwave background radiation. Each of these will be discussed in turn. Our final model of the NIR background spectrum is shown in Fig. 9.

2.1 Hydroxyl emission lines

2.1.1 Emission mechanism

The OH emission (Meinel 1950; Dufay 1951) is by far the largest component of the NIR background at $\lambda \lesssim 2.0$ μm. At longer
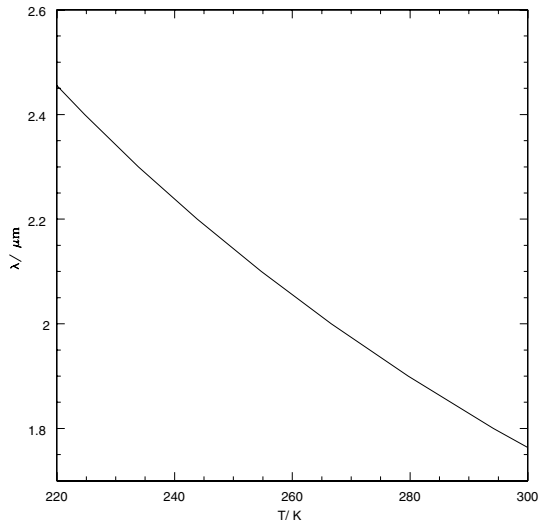
Figure 2. A lower limit on the wavelength at which the OH emission dominates the NIR background as a function of environmental temperature. This was calculated as the wavelength at which the blackbody spectrum from a typical telescope exceeds that of the zodiacal scattered light at $\beta = 60°$.

wavelengths, the background is dominated by thermal emission, from the telescope and dome, etc. The exact wavelength where the thermal emission becomes dominant depends on the temperature of the environment, as illustrated in Fig. 2, which shows the wavelength at which the blackbody spectrum from the telescope exceeds the blackbody spectrum of the zodiacal scattered light, that is, a lower limit to the wavelength at which OH emission dominates the background. Thus, for very cold sites, such as Dome C in Antarctica with an average year round temperature of ≈ 220 K (see Burton et al. 2005) OH suppression would be effective at wavelengths $\lambda \lesssim 2.4 \, \mu m$.

The OH emission arises mainly from the combination of hydrogen and ozone from a layer about 9 km thick at an altitude of around 87 km. This reaction leaves a rotationally and vibrationally excited OH molecule (and an oxygen molecule) with a radiative lifetime of a few milliseconds (Content 1996, and references therein). The emission then results from the vibrational decay of the molecule, that is,

$$H + O_3 \rightarrow OH^* + O_2. \tag{1}$$

The intrinsic width of the lines is determined from the Doppler shift due to the thermal motions of the molecules, from pressure broadening, and from natural broadening. The former effect dominates the inner width of the line and has a Gaussian form, whereas the latter has a Lorentzian profile, giving the lines broad wings, albeit very faint ones.

The Doppler width can be estimated from the following argument. For a gas in thermal equilibrium, the gas particles will have a distribution of speeds, but the most-probable speed will be when the kinetic energy, $mv^2/2$, is equal to the thermal energy kT. Thus, the light emitted from the gas particle will be Doppler-shifted by an amount $\Delta\lambda/\lambda = \pm|v_r|/c$. Thus,

$$\Delta\lambda_{DB} = \frac{2\lambda}{c}\sqrt{\frac{2kT}{m}}, \tag{2}$$

where m is the mass of the particles, and T is the gas temperature. Thus, for OH molecules at a temperature of $T = 200$ K (Rousselot et al. 2000), $\Delta\lambda_{DB} \approx 5 \times 10^{-6} \, \mu m$.

Pressure broadening is the result of interactions or collisions with neighbouring particles in the gas. The effect can be estimated by considering the average time between collisions or encounters, Δt_0, which is approximately equal to the mean free path divided by the average speed of the particles, that is,

$$\Delta t_0 \approx \frac{l}{v} = \frac{1}{n\sigma\sqrt{2kT/m}}, \tag{3}$$

where σ is the collisional cross-section and n is the number density. It is assumed that the shape of the line is Lorentzian, which is the result of an emitting atom undergoing simple harmonic motion. Substituting equation (3) into equation (7) yields

$$\Delta\lambda_{PB} = \frac{\lambda^2 n\sigma_{cross}}{c\pi}\sqrt{\frac{2kT}{m}}. \tag{4}$$

We adopt $\Delta_{\lambda_{PB}} \approx 3 \times 10^{-7} \, \mu m$, an upper limit based on the difference in wavenumber of $<0.002 \, cm^{-1}$ quoted by Turnbull & Lowe (1983).

Natural broadening is due to the Heisenberg uncertainty in the particles energy,

$$\Delta E = \frac{\hbar}{\Delta t}, \tag{5}$$

where Δt is the duration of the change in energy, such that,

$$\Delta\lambda_{NB} \approx \frac{\lambda^2}{2\pi c\Delta t}. \tag{6}$$

The full calculation shows that the full width at half-maximum (FWHM) is in fact twice this value,

$$\Delta\lambda_{NB} = \frac{\lambda^2}{\pi c\Delta t}. \tag{7}$$

Thus, for a vibrationally excited state lasting a few milliseconds, the natural broadening is $\approx 5 \times 10^{-13} \, \mu m$ at $\lambda = 1 \, \mu m$.

From the above estimates, the Doppler broadening is by far the most-dominant contribution to the intrinsic width of the line. However, due to the much broader wings of the Lorentzian pressure broadening, this mechanism is not negligible, as it will contribute much more to the region between the line peaks. The natural broadening can be safely neglected. Therefore, in modelling the OH emission line spectrum we include both the Doppler broadening and the Lorentzian profile, that is, we use a Voigt profile

$$V(\sigma, \gamma, x) = \int_{-\infty}^{+\infty} \frac{e^{(-x'^2/2\sigma^2)}}{\sigma\sqrt{2\pi}} \frac{\gamma}{\pi[(x-x')^2 + \gamma^2]}dx', \tag{8}$$

where $\sigma = \frac{\Delta\lambda_{DB}}{2\sqrt{2\ln 2}}$ and $\Delta\lambda_{DB}$ is the FWHM of the Doppler broadened line given by equation (2), and γ is the half-width at half maximum of the pressure broadened line given by $\frac{\Delta\lambda_{PB}}{2}$ in equation (4).

In our models, we assume a temperature at 87 km of 200 K and $\gamma = 1.5 \times 10^{-13}$ m. The resulting Voigt profile is shown over a range of $\pm 10^{-3} \, \mu m$ in Fig. 3 for the brightest doublets in the H band (which are not resolved). The plot is logged to emphasize the Lorentzian wings which are significant out to $\approx 5 \times 10^{-4} \, \mu m$ either side of the line-centre (cf. zodiacal scattered light described below). However, in practice, the interline continuum is dominated by instrumental scatter. We include this effect in our model, but in practice linewidths are dominated by the instrumental profile (see Section 3) and doublet separations.

Note that although the intrinsic width of any individual line is extremely narrow, as shown in Fig. 3, the OH spectrum is actually composed of close doublets. The doublets are of equal intensity (Osterbrock et al. 1996) and arise due to Λ-type doubling (see Herzberg 1939) from the interaction between the rotation of the O
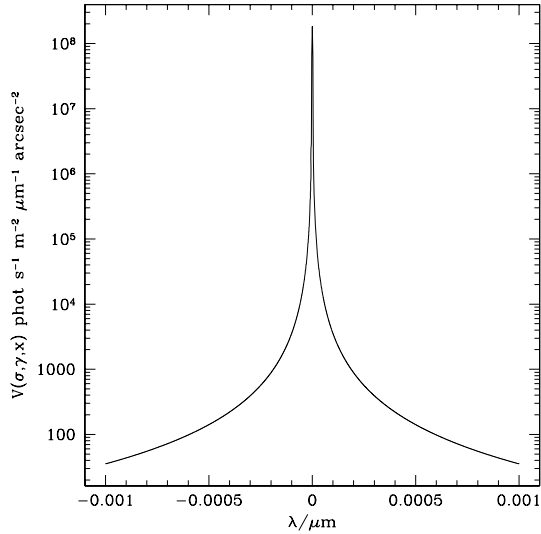
Figure 3. The theoretical logged Voigt line-profile due to Doppler and pressure broadening, of a very bright OH line with an integrated flux of $740 \, \gamma \, \mathrm{s}^{-1} \, \mathrm{m}^{-2} \, \mathrm{arcsec}^{-2}$ assuming $T = 200 \, \mathrm{K}$ and $\gamma = 1.5 \times 10^{-13} \, \mathrm{m}$. The Lorentzian wings are significant out to at least 1 Å from the line-centre.
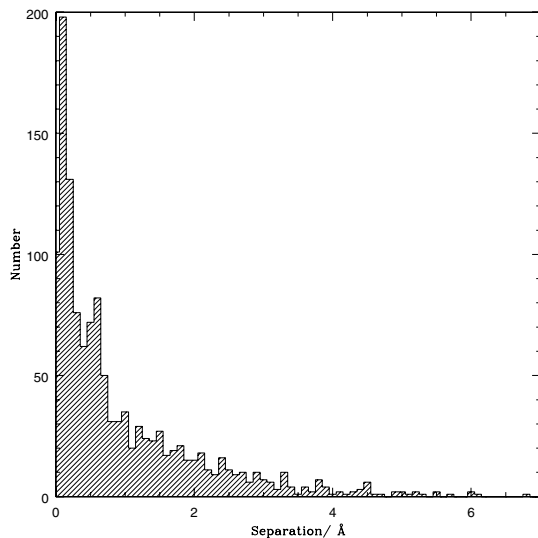

Figure 4. The distribution of doublet separations for OH lines with wavelength 0.9–1.8 µm.

and H nuclei and the orbital angular momentum of the electron. The practical linewidth is determined by the separation of these doublets. Using the models of Rousselot et al. (2000), we have determined the doublet separation for all lines in the wavelength range $0.9 \leqslant \lambda \leqslant 1.8 \, \mu\mathrm{m}$, shown in Fig. 4. The distribution is strongly peaked at separations of 0.1–0.2 Å, which gives a practical limit to the linewidths. Note though that we model the OH spectrum by treating each line individually.

2.1.2 Variability

The OH emission is not only extremely bright, but it is also highly variable in the spatial and temporal domain. This has significant consequences for the removal of the night sky background from observations. Despite its strong variability a few trends with time
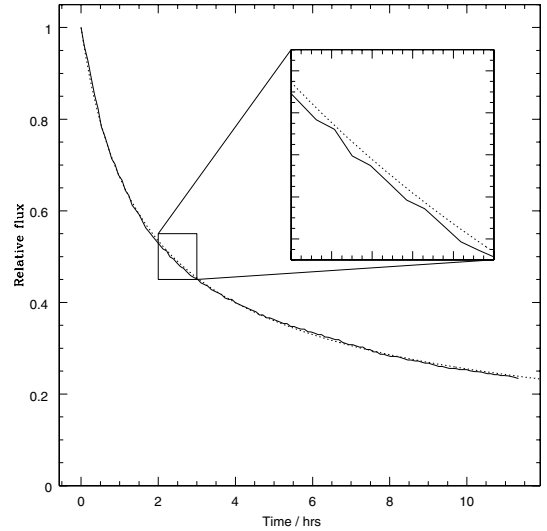

Figure 5. The night time variation in the model of Shimazaki & Laird (1970) by the continuous line, and our analytic approximation by the dotted line.

and space are known. The intensity of the lines increases with zenith distance due to the longer path length through the emitting layer of the atmosphere. The intensity changes as

$$\frac{I(\theta)}{I(0)} = \left[1 - \left(\frac{r \sin\theta}{r + h} \right)^2 \right]^{-1/2}, \tag{9}$$

where θ is the zenith angle, $I(\theta)$ is the intensity at θ, r is the radius of the Earth and h is the altitude of the emitting layer above the ground (Content 1996).

The temporal variation is due to two main causes: diurnal variation in temperature, and density waves in the upper atmosphere.

The diurnal variation has been modelled by Shimazaki & Laird (1970), taking into account reactions between 14 molecular species and diffusion. We show their model in Fig. 5, and summarize their description of the variations. The features result from two emission layers, a broad layer at 45 km altitude and a narrow layer at 80 km. The height of the lower layer rises in the afternoon, and partially disappears around sunset, causing a dip in the OH emission. About 20 min after sunset, both layers rise and the intensity of the emission is greatly enhanced due to an increase in the ozone content at sunset; this causes the sudden rise in the models. After sunset, the lower layer disappears completely because of the removal of hydrogen, and the upper layer decreases slowly due to the decrease in O_3. There is also a rapid dimming and brightening around sunrise. This model is consistent with observations, which show a decrease by a factor of ~ 2 in the OH emission throughout the night (Ramsay, Mountain & Geballe 1992; Content 1996; Glazebrook & Bland-Hawthorn 2001). The model is well approximated by the function $1/(1 + t)^{0.57}$, where t is the time in hours, as shown in Fig. 5. We assume that a typical nights observing lasts for 8 h, so for observations longer than this we reset the OH brightness every 8 h.

Besides this long-term decline in OH line-strength throughout the night, there is also short-term variability caused by density waves (Ramsay et al. 1992; Frey et al. 2000). These are typical periodic variations of length 5–15 min, with an amplitude of about a tenth of the flux.

We assume that all OH lines vary in unison for our simulations. In fact, the real situation is more complicated; lines within the same transition will be correlated, but separate transitions will not.
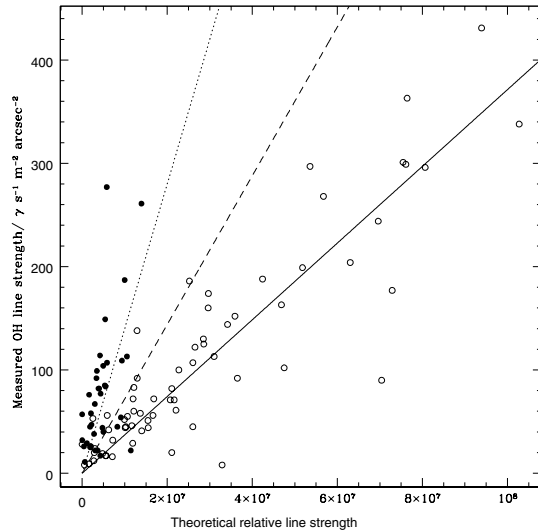
Figure 6. The correlation between the theoretical relative line-strengths of Rousselot et al. (2000) and the measured line-strengths of Maihara et al. (1993b). The closed symbols are for the *J* band and have a slope of 1.4 × 10^{-5}. The open symbols are for the *H* band, with a slope of 3.7 × 10^{-6}. The dashed line between the *J*- and the *H*-band correlations is the geometric mean of 7.2 × 10^{-6}, which was used to calibrate the Rousselot list.

However, to a reasonable approximation, lines from a particular transition all lie within the same region of the spectrum, with very little contribution from other transitions (Davies 2007). Hence the assumption of unison will be true for sections of spectrum ∼0.1 µm long.

2.1.3 Line list

Rousselot et al. (2000) have modelled the emission spectrum of the night-sky OH lines, giving accurate wavelengths for 4732 lines, and approximate relative intensities. In order to use this atlas, we have normalized the theoretical relative intensities to the fluxes measured by Maihara et al. (1993b) (after first removing the O_2 lines). The lines were measured with a resolving power of $R \approx 17\,000$ and remained unresolved. Since the OH lines are close doublets, we combine all theoretical lines which have a separation of $\Delta\lambda < \lambda/17\,000$, before comparing the theoretical and observed line-strengths. Fig. 6 shows the correlation between the theoretical line-strengths and the measured line-strengths. There is a significant difference between the *J*- and *H*-band data, and the *H*-band data show a stronger correlation. This is indicative of the transient nature of the OH lines. We have used the geometric mean of the *J*- and *H*-band correlations to calibrate the Rousselot data.

2.2 Other lines

There are a number of O_2 and N_2 lines in the *J* band that are not included in the models of Rousselot et al. (2000). These lines are self-absorbing and dim rapidly after sunset (Content 1996). We have included line-strengths for these taken directly from Maihara et al. (1993b). For simplicity, we have assumed that they are broadened according to exactly the same prescription as the OH lines.

2.3 Thermal emission from the telescope

There will be significant thermal emission from the telescope and the instrument employed. The instrument design can be optimized
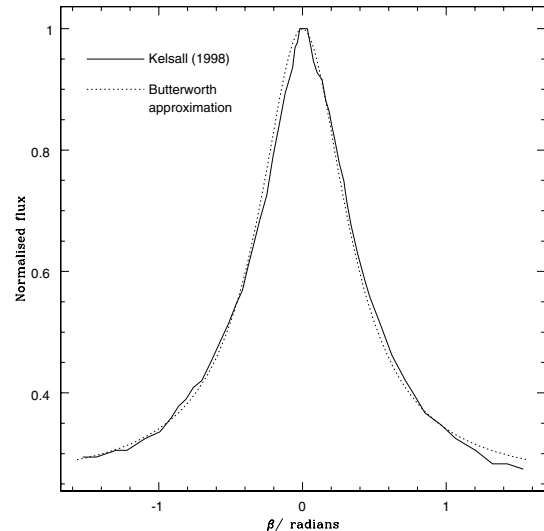

Figure 7. The variation in zodiacal light with ecliptic latitude from Kelsall et al. (1998) shown by the continuous line and our model as given by equation (12) shown by the dotted line.

to cool as many parts as possible thereby minimizing its contribution to the background, but some parts, for example, optical fibres may be external to the main instrument dewar and therefore unable to be cooled. Even so, the main contribution will come from the much larger telescope, and we model the emission as a blackbody spectrum at $T = 273$ K, with an emissivity of $\epsilon = 0.02$ (chapter 7, Cox 2000), that is,

$$B_\lambda = \frac{2hc^2}{\lambda^5(e^{hc/k\lambda T} - 1)},\tag{10}$$

which can be expressed as a surface brightness,

$$\begin{aligned} N_\phi &= \epsilon\lambda_{\mu m}\frac{B_\lambda}{hc}\left(\frac{\pi}{180 \times 3600}\right)^2 \\ &\approx \frac{1.41 \times 10^{16}\epsilon}{\lambda_{\mu m}^4(e^{14\,387.7/\lambda_{\mu m}T} - 1)}, \end{aligned}\tag{11}$$

in units of photons s^{-1} m^{-2} μm^{-1} $arcsec^{-2}$.

2.4 Zodiacal scattered light

Zodiacal light is emission from sunlight scattered by interplanetary dust. As such it has a solar spectrum, multiplied by the scattering efficiency as a function of wavelength. Individual measurements in the optical (e.g. Levasseur-Regourd & Dumont 1980) and the NIR (e.g. Noda et al. 1992; Matsumoto et al. 1996) seem to show a solar spectrum when considered individually, but have different normalizations when compared, with the NIR zodiacal spectrum being twice as bright due to the increased scattering efficiency off larger dust particles at longer wavelengths (Matsumoto et al. 1996).

The brightness of the zodiacal light obviously decreases as a function of ecliptic latitude. Kelsall et al. (1998) have modelled the zodiacal light in detail in order to remove this component when analysing data from the *COBE* Diffuse Infrared Background Experiment (DIRBE). We approximate the Kelsall model with a Lorentzian function, which enables a faster calculation whilst maintaining most of the accuracy. In Fig. 7, we show their model's dependence on ecliptic latitude, β, for 1.25-µm emission and our fit to it

Table 2. Measurements and model predictions of the zodiacal scattered light compiled from the literature.

| Source | Measurement or model | Flux/γ s^{-1} m^{-2} arcsec^{-2} | | Ecliptic latitude |
		J	H	($°$)
Maihara et al. (1993b)	Measurement	–	590	–
Cuby, Lidman & Montou (2000)	Measurement	1200	2300	–
Thompson et al. (2007) (NICMOS)	Measurement	72	150	−45 (Hubble Deep Field)
Thompson et al. (2007) (NICMOS)	Measurement	74	220[a]	57 (*Hubble Deep Field*-North)
Thompson et al. (2007) (IRTS NIRS)	Measurement	–	110	–
Kelsall et al. (1998)	Model[b]	43	–	46 (Lockman Hole)
Kelsall et al. (1998)	Model	99	–	0
NICMOS handbook[c]	Model	130	77	0
NICMOS handbook	Model	50	31	45
Stockman et al. (1998) JWST	Model	–	3	1-au orbit

[a]Includes instrumental background. [b]Model based on *COBE* DIRBE measurements. [c]From an STScI report by M. Stiavelli, see http://www.stsci.edu/hst/nicmos/documents/handbooks/current_NEW/c04_imaging8.html#309385.

given by

$$I = \frac{0.75}{\left[1 + \left(\frac{2\beta}{0.743}\right)^2\right]} + 0.25, \qquad (12)$$

where β is in radians.

For the majority of astronomical applications of FBGs, we anticipate a preference for high *galactic* latitudes, and therefore a wide range of ecliptic latitudes depending on the galactic longitude, which could change the zodiacal brightness by a factor of ∼2.

For the purposes of the model, we normalize a solar spectrum (using equation 11 with $\epsilon = 1.08 \times 10^{-13}$ and $T = 5800$ K) to the Kelsall et al. (1998) model value for the ecliptic plane. Several measurements and models of the zodiacal scattered light exist in the literature. These are compiled in Table 2. Note that while the models predict the zodiacal scattered light should be fainter at longer wavelengths (since it has the form of a $T = 5800$ K blackbody) the measured data disagree with this. Note too that the ground-based measurements are rather higher than the models. In Section 3, we suggest that for the ground-based measurements scattering of OH light may have a significant contribution to these values.

2.5 Moonlight

Moonlight is known to be an insignificant source of background with conventional NIR observations. This is because the OH skylines dominate over scattered moonlight and also because Rayleigh and Mie scattering are both less efficient at longer wavelengths; Rayleigh scattering has a λ^{-4} dependence and Mie scattering has a $\lambda^{-1.3}$ dependence. However, in the regime of OH-suppressed observations these assumptions need to be re-evaluated.

Krisciunas & Schaefer (1991) present a model of the brightness of scattered moonlight that is accurate to 23 per cent. However, this model is only appropriate for observations in the *V* band. Therefore, we have adopted the procedure described in the CFHT Redeye Manual[4] to adapt the model to NIR wavelengths. The principal changes are allowing for the change in colour assuming a solar spectrum, and allowing for the wavelength dependence of the scattering function. We show the results of the model in Fig. 8. Comparing to estimates of the zodiacal scattered light shows that for observation at a full, or nearly full, moon, the scattered lunar light would be significant out to ≈50° from the moon. OH suppressed observations

[4] See   http://www.cfht.hawaii.edu/Instruments/Detectors/IR/Redeye/Manual/chapter7.html

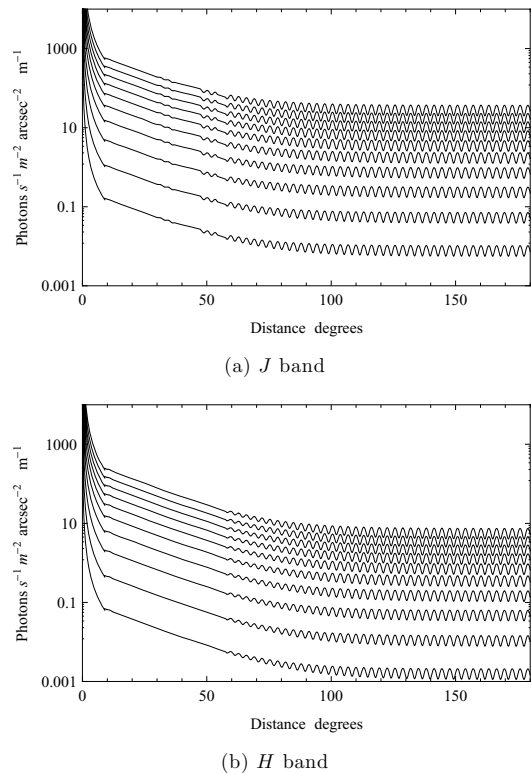

(a) J band

(b) H band

Figure 8. The brightness of scattered moonlight as a function of angular distance away from the moon. The different curves are for different lunar phases from 0 to 180 in steps of 20°. It is assumed that the moon is at the zenith.

should therefore not be performed around the full moon, though 'grey nights' should be acceptable. We note that the effect of scattered moonlight could have had an effect on previous attempts to measure the interline continuum, if these were performed in bright time and the contribution due to moonlight was not taken into account.

2.6 Other components

There are a number of other components that make important contributions to the IR background at longer wavelengths ($>2\,\mu$m), but are negligible in the regime considered here. These are: the
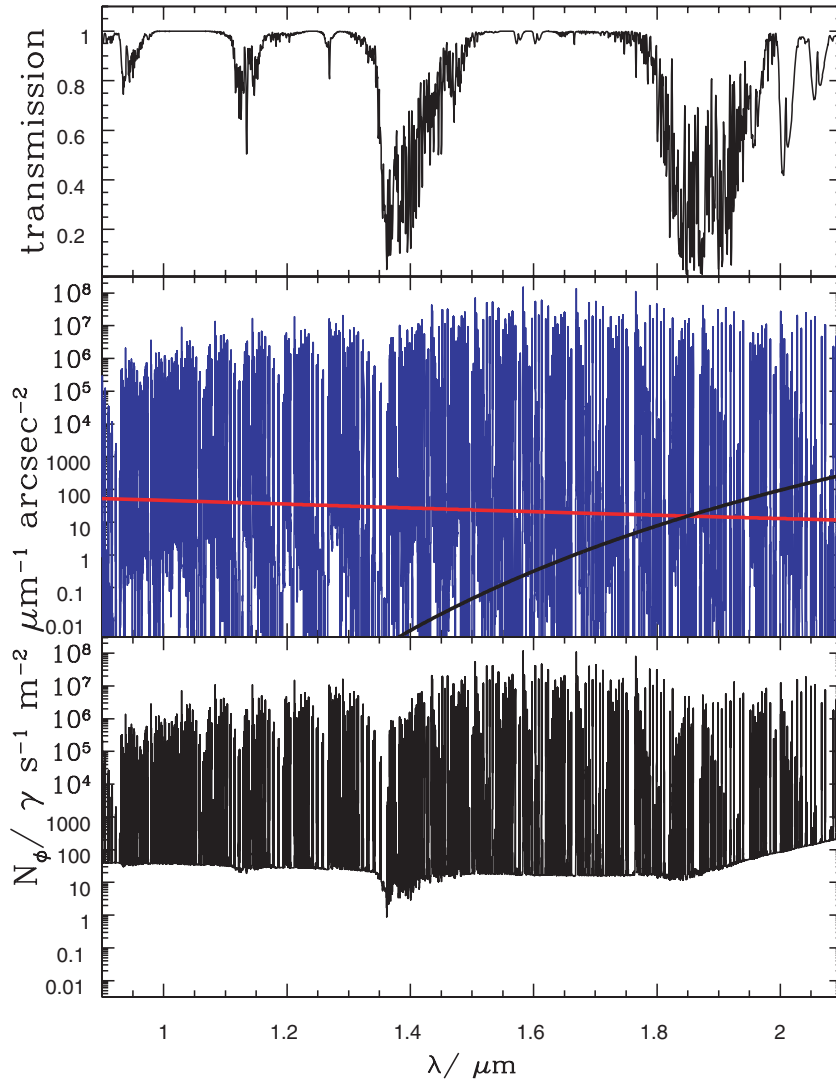
Figure 9. The NIR background model. The top panel shows the atmospheric transmission. The middle panel shows the contributions from the OH spectrum (blue), the zodiacal scattered light (red), and the telescope thermal emission (black). The bottom panel shows the final resulting spectrum.

thermal contribution from the atmosphere, zodiacal emission from interplanetary dust at the ecliptic pole, Galactic emission from from interstellar dust, and the cosmic background radiation.

It is possible that there is also a faint continuum or pseudo-continuum composed of line emission from NO_2 or other atmospheric models (Content 1996). Such emission could perhaps explain some of the discrepancy between measurements of the zodiacal scattered light and the ground-based measurements of the interline continuum. However, there are no reliable measurements of this continuum, since previous estimates do not take into account the instrumental scattering profile (Section 3), therefore we do not include any such continuum in our model.

2.7 Atmospheric transmission

The dominant absorption features of NIR light are those due to water, carbon dioxide, nitrous oxide, methane and ozone in the atmosphere, which define the *J*-, *H*- and *K*-band photometric windows. We have used the model transmission profile at Mauna Kea assuming an airmass of 1.0 and a water vapour column of 1.6 mm in our

default sky background model (Lord 1992, obtained from Gemini Observatories[5]; see Fig. 9).

3 SCATTERING OF DISPERSIVE OPTICS

Because the OH lines are so bright compared to the interline continuum (see Fig. 9), they cannot be cleanly removed through high dispersion spectroscopy. This is because even a small amount of scattered OH light will swamp the true interline continuum. To see this, consider the example of a ruled diffraction grating. The theoretical normalized intensity distribution of a grating with *N* lines is given by

$$I = I_0 \left(\frac{\sin\beta}{\beta} \right)^2 \left(\frac{\sin N\gamma}{N\sin\gamma} \right)^2, \tag{13}$$

[5] http://www.gemini.edu/sciops/ObsProcess/obsConstraints/ocTrans Spectra.html

where

$$\beta = \frac{kb}{2}\sin\theta, \qquad (14)$$

$$\gamma = \frac{kh}{2}\sin\theta, \qquad (15)$$

and k is the wavenumber, b is the slit width, h is the slit separation, and θ is the angle through which the light is diffracted.

Woods et al. (1994) show that equation (13) may be represented as a Lorentzian plus a constant background:

$$Y_{\text{fit}} = \frac{\omega^2}{(\lambda - \lambda_0)^2 + \omega^2} + B, \qquad (16)$$

where Y_{fit} is the ratio of intensity at wavelength λ to intensity at λ_0, B is a constant background term due to Rayleigh scattering and

$$\omega = \frac{\lambda_0}{N_e \pi \sqrt{2}}, \qquad (17)$$

and it is assumed that $\lambda - \lambda_0 \ll \lambda_0$ and $N \gtrsim 100$. Equation (16) is normalized by a factor of $1/(\pi\omega + 2hB)$, where h is the groove spacing.

The most important factor in determining the line contribution is N_e the number of *effective* lines. In theory, the number of effective lines should be the number of lines illuminated. However, Woods et al. (1994) find that N_e is always measured to be much lower than this theoretical limit. In practice, N_e can be taken to be the smallest number of consecutive grating grooves with no imperfections. Woods et al. (1994) report values that are typically 0.25–0.5 times the theoretical upper limit, but occasionally as low as 0.16. Farther from the line-centre, the constant background A_B also becomes important.

Woods et al. (1994) fit profiles to 10 different types of gratings including holographically and mechanically ruled gratings and find that the scattered light is indeed well fitted by a Lorentzian plus constant. Our own measurements with IRIS2 (a grism spectrometer) and AAOmega (a VPH spectrometer) arc spectrum confirm this; we find that the spectral line-profile can be fitted by a function

$$I = I_0 \left\{ \frac{e^{-(\lambda - \lambda_0)^2/2\sigma^2} a}{\sqrt{2\pi}\sigma} + \frac{(1-a)f\sigma\sqrt{2\log(2)}}{\pi[(\lambda - \lambda_0)^2 + 2f^2\sigma^2\log(2)]} \right\}, \qquad (18)$$

which is simply a Gaussian plus a Lorentzian, with a contribution of factor a from the Gaussian and $1 - a$ from the Lorentzian, and the Lorentzian has FWHM f times larger than that of the Gaussian. We find typical values of $f = 4$–6 and $a = 0.89$–0.97. In Fig. 10, we show a fit to the IRS2 and AAOmega arc spectra.

We now show the effects of these scattering wings on the OH spectrum, and in doing so we address the large discrepancy between the ground-based observations of the interline continuum listed in Table 2 and the model values. We show the effect of scattering on the OH lines around $\lambda = 1.665\,\mu\text{m}$, where Maihara et al. (1993b) measured the interline continuum.

We have modelled the scattering from the University of Hawaii 2.2-m Coudé spectrograph, with the 600 lines mm^{-1} grating, and $\lambda_{\text{blaze}} = 1.32\,\mu\text{m}$ used for the *H*-band observations by Maihara et al. (1993b). We have assumed $B = 10^8$, the value for the Bausch & Lomb grating used by Woods et al. (1994).

We use equation (16) to model the scattering profile as it can be normalized exactly. Note that this will provide a slightly conservative estimate of the scattering far from the line-centres, as the wings in equation (16) drop off faster than in the full equation.

In Fig. 11, we show the scattered light from all *H*-band lines reported in Maihara et al. (1993b) in the window 1.660–1.675 (cf. their
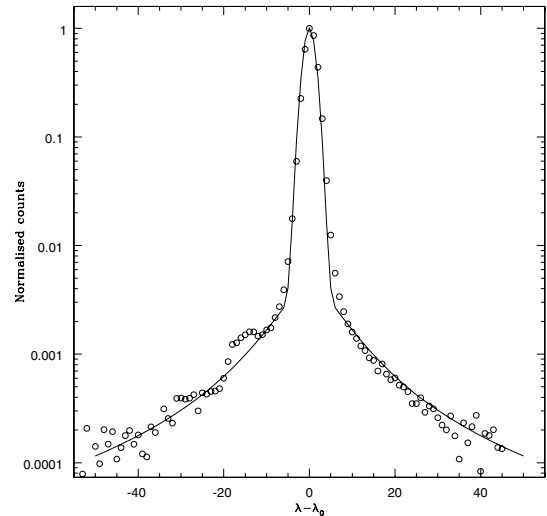
(a) AAOmega, $a = 0.97$ and $f = 5.5$.
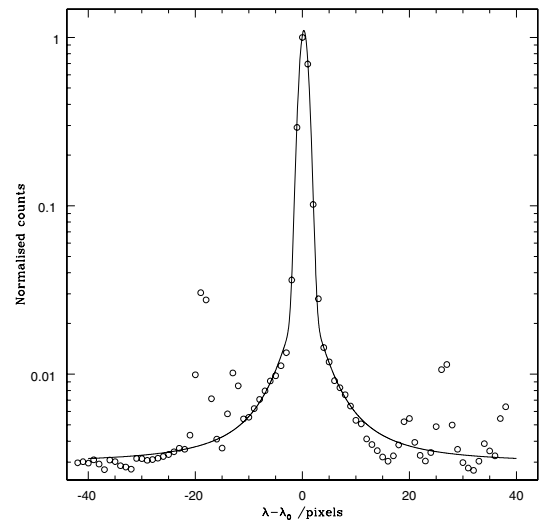
(b) IRIS2, $a = 0.89$ and $f = 4.0$.

Figure 10. The line-profiles of two spectrographs, AAOmega, a VPH spectrograph and IRIS2, a grism spectrograph. The measured data are shown by the open circles, and a fit of the form of equation (18) is shown by the line.
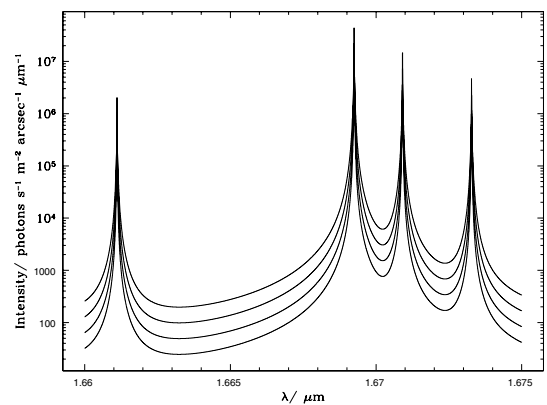

Figure 11. The scattered light due to all measured OH lines in the *H* band. The different curves are for values of $N = 120\,000, 60\,000, 30\,000$ and $15\,000$. The scattered wings are higher for fewer lines.
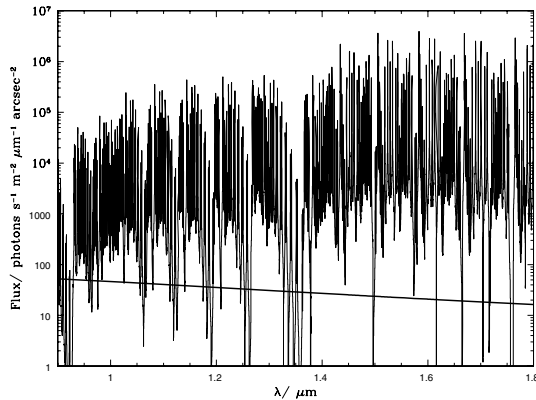
Figure 12. The relative strength of the OH spectrum at $R = 10\,000$ (thin line) compared to the zodiacal scattered light at $\beta = 60°$ (thick line). The scattered OH line light reaches the interline continuum for 11 per cent of the spectrum.

fig. 2) whence the continuum emission was measured. It is unclear what the theoretical value of N_e should be, that is, the number of illuminated lines, but it will be much less than 120 000 (e.g. Woods et al. 1994). The real value of N_e, that is, the smallest number of consecutive lines without imperfections will be smaller again. Therefore, we show plots for four different values of N_e: for $N = 120\,000 = 20\,\text{cm} \times 600$ lines mm^{-1}, that is, the absolute maximum number of lines, and for values of $N_e = 60\,000$, $30\,000$ and $15\,000$. The scattered wings are higher for fewer lines. These estimates give a scattered light intensity of between 35 and 280 photons s^{-1} m^{-2} arcsec^{-2} μm^{-1} at 1.665 μm. Thus, it is possible that scattered light could account for a significant part of the difference between our predicted value of the continuum at $\lambda = 1.665$ μm and that measured by Maihara et al. (1993b), if the number of imperfections in the grating is high.

The scattering behaviour of dispersive optics thus prevents observations reaching the level of the interline continuum regardless of instrument background or spectral resolution. For example, in Fig. 12, we show the relative strengths of the OH spectrum with a resolution of $R = 10\,000$ (using a scattering profile of the form equation 18 with $a = 0.90$ and $f = 4$) and the interline continuum at an ecliptic latitude $\beta = 60°$ over the range $0.9 \leqslant \lambda \leqslant 1.8$ μm. The interline continuum is reached for only 11 per cent of the spectrum between $0.9 \leqslant \lambda \leqslant 1.8$ μm, and 3 per cent between $1.4 \leqslant \lambda \leqslant 1.8$ μm.

4 THE CASE FOR OH SUPPRESSION

By OH suppression we mean the selective removal of OH lines from the incident light whilst allowing light either side of the deleted line to pass through with high efficiency (as opposed to OH avoidance by narrow-band filters, for example, Horton et al. 2004, which selects light between OH lines, and is therefore restricted to narrow bandpasses). We discuss how this may be achieved using FBGs below (Section 5.2). First, however, we address the desired performance of a general OH suppression system.

The OH spectrum contains hundreds of bright lines (Rousselot et al. 2000; see Section 2). An OH suppression system must remove a section of finite wavelength, $\Delta\lambda$, from the incident spectrum for each line to be suppressed. After suppression, these regions will be unsuitable for any scientific analysis. Therefore, OH suppression comes with the penalty of reducing the useful bandpass of

the spectrum, which will vary with the number of lines suppressed and the size of the 'notches', $\Delta\lambda$. In practice, the regions around bright OH lines are not usually suitable for scientific analysis in any case, since the variation of the OH emission hampers the success of sky-subtraction and residual sky lines are usually observed.

In designing an OH suppression system, it is thus necessary to choose how many lines to remove and at what resolution in order to maximize the efficiency of the system. We now address this issue for notches fixed at $\Delta\lambda = 1$, 3 and 10 Å ($\lambda/\Delta\lambda \approx 10\,000$, 3000 and 1000), for varying numbers of lines. The lines are suppressed in order of their relative brightness, with the brightest lines removed first. We measure the efficiency of the system by the fraction of useable spectrum remaining and the brightness of the faintest line suppressed (as an estimate of the expected sensitivity of the observations). Note that in practice many OH lines are very close doublets, and therefore these can be removed with a single 1-Å notch.

Fig. 13 shows the results for up to 200 notches across the *J* and *H* bands. Even with 200 lines of width 1Å, suppressing down to a level of 4×10^{-3} photons s^{-1} m^{-2} arcsec^{-2} μm^{-1} (far fainter than would practically be necessary), only 8 per cent of the *J* band has been removed, and only 7 per cent for 200 lines in the *H* band. Thus, by virtue of the intrinsic slenderness of the OH-emission lines, suppression at high resolution is a highly desirable objective, maintaining a large fraction of scientifically useful wavelength range, whilst vastly increasing signal-to-noise ratio.

5 METHODS OF OH SUPPRESSION

There are several methods which have been proposed, and in some cases employed, to achieve OH suppression. However, to date no instruments exist which can suppress OH light at high efficiency whilst maintaining good interline throughput. In the next section, we will review proposed methods of OH suppression and address their limitations. This will be followed by a discussion of FBGs, in which we will argue that there is now a real case for believing that highly efficient OH suppression will finally be acheived.

5.1 Previous methods of OH suppression

5.1.1 High dispersion masking

Maihara et al. (1993a) present a method of OH suppression using high dispersion and a ruled mirror. The basic idea is to disperse the light at high resolution, whereupon it is reflected from a mirror-mask, which has a series of black lines coincident with the position of the OH lines, which are therefore not reflected. The reflected light, with the OH lines removed, then proceeds back through the camera and grating system in the opposite direction to its initial journey, and hence is recombined into a 'white-light' beam. This final beam may then be dispersed at low resolution as required.

This method of mechanically removing the OH light and recombining is preferable to using software to achieve the same result, for example, by removing the OH lines, and then rebinning the spectrum, because the read-noise becomes much more significant for a highly dispersed spectrum (Iwamuro et al. 1994).

Spectrographs employing this technique have been built and used on the Subaru telescope, for example, OHS + CISCO (Iwamuro et al. 2001; Motohara et al. 2002), and on the University of Hawaii 2.2-m telescope (Iwamuro et al. 1994). These instruments succeeded in reducing the cores of the OH lines to ≈ 4 per cent of their original strength. However, it is the reduction in the background between the lines that is important in assessing the merit of an OH suppression
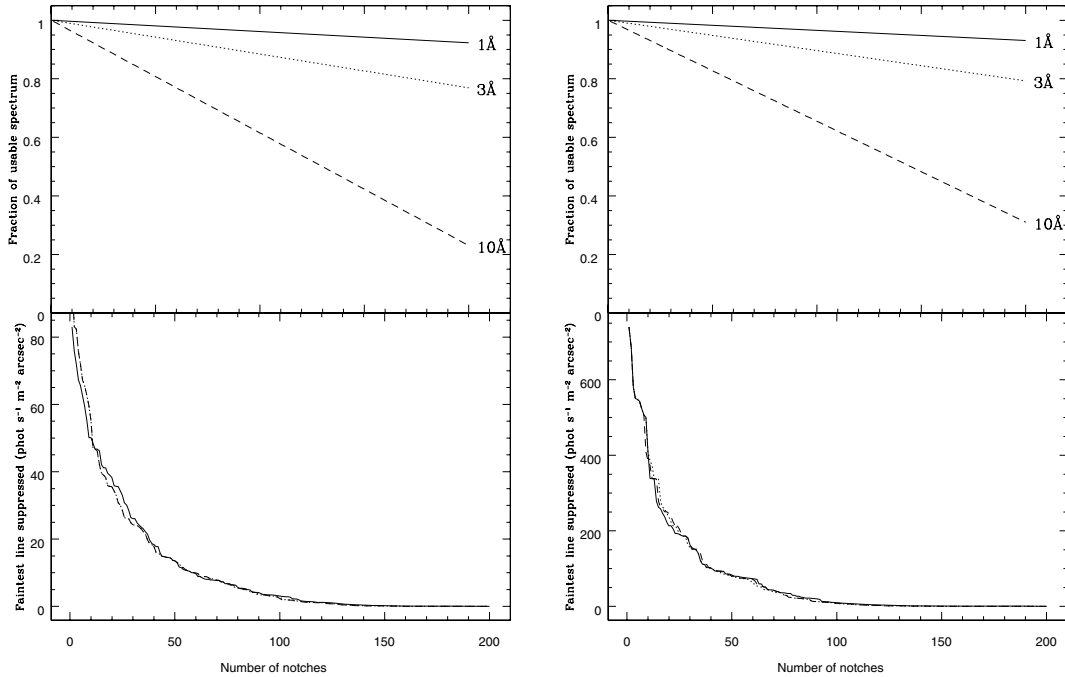
Figure 13. The effect of increasing the number of OH suppressing notches on the incident spectrum. The left-hand panels show results from the *J* band and the right-hand panels show that of the *H* band. The top panels show the fraction of the total passband that remains usable after suppression. The bottom panels show the flux of the faintest line suppressed.

system, since this is where scientific measurements will be made. Thus, the fundamental limit to the performance of OHS-style suppressors is the scattering of the OH light. This process was described in detail in Section 3. Fig. 11 illustrates the problem: the Lorentzian scattering wings prevent the interline continuum ever being reached. This is true even at high resolution as shown in Fig. 12. Furthermore, in a mirror-mask system the problem of scattered light is compounded, since scattering will occur for each pass through the grating system, and from diffraction from the mirror-mask itself. The scattered wings of the OH line remain in the spectrum since only the core of the line is suppressed with the mirror mask; between the lines the gains will be much less. To realize efficient OH suppression, this scattering problem must be avoided; the OH light must be removed prior to dispersion.

The same principle of a mirror-mask design is being used in Fibre Multiobject Spectrograph (FMOS; Maihara et al. 2000), which should begin use on the Subaru telescope in 2008. The performance of the OH suppressing part of the FMOS instrument is expected to be similar to OHS, but with the advantage of multiplexed observations and improved spectrographs.

5.1.2 Rugate filters

Rugate filters have been proposed by Offer & Bland-Hawthorn (1998) as a method of OH suppression applicable to imaging. Rugate filters consist of a transparent substrate on to which is deposited a thin film of dielectric material. By varying the composition of the film as it is deposited, the refractive index can be made to vary throughout the coating. By controlling the refractive index variation, any arbitrary transmittance profile may be designed, for example, by combining various sinusoids using Fourier theory.

The chief obstacle in realizing OH suppression with rugate filters has been the difficulty in manufacturing. Deposition of each layer

of the film requires very tight tolerance in the thickness of the layer, which must be uniform over the entire area of the substrate. These difficulties have so far made manufacturing of satisfactory filters impracticable. A further problem is that even should the tolerance of each layer be met, thousands of layers are required which leads to a coating which is tens of μm thick. The stresses in each layer of so many coatings ultimately makes each layer unstable (Ennos 1966).

5.1.3 Holographic filters

Blais-Ouellette et al. (2004) suggest the use of volume-phase holographic filters (Barden et al. 2000) as a method of OH suppression. These filters use holographic recording techniques to record a grating consisting of fringes of varying refractive index in a volume of transparent material, typically dichromated gelatin. If the fringes are arranged to be parallel to the surface of the filter, no dispersion occurs, and the grating acts as a notch filter. By controlling the fringe pattern, several notch filters may be multiplexed within a single filter. Blais-Ouellette et al. (2004) demonstrate a filter capable of suppressing 10 lines (or close doublets) at 10 dB, with a notch width of 1 Å and an interline throughput of 85 per cent.

These filters have the advantage as a viable method of OH suppression in that they are presently available for purchase from Photon Etc. However, there are some difficulties in their use in a practical situation. The wavelength response of the filter profile shifts as a function of off-axis angle, which restricts the use of the filters to slow beams. In order to use the filters in slow beams, the notch width must be made larger to accommodate the shifting filter profile, reducing their efficiency. Another issue is that a single filter is made up of several glass substrates which are cemented together such that the layers between each substrate align with the optical axis. To avoid scattering between the layers, the boundaries must
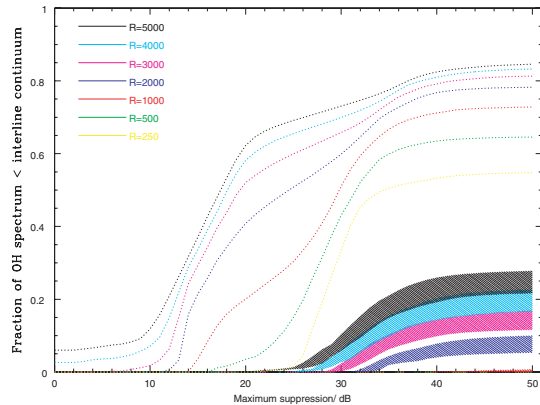
Figure 14. The fraction of the 1.4–1.8 μm window over which observations can reach the interline continuum as a function of maximum suppression. Spectral resolutions of $R = 250, 500, 1000, 2000, 3000, 4000$ and 5000 are shown by the different colours. The hatched curves show the range for the Kelsall et al. (1998) and NICMOS models at an ecliptic latitude of $\beta = 60°$. The dotted lines are for the interline continuum of Maihara et al. (1993b).

be blackened, which therefore divides the beam before it enters the VPH.

A comparison of VPH filters with FBGs strongly favours the latter. The relatively small number of lines per filter compared with FBGs, coupled with the lower interline throughput is disadvantageous. FBGs are able to suppress 150 lines across a 400-nm window with a interline throughput of 88 per cent. The same 150 lines would require 15 of the filters demonstrated by Blais-Ouellette et al. (2004), or five out of the 30 line filters advertised by Photon Etc. This would result in a throughput of 9–44 per cent.

FBGs can suppress the OH lines at 30 dB, whereas the VPH filters suppress filters at 10 dB. This makes a significant difference in the performance of the two devices. In the next section, we show how the suppression factor strongly influences the performance of FBGs (Fig. 14). If the OH lines are not suppressed strongly enough, the residual lines will still dominate the background signal due to the scattering effects described above.

5.2 Fibre Bragg gratings

Bland-Hawthorn et al. (2004) present a radical new approach to OH suppression employing FBGs. FBGs are optical fibres with a slowly varying refractive index in the core. As light propagates down the fibre, it suffers Fresnel reflections due to the varying refractive index. By properly controlling the index variation, it is possible to produce strong reflections at specific wavelengths. Bland-Hawthorn et al. (2004) showed that it was possible to produce a low-loss filter defined by a series of irregularly spaced, narrow notches with suppression depths as high as 30 dB, whilst maintaining high throughput outside the notches (losses ≈ 0.03 per 100 nm wavelength coverage). Their design corrects for 18 OH lines in the H band.

Two steps are required in order for this technology to be useful for astronomical observations: the ability to suppress many lines, and the ability to work with multimode fibres. These two requirements are not easily compatible. When one considers the huge complexity of suppressing up to 150 OH lines within the H band, the amplitude and phase structure of the grating demands that this is printed into a single-mode fibre so that the $\sim 10^5$ cycles are available to produce the filter response along one propagation axis (Bland-Hawthorn et al. 2008). In a multimode fibre, a much larger number of cycles are required to achieve the same grating response to each mode, thereby greatly increasing the complexity of the FBG. Thus, FBGs achieve peak performance within a single-mode fibre (10-μm core), whereas astronomical fibres, with their larger core diameters, are highly multimoded. Therefore, a successful implementation of photonic OH suppression requires that the many modes of the input fibre be transformed into the same number of single modes acting in parallel.

To this end, we developed the photonic lantern (Bland-Hawthorn 2005; Leon-Saval et al. 2005) which decomposes the spatial modes of the input fibre into a parallel degenerate array of single-mode fibres. Identical gratings are printed into each of the fibres. Once the filtered light exits the fibre array, these are recombined by an identical taper back to a multimode output fibre. Horton & Bland-Hawthorn (2007) have demonstrated that fibres with as few as five modes can have high coupling efficiency, paving the way for AO fed FBGs in the near future.

The combination of these two innovative technologies means that the development of FBGs for astronomical purposes is now at a stage that is compatible with astronomical requirements. FBGs are now being developed to remove ~ 150 lines in each of the J and H bands. The brightest lines are suppressed at 30 dB, and fainter lines are suppressed by an amount proportionally less. Photonic lanterns are being developed with 37 fibre tapers. The resulting multimode FBG will be a complex device, and expensive to manufacture, at least in the present incarnation. Initially, only tens to hundreds of fibres are envisaged, but ultimately, the technology will undergo further development to bring down the unit cost.

In light of the discussion in Section 3, an important advantage of FBGs is that they can be incorporated into a spectrograph before the dispersive elements, thus the OH light is strongly rejected with very little scattering, offering for the first time the capability to achieve extremely deep NIR observations without the requirement to put a telescope and instrument into space.

The ultimate goal of OH suppression is to reach the interline continuum over a significant fraction of the J and H bands. The level of suppression required to achieve this depends on the spectral resolving power of the instrument. Fig. 14 shows the fraction of the 1.4–1.8 μm window in which the OH spectrum falls below the zodiacal scattered light as a function of suppression and spectral resolution. A maximum suppression of 40 dB yields significant gains, and thereafter the improvements flatten off. This is true of both the high interline continuum of Maihara et al. (1993b) and the models.

5.3 Simulations

In order to assess the expected performance of FBGs, we have thoroughly modelled an entire observing system from astronomical source to the detector, incorporating an instrument employing FBGs. Three different source types were chosen to illustrate the power of OH suppression with FBGs, a $z = 11$ QSO, a $z = 3$ galaxy and a faint T dwarf.

5.3.1 System model

The simulations work by starting with a fluxed 1d object spectrum and modifying it according to background contributions and losses for each component of the system. A simplified schematic overview of a typical simulation is shown in Fig. 15, showing the sources of emission and losses we have modelled. Many of the components
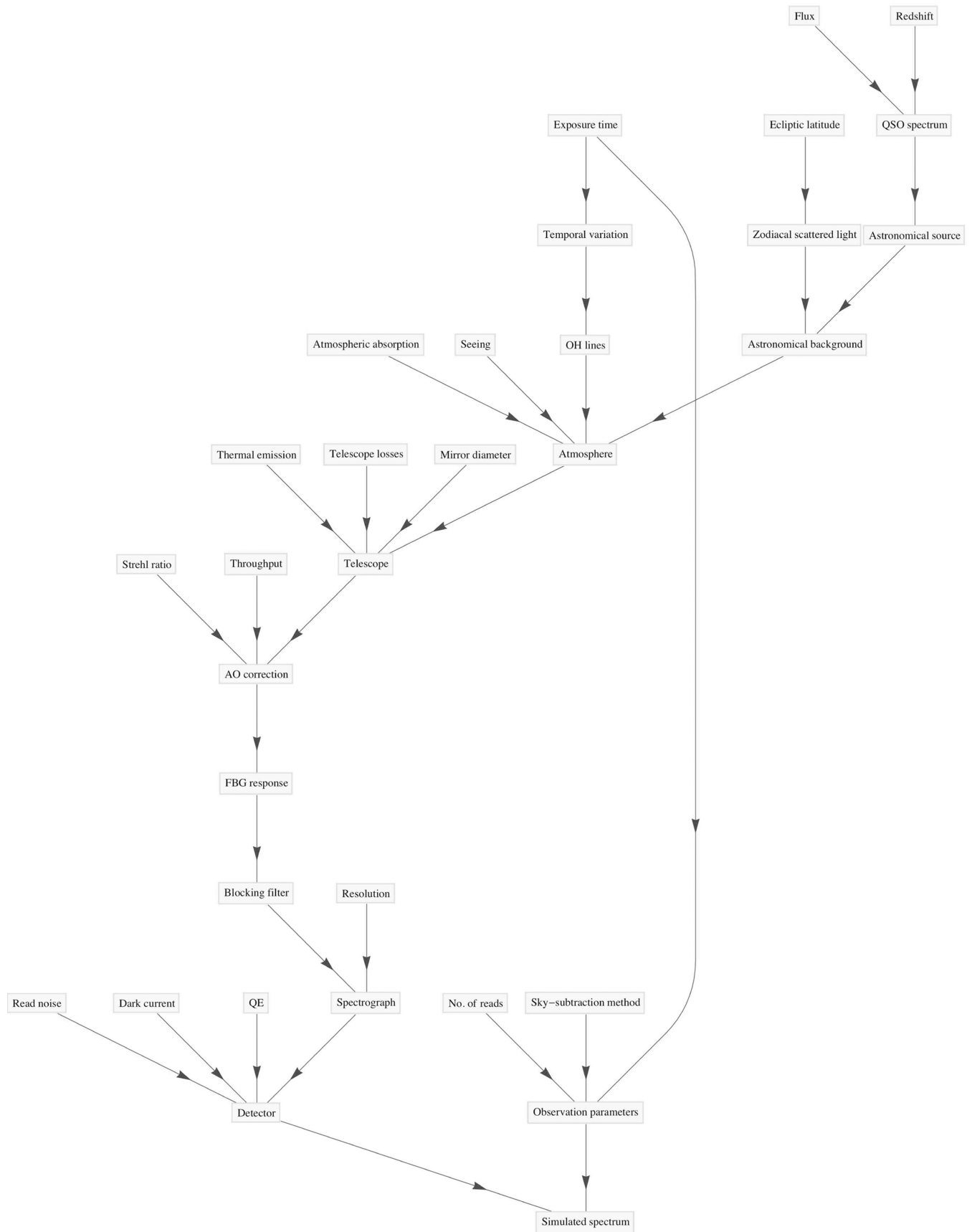
Figure 15. A much simplified schematic overview of a typical simulated spectroscopic observation, showing the main causes of emission and losses, with which we have modelled the performance of the fibres.

shown are optional, for example, we can also model galaxy, and stellar spectra, amongst other sources. It should also be noted that most of the components shown are fully adjustable, for example, OH line-strengths, zodiacal scattered light etc. can all be adjusted according to meaningful parameters. The observation parameters and mode of observation are then applied to generate the final spectrum.

The details of the simulations were as follows. The zodiacal background is described above (Section 2.4), and we modelled observations at $\beta = 60°$. The OH line spectrum was used as described in Section 2.1, including temporal variation. The atmospheric absorption is described in Section 2.7. We assumed that an 8-m telescope with obstruction losses of 5 per cent was employed in Cassegrain focus with a reflectivity of 97 per cent off both mirror surfaces. We modelled the thermal emission from the telescope as a $T = 273$ K blackbody with an emissivity of 3 per cent.

The FBGs are modelled as 150 line gratings of 1 Å width. The fibres are taken to have a core size of 25 μm, that is, few mode fibres, corresponding to an angular diameter of 0.2 arcsec at *f*/5. The fibre coupling losses are $\approx$0.15 for a few mode fibre of this size (Horton & Bland-Hawthorn 2007). These are fed with a generic AO system, achieving a Strehl ratio of 0.3, with a throughput of 0.9. The fibre aperture losses are thus, $\approx$46 per cent, due to lost light, assuming that the PSF can be modelled as a diffraction-limited PSF plus a Gaussian seeing disc with FWHM 0.5 arcsec, where the two components are normalized such that the diffraction-limited part contains a power equal to the Strehl ratio.

The spectrograph is assumed to have a throughput of 35 per cent, and an adjustable resolution. The sampling is assumed to be 3 spatial by 3 spectral pixels. A systematic wavelength calibration error of a tenth of a pixel FWHM has been included. We assume that a blocking filter is used at the entrance to the spectrograph and apply the corresponding transmission. The detector is taken to have an effective read noise of $4e^-$ when up-the-ramp sampling is employed, and a dark present of two thousandths of an electron per pixel per second, typical of Rockwell Hawaii-2RG HgCdTe arrays (Smith et al. 2006). Observations are composed of 30-min reads, with sky-subtraction performed by cross-beam switching between pairs of fibres.

5.3.2 General results

The system described above results in a reduction of the total background by a factor of 42 and 17 in the *H* and *J* bands, respectively, corresponding to 4 and 3 mag. *It is important to realize that this reduction applies across the entire spectrum including the interline regions.* For example, at $\lambda = 1.665$ μm (where Maihara et al. 1993b measured the interline continuum), the background would be dominated by scattered OH light in the absence of FBGs. However, because FBGs suppress the OH lines before scattering, the interline background is also reduced by a factor of 40. We emphasize that it is the magnitude of the interline reduction that is scientifically important, and that the advantage of FBGs over other methods of OH suppression is their ability to suppress the OH light before the light enters the spectrograph, and therefore before any significant scattering surfaces are encountered. Methods to remove OH light after dispersion may achieve a good overall reduction in the overall background due to the removal of the cores of the bright emission lines (as could also be achieved through software suppression), but between the lines, where the science will be done, the reduction will not be as good due to the scattered light.

We compare the reduction in background resulting from OH suppression to that resulting from AO correction in Fig. 16. Changing
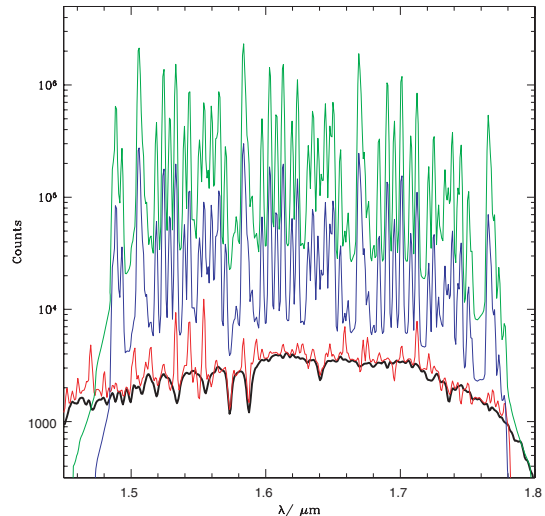

Figure 16. Simulated spectra of a $z = 3$ $H = 19$ mag galaxy, observed at a resolution of $R = 1000$ for 1 h. These spectra are of a very bright galaxy, and have no sky-subtraction in order to emphasize the equivalent width of the OH lines compared to the continuum. The thick black line shows the actual object spectrum; the red line shows the object plus background for the galaxy observed with AO and FBGs as described in Section 5.3.1; the blue line is for AO correction only; and the green line is for natural seeing of 0.5 arcsec.

from a 0.5-arcsec aperture for natural seeing to a 0.2-arcsec aperture for AO correction results in a reduction of the background by a factor of 6.25. Applying OH suppression then results in a *further* reduction of a factor of 42 in *H* or 17 in *J* as described above. Given that diffraction-limited observations would severely oversample a high-redshift galaxy of size 1 arcsec, we argue that the main advantage from AO for cosmology is the reduction in background light. Consequently, we argue that OH suppression is at least as important for cosmology as AO, since it produces a greater level of background suppression. Of course both technologies must proceed together but, given the significant advances promised by OH suppression, we argue that the development of FBGs should be given a very high priority for future instruments on existing telescopes and on the ELTs of the future.

5.3.3 QSO simulation

The typical surface brightness of sources responsible for reionization is AB = 27–29 mag arcsec^2, achievable by a source of 26 AB mag arcmin^{-2} if the number density is $\approx$0.1–10 arcmin^{-2} (Stiavelli, Fall & Panagia 2004). This source density is reasonable considering the number counts of Yan & Windhorst (2004) for $z \approx 6$ galaxies, which have ≈ 0.3 galaxies arcmin^{-2} at AB = 26 mag in approximately the same waveband. We have therefore assumed an unabsorbed magnitude of AB = 26 mag, or $H = 24.6$ Vega mag (using the conversion given by Tokunaga & Vacca 2005).

We used a simulated $z = 10, H = 20$ mag spectrum (provided by Richard MacMahon, private communication), which includes Gunn & Peterson (1965) absorption and metal line absorption. We then redshifted this to $z = 11$ (assuming no further absorption due to neutral hydrogen), and normalized the *H*-band flux to 24.6 mag. Fig. 17 shows a simulation of a 70 h exposure at $R = 1000$. The bottom panel shows the results of an identical system without FBGs, and is dominated by residual sky lines. The use of FBGs results in a re-
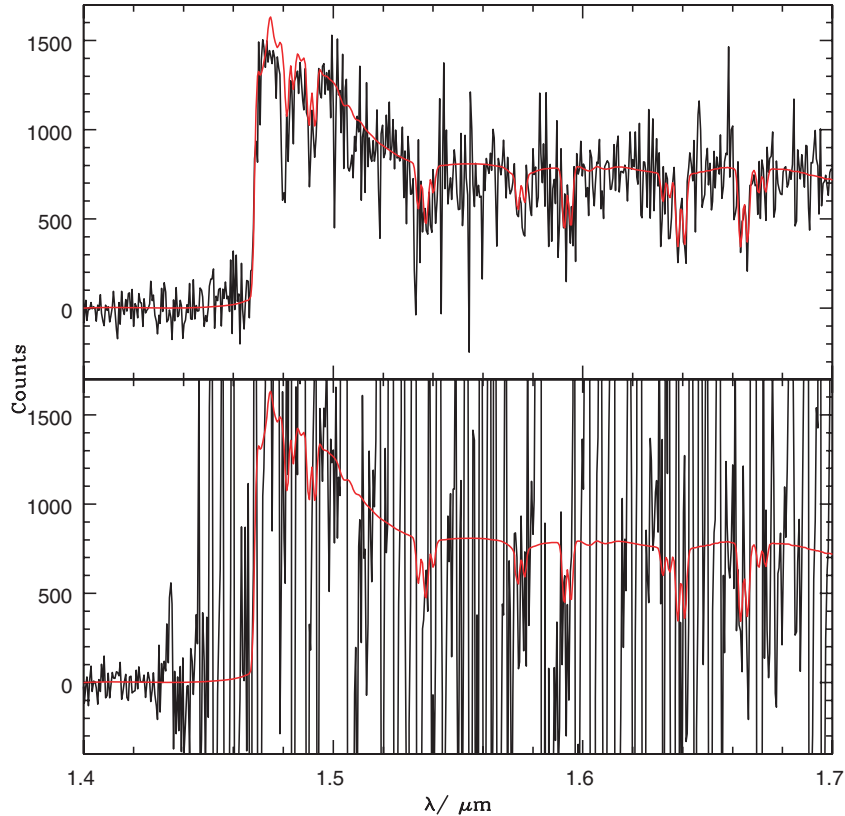
Figure 17. A simulated spectrum of an $H = 24.6$ Vega mag, $z = 11$ QSO as observed by the system described in Section 5.3.1. The exposure time was 70 h, and the spectral resolution was $R = 1000$. The top panel shows a system with FBGs and the bottom panel shows an identical system without FBGs. The red lines indicate the true object spectrum.

markable improvement in the quality of the resulting spectrum. The Gunn & Peterson (1965) absorption trough is clearly visible when FBGs are used, whereas without FBGs the spectrum is completely dominated by noise and residual sky lines. FBGs offer the chance to accurately measure the redshifts of extremely faint objects during the epoch of reionization. This will be discussed further in the summary.

5.3.4 Redshift 3 galaxy simulation

We have simulated an observation of an $H = 23$ Vega mag galaxy at $z = 3$. The exposure time was 8 h and the resolution was $R = 1000$. The galaxy spectrum is from the synthetic spectral libraries of Bruzual & Charlot (2003), and assumes an instantaneous burst of star formation, an age of 1 Gyr and solar metallicity. Fig. 18 shows the resulting spectrum as observed through a system with FBGs and a system without FBGs. The FBGs result in a high-quality spectrum with Ca H and Ca K absorption lines clearly visible. Thus, FBGs offer the chance to obtain precise measurements of galaxies when the Universe was only a fifth of its present age, allowing a detailed study of evolution over almost the entire history of galaxies.

5.3.5 T dwarf simulation

Cool stars emit most of their radiation in the NIR and are identified by molecular absorption features. The faintest T dwarf to date has a magnitude of $J = 20.28$ Vega mag (Zapatero Osorio et al. 2002; see

Adam Burgasser's T dwarf page[6]); however, most known T dwarfs have magnitudes around $J \sim 15–16$ Vega mag, and are located within a few tens of parsecs from the Sun.

Fig. 19 shows a simulated spectrum of a $J = 21$ Vega mag T dwarf observed at $R = 1000$, in J and H simultaneously, for 8 h. The spectrum is that of a T5 dwarf, 2MASS 0559−1404 (McLean et al. 2003[7]) scaled to $J = 21$. The FBG spectrum is of much superior quality to the spectrum employing ordinary fibres. FBGs offer the opportunity to identify much fainter cool stars than ever before.

5.4 Summary

We have demonstrated the possibility to achieve extremely deep NIR observations using FBGs. Hitherto, attempts to obtain deep NIR observations have been frustrated by the extremely bright and variable OH emission from the atmosphere. We have shown that in order to efficiently remove OH lines from astronomical spectra the OH lines must be suppressed prior to the light entering the camera. Scattering from dispersive optics results in Lorentzian line-profiles, the wings of which are much brighter than the true interline continuum, and in most cases are brighter than the sources being observed. The variability of the OH lines makes it extremely difficult to subtract cleanly.

FBGs can suppress OH lines at high resolution, $R = 10\,000$, high suppression (30 dB) and with high throughput out of the suppression

---
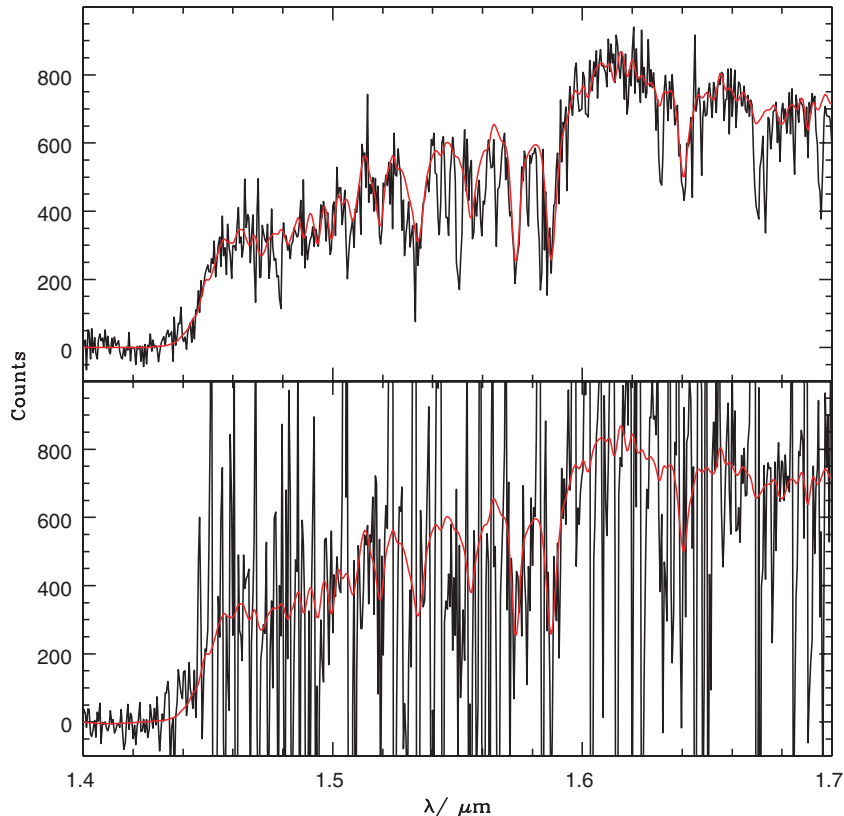
[6] http://web.mit.edu/ajb/www/tdwarf/

[7] http://www.astro.ucla.edu/~mclean/BDSSarchive/

Figure 18. A simulated spectrum of an $H = 23$ Vega mag, $z = 3$ galaxy as observed by the system described in Section 5.3.1. The exposure time was 8 h, and the spectral resolution was $R = 1000$. The top panel shows a system with FBGs and the bottom panel shows an identical system without FBGs. The red lines indicate the true object spectrum.

regions ($\approx$90 per cent). Suppressing 150 lines in the J or H band, with a maximum suppression of 30 dB, decreases the sky background by a factor of $\approx$45, or 4 mag. Achieving this will have the potential to extend NIR observation to objects orders of magnitude fainter than presently possible. The resulting observations will have a great impact on many areas of astronomy, and we highlight a few now.

5.4.1 First light and reionization

FBGs offer the opportunity to study the rest-frame ultraviolet light from objects at the time of reionization. Fig. 17 shows a simulated spectrum of a $z = 11$ QSO with a clearly visible Gunn & Peterson (1965) absorption trough. Spectroscopy with FBGs would therefore be ideally suited to identifying first light sources, performing follow-up spectroscopy from surveys of Lyman α emission (e.g. Taniguchi et al. 2005), continuum drop-out due to the Lyman break (e.g. Stanway et al. 2004) and ultranarrow band surveys (e.g. Horton et al. 2004). A spectrum of a complete Gunn–Peterson trough would confirm the redshift of the source, which would allow estimates of the luminosity function and number-magnitude relations of the first sources of light in the Universe. If the sources are bright enough, then it may be possible to also measure the line-profile of Lyman α which is expected to be distinctively asymmetrical due to resonant scattering bluewards of the line-centre. The intrinsic width of the line itself will depend on the type of source; quasars will have broad Lyman α lines, FWHM $\approx 1500 \, \text{km s}^{-1}$, whereas galaxies will have narrower lines, FWHM $\approx 100 \, \text{km s}^{-1}$. The difference in linewidths leads to the possibility of distinguishing the type of sources respon-

sible for ionization. For the purpose of confirming the presence of a redshifted Lyman α line, it is necessary to resolve the lines sufficiently to measure the asymmetry, requiring resolutions of $R \gtrsim 300\,000/100 = 3000$.

The combination of FBGs and ELTs will make it possible to thoroughly explore the ionization history of the intergalactic medium (IGM) using NIR spectroscopy of first light objects. The detection of complete Gunn–Peterson absorption troughs in the spectra of high-redshift quasars (Becker et al. 2001) shows that the IGM was partly neutral at $z \approx 6.5$. However, a neutral fraction of only $x_{HI} \approx 10^{-4}$ is all it takes for complete absorption due to the high oscillator strength of the Lyman α transition. Thus, to probe further back in time to uncover the details of the ionization of the IGM, we must look to other means.

Oh (2002) suggest that metal absorption lines may provide a useful means of measuring the ionization state of the IGM. O I and Si II have ionization potentials close to that of neutral hydrogen, and thus measuring the neutral fraction of either of these metals could be used to infer the neutral fraction of hydrogen. Furthermore, there are useful lines just longer than Lyman α so a single measured spectrum may incorporate Lyman α 1215.67 Å, O I 1302.17 Å and Si II 1260.42 Å. These metals have previously been observed in the spectra of high-redshift quasars (Pettini 2001) verifying their usability for this purpose. Since it is unlikely that metals will be smoothly distributed through the IGM, the signature will be in the form of a forest of lines rather than a Gunn–Peterson trough. A measurement of the equivalent width of the lines, or even simply the number of lines will provide precious information on the ionization state.
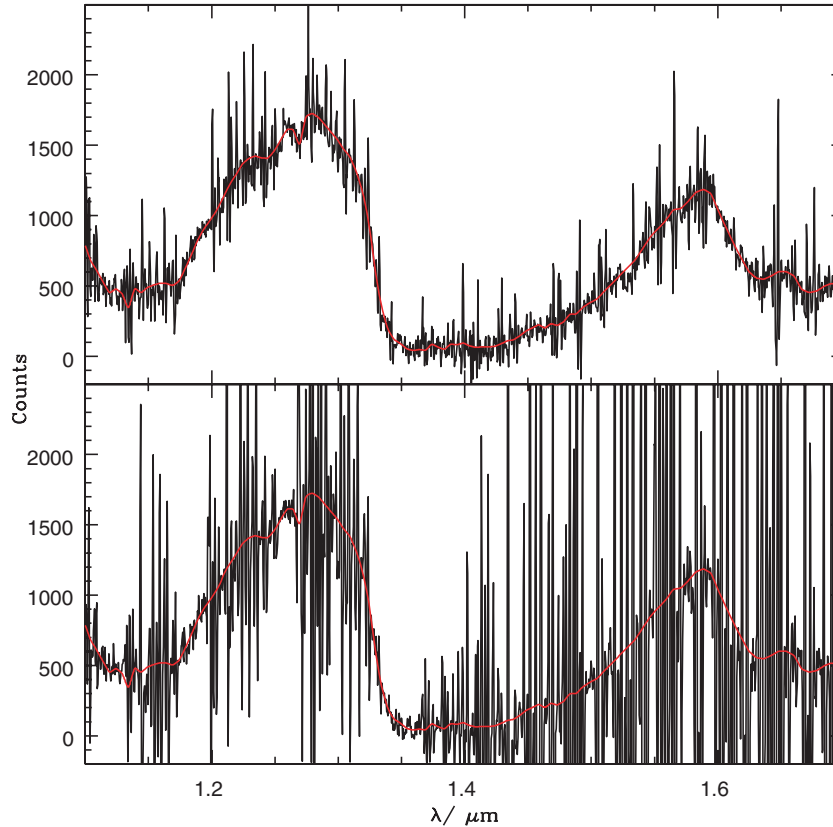
Figure 19. A simulated spectrum of a $J = 21$ Vega mag, T5 dwarf as observed by the system described in Section 5.3.1. The exposure time was 8 h, and the spectral resolution was $R = 1000$. The top panel shows a system with FBGs and the bottom panel shows an identical system without FBGs. The red lines indicate the true object spectrum.

The detection or non-detection of a radiation damping wing has the potential to divulge important information on the ionization state of the IGM. If a radiation damping wing is found, it will provide a very accurate measure of the neutral fraction of hydrogen (Miralda-Escude 1998). If the ionizing sources reside in significant cosmic Strömgren spheres then a damping wing will be impossible to measure (Cen & Haiman 2000; Madau & Rees 2000). However, the line-profile may still be identified as Lyman α due to its asymmetrical shape, and thus the lack of a damping wing can place constraints on the ionizing fluxes of the sources. If the sources are galaxies, then it may also place constraints on the clustering of the objects, since it is unlikely that a single primordial galaxy would produce a significant ionized region on its own. If spectra can be obtained with a sufficiently high signal-to-noise ratio, then the tail of the damping wing may provide detailed information on the size of the Strömgren spheres.

Fig. 20 shows the same simulation as described in Section 5.3.3, except that the telescope was assumed to have a diameter of 30 m. Metal absorption lines are now visible in the spectrum.

5.4.2 Galaxy evolution

Deep NIR imaging and spectroscopy with FBGs will allow galaxy evolution to be studied over most of the history of the Universe. Fig. 18 shows a $H = 23$ Vega mag galaxy at $z = 3$. Absorption lines are clearly visible. It would thus be possible to measure galaxy masses, luminosities and star formation rates from emission lines, and to derive the evolution of the star formation his-

tory, the luminosity function, the Faber & Jackson (1976) relation and if high spatial resolution imaging is available, the Fundamental Plane.

5.4.3 Cool stars

NIR observations are essential for identifying cool stars such as L, T and Y dwarfs. Measuring the number density of such objects as a function of mass is a critical test of our understanding of both star and planet formation processes. NIR spectroscopy is essential for fully characterizing their temperature, surface gravity and atmospheric conditions. The use of FBGs would allow observations of objects orders of magnitude fainter than presently possible (see Fig. 19), thus extending the survey volumes and numbers of such objects, and placing strong constraints on the low mass end of the initial mass function.

6 REALIZATION

The 8–10 m telescope generation has been with us for more than a decade, and astronomy must now plan for a new generation of facilities if the field is to advance in the years to come. For ground-based observations, the Earth's atmosphere provides a major challenge to scientific progress on two fronts: (i) the movement of air cells blurs the cosmic signal; and (ii) hydroxyl emission from the upper atmosphere produces a perpetual twilight, thereby greatly increasing the background signal in astronomical observations. At
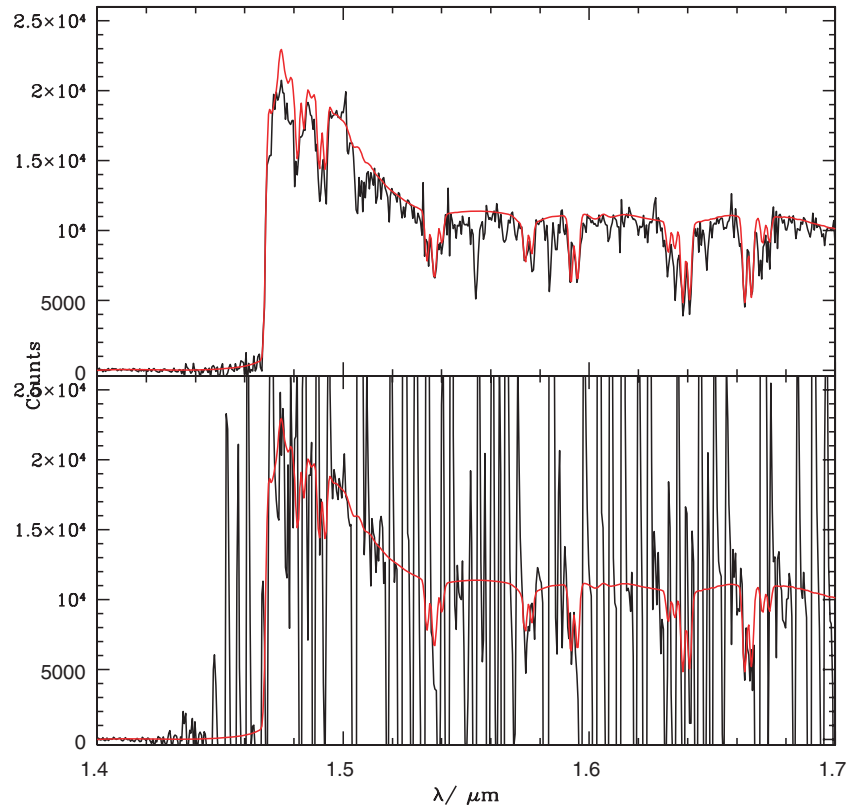
Figure 20. A simulated spectrum of an $H = 24.6$ Vega mag, $z = 11$ QSO as observed by the system described in Section 5.3.1 with a 30-m telescope. The exposure time was 70 h, and the spectral resolution was $R = 1000$. The top panel shows a system with FBGs and the bottom panel shows an identical system without FBGs. The red lines indicate the true object spectrum.

NIR wavelengths, there are clear indications that adaptive optics is overcoming the problem of seeing and is starting to deliver unprecedented resolution and sensitivity at NIR wavelengths (Liu 2006). In contrast, the extreme brightness of the IR sky remains an unsolved problem.

The JWST (Gardner et al. 2006) will be launched sometime in the next decade, bypassing altogether the problems associated with observing through the atmosphere. However, it will have a 6-m aperture compared to apertures of 20–42 m of the next generation of ground-based telescopes. The satellite will be inaccessible and will not be serviced by astronauts, thus the instrument suite will have restricted functionality compared to what is possible today on ground-based sites. For these reasons, it is important to continue to address the issue of the IR night sky background for ground-based telescopes.

The AAO has spearheaded the development of a new technology that promises to completely solve the IR twilight problem, that is, by removing 99 per cent of the OH line emission with an overall efficiency of 80 per cent or better. The concept exploits two recent advances in photonics: (i) broad-band FBGs; and (ii) multimode to single-mode photonic lanterns. While the development remains challenging, there are indications that the development is on track to solve the OH suppression problem once and for all. We have demonstrated the extraordinary scientific potential of these fibres in this paper. Note that all simulations include correction by adaptive optics, for both cases, with and without FBGs. The observations we have presented and discussed would not be possible through use of adaptive optics alone. The improvements over the AO-only observations demonstrate a compelling need for FBGs, if such deep observations are to be achieved. The development and implementation of FBG-fed instruments on existing 8 m, and especially on future ELTs, deserves equal attention to that which has been invested in the development of adaptive optics.

In subsequent papers, we will discuss the relative merits of OH suppression in natural seeing compared to corrected seeing through adaptive optics, and describe instrument concepts that are under development to capitalize on the photonic OH suppression development.

REFERENCES

Barden S. C., Arns J. A., Colburn W. S., Williams J. B., 2000, PASP, 112, 809
Becker R. H. et al., 2001, AJ, 122, 2850

Blais-Ouellette S., Artigau É., Havermeyer F., Matthews K., Moser C., Psaltis D., Steckman G. J., 2004, in Atad-Ettedgui E., Dierickx P., eds, Proc. SPIE Vol. 5494, Optical Fabrication, Metrology, and Material Advancements for Telescopes. SPIE, Bellingham, p. 554

Bland-Hawthorn J., 2005, Anglo-Australian Observatory Newsletter, 108, 4

Bland-Hawthorn J., Buryak A., Kolossovski K., 2008, J. Opt. Soc. America A, 25, 153

Bland-Hawthorn J., Englund M., Edvell G., 2004, Optics Express, 12, 5902

Bruzual G., Charlot S., 2003, MNRAS, 344, 1000

Burton M. G. et al., 2005, PASA, 22, 199

Cen R., Haiman Z., 2000, ApJ, 542, L75

Committee On The Physics Of The Universe B. O. P., Astronomy D. O. E., Physical Sciences N. R. C. O. T. N. A., 2003, Connecting quarks with the cosmos: eleven science questions for the new century (Connecting quarks with the cosmos: eleven science questions for the new century/Committee on the Physics of the Universe, Board on Physics and Astronomy, Division on Engineering and Physical Sciences, National Research Council of the National Academies. Washington D. C.: National Academies Press

Content R., 1996, ApJ, 464, 412

Cox A. N., 2000, in Arthur N. C., ed., Allen's Astrophysical Quantities, 4th edn. AIP Press; Springer, New York

Cuby J., Lidman C., Moutou C., 2000, ESO Messenger, 101, 2

Davies R. I., 2007, MNRAS, 375, 1099

Dufay M., 1951, Publications of the Observatoire Haute-Provence, 2

Ennos A. E., 1966, Applied Optics, 5, 51

Faber S. M., Jackson R., 1976, ApJ, 204, 668

Frey H. U., Mende S. B., Arens J. F., McCullough P. R., Swenson G. R., 2000, Geophys. Res. Lett., 27, 41

Gardner J. P. et al., 2006, Space Sci. Rev., 123, 485

Glazebrook K., Bland-Hawthorn J., 2001, PASP, 113, 197

Gunn J. E., Peterson B. A., 1965, ApJ, 142, 1633

Herzberg G., 1939, Molecular Spectra and Molecular Structure. I. Diatomic Molecules. Prentice-Hall, Inc., New York

Horton A., Bland-Hawthorn J., 2007, Optics Express, 15, 1443

Horton A., Parry I., Bland-Hawthorn J., Cianci S., King D., McMahon R., Medlen S., 2004, in Moorwood A. F. M., Iye M., eds, Proc. SPIE Vol. 5492, Ground-based Instrumentation for Astronomy. SPIE, Bellingham, p. 1022

Iwamuro F., Maihara T., Oya S., Tsukamoto H., Hall D. N. B., Cowie L. L., Tokunaga A. T., Pickles A. J., 1994, PASJ, 46, 515

Iwamuro F., Motohara K., Maihara T., Hata R., Harashima T., 2001, PASJ, 53, 355

Kelsall T. et al., 1998, ApJ, 508, 44

Krisciunas K., Schaefer B. E., 1991, PASP, 103, 1033

Leon-Saval S., Birks T., Bland-Hawthorn J., Englund M., 2005, Optics Letters, 30, 19

Levasseur-Regourd A. C., Dumont R., 1980, A&A, 84, 277

Liu M. C., 2006, in Ellerbroek B. L., Bonaccini C. D., eds, Proc. SPIE Vol. 6272, Advances in Adaptive Optics II. SPIE, Bellingham, p. 14

Lord S. D., 1992, NASA Technical Memor., 103957

McLean I. S., McGovern M. R., Burgasser A. J., Kirkpatrick J. D., Prato L., Kim S. S., 2003, ApJ, 596, 561

Madau P., Rees M. J., 2000, ApJ, 542, L69

Maihara T., Iwamuro F., 2000, in van Breugel W., Bland-Hawthorn J., eds, ASP Conf. Ser. Vol. 195, Imaging the Universe in Three Dimensions. Astron. Soc. Pac., San Francisco, p. 585

Maihara T., Iwamuro F., Hall D. N., Cowie L. L., Tokunaga A. T., Pickles A. J., 1993a, in Fowler A. M., ed., Proc. SPIE Vol. 1946, Infrared Detectors and Instrumentation. SPIE, Bellingham, p. 581

Maihara T., Iwamuro F., Yamashita T., Hall D. N. B., Cowie L. L., Tokunaga A. T., Pickles A., 1993b, PASP, 105, 940

Maihara T. et al., 2000, in Iye M., Moorwood A. F., eds, Proc. SPIE Vol. 4008, Optical and IR Telescope Instrumentation and Detectors. SPIE, Bellingham, p. 1111

Matsumoto T., Kawada M., Murakami H., Noda M., Matsuura S., Tanaka M., Narita K., 1996, PASJ, 48, L47

Meinel I. A. B., 1950, ApJ, 111, 555

Miralda-Escude J., 1998, ApJ, 501, 15

Motohara K. et al., 2002, PASJ, 54, 315

Noda M., Christov V. V., Matsuhara H., Matsumoto T., Matsuura S., Noguchi K., Sato S., Murakami H., 1992, ApJ, 391, 456

NRC, 2001, Astronomy and Astrophysics in the New Millennium (Astronomy and Astrophysics in the New Millennium, Astronomy and Astrophysics Survey Committee, Board on Physics and Astronomy-Space Studies Board, Division on Engineering and Physical Sciences, National Research Council. Washington DC: National Academy Press, 2001

Offer A. R., Bland-Hawthorn J., 1998, MNRAS, 299, 176

Oh S., 2002, MNRAS, 336, 1021

Osterbrock D. E., Fulbright J. P., Martel A. R., Keane M. J., Trager S. C., Basri G., 1996, PASP, 108, 277

Pettini M., 2001, in Pilbratt G. L., Cernicharo J., Heras A. M., Prusti T., Harris R., eds, ESA SP-460, The Promise of the Herschel Space Observatory. ESA, Noordwijk, p. 113

Ramsay S. K., Mountain C. M., Geballe T. R., 1992, MNRAS, 259, 751

Rousselot P., Lidman C., Cuby J.-G., Moreels G., Monnet G., 2000, A&A, 354, 1134

Shimazaki T., Laird A., 1970, J. Geophys. Res., 75, 3221

Smith R. et al., 2006, in Dorn D. A., Holland A. D., eds, Proc. SPIE Vol. 6276, High Energy, Optical, and Infrared Detectors for Astronomy II. SPIE, Bellingham, p. 24

Stanway E. R., Bunker A. J., McMahon R. G., Ellis R. S., Treu T., McCarthy P. J., 2004, ApJ, 607, 704

Stiavelli M., Fall S. M., Panagia N., 2004, ApJ, 600, 508

Stockman H. S., Stiavelli M., Im M., Mather J. C., 1998, in Smith E. P., Koratkar A., eds, ASP Conf. Ser. Vol. 133, Science With The NGST. Astron. Soc. Pac., San Francisco, p. 24

Taniguchi Y. et al., 2005, PASJ, 57, 165

Thompson R. I., Eisenstein D., Fan X., Rieke M., Kennicutt R. C., 2007, ApJ, 657, 669

Tokunaga A. T., Vacca W. D., 2005, PASP, 117, 421

Turnbull D. N., Lowe R. P., 1983, Can. J. Phys., 61, 244

Woods T. N., Wrigley R. T. III, Rottman G. J., Haring R. E., 1994, Appl. Opt., 33, 4273

Yan H., Windhorst R. A., 2004, ApJ, 612, L93

Zapatero Osorio M. R. et al., 2002, ApJ, 578, 536

This paper has been typeset from a TeX/LaTeX file prepared by the author.