Design and Implementation of Data Reduction Pipeline Software for the MOSFIRE Near-Infrared Spectrometer

Robert Lasenby*
St John's College, University of Cambridge, Cambridge, UK, CB2 1TP

Mentor : Professor C. Steidel†
California Institute of Technology, Pasadena, California 91125, USA
(Dated: September 25, 2009)

Infrared spectroscopy is a vital tool in modern astronomical research, providing important data on phenomena including star formation, galactic centre objects and high redshift galaxies. MOSFIRE, the Multi-Object Spectrometer For Infra-Red Exploration, is a near-infrared imager/spectrometer, currently under development at Caltech and UCLA, and intended for installation at the W.M. Keck Observatory in Hawaii in early 2010. Using a unique configurable slit mask, it will be capable of simultaneously obtaining up to 46 object spectra in a single exposure, allowing astronomers to acquire the large numbers of spectra necessary for detailed understanding of object populations.

In addition to the instrument itself, the MOSFIRE team will also provide a data reduction pipeline software package, enabling users to obtain accurate spectra from the raw instrument data. This process involves carefully characterizing and compensating for the properties of the instrument and of the atmosphere, as far as possible in an automated manner. Here, we describe a prototype implementation of portions of this pipeline, along with methods to simulate the output of the still-to-be-completed instrument to the degree necessary for testing this.

I. THE MOSFIRE PROJECT

The MOSFIRE instrument [1] is intended to fulfil an important strategic role at the Keck Observatory. With the ability to obtain up to 46 object spectra simultaneously, it will enable the investigation and deeper understanding of object populations, in areas including

- Young stars
- Galactic centre objects
- Extremely low mass stars in star forming regions
- Low mass end of the stellar initial mass function
- Stellar populations in nearby galaxies
- Surveys of star-forming galaxies and active galactic nuclei at high redshifts
- Surveys of passively evolving galaxies at high redshifts
- Star formation and populations of dusty galaxies
- Searches for extremely high redshift objects ("first light" surveys)

In addition, its location at the Cassegrain focus of the Keck I telescope will aid in balancing observations between the two telescopes[1] — since the existing state-of-the-art near-infrared instruments are installed on Keck II, which also possess adaptive optics capabilities, an advanced near-infrared instrument on Keck I will ease scheduling competition. Furthermore, MOSFIRE will complement the imaging and spectroscopy capabilities of the LRIS [2] optical instrument on Keck I, contributing to the large number of scientific programs that require observations of the same objects at multiple wavelength ranges.

II. INSTRUMENT DETAILS

Fig. 1 shows a cut-away diagram of the MOSFIRE instrument, which produces spectroscopic-mode images by the following process:

- Light from the telescope enters the instrument through the dewar window. Since thermal radiation from objects at room temperature has significant power in the infrared, it it necessary to keep the internal components of the instrument at low temperatures inside a cooling vessel (dewar).
- The beam passes through the Configurable Slit Unit (CSU). This is a set of 46 bars, each split into two parts — these parts can be moved independently and with high precision to form a slit of the desired width at any horizontal position. The purpose of the slit unit is to block out all of the light apart from that coming from the desired target objects. Fig. 2 illustrates this masking process. In comparison to previous masking techniques — for example, some instruments require a custom metal mask to be machined and installed for each observing session — the CSU should increase the efficiency of observing, and allow for on-the-fly changes of the mask configuration.

*Electronic address: `rnl22@cam.ac.uk`
†Electronic address: `ccs@astro.caltech.edu`

[1] The W.M. Keck Observatory consists of two 10m optical/infrared telescopes, Keck I and Keck II

- The Flexure Control System (FCS) mirror attempts to distort the beam so as to cancel out any changes in the optical properties of the instrument, e.g. due to very slight bending of components under gravity.

- The beam goes through a band-pass filter, which allows through only those frequencies we intend to study. MOSFIRE can observe in the Y, J, H and K bands, which together span the $\sim 0.9 - 2.4\mu$m range of the electromagnetic spectrum.

- With the instrument in spectroscopic mode, the beam then hits the reflection grating, which spreads the incoming radiation horizontally according to its wavelength. This process is shown in the second stage of Fig. 2.

- Finally, the beam is focussed by the lenses of the camera onto the detector, which is a 2048×2048 pixel infrared-sensitive array. The detector used in MOSFIRE is of a more modern generation than those employed in previous instruments such as NIRSPEC [3], and offers superior sensitivity, lower noise and other desirable properties.

Fig. 2 gives a visual overview of the "transformations" the instrument applies to the incoming beam — the result, as displayed in the rightmost panel, is an image on the detector that consists of horizontal strips for each slit, with different wavelengths spread horizontally, and the vertical direction still corresponding to one axis of 'sky position'. In Fig.4, we can see a close-up of the detected image — this illustrates the fact that multiple slits can be lined up to form a single longer slit, as in the figure's top double-length slit. The bright, almost-vertical lines are OH emission features, caused by the excitation of water molecules in the atmosphere. Also visible are faint, horizontal lines — these correspond to the radiation from the target objects, which in this case have continuum (smooth) spectra. As in this figure, it will typically be the case that the sky background radiation is much brighter than that from the astronomical objects being studied — it is the job of the data analysis process to extract this weak signal as effectively as possible.

III. SIMULATION

Since the MOSFIRE instrument is currently under construction, it was not possible to obtain data from it to use in the design and testing of the data reduction pipeline. The first stage of this project was therefore to implement a program simulating the spectroscopic-mode output of MOSFIRE, given the instrument configuration, sky state and target spectra. Figures 2 and 4 show outputs from this simulation program.

The first goal of the simulation process is to transform the spectral irradiance 'sky image' — the incoming power per unit area (of the telescope's focal surface, say, corresponding to 'area' on the sky) per unit wavelength — into the irradiance 'detector image', the incident power (at the detector surface) per unit area (we don't need to know the spectral information at the detector, since to a good approximation the detector is equally sensitive to frequencies within the range transmitted by the filters). Given the irradiance image, we can use the known properties of the detector (sensitivity, read noise etc.) to simulate the output produced for an exposure of a specified time.

We can construct the 'sky image' by assuming that the spectral irradiance due to atmospheric emission is roughly constant as a function of position, and using a theoretical model of the emission features [4] to determine this spectrum. Similarly, we can obtain theoretical / observed spectra and spatial distributions for our target objects (the simulated objects in figures 2, 3 and 4 have black-body spectra and point-like spatial distributions, to within atmospheric seeing).

To transform the 'sky image' into the 'detector image', we need to know the 'transfer function' defined by the optics of the instrument. Since the instrument is, optically, a linear system, this can be summarised by the 'impulse response function': the detector irradiance that results from unit input power being concentrated at a given sky position and wavelength. During the design of the instrument, a computerised optical model was constructed — by raytracing this model, the response function can predicted, and used in the simulator.

Though the software implementation of the simulator uses a number of approximations [5], it is written in a fairly generic fashion, and with feasible extra work may prove to be useful in the simulation of similar instruments (e.g. the MOBIE spectrograph [6] proposed for the Thirty Meter Telescope (TMT)).

IV. DATA REDUCTION PIPELINE

The overall aim of the data reduction pipeline is to go from raw instrument outputs to object spectra. To accomplish this, we need to account for the effects of the atmosphere, the instrument optics and the detector.

A. Instrument optics

Compensating for the effects of the instrument optics is similar to 'inverting' the simulation process — given a 'detector image', we want to infer the 'sky image' that produced it (at least in the positions picked out by the slit mask). We could proceed, as with the simulator, by using a pre-computed model of the instrument's transfer function and the properties of the detector. However, due to slight variations of the transfer function that may occur over time (either short-term, e.g. due to instrument flexure, or longer-term, e.g. due to degradation of mirror coatings), it is sensible to infer whether the transfer function has changed using data from the instrument.

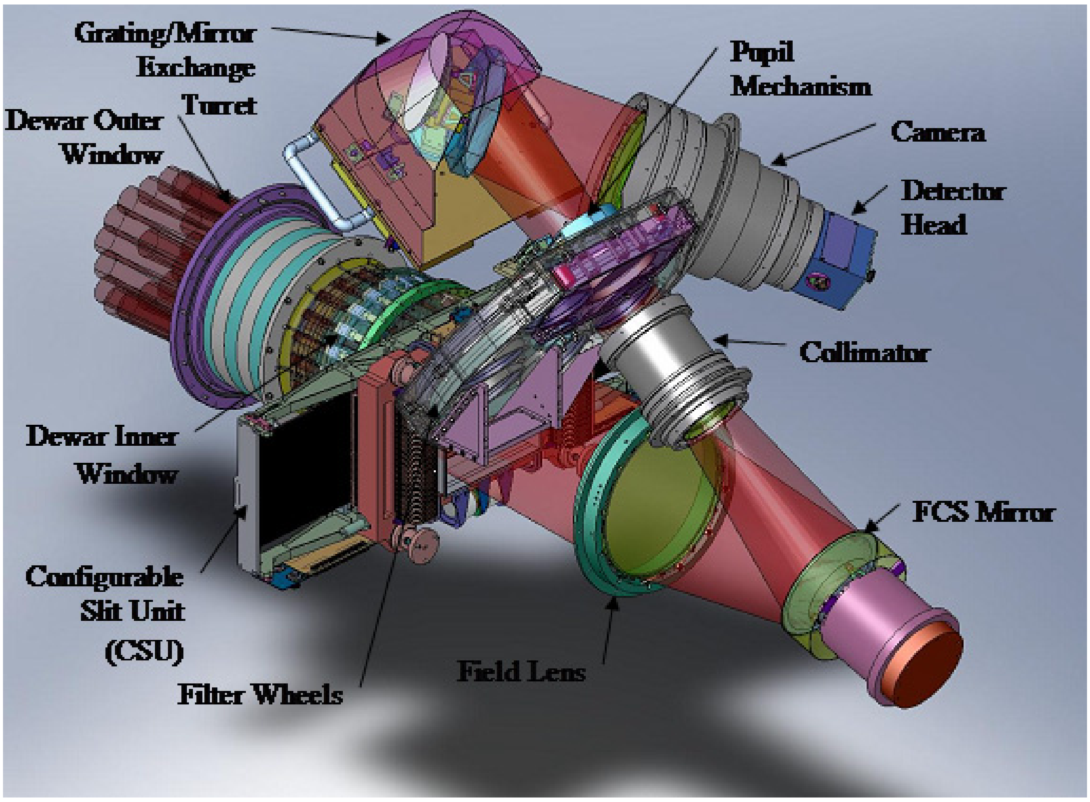

FIG. 1: Cut-away diagram of MOSFIRE.

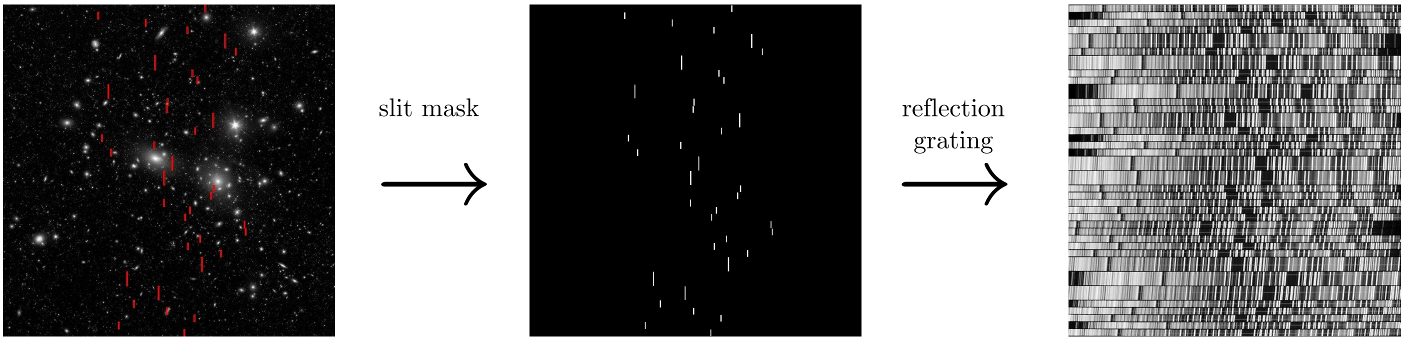

FIG. 2: Stages of image formation in the instrument. Left is a sample target field, with slit positions marked in red. Middle is the result after the slits have blocked out the light from the rest of the image, apart from the target objects, and right is the image after the reflection grating has dispersed different wavelengths of light horizontally (images are simulated).

When discussing the simulator, we observed that the transfer function describes the detector irradiance function produced by a unit power input at a specific wavelength and sky position. As with most such optical systems, MOSFIRE's response function is basically a concentrated peak — it therefore makes sense to separate the response function into 'peak location', 'peak height' and 'peak shape' components. The peak location function has the form $f : \lambda, x, y \mapsto p_x, p_y$, where x, y are the 'sky coordinates' and p_x, p_y the 'detector coordinates'. In addition, since the slit widths will typically be small, for a given y a slit will effectively pick out a specific x value, so giving a bivariate mapping $p_x, p_y = g(\lambda, y) = f(\lambda, x(y), y)$. When calibrating the transfer function using instrument data, we deal with these simplified mappings (one for every slit).

Breaking this down further, the optics of the instrument are such that we can split $\lambda, y \overset{g}{\mapsto} p_x, p_y$ into the three mappings $\lambda, y \mapsto \eta, y \mapsto p_x, y \mapsto p_x, p_y$ (where $\eta(\lambda)$ is some arbitrary strictly monotonic function of λ). Fig. 3 shows these stages graphically — the $p_x, p_y \mapsto p_x, y$ mapping removes any vertical 'distortion' introduced by the optics, the $p_x, y \mapsto \eta, y$ mapping removes any 'tilt' caused by the optics or by the tilt of the slit itself, then the $\eta, y \mapsto \lambda, y$ mapping aligns the spectrum with the physical wavelength scale. The advantage of this decomposition is that the different mappings can be inferred

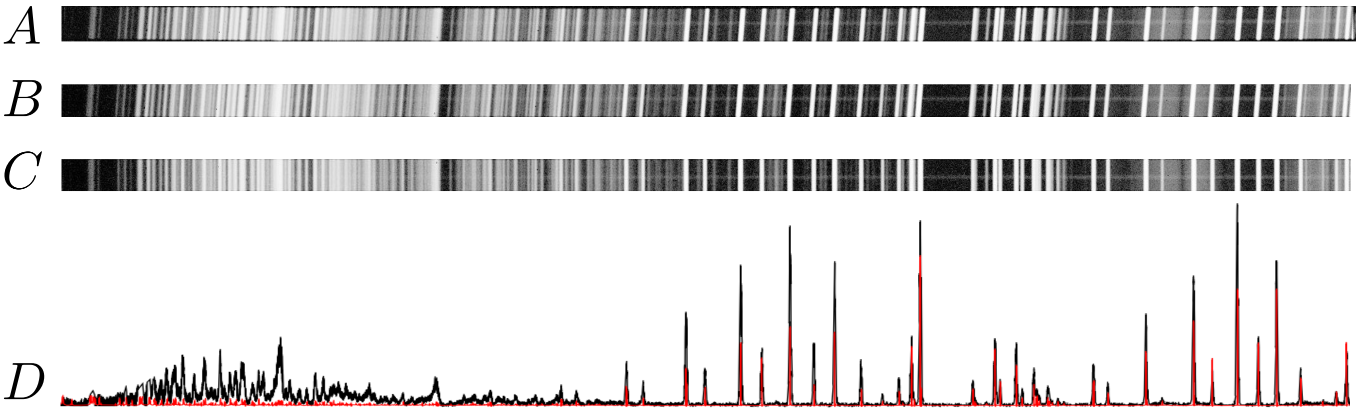

FIG. 3: Stages of calibration (where η is a strictly monotonic function of λ, with $\eta(\lambda) = p_x(\lambda, y_c)$ for y_c at the centre of the slit):
A : raw slit image in p_x, p_y (note that the top and bottom edges are slightly u-shaped)
B : resampled to p_x, y
C : resampled to $\eta(\lambda), y$
D : plot of pixel values versus $\eta(\lambda)$ (black), compared to list of theoretical OH spectral lines [4] (red).

from different parts of the data, increasing the efficiency as compared to trying to infer the whole function at once.

The above deals with the 'peak location' part of the transfer function. For the 'peak height' component (corresponding to the 'efficiency' of that transfer path), we make use of flat field exposures. These are produced by artificially illuminating the interior of the telescope dome, and pointing the telescope at a patch of this, thus obtaining an exposure corresponding to spatially uniform (and spectrally approximately-known) irradiance (a simulated example is shown in Fig. 5). If we divide the sky exposure by the flat field, we compensate for differences in the overall efficiency of transfer paths, where this also takes into account differing detector pixel efficiency. In addition, the well-defined slit-edges in a flat field exposure (e.g. Fig. 5) are a useful source of information on the $p_x, p_y \mapsto p_x, y$ mapping discussed above.

Accurately determining the 'peak shape' component of the transfer function (basically the point spread function of the optics) is slightly less important for the data reduction pipeline, as we do not attempt to compensate for it. While it would be possible to attempt to deconvolve the effect of the PSF, doing so in the presence of noise demands assumptions about the expected image, which can introduce very difficult-to-determine biases into the results.

In this project, prototype implementations of these calibration stages were constructed, and tested using the outputs from the simulator (Fig. 3 shows intermediate stages from the software processing a slit from the simulated image of Fig. 4).

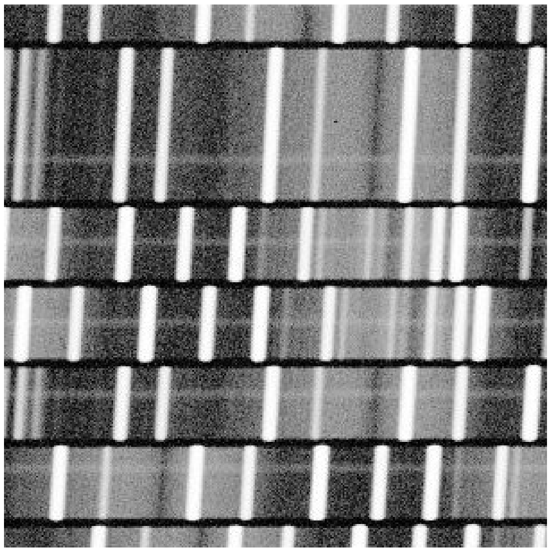

FIG. 4: Close up of simulated spectroscopic-mode detector image, K-band, showing atmospheric OH emission lines and continuum sources (image is histogram-equalised).

B. Detector

As discussed above, the use of flat field exposures should compensate for variations in sensitivity between detector pixels. However, there is also the issue of 'cosmic rays' — the generic term for events which deposit (or sometimes remove) unpredictable amounts of charge into one or more pixels. These can be caused by genuine cosmic rays (i.e. energetic particles originating from space), or by events closer to the detector.

There are a number of approaches to dealing with cosmic rays. Since MOSFIRE's detector supports frequent read-outs, it would be possible to monitor the sequence of values from each pixel, noting any sudden jumps likely to have resulted from cosmic rays and disregarding the contaminated measurements. Similarly, if multiple exposures of the same object are available, the low probability of the same pixel being hit in multiple exposures means

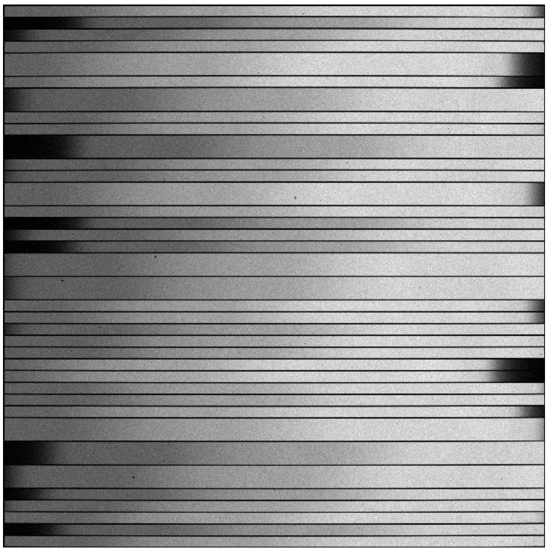

FIG. 5: Simulated spectroscopic-mode flat field.

that an approach similar to median-filtering can eliminate cosmic ray artifacts. Alternatively, morphological operators can be applied to a single image to detect such artifacts. Since the precise structure of a MOSFIRE observing session has not yet been fully defined, software for this stage has not yet been implemented.

C. Atmosphere

As discussed in section II and shown in figure 4, atmospheric OH emission lines are a strong feature in near-infrared spectra, and need to be removed in order to obtain clean object spectra. If a series of temporally close exposures are available, with the telescope moved so that the object occupies different positions between exposures, and the relative strengths of sky OH lines change sufficiently slowly, then taking the differences between temporally adjacent exposures (or some more sophisticated version of this) should remove the sky background, leaving positive and negative copies of the object spectrum. Alternatively, if sufficiently close exposures are not available, the sky background can be modelled on the basis that it is spatially approximately constant, unlike the localised target objects, and so subtracted with (hopefully) little error.

In addition to emission features, the atmosphere also has absorption features — that is, it transmits different wavelengths with different efficiencies. If it is necessary to compensate for these effects, an object whose spectrum is well-known (e.g. a solar-type star) can be observed, and the theoretical spectrum used to compensate for the atmospheric (known as telluric) absorption features. Again, due to uncertainty about the eventual operating procedure and lack of project time, these parts of the pipeline were discussed but not implemented.

V. DISCUSSION

This project has produced a realistic simulator for spectroscopic-mode MOSFIRE observations, along with a prototype implementation for the optical calibration stages of the data reduction pipeline, and investigations of algorithms for other parts of the pipeline. These will hopefully form the basis for a fully-fledged, mostly-automatic pipeline, to be implemented by the MOSFIRE team once physical construction of the instrument is close to completion.

Methods

Detailed documentation of the algorithms implemented, along with the source code itself, is available from the project's home page [5]. The source code is written in Python, using the SciPy and PyFits libraries.

Acknowledgments

This work was carried out under the Summer Undergraduate Research Fellowships (SURF) program at the California Institute of Technology, as part of the Caltech-Cambridge Exchange scheme. The author would like to thank his mentor, Professor C. Steidel, for valuable discussions and assistance, and Dr N. Konidaris and other members of the MOSFIRE team for their collaboration. He also wishes to thank Lauren Stolper, the ISP/SFP staff at Caltech, and all those who made the Cambridge-Caltech exchange possible.

[1] I. S. McLean, C. C. Steidel, K. Matthews, H. Epps, and S. M. Adkins, SPIE conference (2008), MOSFIRE: a multi-object near-infrared spectrograph and wide-field camera for the Keck Observatory.

[2] J. B. Oke, J. G. Cohen, M. Carr, J. Cromer, A. Dingizian, F. H. Harris, S. Labrecque, R. Lucinio, W. Schaal, H. Epps, et al., Publications of the Astronomical Society of the Pacific **107**, 375 (1995), The Keck Low-Resolution Imaging Spectrometer.

[3] I. S. McLean, in *Proceedings of SPIE, Volume 3354* (1998), p. 566, Design and development of NIRSPEC: a near-infrared echelle spectrograph for the Keck II telescope.

[4] *Observing Condition Constraints: Sky Background*, `http://www.gemini.edu/sciops/telescopes-and-sites/observing-condition-constraints/sky-background`.

[5] *MOSFIRE Data Reduction Pipeline — project page*, `http://code.google.com/p/mosfire/`.

[6] R. A. Bernstein and B. C. Bigelow, in *Proceedings of the SPIE, Vol 7014* (2008), An optical design for a wide-field optical spectrograph for TMT.