Spectroscopy

1. Introduction

Spectrometers divide the light centered at wavelength $\lambda$ into narrow spectral ranges, $\Delta \lambda$; if the resolution $R = \lambda/\Delta \lambda > 10$, the goals of the observation are generally different from those in photometry. In making this statement, we are pretending that the definition of $\Delta \lambda$ is obvious, whereas it can depend on the characteristics of the instrument and the resulting profile of the response as a function of wavelength to a monochromatic input signal. To be more rigorous, we can define this instrumental profile (analogous to the point spread function for imaging) to be $P(\lambda)$ and then

$$\Delta \lambda = \int \frac{P(\lambda) d\lambda}{P(\lambda_0)}$$  \hspace{1cm} (1)

In cases where $P(\lambda)$ is well-determined and stable, spectral information can be extracted to finer resolution than $\Delta \lambda$ if the signal to noise is sufficient. As in astrometry, line centers can be measured to an accuracy approaching $\Delta \lambda$ divided by the signal to noise. By modeling or deconvolution it is also possible to probe whether a measured profile is produced by a number of lines at slightly different wavelengths and blended by the limited resolution.

There are three basic ways of measuring light spectroscopically:

1.) Differential-refraction-based, in which the variation of refractive index with wavelength of an optical material is used to separate the wavelengths, as in a prism spectrometer.

2.) Interference-based, in which the light is divided so a phase-delay can be imposed on a portion. When the light is re-combined, interference between the two components is at different phases depending on the wavelength, allowing extraction of spectral information. The most widely used examples are diffraction grating, Fabry-Perot, and Fourier spectrometers. Heterodyne spectroscopy also falls into this category, but we will delay discussing it until we reach the submillimeter regime in Chapter 7.

3.) Bolometrically, in which the number of charge carriers generated when a photon is absorbed in the detector, or the energy of that photon, are sensed. These methods are widely applied in the X-ray and will be discussed in Chapter 9.

In addition to spectral information, most spectrometers also provide at least some information about the distribution of the light on the sky. Where a full image can be obtained along with the spectral data for each point in the image, we imagine a three dimensional space called a data cube, with spectra running in the $z$ direction and such that any slice in an $x,y$ plane produces an image of the source at a specific color. The ability of spectrometers to produce data cubes varies. Simple grating or prism spectrometers usually use a slit to avoid overlap of the dispersed beam sections and therefore only provide spatial information in one direction, that is they produce the $x,z$ part of the data cube directly. To some extent this shortcoming can be mitigated with an image slicer or integral field unit (IFU) that
rearranges the image of the source so part of it in the direction perpendicular to the slit can enter the spectrometer and be dispersed. For a Fabry-Perot spectrometer, the x, y slices of the cube are produced directly, that is the instrument images in a narrow spectral band (actually a series of such bands, but all but one are usually eliminated to avoid spectral overlap). The Fourier Transform spectrometer, if used with a detector array, yields a value of the Fourier transform of the z axis and the x and y axes simultaneously; by scanning the path difference the full transform is obtained over the field of view. This output can be inverted to provide the full spectral data cube.

As discussed in the introduction to Chapter 2, the etendue, $A\Omega$, is preserved in any perfect optical system. For a spectrograph, we have the further restriction that the achievable spectral resolution goes as the etendue. Thus, to operate efficiently, a spectrograph must match the etendue of the telescope. As a result, spectrographs of a given resolution and field of view must scale in size with the telescope; spectrographs on 10-m class telescopes are huge, very expensive instruments.

2. Prism Spectrometers

Light passing through a prism is deflected as in Figure 1. The input and output beams are bent by Snell’s Law, and the refractive index determines the net bending of the beam. An empirical formula due to Sellmeir for the wavelength dependence of the refractive index is

$$n^2(\lambda) = A + \frac{B_1\lambda^2}{\lambda^2 - C_1} + \frac{B_2\lambda^2}{\lambda^2 - C_2} \quad (2)$$

where $A$, $B_1$, and $C_1$ are constants. As a result the net deflection of the beam depends on the wavelength – the output beam deflections separate into the colors of the light.

This behavior allows construction of perhaps the simplest form of spectrograph, Figure 2, which illustrates many of the essential features of dispersive spectrographs in general. Starting at the detector and working forward, the camera lens maps the range of dispersion angles emerging from the prism into spatial positions on the detector, so each position corresponds to a specific wavelength. The prism must be put in a collimated beam, otherwise the range of emergent positions from the prism will cause these positions to overlap for different wavelengths, as can be seen from Figure 1 compared with Figure 2. The collimator and camera lenses together act as a simple relay optic, so if the prism were...
removed the instrument would map positions onto the sky onto the detector. Here again, we have to restrict this mapping so positions parallel to the dispersion direction of the prism do not get mapped onto other positions but at different wavelengths. There is no problem, however, in mapping sky perpendicular to the dispersion onto the detector. Thus, a slit is needed to restrict the amount of the sky that is imaged through the instrument.

Another solution to the need for collimated light into the prism is to place it over the entrance to the telescope, creating an objective prism spectrometer. It is not possible to interpose a slit into this arrangement, so it has the disadvantage that the sky is imaged on top of the spectrum of the source. For example, if there are 100 spectral elements in the spectrum of a source then 100 sky positions at different wavelengths will be imaged to the same place on the detector, raising the background by a factor of 100. This approach is a way to get spectra of a huge number of sources simultaneously (e.g., the stellar spectra Annie Jump Cannon used to establish the spectral classes). However, for modern work it has largely been supplanted by multi-object spectrographs that replace the conventional slit with a large number of individual small fields and arrange their light to enter the spectrograph in a way equivalent to a slit spectrograph.

3. Interferometric spectrometers

3.1 Diffraction gratings

Prisms have limited design flexibility, since the spectrograph must utilize the spectral dispersion properties of optical materials that have good transmission, can be manufactured in large sizes and to high quality, and so forth. A more flexible family of spectrometers that dominates in astronomical applications is based on diffraction gratings. The simplest form of these instruments is similar to the design of Figure 2, with the substitution of a grating for the prism. Therefore, we begin the discussion by concentrating on the properties of a diffraction grating.

The principle behind a grating is to slice the collimated beam in the spectrometer spatially, in narrow strips parallel to the slit. Although gratings can be made that do so in transmission, a broader range of possibilities is possible if the slicing occurs in reflection. We therefore redraw the spectrograph as in Figure 3, with a layout that uses mostly reflecting optics (except for the corrector plate).

If we can understand how the grating works, then the rest of Figure 3 becomes easy to understand! It is conventional to discuss a grating in terms of an analogy with Young’s experiment (diffraction and interference through two slits) or, as in

![Figure 3. Schematic layout of a grating spectrometer.](image)
The incoming beam propagates through the slits, emerging from each as a pattern of Huygens wavelets that interfere at the screen to the right. For the light that proceeds to the screen in the same direction as the incident beam, all the wavelets for a given wavelength have the same path from slit to slit and interfere constructively. This gives the order $m=0$, and no spectral information, so it is sometimes called the white light peak. As we move down the screen, first we encounter a region where virtually all the interference is destructive, and then the first order peak, $m = 1$, where the path difference from slit to slit is one wavelength. However, now since the relevant measure of path lengths from the slits is in units of wavelength, we reach the constructive interference peak for $\lambda_1$ before the one for $\lambda_2$. As we proceed further down the slit, progressively encountering peaks for 2, 3, and 4 wavelengths path difference slit-to-slit, the spread between $\lambda_1$ and $\lambda_2$ grows, i.e., the spectral resolution gets larger.

We can understand the behavior mathematically by observing that the path difference between successive slits is

$$ p = d \sin i + d \sin \theta \quad (3) $$

leading to a phase difference of

$$ \delta = \frac{2 \pi p}{\lambda} = \frac{2 \pi d (\sin i + \sin \theta)}{\lambda} \quad (4) $$

Figure 4, an expanded version with more slits.

Figure 4. Diffraction and interference through multiple slits, indicated for two wavelengths ($\lambda_2$ shaded and $\lambda_2$ clear, $\lambda_1 < \lambda_2$).
The peaks of intensity occur when \( p = m\lambda \), that is the path difference between successive slits is an integral number of wavelengths, with the integer the order, \( m \). If \( a \) is the amplitude diffracted by a single slit, then the amplitude, \( A \), for \( N \) slits is the sum:

\[
A e^{i\phi} = a(1 + e^{i\delta} + e^{i2\delta} + \ldots + e^{i(N-1)\delta}) \\
= a\frac{(1 - e^{iN\delta})}{(1 - e^{i\delta})} \\
\tag{5}
\]

where \( \phi \) is the phase. If we assume the beam is incident perpendicular to the row of slits (just to simplify the final result), the intensity is

\[
I(\theta) = AA^* = \frac{a^2 \sin^2(N\delta/2)}{\sin^2(\delta/2)} = a^2 \left[ \frac{\sin^2 \left( \frac{N \pi d \sin \theta}{\lambda} \right)}{\sin^2 \left( \frac{\pi d \sin \theta}{\lambda} \right)} \right] \\
\tag{6}
\]

where \( A^* \) is the complex conjugate of \( A \). The term in square brackets gives the profile of the peak of the intensity – the instrumental profile, \( P(\lambda) \). The intensity of the light diffracted by a single slit, \( a^2 \) is the sinc function,

\[
I_{\text{slit}}(\theta) = a^2 = I_0 \frac{\sin^2 \left( \frac{\pi w \sin \theta}{\lambda} \right)}{\left( \frac{\pi w \sin \theta}{\lambda} \right)^2} \\
\tag{7}
\]

where \( w \) is the slit width. The envelope of the peaks at all the orders has this form (see Fig. 4)

Although we have a way to distribute the wavelengths in the incoming beam, progressively less light gets through as we go to higher orders and larger \( \theta \), as shown by Figure 4 and equation (7). Therefore, we need to modify the approach if we are to get good efficiency (plus, a field of slits throws away most of the incident light before we even get to discuss it).

To do so, we replace the slits with long thin mirrors – other than the nuisance of keeping the input and output beams separate, this change should make no difference to the physics of the multiple slit device. If we in addition tilt the mirrors so the specularly reflected direction is just toward one of the orders \( m > 0 \), then we peak the single slit response in equation (7) on that order. We have just built ourselves a diffraction grating blazed for the order we have selected for the mirror tilt.

Figure 5 shows some of our little tilted mirrors. We can adapt equation (3) to this figure to obtain (making the substitution of \( m\lambda \) for \( p \))

\[
m\lambda = d(\sin \alpha + \sin \beta) \\
\tag{8}
\]

where \( \alpha \) is the angle of the incoming beam to the grating normal and \( \beta \) is the angle of the outgoing beam (see Figure 5). The sign convention is that \( \alpha \) and \( \beta \) have the same sign if they are on the same side.
of the grating normal (as in Figure 5); $m$ can be either positive or negative with no difference in meaning. Equation (8) is known as the grating equation and will be used for a number of derivations below. In Figure 5 $\delta_0$ provides the path difference between mirrors to put the center of $l_{slit}$ (equation (7)) on the desired order, e.g., $\delta_0 = \lambda/2$ for $m=1$. That is, the blaze is optimally effective for just one wavelength, and we say the grating is blazed at first order for this wavelength. Fortunately, the sinc function in equation (7) falls off slowly with $\theta$, so the grating can have good efficiency for a substantial range of wavelengths around the blaze wavelength.

Figure 6 shows the efficiency of a real grating, in this case one ruled at 600 grooves/mm and blazed at 1.2 $\mu$m. As might be expected from our discussion of wire grid polarizers in the preceding chapter, gratings reflect different polarizations somewhat differently. We show the first four orders. In first order, the range of good efficiency is about 80% of the blaze angle. From equation (7), when operated at higher orders and proportionately shorter wavelengths, the fractional region of high efficiency should decrease in proportion to the order, as seen.

We now consider the resolution of a grating in more detail. We get the angular dispersion of the grating by differentiating equation (8) with regard to $\beta$:

$$m \Delta \lambda = d \cos \beta \Delta \beta \quad (9)$$

or

$$\frac{\Delta \beta}{\Delta \lambda} = \frac{m}{d \cos \beta} = \frac{\sin \alpha + \sin \beta}{\lambda \cos \beta} \quad (10)$$
Figure 6. Efficiency for the first four orders of a 600 g/mm grating blazed at 1.2μm, from Newport/Richardson Labs.
We will now do a similar calculation for the change in slit width on the sky, $\Delta \theta$. We start with the change in $\Delta \beta$, the resolution, with changes in the slit width as viewed from the grating, $\Delta \alpha$. We determine this relationship by differentiating equation (8) keeping $\lambda$ fixed:

$$
\Delta \beta = \frac{-\cos \alpha}{\cos \beta} \Delta \alpha \\
= - \frac{F \cos \alpha \, \Delta \theta}{F_1 \cos \beta} \\
= - \frac{D \cos \alpha \, \Delta \theta}{D_1 \cos \beta}
$$

(11)

where $F$ is the focal length of the telescope and $F_1$ that of the collimator, so $F \Delta \theta = F_1 \Delta \alpha$. The last step arises because the f/# of the beam does not change passing through the slit – see Figure 7. The resolution of the spectrograph is then

$$
R = \frac{\lambda}{\Delta \lambda} = \frac{D_1 (\sin \alpha + \sin \beta)}{D \cos \alpha \, \Delta \theta}
$$

(12)

The best spectrograph performance is often obtained with $\alpha = \beta$. Then we have

$$
R = \frac{2 D_1 \tan \alpha}{D \Delta \theta}
$$

(13)

Equation (13) makes spectrograph design very simple. It says that the resolution of a spectrometer on a telescope of diameter $D$ depends only on the diameter of the collimated beam, $D_1$, the width of the slit projected onto the sky, $\Delta \theta$, and the tilt of the grating.
Some improvement over this situation can be achieved by illuminating the grating at large angles off the normal, see Figure 8. To achieve reasonable efficiency, the illumination needs to be from the direction where the cuts between grating facets are not seen, “no shadowing” in the figure. At large angles, a given input collimated beam can spread over a much larger area of the grating than the simple cross section of the beam, a situation called anamorphic magnification. If the camera optics are large enough to capture all this light the spectral resolution can be enhanced in proportion. This gain can be seen most easily in equation (12), where large values of $\alpha$ can significantly increase the resolution through the $\cos \alpha$ term in the denominator.

![Figure 8](image.png)

Does this make spectrographs complex? No, they remain simple if we say the resolution: 1.) goes inversely as the size of the telescope; 2.) increases with increasing tilt of the grating; 3.) goes inversely as the field projected onto the sky, i.e., the slit width; and 4.) goes in proportion to the diameter of the beam delivered to the camera.

There is another detail to be dealt with before pronouncing our simple spectrograph ready to go into use (and ourselves ready to abandon it). The various orders reflected off the grating will all be at the same angle. From equation (8), they are a result of freezing $\alpha$ and $\beta$ and then letting $\lambda$ change inversely with $m$ to preserve the equality. Leaving all these wavelength ranges piled on top of each other would be disastrous scientifically. The simplest cure is to superpose an “order blocking filter” in the spectrometer optical train. In Figure 7, we leave the possibilities for something more complex open and call it an order sorting device.

3.2 Gratings

The principles described above can be implemented in a variety of grating types.

3.2.1 Ruled, holographic, and volume phase gratings

Most astronomical gratings are ruled. This term refers to a process in which a fine tool (diamond so it does not wear) controlled by a screw drive is drawn over a softer substrate to carve the grooves. The initial master grating is very expensive to manufacture, so one normally uses a grating replicated from it as an epoxy resin casting. Master gratings can also be manufactured holographically. In this case, an interference pattern is projected onto a light sensitive coating on the grating substrate. After exposure, this coating is developed to remove the unexposed regions, leaving a series of grooves. Because there is no mechanical removal of material, holographic gratings have low levels of scattering. However, it is difficult to blaze them as effectively as for ruled gratings and consequently they generally have lower efficiency. Gratings can also be made for the optical and near infrared by imposing a sinusoidal variation of the refractive index in a material by holographic techniques. These volume phase holographic (VPH)
gratings are made in gelatin that is sandwiched between plates of glass for protection. Whether the grating operates in transmission or reflection depends on the orientation of the diffracting fringes. Because they are encased in glass covers, these gratings are more rugged than other types.

3.2.2 Grisms

A convenient disperser might be placed in a filter wheel of an imager and could be rotated into the beam to convert from imaging to spectroscopy. However, this application requires that the dispersion occur more or less without deviating the beam, a condition not satisfied by a prism or a transmission grating. A solution is to replicate a transmission grating on the hypotenuse of a right-angle prism as shown in Figure 9. The deflection of the former is canceled by the latter for a specific wavelength, which can be set to the center of the desired spectral range. The performance of the grism is described by the equivalent to the grating equation:

\[ m\lambda = d(n - 1) \sin \phi \quad (14) \]

where \( \alpha = -\beta = \phi \), which is the angle the hypotenuse of the prism makes to the side normal to the beam, \( n \) is the refractive index of the prism and grating material (the grism is assumed to be immersed in a medium with a refractive index close to 1), \( m \) is the order the grating operates at, and \( d \) is the groove spacing.

3.2.3 Echelle gratings

Our discussion of gratings implicitly assumed that the spectrometer would be operated at a low order. From our statements at the end about spectrometers being simple, if we want high spectral resolution our choices are to use a small telescope, a large grating tilt, a narrow slit, or a large diameter beam in the spectrometer camera. If we want to look at very faint sources with reasonable efficiency, the first and third options are not palatable, and the size and cost of the spectrometer grow rapidly if we make the camera beam large. Therefore, our first approach should be to operate the grating at a large tilt angle, with a large but not Federal-Reserve-threatening beam diameter.

Operating a low-order grating at a large tilt requires a very high groove density. Large gratings can only be made in high quality up to about 1500 grooves per millimeter (Richardson Grating catalog). As the groove spacing becomes finer than about 2000 per mm, the efficiency starts to drop. Although holographic gratings can be made with very high groove densities, in general they fall short of ruled gratings in both size and efficiency. The solution is to use an echelle grating, with coarse groove spacing and used at high order (Figure 10). To avoid overlap of the many orders delivered superimposed by this type of grating, another disperser (grating or prism) is used
operated to disperse the spectrum orthogonally to the echelle grating.

3.3 Grating spectrometers

We now describe a number of alternative spectrometer designs.

3.3.1 Standard Spectrometers

Although the demands of astronomical applications on giant telescopes drive spectrometer design to new and creative solutions to some of the optical problems, it is useful to see some of the basic and classic designs. Figure 11 shows the Ebert-Fastie design. The entrance slit is placed at the focus of a spherical mirror. The mirror will collimate the light that passes through the slit and direct it to the grating, which is plane so it disperses the light and deflects it still collimated (except for the spread of angles due to the dispersion) back to the spherical mirror. This mirror images the light onto the detector array in the form of the spectrum of the light that passed through the slit. The spectrum can be shifted on the detector array by rotating the grating (if the grooves run perpendicular to the page, then the grating is rotated in the plane of the page). For good optical performance, the angle $\theta$ should be made as small as possible so the two critical focal areas (slit and detector array) are as close as possible to the optical axis of the mirror.

This restriction is removed and the performance improved with the Czerny-Turner design (Figure 12). The light passes through the slit at B to an off-axis parabolic mirror, where it is collimated and directed to the grating or prism disperser. Another off-axis paraboloid focuses the dispersed light into a spectrum on the detector array at F. Again, the spectrum can be shifted on the detector array by rotating the grating.

A Littrow configuration uses the same mirror for collimator and camera (Figure 13). In theory, this arrangement allows
good performance from the mirror and the grating by compensating them in the double pass – however, aberrations are introduced by the necessity to offset the entrance slit from the detector array.

### 3.3.2 Some examples of astronomical spectrographs

DEIMOS, the multi-object spectrograph for Keck (Figure 14) looks complex, but there is a multi-object focal plane where up to 130 slit-lets are used to select individual objects from a field of area 81.5 arcmin². Their light then goes to a collimator, a grating, a camera (a second camera may be built in the future), and the detector array. The instrument is on the Nasmyth platform of the Keck II Telescope, on a rotator to compensate for field rotation (due to the Alt-Az mount of Keck). There is substantial flexure as the instrument rotates, which is removed to the 0.5 pixel level by a closed-loop control. The collimated beam is ~ 15 cm in diameter and the final camera is f/1.29 to provide an appropriate scale onto the detector array. A number of gratings are available (as well as an imaging mode), providing spectral resolution in the few thousand range.

Figure 15 shows the optical layout of the Gemini Multi-Object Spectrograph (GMOS). It has been designed with a different philosophy, emphasizing use of lenses. However, the basic spectrograph architecture is readily apparent. The first set of elements in GMOS correct for atmospheric dispersion. The instrument allows multi-object slit-lets over a 30 arcmin² field of view, plus a fiber-optic based integral field unit covering 50 X 50 arcsec. The collimated beam is 9.8 cm in diameter; although it uses all-refractive optics, the instrument is well corrected over the 0.35 to 1.5μm range. Flexure is compensated with an open loop correction system. Spectral resolutions of a few thousand are typical, as well as an imaging mode.
Figure 15. GMOS optical train.

Figure 16. Layout of the Keck HIRES echelle spectrometer.
Figure 16 shows HIRES, a cross-dispersed echelle grating spectrometer for the Keck Telescope. The atmospheric dispersion corrector (ADC) corrects the dispersion from earth’s atmosphere so an achromatic image is delivered to the slit. The calibration lamps and iodine cell together are used for wavelength calibration. The deckers are reflective masks that define the slit length. Behind the slit are filter wheels. Figure 17 shows the resulting spectral format.

Figure 17. Cross-dispersed echelle spectrum from the HIRES instrument
The High Accuracy Radial Velocity Planet Searcher (HARPS) at ESO is optimized for measurements of very accurate radial velocities. It is fed by two optical fibers that convey the light from the telescope focus to the input of the spectrometer optics, so those optics can be fixed and there is no flexure associated with different viewing angles. The spectrometer is inside a vacuum vessel that allows very accurate control of its temperature and also avoids calibration changes associated with variations in atmospheric pressure. It has a 208 mm diameter collimated beam that illuminates an echelle grating blazed at 75° (spectral orders near 100), with a grism cross-disperser to achieve a final spectral resolution of 120,000 over the 378 – 691 nm spectral range.

The use of fibers to convey the light to the spectrometer is a key feature of HARPS. Figure 18 shows the operation of a fiber. The fiber has two zones, with the refractive index in the cladding lower than that in the core. Over a range in input angles (within the acceptance angle, $\theta_{\text{max}}$), the light is held within the fiber by total internal reflection, while light entering outside $\theta_{\text{max}}$ is not reflected and escapes. The acceptance angle is given by

$$NA = \sin \theta_{\text{max}} = \sqrt{n_1^2 + n_2^2}$$  \hspace{1cm} (15)$$

where NA is the numerical aperture. Typical fibers have a NA of 0.22, corresponding to an input beam of f/2.

Fibers can carry light virtually losslessly for many meters, particularly in the visible to about 1.5\( \mu \text{m} \). However, as the light is conveyed down the fiber, the spread of angles is increased, a process termed focal ratio degradation. Thus, a price to be paid for using fibers is an increase in the etendue, that is the
spectrograph has to be built to accommodate a larger AΩ of input light than would be needed without fibers, requiring a larger spectrograph or one more demanding optically. By making the input solid angle relatively large, the additional degradation can be kept small, as shown in Figure 19 (Ramsey 1988). F/ ratios between 3 and 6 allow conveying the light with relatively little loss of etendue for the system, and can provide the major advantage of allowing the light to be injected into the spectrometer remotely, as in HARPS.

One of the more difficult aspects of the optics for a large spectrograph is to match the scale of the spectrum to the detector array for optimum performance. Too coarse a scale undersamples the spectral information and makes it difficult or impossible to extract it. With a CCD detector, too fine a scale requires that the pixels be combined before readout, which can preserve the intrinsic signal to noise but destroys information that is potentially available with the full array format. With an infrared array, since on-chip binning is not possible, too fine a scale reduces the signal to noise achievable in the spectrum, since more pixels need to be read out and combined to make a measurement.

In general, matching the pixel scale is a job that falls on the camera optics and can be a challenging design problem. Figure 15 shows the camera for GMOS, which uses refractive optics, as does Deimos. This option is attractive because a small amount of chromaticity in the camera optics is not a serious problem – if it is in the dispersion direction, it can just be subsumed in the corrections required to calibrate the spectra, while most spectrographs do not have large fields in cross-dispersion for individual spectra. The refractive optics have the advantage that the optics maintain the beam on-axis, allowing for a fast f/number with relatively uniform imaging qualities over the full camera field, along with modest distortion. The MMT red-channel camera in Figure 20 is of Schmidt design – with a corrector plate to compensate aberrations. The HIRES camera is of similar overall design. In the case of the red channel, the camera also provides an on-axis beam, in this case by folding it off a flat mirror and then imaging through a hole in the center of the flat. It is designed so this hole lies where the obscuration due to the secondary mirror of the telescope removes light anyway, so little or none is lost. This concept is called a Pfund design.

3.3.3 Spectrograph Optical Behavior

Grating spectrographs are, of course, subject to all the standard aberrations: spherical, coma, astigmatism. In many cases, slight degradations of the images are hidden because of the relatively large pixels and the effects of the slit and spectral dispersion. Distortion, however, is a particularly critical parameter. The extreme f/numbers required for good matching of the pixels of the detector to the projected slit can result in a substantial level of distortion, and at some level there is distortion in most other aspects of spectrograph optics. This aberration must be corrected very carefully in data reduction,
since it would otherwise result in shifts of the apparent wavelengths of spectral features and would undermine many of the applications for spectrographs. As a result, a key step in the analysis of spectroscopic data is to conduct a fit to the apparent wavelengths of lines from a calibration source and to correct the spectra for the indicated errors in the wavelength scale.

There can also be optical issues associated directly with the grating. Periodic errors in the groove spacing produce spurious lines offset from the real one and that are called ghosts. Rowland ghosts result from large-scale periodic errors, on the scales of millimeters. They are located symmetrically around the real line, spaced from it according to the period of the error and with intensity that increases with the amplitude of the error. Lyman ghosts are farther from the real line and result from periodic errors on smaller scales, just a few times the groove spacing. Satellite lines are close to the real one and arise from a small number of randomly misplaced grooves. Fortunately, modern ruling engines are often controlled through interferometers that generate signals to allow correction of ruling errors, and thus these various spurious lines are minimized in intensity. The relative intensity of some forms of ghosts grows fairly rapidly with increasing order of the real line, so although they are unimportant in low-order instruments they may be significant with high-order echelle gratings, unless the rulings are of very high quality.

Spectrometers are also subject to scattered light. This defect may arise through roughness in the optical surfaces anywhere in the instrument. Grating manufacturing flaws are additional potential sources of scattered light. Examples are roughness in the groove surfaces, dirt on the grating, or irregular groove spacing. Unlike imagers, where the two-dimensional character of the data allows removing scattered light as a natural process during data reduction, spectrometers are basically one-dimensional instruments and it can persist into the final reduced spectra and be difficult to identify. A spectrograph with significant scattered light in its spectra can give erroneous readings for fundamental properties such as equivalent widths of spectral lines; the scattered light will be removed from the lines and spread into a pseudo-continuum. Scattered light can be measured by putting a filter into the beam that blocks all light short (or long) of some wavelength and then measuring any residual spectrum in the blocked range.
3.4 Imaging Spectrographs

As described so far, grating (and prism) spectrographs are best suited for measurements of point sources that fit within their slits, or of long thin sections of a source, e.g. to determine a velocity gradient. There are two approaches to improve the spatial capabilities of these instruments: 1.) multi-object front ends that allow simultaneous spectra to be obtained of a number of sources at various positions within the field of the instrument; and 2.) integral field units that provide full imaging information simultaneously with spectroscopy.

3.4.1 Multi-object spectroscopy

There are two basic forms of multi-object spectrograph. In the first, a plate is put at the telescope focal plane with holes (or short slits) at the positions of sources. If this mask is properly laid out, it allows for the dispersed spectrum of each source without overlaps with the spectra of other sources. The spectrometer then operates in a normal fashion, except that rather than one long slit with a series of spectra for each position along this slit, it obtains individual spectra for each of the small slits in the mask.

Usually, the masks discussed above need to be prepared well in advance, and they also require the spectrograph to be mounted in the usual way on the telescope. A second approach relaxes both of these requirements – feeding the spectrograph with optical fibers. Figure 21 shows the fibers and fiber positioners for Hectospec on the MMT.
positioners for Hectospec, a multi-object spectrometer for the MMT. In this example, the fibers are located mechanically by motor driven “fishing rods” and can be reprogrammed readily. Other designs are driven by piezo-electric crystals. A hybrid approach uses a mask, with fibers installed by hand into this mask (although this is slow and inefficient). Yet another approach is to use a robotic positioner and magnets to hold the fibers in place after they have been moved to the correct position. In any case, the fibers can convey the light to a spectrograph either on or off the telescope, allowing a much broader selection of mounting arrangements. Because high-performance fibers are only available for the optical and near infrared, fiber-fed spectrographs are limited to these wavelength regimes.

3.4.2 Integral Field Units

Integral field units (IFUs) are devices that take the two-dimensional image at the focus of a telescope and re-arrange it into a format that can be received by a traditional grating spectrometer, that is arrange it to fall along a line (emulating a slit) or perhaps a set of small slits lying almost in a line except for small offsets. There are three basic types of IFU: (1.) ones using a bundle of optical fibers to sample the image and rearrange it; and (2.) ones that divide the image using mirrors or (3.) lenses. They are illustrated conceptually in Figure 22. There are a variety of implementations; we discuss one example each of types 1 and 2.

An example of the first kind has been built for GMOS on Gemini. It has a total of 1500 fibers, organized into a 5 X 7 arcsec object field and a 3.5 X 7 arcsec sky field, with each fiber sampling a 0.2 arcsec diameter part of the image and covering the 0.4 – 1.1μm spectral range. To avoid losses due to focal ratio degradation (and fiber cladding), each fiber is terminated at both ends with a tiny lens, speeding up the beam to the range that fibers can transmit without significant AΩ degradation (see Figure 23).

The second widely used type of IFU is based on lenses, or more commonly mirrors, that divide the field up and re-direct it into a format compatible with a grating spectrometer. This class of device was first
demonstrated by Bowen in 1938. The primary challenge is to fabricate and align the tiny mirrors with the necessary accuracy. The UIST IFU demonstrates a solution to this problem, and a similar unit is being implemented for the mid-infrared instrument on JWST. Referring to Figure 24, the light from the telescope enters from the right. It is deflected off the roof mirror to a small optic that substantially slows the beam down, and enlarges the telescope focal plane scale. At this large scale, it is relatively easy to make the necessary array of mirrors that slice the image into pieces and deflect it for each one in the direction required for the IFU output. These image slices are deflected off the other side of the roof mirror and through an array of baffles to eliminate stray light from one slice to another and then to the spectrometer input.

3.5 Data reduction

In general, there should be detailed instructions on reducing spectra for any instrument you might use. They will include a lot of important details individual to that instrument. Here we can only give an overview of typical procedures, assuming a simple slit spectrograph.

The reduction of data from such an instrument can start with the identical steps described in Chapter 4 for imagers, and the product will be a clean two-dimensional image from which the spectrum must be extracted by additional steps. Thus, it is necessary to obtain data, response, and dark-current frames, subtract the dark current from the other two, and divide the resulting data frame by the resulting response (flat field)
frame. Typically, the response frame will have been measured with lamps shining off a diffuse scattering screen. In any case, it will have its own spectral character. To compensate your spectrum for the spectrum of the lamps, you can observe a star of known well-behaved type and reduce it in the same way. We will call this the spectrum of the reference star.

At this point, there may be a standard set of corrections for the distortion imposed by the spectrograph optics. Next, one locates the spectrum and suitable regions for measuring the sky and extracts the measured values from each. A number of specific methods may be used for the extraction. For example, one can just take the signal in a rectangle centered on the spectrum and subtract sky as the signal from a similar rectangle (or rectangles) without the spectrum (analogous to aperture photometry for photometry). Alternatively, there may be a program that fits a profile of the appropriate cross-dispersion shape to the spectrum, analogous to PSF-fitting photometry.

Then, assuming a calibration lamp exposure or equivalent data are available, one does a similar extraction of these data (in the near infrared, the OH airglow lines can be used for a wavelength calibration without a specific lamp exposure). Next, the observed wavelengths for the calibration lines are determined by fitting a line profile to them and compared with the laboratory measured wavelengths. The wavelength errors are determined and removed from the spectrum of the object, e.g. through a polynomial fit to their run with wavelength. One should subject the reference star spectrum to the same steps. At this point, divide the target object spectrum by one of the reference star and multiply the result by a standard template spectrum of the reference star (i.e., a library spectrum that is free of all telluric atmospheric absorptions). This step removes the spectrum of the flat field lamp.

Finally, if data of standard stars were obtained, they should be subjected to the identical steps and the flux level of the new spectrum determined by comparison with the signals from the standards. Spectral lines can then be identified and extracted by fitting them with an appropriate computer code and line profile. If the spectrum is not flux calibrated, the information can be used in a relative sense, e.g. as line equivalent widths, or measuring the strength of one line compared with that of another.

To support these reduction steps, a number of precautions should be observed while taking the data. First, it is desirable to measure the spectrum on more than one set of pixels on the detector array – just as in imaging, this allows substitution of real data for data lost due to bad pixels. This goal can be achieved with a slit spectrometer by taking a series of spectra moving the source along the slit between integrations. This procedure also gives a set of sky measurements taken with the same pixels as the source spectrum that may prove to be a useful way to subtract the sky signal (as opposed to the subtraction based on neighboring regions in a single integration assumed above). If the spectrograph is subject to flexure, calibration lamp exposures and other calibration frames may need to be taken at the same elevation as the object being observed. If the spectrograph does not have atmospheric dispersion correction, one must be aware of the position angle of the slit when spectra are measured. If the atmospheric dispersion is along the slit, then the result will be a distortion of the spectrum, and the full information can often be recovered by fitting this effect. However, if the atmospheric dispersion is perpendicular to the slit, light will be lost at one end of the spectrum or the other (or both!). Additional issues arise with fiber-fed spectrometers. There is no long slit to obtain sky measurements automatically, and for very deep observations using some of the fibers to do so imposes possible mismatches with the performance of the fibers dedicated to the source photons. The safest procedure is to use pairs of fibers, one for source and the other for sky, and periodically to move the telescope to switch their roles. This allow the sky to be measured in exactly the same configuration as is used for the source.

4. Fabry Perot spectrometers

4.1 Principles of operation
Another approach to obtaining spectral information through interference is to divide the photons into two beams using partially reflecting mirrors, delaying one of these beams relative to the other, and then bringing them together to interfere. An implementation of this approach is the Fabry Perot spectrometer. This device is based on two parallel plane plates with reflecting surfaces.

Figure 25 concentrates on the region between these surfaces. At the surface to the right, a portion of the input beam is transmitted and a portion reflected, and the reflected portion is again partially transmitted and reflected at the left surface. However, interference modifies this simple picture. If the spacing, $l$, causes a $180^\circ$ phase shift between the incoming ray and $R_1$, then there will be no light escaping to the left. Under this condition, the phase shift at $T_2$ will be $360^\circ$ and the interference at $T_2$ will be constructive, so the light will escape there. As the spacing between the surfaces is changes, the amount of light escaping to the right will be changed as varying amounts of constructive interference occur at the right side reflective surface.

We now treat this process more quantitatively. First, consider a single reflective surface as in Figure 26. To the left we show a wave of amplitude $a$ incident on a surface that reflects $r$ of the amplitude and transmits $t$ of it. The wave is partially reflected, $ar$, and partially transmitted, $at$. By the principle of reversibility (Edwards 2004), if we reverse the directions of all the rays, there should be no change in the amplitudes. We show this situation to the right, where the reflection and transmission in the reverse direction are $r'$ and $t'$. We then require

$$att' + arr = a$$
$$art + atr' = 0$$

Figure 26. Reflection and transmission at a surface.

from which we conclude $r' = -r$ and $t' = 1 - r^2$. Now consider the reflection and transmission between two such surfaces (Figure 27). The amplitude for the emerging light is the sum of the contributions of each emerging ray:
\[ Ae^{i\phi} = att' + att' r^2 e^{-i\delta} + att' r^4 e^{-i2\delta} + att' r^6 e^{-i4\delta} + \ldots \]  

(17)

where \( \delta \) is the phase difference imposed by the reflections,

\[ \delta = \frac{4\pi d n \cos \theta}{\lambda} \]  

(18)

where \( n \) is the refractive index of the material between the reflecting surfaces. Applying the results derived from equations (16), we obtain

\[ Ae^{i\phi} = a (1 - r^2) (1 + r^2 e^{i2\delta} + r^4 e^{i4\delta} + r^6 e^{i6\delta} + \ldots) \]

\[ = \frac{a (1 - r^2)}{1 - r^2 e^{i\delta}} \]  

(19)

We get the emerging intensity by multiplying the complex amplitude by its complex conjugate, yielding

\[ I_{out} = \frac{a^2 (1 - r^2)^2}{1 - r^2 (e^{i\delta} + e^{-i\delta}) + r^4} \]

\[ = \frac{I_0 (1 - r^2)^2}{1 - 2r^2 \cos \delta + r^4} \]

\[ = \frac{I_0}{1 + \frac{4r^2}{(1 - r^2)^2} \sin^2 \left( \frac{\delta}{2} \right)} \]

\[ = \frac{I_0}{1 + \frac{4R}{(1 - R)^2} \sin^2 \left( \frac{2\pi d n \cos \theta}{\lambda} \right)} \]  

(20)

where the last step depends on \( \cos(2\delta) = 1 - 2\sin^2(\delta) \) and \( R = r^2 \) is the reflected intensity.

Equation (20) shows that \( I_{out} \) has maxima when

\[ m\lambda = 2d n \cos \theta \]  

(21)

where \( m \) is the order. It is also apparent that these maxima are narrower in spectral range the closer \( R \) is to 1 (the closer the reflectivity of the surfaces is to being complete). The finesse of the device is
\[ \zeta = \frac{\pi \sqrt{R}}{1 - R} \quad (22) \]

(valid for \( R > 0.5 \)). The spectral resolution to full width at half power of the transmission profile is

\[ R_{\text{res}} = \frac{\lambda}{\Delta \lambda} = m \zeta \quad (23) \]

and the free spectral range between transmission orders is the finesse times \( \Delta \lambda \). Figure 28 shows the changes in spectral performance with increasing values of \( R \) and hence of finesse.

Because of the angle dependence of the response (the \( \theta \) dependence above), the output off-axis has a ring-like appearance as different orders appear with increasing distance off-axis, see Figure 29. The performance description we have provided applies on axis. When using such a spectrometer to image a wide field, data need to be obtained at a number of wavelength settings and corrected to be combined into the final image.

### 4.2 Practical spectrometers

The natural advantage of Fabry Perot spectrometers is that they provide images in a very narrow spectral band, the appropriate data format if the science objective is to trace the distribution of an emission line on the sky. Another advantage is that they can be operated at very high order, e.g., \( m \sim 500 \), significantly higher than is practical with an echelle grating. As a result, a Fabry Perot can provide high spectral resolution in a compact package.

In the optical and infrared, Fabry Perot etalons are manufactured by evaporating suitable reflective coatings on substrates of optically transmitting material. Etalons can also be made by evaporating the reflective layers on opposite sides of a plane-parallel transmissive plate. In the far infrared, they can be built using free standing metal meshes as reflectors. The etalon needs to operate in a collimated beam, so the basic optical layout is similar to that for a dispersive spectrometer. Figure 30 is an example, the FP channel of the Robert Stobie spectrograph for the SALT. This instrument provides imaging spectroscopy over a 3’ field and at spectral resolution from 2500 to 250,000.

![Figure 29. Field properties of a Fabry Perot spectrometer.](http://www.optics-as.com/images/ch4.jpg)

Figure 29. Field properties of a Fabry Perot spectrometer.

![Figure 30. A Fabry Perot spectrometer.](http://www.optics-as.com/images/ch4.jpg)

Figure 30. A Fabry Perot spectrometer.
Fabry Perot etalons are scanned and set in wavelength either by controlling the separation of the reflecting plates, or by varying the gas pressure so the refractive index of the medium between the plates changes. Solid etalons are scanned in wavelength by tilting them. In any case, it is necessary to isolate a single spectral order so the observations refer to a single wavelength range. This isolation can be provided by narrowband filters, or by a second etalon of low spectral resolution.

We have so far discussed the advantages of Fabry Perot spectrometers for imaging in spectral lines. They have significant disadvantages also. The first is that the different wavelengths must be measured in a time sequence, so any change in the measurement conditions can affect the calibration of the measurements. As a result, they are not suitable for measuring spectral features of very small equivalent widths; such features are better measured with a dispersive spectrometer where all the surrounding wavelengths are measured simultaneously. Fabry Perot spectrometers are also relatively finicky devices, requiring a high degree of temperature control and a vibration-free environment to operate correctly. Errors in control, deviations from collimation of the incoming light, departures from perfect flatness of the reflecting plates all appear as degradation in the finesse and hence a reduction in the resolution.

5. Fourier Transform Spectrometers

5.1 Principle of Operation

There is yet another way to obtain spectra by introducing a phase difference in a beam of light. The principle behind the Fourier Transform Spectrometer (FTS) is shown in Figure 31. The beam of light is divided, ideally 50/50, at the beam splitter. One of the resulting beams is reflected off a stationary mirror and back through the beam splitter to the detector. The second beam goes to a moving mirror and from there to the beam splitter and the detector. The two beams interfere at the beam splitter as they are returned by the two mirrors – as the moving mirror is scanned, the phase of the interference changes. Thus, if monochromatic light is put in, the signal received at the detector will look like the example in Figure 31. Each wavelength in the input beam will produce a similar sinusoidal response in units of wavelength as the mirror is moved. Thus, if the moving mirror is scanned at a constant rate, there will be a unique frequency at the detector for each wavelength in the input beam. Since Fourier transformation isolates the
frequencies in a signal, carrying out this mathematical operation on the detector output converts it to a spectrum in frequency units.

To analyze this behavior we start with the amplitude of the two beams after they have been divided and re-combined:

$$E = h_1 e^{i\phi_1} + h_2 e^{i\phi_2}$$  \hspace{1cm} (24)

Assuming that the beam splitting is perfectly 50/50 ($h_1 = h_2 = h$), the resulting intensity is

$$I = EE^* = 2h^2 (1 + \cos(\phi_1 - \phi_2)) = 2h^2 \left(1 + \cos\left(\frac{2\pi\Delta}{\lambda}\right)\right)$$  \hspace{1cm} (25)

where the phase difference has been converted to an expression involving the path length difference, $\Delta = \lambda(\phi_1 - \phi_2)/2\pi$. Now we scan the moving mirror at velocity $v$, $\Delta = vt$. We then have that the phase difference $\phi_1 - \phi_2 = 2\pi ft$, where $f=2v/\lambda = 2v(v/c)$, and $v$ is the frequency of the photon. We can set the frequency $f$ by controlling the speed of the moving mirror. Thus, we can down convert the photon frequency to a range where our detector can track it.

From equation (25), the modulation sits on top of a constant signal, $2h^2$, which contains no useful information other than the intensity of the source. We discard it and then the signal for $n$ wavelengths becomes

$$I(t) = \sum_n 2h_n^2 \cos(2\pi f_n t)$$  \hspace{1cm} (26)

This has a form that can be inverted by discrete Fourier transformation to yield the spectrum.

5.2 Practical Applications

![Figure 32. The Herschel/SPIRE imaging Fourier Transform spectrometer.](image)
Fourier transform spectrometers were popular in the era of low-performance infrared detectors because of the “Fellgett advantage” (Fellgett 1971). Information on all the wavelengths is collected simultaneously. If the system is background-noise-limited, then the fact that the detector is exposed to all the wavelengths results in a corresponding high level of noise. However, if the system is limited by the detector noise, then the net noise is not increased and the resulting spectrum can have a significantly greater ratio of signal to noise than would have been obtained with the same detector measuring one wavelength at a time. Thus, in the detector-limited situation, the signal to noise is improved in principle in proportion to the square root of the number of spectral elements using the FTS (Saptari 2003). Like the Fabry Perot spectrometer, the FTS is capable of imposing very large phase differences in a relatively compact package and hence is useful for very high spectral resolution. Another advantage is that a FTS provides images simultaneously with spectra if it is used with an array of detectors. This can be described as a throughput advantage, since a grating spectrometer must use one dimension of its detector array for spectral information while a FTS can use both for imaging and encodes the spectral information in the time dependence of the signal. For this reason, the FTS is still a useful type of instrument, as is demonstrated by its use in the SPIRE instrument on Herschel (Figure 32).

There are some cautions about the performance of an FTS. The throughput and Fellgett advantages may be difficult to realize in practice (Saptari 2003). The Fellgett advantage vanishes if the system is limited by background noise, as is generally true for modern instruments. In fact, it can be replaced by the “Fellgett disadvantage” if the noise varies significantly across the free spectral range – for example if there is a large change in the background level with wavelength. If this is the case, then the higher level of noise for part of the spectrum will degrade the entire measurement. The operation of the FTS also requires that the detector be read out frequently, which can degrade their noise if they are integrating devices. In addition, the encoding of the spectrum in a FTS results in a degradation in signal to noise compared with an equivalent multi-channel spectrometer (Treffers 1977).

There is a sharp limit on the phase difference that can be obtained with a FTS, set by the constraints on the distance the moving mirror can go. As a result, the measured profile of an unresolved spectral line is a sinc function, see Figure 33. The ringing can be reduced by reducing the weight of the measurements made when the moving mirror is far from the equal-path position, at the expense of spectral resolution. This process is called apodization. A triangle weighting is perhaps the most popular but there are many other options (e.g., Wolfram Mathworld, http://mathworld.wolfram.com/ApodizationFunction.html).

6. Interference Filters

Spectral bands are defined for photometry, order blocking in spectrometers, and other applications either with colored glasses and crystal absorptions, or with interference filters. Colored glasses were the basis for Johnson’s UBVR bands, and in the far infrared a similar approach can be taken using the transmission bands of crystals. However, interference filters are more widely used because they have
better-defined transmission characteristics (sharp onset and conclusion of transmission) and can be designed to have central wavelengths and wavelength ranges as desired.

Figure 34 shows the basic building block of an interference filter, called a cavity. It is a low-order (in this case m=1) solid Fabry Perot interferometer. The layers are built up by evaporation onto a transparent substrate that provides the mechanical strength for the filter. The gap is determined by a spacer layer, which can be m/2 waves thick, with m generally small. The reflective surfaces are provided by ¼ wave layers alternating between low and high refractive index and designed to provide the desired finesse. The performance of the filter can be improved by using a series of cavities, as in Figure 35. Additional blocking of wavelengths far from the filter passband is usually provided by other filter elements based on the same principles.

It is relatively difficult to block well wavelengths more than twice the passband of an interference filter and often absorption by the substrate is used to supplement that provided by interference to assist the blocking.

As would be expected from the behavior of Fabry Perot spectrometers in general, the spectral properties of an interference filter depend on the angle it makes with a beam of light. The central wavelength moves from \( \lambda_0 \) to shorter wavelengths, \( \lambda_c \) as the filter is tilted by \( \theta \) according to

\[
\lambda_c = \lambda_0 \left( 1 - \frac{\theta^2}{2n^*} \right) \tag{27}
\]

where \( n^* \) is the effective refractive index of the filter. At large angles the transmission tends to be reduced and passband broadened. However, at small angles the behavior can provide a convenient way to tune the wavelength of the passband. The behavior also results in a degradation of the purity of the passband when a filter is placed in a beam with a large angle of convergence (small f/#). In addition, the spectral characteristics of interference filters shift slightly toward shorter wavelengths as they are cooled.

http://www.fluorescence.com/tutorial/int-filt.htm
Additional Reading


References

Volume phase gratings:

http://www.astro.ljmu.ac.uk/~ikb/vph.html


