1

The Spectroscopic Toolbox

1.1 Introduction

Present spectroscopic instruments use essentially bulk optics, that is, with sizes much greater than the wavelengths at which the instruments are used (0.3–2.4 μm in this book). Diffraction effects – due to the finite size of light wavelength – are then usually negligible and light propagation follows the simple precepts of geometrical optics. A special case is the disperser or interferometer that provides the spectral information: these are also large size optical devices (say 5 cm to 1 m across), but which exhibit periodic structures commensurate with the working wavelengths. The next subsection gives a short reminder of geometrical optics formulae that dictate how light beams propagate in bulk optical systems. It is followed by an introduction of a fundamental global invariant (the optical etendue) that governs the 4D geometrical extent of the light beams that any kind of optical system can accept: it is particularly useful to derive what an instrument can – and cannot – offer in terms of 3D coverage (2D of space and 1 of wavelength).

1.1.1 Geometrical Optics #101

As a short reminder of geometrical optics, that is, again the rules that apply to light propagation when all optical elements (lenses, mirrors, stops) have no features at scales comparable or smaller than light wavelength are listed:

1) Light beams propagate in straight lines in any homogeneous medium (spatially constant index of refraction n).
2) When a beam of light crosses from a dielectric medium of index of refraction n_1 (e.g., air with n close to 1) to another dielectric of refractive index n_2 (e.g., an optical glass with refractive index roughly in the 1.5–1.75 range), part of the beam is transmitted (refracted) and the remainder is reflected. The normal to the surface, the input ray, and the reflected and transmitted rays are in the same plane, called the incidence plane. For a ray at incidence angle (the angle between the ray and the normal to the surface) θ_i, the transmitted angle θ_t and the reflected angle θ_r are given by the very simple formulae: $\theta_r = \theta_i$ and $n_2 \sin\theta_t = n_1 \sin\theta_i$ (see Figure 1.1).

   As indexes of refraction vary with wavelength, a multi-wavelength single light ray is transmitted as a multicolored fan. Note that the transmission

Optical 3D-Spectroscopy for Astronomy, First Edition. Roland Bacon and Guy Monnet.

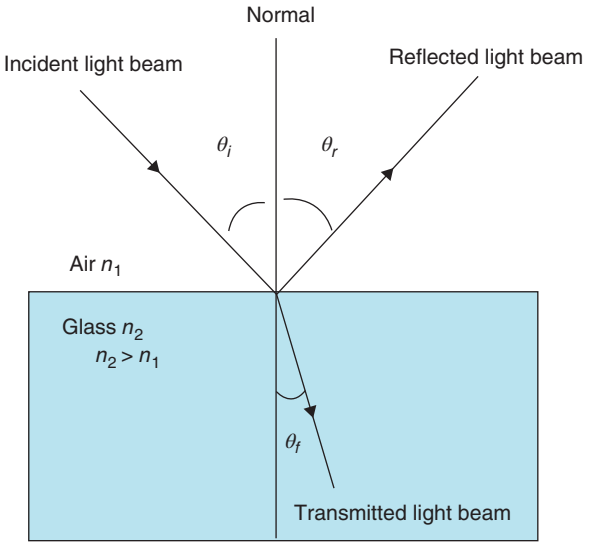


Figure 1.1 Light propagation at an air–glass interface.

formula above does not give a real θ_t value when $(n_1/n_2)\sin\theta_i$ is greater than 1. Hence, for an incidence angle greater than $\arcsin(n_2/n_1)$, there is no transmitted light and the beam undergoes total reflection inside the high refractive index medium n_1. This is a useful trick when applicable, since this is the only way to get reflection of a beam of light with 100% efficiency, provided all rays have incidence angles greater than the critical value and the glass surface is superclean.

To extend the above formulas to mirrors in an index of refraction n_1 medium, one just uses n_1 before the mirror and $n_2 = -n_1$ after (1 and -1 when the mirror is in vacuum).

3) Light is actually an electromagnetic wave that carries two orthogonal so-called polarization states, the p-state with the electric field parallel to the incidence plane and the s-state with the electric field perpendicular to the incidence plane. The laws of geometrical reflection and transmission of light are exactly the same for both polarizations, except when using the few so-called anisotropic crystals. On the other hand, the reflection coefficients R and the transmission coefficients T at the interface between two dielectrics are different for the p and the s components except for normal incidence, that is, for $\theta_i = \theta_t = 0$. They are given by the Fresnel equations:

$$R_p = \frac{(n_1\cos\theta_t - n_2\cos\theta_i)^2}{(n_1\cos\theta_t + n_2\cos\theta_i)^2} \quad R_s = \frac{(n_1\cos\theta_i - n_2\cos\theta_t)^2}{(n_1\cos\theta_i + n_2\cos\theta_t)^2} \tag{1.1}$$

From energy conservation, the transmission coefficients are $T_\mathrm{p} = 1 - R_\mathrm{p}$ and $T_\mathrm{s} = 1 - R_\mathrm{s}$.

Nearly unpolarized input light, that is, with an equal mix of p and s states is actually the most common case for artificial light sources, with the notable exception of many lasers. This is true also for most natural astrophysical sources with a few exceptions (active galactic nuclei in particular). In the unpolarized case, the

reflection coefficient is the mean value of R_s and R_p. For common optical glasses and small incident angles, this gives an about 4% light loss (percentage of reflected light) when crossing from air to glass. Many IR glasses or crystals, however, have indexes of refraction as high as 2.5, giving much higher reflection losses ($\sim$ 18%). For reasonable angles, light beams inside spectrographs remain largely unpolarized, except when a high blaze angle grating is used as seen in Section 1.4.5, unless the instrument is dedicated to spectropolarimetric investigations, using its own internal polarization device to separate the p and s beams.

It is easy to see that R_s is never equal to zero; on the other hand, $R_p = 0$ at the so-called Brewster incidence angle θ_B given by $\tan\theta_B = n_2/n_1$. For $n_1 = 1$ (air) and $n_2 = 1.5$ (typical low index glass), this gives $\theta_B = 56°$. Light rays striking a glass at Brewster incidence angle are thus fully s-polarized. Finally, at the critical incidence angle $\arcsin(n_2/n_1)$, with $\cos\theta_t = 0$, Fresnel equations give $R_p = R_s = 1$, indicating indeed total reflection of the rays for the two polarizations.

1.1.2 Etendue Conservation

Let us remind first that the solid angle of a cone of any shape is the area it subtends on a sphere of unit radius; it is thus a dimensionless quantity. In particular, the solid angle of a circular cone of light with half apex angle α is $\Omega = 2\pi(1 - \cos\alpha)$. For small values of α, this gives approximately $\Omega = \pi\alpha^2$.

The etendue or optical throughput expresses quantitatively how much a beam of light is spread out in area and solid angle. Taking an infinitesimal surface element dS immersed in a medium with refractive index n, emitting light inside an infinitesimal solid angle $d\Omega$ and at an angle θ from the normal to the surface, the resulting etendue is $d^2E = n^2 dS\cos\theta\ d\Omega$ (see Figure 1.2). Solid angles being dimensionless quantities, the etendue has the dimension of area. For a full light beam, it is obtained by integrating d^2E over area and solid angle, giving

$$E = n^2 S\Omega \tag{1.2}$$

In particular, for a light cone of half-angle α orthogonal to the surface S, we get

$$E = n^2 S\int_0^{\alpha} \cos\theta\ d\Omega = n^2\pi S\sin^2\alpha \tag{1.3}$$

This computation can be carried out in principle at any location along the optical path; in practice, for imaging systems, it is usually done either at the level of the light source itself (or any of its image) or at the level of the pupil (or any of its image). Seen from the source of light (e.g., from the telescope focal plane for astronomical purposes), this is essentially the product of the sky field area by

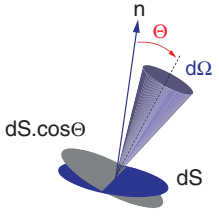


Figure 1.2 Visualization of the infinitesimal etendue component $d^2E = n^2 dS\cos\theta\ d\Omega$.

the solid angle subtended by the pupil (telescope primary mirror). Seen instead from the pupil, this is as well the product of the pupil (telescope mirror) area by the solid angle subtended by the sky field. Note that the light flux Φ carried by a beam of radiance L and etendue $n^2 S\Omega$ is $\Phi = LS\Omega$: consequently, while it is often useful to consider optical systems illuminated by a point light source ($S = 0$) or a parallel beam ($\Omega = 0$), both have no physical meaning as they carry zero energy. Actually, there is a minimum beam etendue that is set by finite light (wave)length λ. For a circular source of diameter d in air, the minimum beam half-angle set by diffraction is $\alpha \sim \lambda/d$, or a minimum etendue:

$$E = \left(\pi^2 \lambda^2\right)/4 \tag{1.4}$$

Note that this is the etendue of a diameter λ disk uniformly emitting in half-space.

The etendue concept is a fundamental and highly useful tool, because as a beam propagates inside any optical system, its etendue never decreases, being at best constant, the so-called $S\Omega$ conservation. The crucial point here is "any optical system": the light beam can be, for example, transmitted through a bundle of tapered (conical) optical fibers, sliced with multi-mirrors and then recombined; in fact it can go through any imaging and non-imaging combination you care to consider, and still at best its original etendue is conserved. For an extreme example, launch a low-etendue beam from a He–Ne laser and put a diffuser on the beam trajectory: the etendue can easily increase by a factor of 10^6 or more as almost fully collimated laser light is diffused almost uniformly over a whole half-space (a 2π solid angle). On the other hand, for imaging systems with small optical aberrations both in the source and pupil planes, the etendue computed either from pupil images or from source images is the same and nearly conserved, easily by better than one part in one thousand, as the light beams propagate inside the optical system: following the etendue along the light path, ultimately down to the detector plane, is thus a simple and powerful way to size up the optical components and the detector.

Derivation of this fundamental invariance from the two principles of thermodynamics is quite straightforward from the following thought experiment: A blackbody source of area S and radiance L is immersed in a medium of index of refraction n and emits light in a solid angle Ω, as per Figure 1.3, upper part. Light goes through a non-absorbing (perfect light transmission) arbitrary optical system and emerges through a surface S' immersed in a medium of index of refraction n', with a solid angle Ω' and radiance L'. The total flux

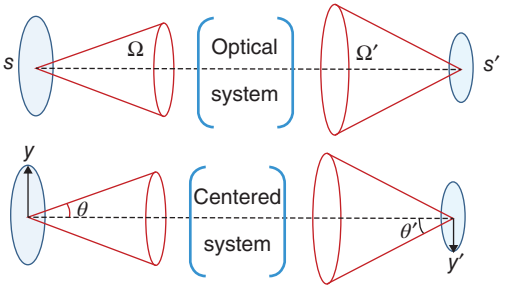


Figure 1.3 Schematic illustration of the 2D etendue conservation $S\Omega = S'\Omega'$ for any optical system and the 1D etendue conservation $y \sin\theta = y' \sin\theta'$ for any centered system.

emitted by the source is $\Phi = LS\Omega$. The total flux collected at the output is $\Phi' = L'S'\Omega'$. From the first principle of thermodynamics (flux conservation), $L'S'\Omega' = LS\Omega$. From the second principle of thermodynamics (non-decreasing entropy), $L'/n'^2 < L/n^2$; otherwise, for example, a thermocouple connecting S and S' would give an electric current with only one source of heat (the blackbody source), in clear violation of the second law. Finally, for any optics $n'^2S'\Omega' > n^2S\Omega$. Q.E.D.

Most optical imaging systems actually use centered optics, that is, with all powered (non-flat) optical surfaces of lenses and mirrors having the centers of curvature aligned along a common axis, called the optical axis. Often, all surfaces are on top rotationally symmetric along this axis, but this is not the case when, for example, astigmatic or toroidal lenses and/or mirrors are used, as for prescription glasses used to correct eye's astigmatism. For such centered optical systems and negligible aberrations in the pupil and field images, the etendue conservation actually works in two dimensions in any plane section containing the optical axis, as established below. This can be derived from Fermat's principle, namely, that light follows trajectories for which the optical path $\int n\, dl$ is an extremum: the end result is that for a small 1D source of half-length y perpendicular to the optical axis emitting light in a cone of half apex angle θ (which, on the other hand, can be very large), and any centered optical system with small aberrations, the image of the source has a half length y' and emits light in a cone of half angle θ' with $ny \sin\theta = n'y' \sin\theta'$ (the so-called Abbe's sine condition). Given that the general 2-D etendue conservation gives in that case $n^2y^2\sin^2\theta = n'^2y'^2\sin^2\theta'$, one sees that for centered systems, there is, in addition, conservation of the 1-D "linear" etendue $y \sin\theta$ in any section along the optical axis (see Figure 1.3, lower part).

OPTICAL ETENDUE #101 TOOLBOX

Optical medium refraction index n
Source circular area $S = \pi y^2$
Light cone solid angle Ω, half-angle θ.

Follow beam propagation over each source/pupil image

- For any optical system: $n^2S\Omega$ (at best) invariant
- For any centered optical system: $ny \sin\theta$ (at best) invariant

One telling illustration of etendue conservation concerns optical fibers (Figure 1.4). They are commercially available as almost unlimited length cables with three cylindrical components from center to edge: a high-index glass core of up to a few 100 µm diameter, a lower index glass cladding, and a protecting envelope. For a beam angle smaller than the fiber critical angle, θ_M, light injected in the core is trapped by total reflection and propagates to the other end, with essentially zero energy loss from the many reflections at the core-cladding interface. In practice, this angle limitation is expressed by the fiber maximum

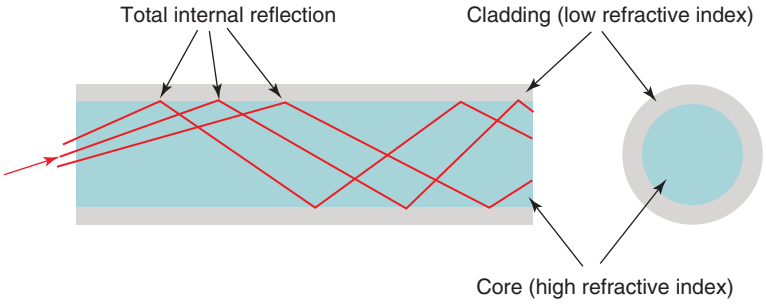


Figure 1.4 Principle of the optical fibers. Light entering the fiber core is trapped by total reflection at the core-cladding interface and propagates to the fiber end.

numerical aperture (NA) $n_1 \sin\theta_M$. It is easy to show that NA $= \sqrt{{n_1}^2 - {n_2}^2}$, where n_1 is the core refraction index and n_2 the cladding refraction index. A typical value is NA ~ 0.22, corresponding to an acceptance light cone of half-angle 12.7° in air. For astronomical applications, fiber length is relatively short, 50 m at most, and the fiber's internal light transmission is extremely good, from roughly 0.4 to 1.7 μm.

For fiber core diameters greater than $\sim$ 10 μm, geometrical optics applies, and owing to etendue conservation, beam linear etendue $n \sin\theta$ is in principle perfectly conserved as the beam propagates through and exits the fiber. In real life, it increases in case of even low fiber stress due to cable handling, and/or even gentle bending applied to carry the light beams to their required location. One typical application uses hundreds of 30-m long fibers to pick astronomical objects at the moving prime focus of a telescope some 15 m up and carry the light to a number of spectrographs conveniently located on the floor. A rule of thumb is then a $\sim$ 15% linear etendue degradation for an input beam close to maximum acceptance angle, and more for smaller angles.

1.2 Basic Spectroscopic Principles

1.2.1 The Spectroscopic Case

In the IR-optical domain covered in this book, individual photon frequencies ν are much too high (3×10^{14} Hz at 1 μm wavelength) for present technology to be directly measured with a coherent detector, as routinely done in the radio to far infrared domain. A separate coherent device is thus required to sort out the photons according to their frequency before they are sent to a non-coherent 2D detector. The detector then registers the total number of photoelectrons generated at each of its pixels during the exposure. The main figure of merit of such a spectrographic instrument is its spectral resolution $\mathfrak{R} = \lambda/\delta\lambda$, where $\lambda = 1/\nu$ is the wavelength and $\delta\lambda$ is the smallest wavelength variation that can be detected by the instrument.

In practice, this sorting out can be done either by using a filter or by using a disperser. As the name implies, a filter lets out only one wavelength slice at a time; to

get a number of wavelength bins, it is thus necessary to use a filter whose bandpass can be shifted at will (not a simple endeavor though) and make successive exposures. As the name also implies, a disperser (a grating or a prism) receives, say, a parallel beam of light and sends back dispersed parallel beams (i.e., with different inclinations for different wavelengths), which are imaged on the detector. In the astronomical domain, exchangeable interference filters coupled to an imager are widely used for multi-wavelength imaging with spectral resolutions of at most ~ 50. On the other hand, dispersers are by far the most common device used for bona fide spectrography, loosely defined as delivering a minimum spectral resolution of ~ 300.

Irrespective of their design, spectral properties of spectrographic instruments are characterized by a set of three generic values, their central wavelength λ_c, spectral range $\Delta\lambda$, and resolved spectral width $\delta\lambda$. This set gives two unitless parameters defining the instrument spectral grasp, namely, its free spectral range $R_c = \lambda_c/\Delta\lambda$ and its mean spectral resolution $\mathfrak{R}_c = \lambda_c/\delta\lambda$. The wavelength domain covered by large 'optical' telescopes (as opposed to radio-telescopes), that is, a whopping 0.3–24 μm range, is usually split in four domains: the so-called optical domain (0.3–0.95 μm); the near-IR (0.95–2.4 μm); the thermal IR (2.4–7 μm); and the medium IR (7–24 μm). They correspond to quite different instrument technologies and even science goals, with most ground-based astronomical observations performed in the first two spectral regions. In terms of spectral resolution, there are essentially four regimes: low spectral resolution (500–1500) for large surveys of distant galaxies; medium resolution ($3{,}000 - 6{,}000$) for most galaxy studies, high resolution (15,000–30,000) for precise radial velocities and/or abundance studies of individual stars or ionized gas regions, and very high spectral resolution (>100,000) for ultra-precise abundance determination in stars or in the interstellar/intergalactic medium, plus search for exoplanets. The 3D line of work explored in this book is mainly concerned with the first two regimes, that is, low and medium spectral resolution.

It is essentially impossible to build a single spectrograph that could cover efficiently the full 0.3–2.4 μm optical to near-infrared range, and most spectrographs are in fact limited to one octave at best, that is, a factor of 2 in wavelength breadth. This corresponds to a maximum free spectral range $R_c = 1.5$. Nevertheless, to cover the full optical-near infrared spectral range simultaneously, one can build, for example, a three-arm instrument with a combination of two dichroic beamsplitters[1] sending three selected spectral windows of manageable widths (e.g., 0.3–0.5 μm, 0.5–1 μm, 1–2 μm) to three optimized spectrographs; one example is the X-shooter at the European Southern Observatory (see the corresponding ESO web pages). One advantage of that multiarm approach is that short-lived phenomenas, such as γ-ray burst remnants (resulting from one of the most powerful known explosions in the Universe), can be identified over this wide spectral range in, say, a single 30-min exposure, before fading below detectivity limit.

1 A plane parallel plate with a complex set of dielectric coatings on one surface, which reflects the short wavelengths and transmits the longer ones. It is usually put at a large angle – for example, 45° – with respect to the optical beam to separate the reflected and transmitted components.

1.3 Scanning Filters

1.3.1 Introduction

The most commonly used scanning filter in astronomy is the Fabry–Pérot interferometer. This is a resonant cavity made of two parallel plane glass plates facing each other and with highly reflective ($R \geq 90\%$) internal surfaces. Cavity spacing between the two plates is usually between a few hundred microns and a few millimeters for astrophysical applications. The two outer surfaces of the plates are antireflection coated and with a $\sim 0.5^{\circ}$ wedge to prevent troublesome artifacts. As shown below, the depth of the cavity must be controlled to a very small fraction, easily 1% of the light wavelength λ, by no means a trivial endeavor at optical to near-IR wavelengths. Plates are usually made of fused silica to take advantage of its very low thermal expansion coefficient and ability to be accurately figured and exquisitely polished. Very high reflectivity coatings with low absorption are obtained in the optical region above ~ 450 nm by vacuum deposition of alternative high and low refractive index interference layers on the plates' inner surfaces; a thin gold layer is used instead above ~ 1000 nm.

A light ray of wavelength λ entering the Fabry–Pérot (FP) cavity – also called etalon – at an angle θ with respect to the normal to the cavity undergoes multiple reflections inside its cavity of optical length ne (cavity depth e & refractive index n), as shown in Figure 1.5. Parallel rays exiting the etalon interfere with each other. In the idealized case of perfect parallel plates with 100% reflectivity, that interference is perfectly constructive when the optical path difference between two successive rays is any multiple of the wavelength. This gives the canonical FP phasing equation:

$$2ne\cos\theta = p\lambda \tag{1.5}$$

Here p, the etalon order, is a positive integer, usually in the 200–2000 range. For such rays, 100% of the light is transmitted, while in all other cases the rays are reflected back to the light source. Transmitted light versus wavelength for any fixed angle θ is then a Dirac comb function with zero width peaks separated by the (slowly varying) etalon free spectral range $\Delta\lambda = \lambda/p$. In most cases, one needs to get only one wavelength peak, and the many others transmitted by the cavity need to be filtered out.

When illuminated uniformly by a monochromatic source, transmitted light appears as a series of concentric rings at infinity (Figure 1.6) of angular radii θ_p given by the phasing equation. Different wavelengths give ring patterns of different sizes; this is the basis of the two most common astronomical filters, the

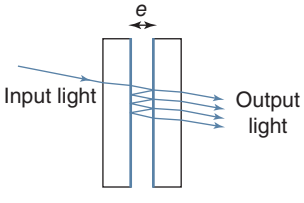


Figure 1.5 Principle of the Fabry–Pérot interferometer. Light rays are trapped by multiple reflections inside a cavity of depth e. The cavity acts as a spectral filter as only rays undergoing wavelength-dependent constructive interference are transmitted; all the others being reflected by the etalon.

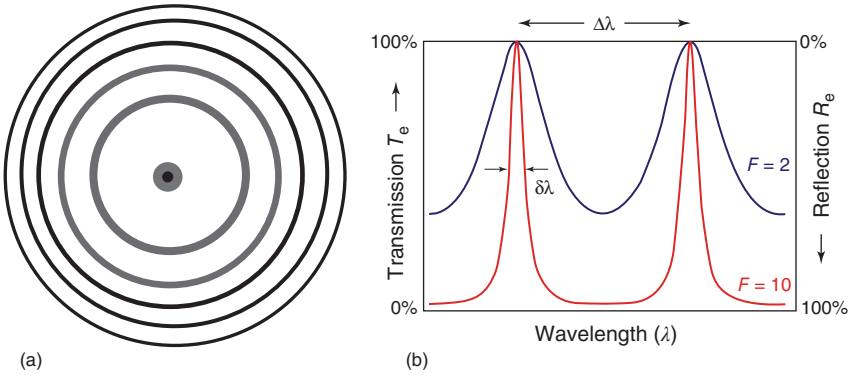


Figure 1.6 (a) Set of k rings ($k = 0$ to 5) from monochromatic light uniformly illuminating a Fabry–Pérot etalon, with exact phasing at the center ($k = 0$). The successive ring radii obey the canonical equation $2ne\cos\theta_p = p\lambda$, which for small angles θ_p translates into ring radii $r_k \propto \sqrt{k}$. With rings FWHM $\propto 1/\sqrt{k}$, rings etendues are all the same. (b) Transmission function for two etalons of respective finesses 2 (blue) and 10 (red). (Credit DrBob, Wikipedia.)

non-scanning interference filter with a fixed optical depth cavity and the scanning Fabry–Pérot interferometer with a variable depth cavity.

In real life, however, interferences are not perfectly constructive and the crucial etalon quality factor is its finesse N, the optical equivalent of the resonant cavity quality factor Q in the radio domain. This is the unitless ratio between the transmission peaks separation (aka the free spectral range) and their full width at half maximum (FWHM). At the cavity level, N can be seen as the effective number of parallel exit rays coming from a single entrance ray and interfering with each other. At the data set level, this is the number of independent spectral bins that can be distinguished. Ideally, N would be set solely by the reflectance factor R of the plates, with $N_r \sim \pi\sqrt{R}/(1-R)$. In that case, normalized transmitted light intensity T versus wavelength λ has an Airy shape (Figure 1.6), with

$$T_\lambda = \frac{1}{1 + {N_r}^2 \sin^2 \pi \frac{(\lambda - \lambda_0)}{\Delta\lambda}} \tag{1.6}$$

In this formula, $\Delta\lambda$ is the free spectral range and λ_0 the peak wavelength.

Moreover, reflective coatings slightly absorb light (absorption factor A), and the normalized peak transmission is not 1 any more, but $\sim (1 - N_r A)$ instead. For reasonable values of the reflective finesse ($N_r \leq 50$), this gives a few percentage light loss above $\sim$ 480 nm. Moreover, the effective finesse N is always smaller than N_r, owing to the cavity residual optical defects: this gives a small range for e in the canonical equation above, and hence again non-perfect constructive interference. It is also necessary to allow for a finite range of θ to fall on any detector pixel, since a beam with a zero width θ angle would have zero etendue, and hence would carry a zero photon flux. This again spoils the light beam's constructive interference and lowers N.

The end result is an overall finesse N and a global transmission τ, with a comb-like spectral transmission curve $T \sim \tau/[1 + N^2 \sin^2 \pi(\lambda - \lambda_0)/\Delta\lambda)]$.

$\lambda - \lambda_0 = 0$ corresponds to the transmission peak $T = \tau$; $\lambda - \lambda_0 = 1/N$ to a halved transmission $T = \tau/2$; $\lambda - \lambda_0 = \Delta\lambda/N$ to the minimum transmission $T = \tau/(1 + N^2)$.

The insert below gives quantitative estimates of the overall finesse, spectral resolution, and transmission of a Fabry–Pérot etalon.

FABRY–PÉROT ETALON COOKING BOOK

Plates reflectivity R, absorption A; cavity depth e & r.m.s. defects d
Beam incidence angle θ & radial angular width $\delta\theta$
Cavity order $p = 2e\cos\theta/\lambda_0$, free spectral range $\Delta\lambda = \lambda_0/p$

Reflective finesse $N_r \sim \pi\sqrt{R}/(1 - R)$; efficiency $\tau_r \sim 1 - N_r A$
Defect finesse $N_d \sim \lambda/2d$; efficiency $\tau_d \sim (N_d/N_r)\ \arctan(N_r/N_d)$
Beam finesse $N_\theta \sim 1/p\theta\ \delta\theta$; efficiency $\tau_\theta \sim (N_\theta/N_r)\ \arctan(N_r/N_\theta)$
$[\theta = 0\ \Rightarrow\ N_\theta \sim 2/p\ (\delta\theta)^2]$

Overall finesse N with: $1/N^2 \sim 1/{N_r}^2 + 1/{N_d}^2 + 1/{N_\theta}^2$

Spectral resolution $\mathfrak{R} = pN$; efficiency $\tau = \tau_r \times \tau_d \times \tau_\theta$

Transmission $T_\lambda \sim \tau/[1 + N^2 \sin^2\pi(\lambda - \lambda_0)/\Delta\lambda]$

To get a feel of what this threatening formulae mean, let us take a generic order p etalon with the reflective finesse of plate coatings N_r, and hence with a potential spectral resolution $\mathfrak{R}_r = pN_r$. It is always a good idea to get the cavity root mean square defects (due to plates figuring, polishing, and parallelism errors) as small as possible, with at least $N_d = 2N_r$. Similarly, the working field and the resolved angular size (set up by the detector pixel size) should be small enough to get at least $N_\theta = 2N_r$. With those minimum values, the insert formulas give $\mathfrak{R} \sim 0.82\ \mathfrak{R}_r$ and $\tau \sim 0.85\ \tau_r$. These are reasonably good results, which justifies a $N_d \geq 2N_r$ & $N_i \geq 2N_r$ rule of thumb to get a decent etalon performance. In practice, this means matching the etalon figuring and adjustment requirements with the plates reflectivity specification, and limiting the field of view to the maximum value compatible with the required beam finesse, and possibly less.

1.3.2 Interference Filters

Many spectroscopic systems require a spectral filter that lets through light in a fixed spectral band toward the instrument. A classical Fabry–Pérot etalon can provide any required spectral band, even exceedingly narrow ones, but cannot foot the bill because of the many other spectral bands getting through the different etalon orders. What is used instead is an interference filter, an avatar of the classical etalon, with an extremely low order ($p = 1$ or 2) solid cavity sandwiched between multilayer dielectric stacks, all created by vacuum deposition. A huge reflective finesse of up to ~ 800 can be obtained. The resulting spectral

resolution can then be up to ∼ 1000 in the optical range above about 480 nm, with a minimum transmission peak around 50%. These performances would be all but impossible to attain with the classical etalon and its maximum defect finesse ∼ 100. Such high performance interference filters are commercially available from a number of vendors and are widely used in a large variety of optoelectronics systems.

These are on top very rugged and highly stable devices, with lifetimes of usually over a decade, provided they are kept in a reasonably dry environment. One proviso is their significant temperature-related bandpass shift. The temperature coefficient is usually positive, that is, the central wavelength transmitted by the filter shifts to a higher value when temperature increases, by very roughly +0.01 nm per degree: this could in theory be used to shift the filter bandpass at will; this is, however, not practical because of the many hours required to change the filter temperature significantly in a homogeneous manner. Performance (spectral resolution and peak transmission) drops rapidly below ∼ 480 nm, because of rising internal absorption from any known high-index material. Interference filters above ∼ 1800 nm central wavelength are easily subject to delamination of the thicker coating layers, especially as they are usually used in a cryogenic environment (operating temperature of ∼ 77°K) to avoid excessive thermal emission from the filter itself.

The spectral bandpass of a simple interference filter, made with a central $\lambda/2$ optical depth cavity, sandwiched between alternative $\lambda/4$ optical depth high and low index layers, has the classical strongly peaked etalon Airy shape. By using a more complex layer set, it is in fact possible to almost get the optimum square shape, if at the expense of some peak transmission loss. Note that unwanted light, that is, outside of the filter spectral bandpass, is reflected rather than absorbed as with color filters: on the negative side, extra care must be taken to avoid that it is reflected back toward the instrument by any optical surface behind the filter; on the positive side this light might be used if needed, for example, to monitor atmospheric transmission changes.

Interference filters owe their usefulness to their large accepting etendue. Their linear optical etendue in any direction is $y \sin\theta$, where y is a circular filter half-size, and θ the maximum beam cone half-angle. Standard interference filter sizes are usually up to 2 in. in diameter (50.8 mm), but a few vendors can provide up to $\sim 185 \times 185$ mm filters, which can be mosaicked to get even larger areas. The beam half-angle θ for a filter with spectral resolution (for a parallel orthogonal beam) $\mathfrak{R}$ is readily computed from the cooking book insert as $n/\sqrt{\mathfrak{R}}$. Here, n is the cavity refractive index, about 1.5 if it is made of a low index material, and 2.4 if it is made of a high index one. Even at its maximum spectral resolution of ∼ 1000, a high-index interference filter can still accept a rather big cone of light, with a half-angle $\theta \sim 0.1$ rad. or 5.7°. For filters with a wider bandpass, the acceptance angle is larger, but limited anyway to roughly 12° because of polarization effects in the coating layers at too large angles. Note that the acceptance angle decreases only with the square root of the spectral resolution, and not with the spectral resolution when using a grating to filter the light.

There are two basic ways to insert an interference filter in an astronomical instrument, namely, either on an image of the pupil (pupil mounting) or on an image of the sky field (field mounting).

a) *Pupil mounting*. This gives the best peak transmission at the cost of a gradual shift of the central wavelength to lower values for field positions away from the field center. It is in particular preferable when the input light etendue is much smaller than the available filter etendue (say, by at least a factor of 3 in linear etendue), since the shift effect above becomes fairly negligible. Note that the optical quality of the filter then needs to be as good as that of the main instrument optics, with the corresponding technical specification as required to the vendor.
b) *Field mounting*. This is often the preferred choice since it gives the same bandpass for all points in the field (provided the filter bandpass is the same over the useful area of the whole filter, an important technical specification), if at the cost of significant light loss, $\sim 30\%$ when the input etendue is equal to the filter etendue. With this mounting, the filter figuring error budget is much relaxed; on the other hand, extra care must be taken to minimize dust particles on the two outside faces of the filter that would give artifacts on the final field image. To minimize this effect, the filters are generally put slightly out of the exact field position in order to defocus any remaining dust image on the detector.

1.3.3 Fabry–Pérot Filter

As for an interference filter, there are two ways to insert an etalon, the pupil mounting and the field mounting. Typical useful size of an etalon is 50 mm diameter, but up to 150 mm diameter plates can be manufactured.

a) *pupil mounting*. This leads to the classical Fabry–Pérot spectrograph, presented in Section 3.3, which gives spectral information over a wide field of view. Its accepting etendue is indeed extremely large, according to the cooking book recipe $\theta\delta\theta = 1/2\mathfrak{R}$ (to get the beam finesse N_θ twice larger than the final finesse N). Here, θ is the beam cone half angle and $\delta\theta$ the ring width at θ, which must cover at least 1 pixel on the detector. Taking a typical 50 mm diameter etalon, an F/1.4 camera, and a 4k × 4k detector with 12.5 μ pixels, this gives a huge beam cone half angle $\theta = 21°$ for an already large maximum spectral resolution $\mathfrak{R} = 7{,}656$. For a large 8-m diameter telescope, this means a quite sizeable $7.9'$ diameter working field. On the other hand, requirements in terms of plates figuring and parallelism quality are tough, since the $N_d \geq 2N_r$ cooking book recipe applies to the full pupil area.
b) *Field mounting*. This is rarely used, as the beam half-angle on the etalon must be at most $\sim 1/\sqrt{\mathfrak{R}}$. This leads to a much smaller field of view than with pupil mounting. Accepting linear etendue of a $2y$ diameter etalon is $y/\sqrt{\mathfrak{R}}$, obviously that of a same resolution and same diameter interference filter with cavity refractive index equal to 1. On the other hand, plates figuring and parallelism requirements are much relaxed since they now apply at the very small size of a detector pixel projected back on the etalon ($y/2048$ when using a 4k × 4k detector).

1.4 Dispersers

As their name implies, dispersers act by dispersing, that is, by changing the inclination of incoming light beams as a function of light wavelength in the so-called dispersion plane. To separate the output beams according to wavelength, the incoming beam must have a narrow angular width in the dispersion direction. On the other hand, the angular length in the orthogonal direction can be very large, as it is only limited by the field of view of the camera. The usual arrangement is then to limit the beam with a narrow, but possibly long, slit located at infinity with respect to the disperser. Note that while the slit is nearly always put on a sky image and the disperser on an image of the telescope primary or secondary mirror, the opposite combination works too. Actually, this exotic variant is used for the lenslet-based integral field spectrograph (Section 4.2).

1.4.1 Prisms

Since Isaac Newton's seminal experiences and for about two centuries, prisms have been *the* disperser used for all spectroscopic observations, including of course astronomical ones. Prisms' principle is the essence of simplicity: since the index of refraction n of all transparent materials varies with wavelength λ (actually decreasing for increasing wavelengths), the output light from a glass wedge of apex angle A illuminated with a polychromatic parallel beam (incidence angle θ) consists of a fan of parallel monochromatic beams deviated toward the prism's base and with emergent angles θ' nicely sorted out according to λ.

As illustrated in Figure 1.7, prisms are typically used in their minimum deviation (symmetrical) configuration, with the emergent beam angle θ'_c at the central wavelength λ_c equal to the entrance angle θ. This happens for $\sin\theta = n_c \sin(A/2)$, where n_c is the glass refractive index at λ_c. It is easy to show that the central angular deviation $\delta\theta'/\delta\lambda$ is then equal to $K\delta n/\delta\lambda$, with the constant K given by

$$K = \frac{2\sin(A/2)}{\sqrt{1 - n_c{}^2 \sin^2(A/2)}} \tag{1.7}$$

K can be quite high, for example, $K \sim 1.7$ for a quite typical apex angle of 60° and central index $n_c = 1.62$, but the glass dispersion $\delta n/\delta\lambda$ is always small for transparent glasses, with typically only a 2% refractive index variation from the blue (486 nm) to the red (656 nm). Near strong absorption bands, glasses might in fact exhibit large dispersions, but then would be hugely variable and associated with large light losses.

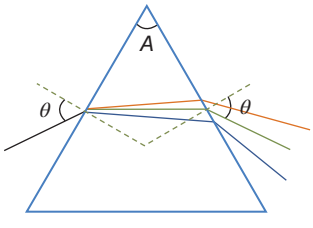


Figure 1.7 Prism's principle: The figure shows a prism used at minimum deviation for the central wavelength (green ray). This is a symmetrical configuration with similar beam incident and emergent angles θ. The extreme wavelength beams are shown in red and blue.

Linear etendue conservation in the dispersion direction between the grating plane and the detector plane can readily be used to set up the basic spectrograph parameters as per the following insert. In the orthogonal direction, the available etendue is very large and the slit length limited only by the spectrograph optics field and/or the detector length. In the dispersion direction, on the other hand, it is limited by the prism's low dispersion for any required spectral resolution $\mathfrak{R} = \lambda/\delta\lambda$. Linear etendue conservation between the prism plane and the detector plane gives the insert formulae. The prism figure of merit $K\lambda\ \delta n/\delta\lambda$ is a dimensionless number that expresses its dispersion efficiency. For the typical glass selected above, the figure of merit is ~ 0.11, an order of magnitude less than the corresponding figure for a grating (see Section 1.4.2).

Prism Spectrograph Cooking Book

Prism apex angle A; diameter d; glass relative dispersion $\Delta = \lambda\ \delta n/\delta\lambda$
$K = 2\sin(A/2)/\sqrt{1 - n^2\sin^2(A/2)}$; unitless figure of merit: $K\Delta$

Telescope diameter D, on-sky slit width α

Spectral resolution $\mathfrak{R} = (K\Delta)\ d/D\ \alpha$ (at minimum deviation)

Prism diameters can be very large, 1 m or more for a very few common optical glasses including fused silica. For diameters ≤ 30 cm, a large palette of glasses can be produced, including expensive UV and/or IR transparent crystals. Care must be taken to subject prisms to only slow homogeneous temperature changes, as glass refractive indexes are temperature dependent.[2] This can be important for large prisms, given their huge thermal inertia.

Prism transmissions are usually quite high, 90% or more if its two surfaces are antireflection coated for the mean optical ray's incidence/reflective angle. Furthermore, the spectral range potentially covered is in principle limited only by the transparency of the prism material. This means that it is possible to cover easily an 1 octave spectral range (i.e., a factor of 2 between the lowest and the highest wavelength) in one go, or even much more, if at the cost of significant reflection losses. For example, most glasses are transparent from about 0.4 to 1.5 μm, and some crystals from, for example, 0.15 to 9 μm, a whopping factor of 60. Another very attractive property is that all the light is concentrated in a single spectrum, not in a number of different "orders" as for diffraction gratings (see Section 1.4.2).

Nevertheless, as pointed out above, a big limitation of prisms is their very small angular dispersion $\delta\theta'/\delta\lambda$ (see Exercise 3): to get the relatively high spectral resolutions (a few thousands) most often required for astronomical purposes, one would need to put an exceedingly narrow slit on the object under study, thus rejecting most of its light. In addition, glass dispersion significantly decreases with increasing λ, typically by a factor of 5 over 1 octave in wavelength, with a nonideal corresponding variation of 2.5 in spectral resolution. As a result of

2 At room temperature, fused silica 633 nm refractive index temperature dependence is $+10^{-5}/°K$, equivalent to a 0.4 nm wavelength shift per °K.

these two shortcomings, prisms are now used only for a few specific cases, and diffraction gratings with much higher and more uniform angular dispersion (see Exercise 4) are chosen instead for most spectrographic applications.

1.4.2 Grating Principle

A diffraction grating is a mirror or window with a periodic structure (often called grooves), which diffracts polychromatic (i.e., made of many wavelengths) input light into monochromatic beams traveling in different directions. In the canonical case of a plane parallel beam of wavelength λ impacting at incidence θ a plane diffraction grating with a parallel straight grooves per unit length, the various output light directions θ'_λ are restricted to the values for which light scattered from adjacent elements of the grating are in phase (see Figure 1.8). The relationship between the incidence and diffracted angles in media of respective refractive indexes n and n′ can be obtained easily. This is the grating fundamental equation:

$$n \sin\theta + n' \sin\theta' = ka\lambda \tag{1.8}$$

Here, the order k is any integer – positive, negative, or null.

For $k = 0$, we recover the classical reflection law for a normal mirror: this nondispersed zero order "white" light beam is mostly a nuisance, giving bright parasitic light on the detector. In a few cases, however, it is useful, for example, to monitor sky transparency in real-time.

At optical-NIR wavelengths, it is quite convenient to express λ in micrometer: a is then the number of grooves per micrometer, with $a\lambda$, as it must be, a dimensionless number. The grating equation shows that the groove period $1/a$ cannot be smaller than $\lambda/2$: for this limiting value, θ and θ' are respectively equal to +90° and −90°, that is, the incident and the order 1 diffracted beams are both parallel to the grating surface.

The periodic structure (or grooves) that transforms a mirror in a reflection grating is made by modulating its surface shape (amplitude grating). This can be done by pushing aside a metal coating on a glass surface (surface relief gratings) or by etching a light-sensitive material illuminated by interfering laser beams (holographic gratings). Efficient transmission gratings are made by modulating the refractive index of a thin gelatine layer on top of an optical window (volume phase holographic gratings, VPHG), also through illumination by interfering laser beams.

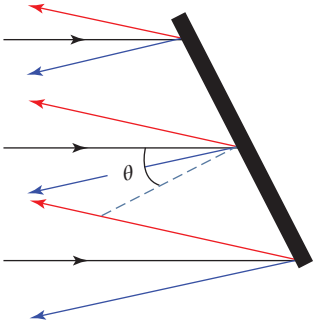


Figure 1.8 Grating's Principle's: A parallel polychromatic light beam (black rays) falls on a plane reflection grating at incidence angle θ. In that illustration, first order green rays (not shown) are diffracted back at the same angle θ along the input beam (Littrow condition), while extreme wavelength beams are shown respectively in red and blue.

Most astronomical applications require covering a wide spectral range at once. This usually leads to using diffraction gratings in their first order ($k = +1$): the grating equation above shows that one can then cover almost one octave without mixing first order diffracted light with any of the other orders, including the non-dispersed, thus always bright, zero order. Of paramount importance is then getting the best possible efficiency, that is, with most of the diffracted light concentrated in the first order. For amplitude gratings, this can be obtained by manufacturing nearly triangular grooves with facets at the so-called blaze angle ϕ with respect to the normal to the grating. Peak efficiency close to 1 is then obtained at the blaze wavelength λ_c, given by $2 \sin \phi = k a \lambda_c$. At this wavelength, the incidence angle and the first order diffraction angle are the same (Littrow condition) and equal to the blaze angle. In practice, for blaze angles up to $\sim 30°$, first order efficiency is still good (say $> 70\%$) over 1 octave. For higher blaze angles, groove size becomes comparable to the light wavelength: this results in efficiency curves for the two polarizations that are still highly peaked, but at different wavelengths. The end result for natural (unpolarized) light is smaller efficiencies, say, below 50% for a 60° blaze angle over 1 octave.

"Normal" optical systems, that is, combinations of mirrors and lenses, obey the principle of inverse return of light, both for light path and light efficiency, and that for each of the two linear polarizations. This is still true with a grating along the optical path.

1.4.3 The Grating Spectrograph

The canonical plane grating long-slit spectrograph uses the following components (see Figure 1.9): (i) a long but exceedingly narrow width slit, (ii) a collimator imaging the slit of at infinity (angular width α), (iii) a plane grating located on the exit pupil (diameter d), and (iv) a camera (aperture ratio Ω) to image the dispersed slit on the detector (projected slit width w). Note that an image of the sky is usually put on the slit with an image of the telescope primary (or secondary) mirror

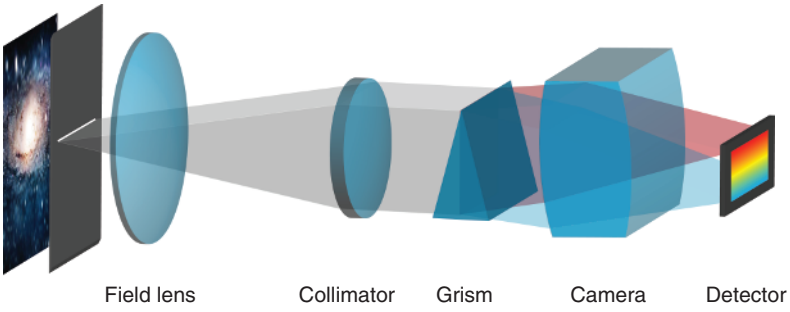


Figure 1.9 3D view of the long-slit grating spectrograph concept. (a) The 2D image at telescope focus is sliced by a long narrow horizontal slit. (b) Light on the detector is dispersed in the vertical direction. Note that we follow in this book the two usual conventions for figures showing light propagation inside an optical system: (i) whenever possible, light goes from left to right (top to bottom for a vertical optical axis) and (ii) the horizontal and vertical scales are generally not the same, usually exaggerating the angles of the optical beams for better clarity.

on the grating, but the other way would work too and has been occasionally done.

The long narrow slit, typically 3×4096 pixels at the detector level, or $0.6'' \times 15'$ on the sky or with a 4-m class telescope, is dictated by the dispersive geometry of the grating. This extremely thin shape is far from optimum though, as astronomical objects other than stars basically always have much more roundish shapes, and most of their light is just wasted when cut by the slit. On large telescopes, the light of even originally point-like stars is not always fully collected by long-slit spectrographs: star focal images have a disk shape of roughly $0.4''$ to $1.8''$ diameter depending on atmospheric turbulence, with a median value around $0.7''$ at the best sites in the world. In the example given above, more than half of the time a significant fraction of the precious starlight would be wasted.

2D etendue conservation does not however prevent sending an initially round object through a narrow rectangular field, while keeping a round pupil at infinity, provided the slit and object areas are equal, or the object is of diameter up to $24''$ for the $0.6'' \times 15'$ slit referred above. Unfortunately, this transmogrification does not conserve the 1D linear etendues and thus cannot be done with simple centered optical systems, for example, by using a set of cylindrical lenses. This necessitates developing instead complex multimirror/lenses systems that are notoriously difficult to fabricate and align (see Chapter 4 for much more on these so-called image slicers).

As seen above, for an input plane parallel light beam at angle θ with respect to the normal of the grating, an exit plane parallel beam at angle θ' for the wavelength λ will be in phase if and only if $\sin\theta + \sin\theta' = k\,a\,\lambda$; k, the spectrum order, is by principle an integral number, and a is the number of lines per micrometer on the grating plane, with λ the wavelength in micrometer.

In the $k = 0$ case, we have $\theta = -\theta'$ independent of the wavelength of light. This so-called zero-order or white light thus follows a simple reflection on the grating plane, collecting light at all wavelengths. This is not only a waste of light, but also adds a bright line imprinted on the detector. It is thus particularly important to get a grating with no more than a small percentage of light in zero order.

1.4.4 Grating Species

There are essentially two grating species, the surface relief grating (mechanically ruled or holographically imprinted) for which the periodic wavefront modulation is governed by periodic grooves depth, and the VPHG with periodic modulation of the index of refraction of a thin gelatine dichromate layer deposited on an optical plate. The first variant is now used solely in reflection, with an aluminum or gold (for the NIR) layer deposited on the front surface. The second works in transmission only, with a better efficiency than the equivalent surface relief transmission grating, $\sim 85\%$ for a $45°$ blaze angle instead of $\sim 30\%$ only. VPHGs are often sandwiched between two identical prisms so that light at the central wavelength goes directly through the disperser with zero deviation.

1.4.5 Grating Etendue

For optimum sensitivity, gratings are used mostly at or close to zero deviation (Littrow condition), that is, with $i = i' = \phi$ (blaze angle) for the central wavelength λ_c. In order to fully accept the usually circular pupil, they have in general a rectangular shape with an aspect ratio (length L over height d) at least equal to $1/\cos\phi$. Differentiating the canonical equation with respect to wavelength gives the angular width of the slit β as $\beta = \delta i = 2\ \tan\phi\ /\ \Re$, where $\Re$ is the spectral resolution $\lambda_c/\delta\lambda$. Note that in first approximation and according to the canonical equation, the spectral resolution over the spectral domain covered on the detector is proportional to λ/λ_c. This is not perfectly constant, but better than when using prisms.

Linear etendue conservation in the dispersion direction between the grating plane and the sky plane can be used to set up the basic spectrograph parameters as per the following insert. In the orthogonal direction, the available etendue is very large and the slit length limited only by the spectrograph optical field and/or the detector length. Note that the fundamental relationship between the spectral resolution and the slit linear etendue has exactly the same shape as for a prism, the only difference being the dimensionless figures of merit of different dispersers ($2\tan\phi$ for a grating).

Grating Spectrograph Cook Book

Grating height d (pupil diameter); blaze angle ϕ
Telescope diameter D, on-sky slit width α

Spectral resolution $\Re = 2d\ \tan\phi/D\alpha$ (at zero deviation)

A number of manufacturers provide off-the-shelf reflective gratings with a large variety of groove periods/blaze angles and sizes, a few up to about 30 cm height. They all are actually replicas molded in the thousands from very expensive ruled masters. Larger sizes can be achieved by mosaicking a few identical gratings, if with stringent alignment requirements. Reflective gratings can be produced for virtually any working wavelength.

VPHGs are on the other hand mostly custom made, with heights (size perpendicular to dispersion) up to about 25 cm. Mosaicking is rarely done since it is very difficult to produce two closely identical VPHG. Due to the gelatine dichromate bandpass, VPHGs can work roughly from 370 to 1700 nm. They can be operated at cryogenic temperatures when needed.

Maximum blaze angle contribution $\tan\phi$[3] is usually ~ 0.7 for a VPHG covering one octave (with simple inline optics); ~ 1.7 for a ruled or holographic reflection grating (with larger, more complicated optics however); ~ 4 for a high-order echelle grating (which however requires an additional disperser used as an order sorter). The bottom line is that the only free construction parameter is the grating diameter d, leading automatically to very large gratings (hence big

3 Higher values are technically feasible, but entail up to 50% light loss as the peak efficiency wavelengths as the p and s polarization components become widely separated.

expensive instruments) when wanting either a large spectral resolution or a substantial on-sky slit width and especially when requiring both.

w, the width of the instrument slit projected on the detector by the camera of aperture ratio Ω, usually corresponds to ~ 2 detector pixels (so-called Nyquist condition); it can occasionally be set up as large as 3–4 pixels when a high signal-to-noise ratio is required: this is in practice the province of high ($\mathfrak{R} \sim 3.10^4$) and very high spectral resolution ($\mathfrak{R} \sim 10^5$) astrophysics for which reaching minimum signal to noise ratios of 50 and 500 per resolved spectral pixel respectively is the norm. Note that because of linear etendue conservation between the grating plane and the detector plane, w is given by $w\,\Omega = D\,\alpha$.

In most cases, the spectral resolution is thus directly set up by the slit width, itself imposed by the need for collecting enough of the science object and/or the use of a high aperture ratio on the detector to get enough sensitivity. In the case of bright objects, a very narrow slit can be used in principle: getting a substantial spectral resolution with a small grating (hence a small instrument) then becomes easy. There is a limit, however, as part of the light going through the narrow slit is diffracted and begins to overfill the instrument pupil. The theoretical limit is given by a simple formula, namely, $\mathfrak{R}_M = kN$, where k is the grating order and N is the total number of grooves covered by the pupil. Settling at that limit though would result in $\sim 50\%$ light loss and a practical upper limit for the actual spectral resolution of a spectrographic instrument is in fact $\sim kN/2$, unless the grating/camera combination is enlarged (at a cost) to collect a significant part of the diffracted light.

The optimum spectral resolution $\mathfrak{R}$ for an astronomical spectrograph is very much science dependent, for example, varying from 10^3 or so for optimum detection of extremely faint objects in the near-UV to yellow region to 4.10^4 for determining abundances of key chemical elements in Galactic stars, to more than 10^5 for the indirect detection of exoplanets from minute radial velocity variations of their parent stars. See Exercise 12 for the specific analysis of absorption line radial velocity accuracy versus signal to noise ratio.

1.4.6 Conclusion

The vast majority of astronomical spectrographic instruments uses the disperser approach, almost always a plane grating, either in transmission or in reflection. This is the basic building block for the many variants presented later in this book, which differ in the shape of the sky field selected at the telescope focal plane. This ranges from a simple narrow slit (long-slit spectrography) to multislits/holes on well-separated targets (multiobject spectrography) to a single squarish field (integral field spectrography).

1.5 2D Detectors

1.5.1 Introduction

A suitable 2-D digital detector (for lack of an even better 3-D ones, see Section 7.3) is the key element of any imager and particularly the spectroimagers covered

in this book. There is a long list of perquisites for such detectors, and particularly, (i) high, ideally up to 100% quantum efficiency, that is, one electron "created" by each incoming photon, in a large wavelength range; (ii) very low dark noise (spurious electrons created during integration time) and readout noise (spurious electrons created by the detector readout electronics); (iii) large format, up to billions of detector pixels, and (iv) high linearity and high dynamical range, that is, output signals that are precisely proportional to the incoming photon flux over a wide flux range. There are also qualitative, but still crucially needed, additional features, such as being rugged (idiot/astronomer-proof), easy to use, and with few if any troublesome artifacts, such as charge blooming around overexposed pixels. This section covers the phenomenal progress enjoyed during the last 50 years and the current state-of-the-art 2D detectors landscape.

1.5.2 The Photographic Plate

Very large 2D integrating detectors were already available at the turn of the nineteenth century, namely, photographic plates or films, sensitive from the near-UV to the red domain. A single 50 cm × 50 cm photographic plate with about 12 μm spatial resolution (a typical value for plates optimized for low-level flux measurements) actually offered a whopping 1.6 billion spatial pixels, a format now just attained by the largest detector mosaics being built; see below. Besides, photographic plates are cheap, rugged, easy to use (just open and then close a shutter for an exposure followed by a few hours to develop, fix, wash, and dry), are operated at room temperature, and have the nice feature of doubling as their own data archive. Unfortunately, they also get a long list of dire shortcomings: extremely low quantum efficiency, possibly as "high" as a few 10^{-3}, but then with a very small dynamical range, plus extreme nonlinear behavior not only relative to the object flux but also with respect to integration time. Data extraction was terribly slow and of limited accuracy, adding to the overall spectacular inefficiency of this purely analog device.

Digital detectors with a much higher quantum efficiency (up to 10% in the blue), near perfect linearity, and high dynamical range were actually available soon after World War II, with the photomultiplier tubes. An incoming photon strikes a photocathode, ejecting one electron through photoelectric effect. These electrons are accelerated inside a vacuum tube, striking multiple dynodes. As a result, a single primary electron ultimately gives a hundred million electrons, which are easily detected at the end of the tube as a very short current pulse. Note that this is not an integrating detector but works by counting incoming photons on the fly one by one. Unfortunately, photomultipliers are essentially 0-D detectors, that is, offer only one spatial pixel, and thus took only a small part of the astronomical pastures, mostly to measure the integrated light flux of stars or central parts of galaxies.

1.5.3 2D Optical Detectors

The first 2D electronic optical detectors (as well as some 1-D ones) were introduced in the late 1960s, with many variants of video cameras of increasing sensitivity, ultimately up to individual photons detection. Essentially 2D versions

of the photomultiplier, the latter featured single photoelectrons counting, quantum peak efficiency ~ 10% in the blue (typically 4% in the red) domain, and respectable, but not huge, dynamics with recordable fluxes from around 1 to a few thousand photoelectron per pixel per hour. Despite their small formats (a few 10^5, typically 40×40 μm, pixels), they quickly displaced photographic plates for all astronomical work, except for completing the few long-term imaging surveys of large fractions of the sky carried out at the time.

However, soon after, in the late 1970s, these 2D photon counting detectors were themselves quickly and almost entirely replaced by the charge coupling devices, or in short CCDs, invented a decade earlier [13]. These are essentially integrating detectors, such as photographic plates but unlike photon counting ones. They thus suffer not only from dark noise but also from readout noise. Initially, CCD formats were also quite small, quite comparable to photon counting formats at the time.

Somewhat ironically, their introduction resulted initially in a significant loss in low photon flux detectivity compared to photon counting detectors, as CCDs then featured huge readout noises, around 100 electron r.m.s. per pixel (versus essentially zero with photon counting!). But, they already have much better quantum efficiencies, peaking around 50% in the blue–green, high dynamics from a few to up to 32,000 electrons per pixel, and most importantly turned out to be extremely rugged and easy to use. Their sensitivity loss for extremely low photon fluxes had actually a much smaller impact than could have been imagined: the vast majority of astronomical (and non-astronomical) observations relies on detecting a small variation in a high photon flux, rather than a very small photon flux over a negligible background. Canonical examples of the former are (i) all wide spectral band imaging for which the photon flux is dominated by the night sky background and the goal is to detect the very small flux variation due to, for example, a distant galaxy, and (ii) most absorption-line spectroscopy of galaxies for which the goal is to measure relatively slight flux dimming (due to atomic or molecular absorption) of the strong object stellar continuum. On the other hand, the canonical example of the extremely low photon flux case is high spectral resolution (typically 12,000) spectroscopy of emission lines in extremely faint ionized gas regions, a respectable scientific domain, but representing a very small fraction of all astronomical observations.

A mid-2010s top-of-the-art CCD (see Figure 1.10) is an almost ideal detector, featuring >80% QE over one wavelength octave, very small dark current ($1e^-$/px/hr or less), and readout noise ($1.5 - 3e^-$/px/readout). The dynamical range is huge – a factor of 30,000 or so, and data output is immediately available in a digital format. Note that while off-the-shelf CCDs for laymen applications (video cameras) are dirt cheap, a so-called science-grade 4096×4096 ($4k \times 4k$ in detectors lingo) 12.5 μm pixel CCD costs ~ 50,000 Euro/USD. A major cost escalation driver is the thinning of the detector: this is a time-consuming low-yield process, but one which effectively doubles the detector QE. Another is the on-chip implementation of extremely low-noise electron amplifiers to drastically reduce the readout noise of standard CCDs. Also, while commercial CCDs operate at room temperature with very short integration times (50–100 Hz

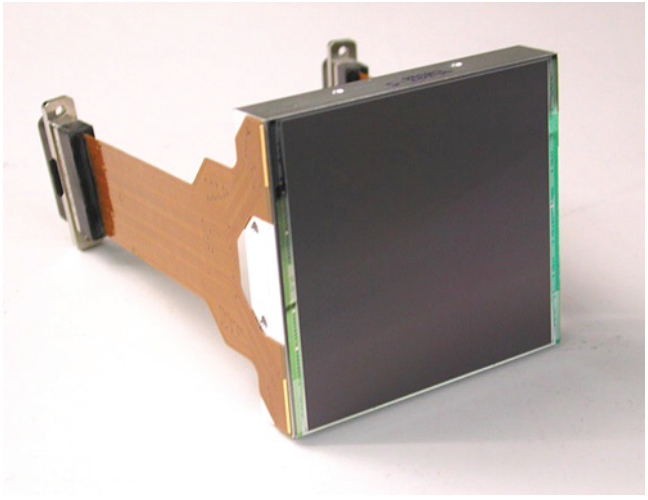


Figure 1.10 This shows the state-of-the-art 4096 × 4112, 15 μm pixels CCD231-84 from e2v, one of the leaders in the field. Since this high performance device is four-side buttable, it can be used as a building block for the development of extremely large mosaics. Credit Paul Jordan [14], e2v, the UK. (Reproduced with permission of Paul Jordan.)

video frame rate), astronomical CCDs are operated with typically 30–45 min integration times (a factor of 2.10^5!). To lower dark current to at most a few electrons per pixel per hour, detectors are cooled to around −90 °C, a significant design and operating complication: in particular, the detector is housed in vacuum inside a cryostat, with a thick entrance window, often doubling as the last lens element of the camera. A gentle dry nitrogen flow is often used to prevent frost formation on the window front face.

CCDs can be fully buttable on their four sides with very little dead space between them, and mosaics of more than a billon pixels have been built, with 4k × 4k single CCDs used as building blocks. This is essential for most imaging instruments, covering large fields on the sky, but is also increasingly required for spectrographic 3D instruments, as builder teams get more and more ambitious in terms of simultaneous spatial and spectral coverage.

CCDs "natural" spectral coverage starts from the near-UV to roughly 0.8 μm: as the light wavelength increases, the detector thin silicon material gets transparent, leading to free-falling quantum efficiency and parasitic fringing at the level of a small percentage as part of the light going through the detector is reflected back and interferes with itself. The so-called red-optimized thicker CCDs retain good quantum efficiencies up to ∼0.98 μm and exhibit much less fringing (∼0.4%). Note that most fringes arise from the highly variable upper atmosphere emission lines strongly prominent at these wavelengths and are not easily calibrated out. Another annoying feature comes from cosmic rays impacting the detectors and creating point-like bright spots on the final data: typical spot counts are ∼4 per second integration time with a standard 4k × 4k CCD, ∼20 for the thicker red-optimized one. This means some 15,000–75,000 impacts recorded for a typical 1-h exposure. Fortunately, their distinct highly peaked shape helps much to

get highly efficient rejection algorithms in the subsequent data processing phase (see Section 8.3).

1.5.4 2D Infrared Arrays

Until the early 1980s, for lack of the equivalent of photographic plates in the optical range, near IR astronomical observations above 1 μm were painstakingly performed with single-pixel semiconductor or bolometer detectors. Development for the US Air Force of 2D IR Arrays based on HgCdTe or InSb semiconductors that now feature up to 4k × 4k pixels (e.g., the Teledyne H4RG array) has been a tremendous bonus, as soon as these (at first much smaller) detectors hit the civilian market. They remain relatively expensive, with a cost per pixel one order of magnitude higher than for optical CCDs. Like CCDs, they are fully buttable and large arrays featuring up to 100 million pixels are currently in operation. Unlike CCDs, it is possible – and almost always advisable – to perform multiple nondestructive readouts during the integration time as the IR photons slowly build the spectra (or the image) on the detector.

Performance is splendid with, in particular, about 80% quantum efficiency in a large domain, starting around 0.9 μm and reaching up to 5.5 μm. Actually, with recent improvements in the manufacturing of the HgCdTe material, the long wavelength cutoff can be tailored anywhere between 1.6 and 5.5 μm by tweaking the chemical element ratios. Near IR arrays are generally cooled around 70 °K to reduce integration noise to about 0.02 e^-/px/s; this relatively high level compared to CCDs usually limits individual integration times to a few minutes at most. With low speed readout around 50,000 pixels per amplifier per second and special reading tricks, readout noise is of the order of 7 e^-/px (again almost an order of magnitude higher than with CCDs): to avoid excessive total reading time, an individual 4k × 4k array holds up to 64 amplifiers working in parallel and the detector controller continuously performs low-speed nondestructive readout of the array until the end of the integration time.

IR arrays are less sensitive than CCDs to cosmic ray impact and, besides, their multiple nondestructive readout schemes can be used for real-time rejection. On the other hand, they are more "touchy" than CCDs, with possible artifacts, such as hot (high-noise) pixels, or parasitic light emission from the amplifiers that can reach the array corners.

1.5.5 Conclusion

The optical/NIR 2D-detector astronomical landscape is currently dominated by two commercial solid-state integrating detectors, CCDs for the optical range (up to 0.98 μm for the so-called deep depletion CCDs) and charge injection devices (CIDs) for the NIR, with long wavelength cutoffs that can be tailored to the instrument requirements in the 1.6–5.5 μm range. Pixel size is usually around 12 μm for CCDs and 15 μm for near IR arrays. Very large mosaics can be built, covering any spectrographic need. Their low readout and integration noise and high quantum efficiency (actually close to 100%) make them almost ideal detectors, except for a few photon-starved cases such as high-resolution spectroscopy of

fast transient sources or real-time measure of atmospheric turbulence to correct image blurring (Section 9.4).

One worrisome recent commercial development is the increasing replacement of CCDs by cheaper optical charge injection devices for most laymen applications (surveillance, mobile phones, etc.), except the really high-end ones. This trend might well someday make science-grade CCDS no more available for astronomical purposes, at the cost of a significant hit in detectivity, owing to the intrinsically higher CID readout noise.

A recent emerging detector is the avalanche photodiode 2D array. This is a return to the short-lived photon-counting era, but with much higher QE (typically 60% peak), more rugged devices, and a larger wavelength range covering both the optical and NIR domains, for example, up to 1.65 μm, but with extremely small formats (typically up to 8 × 8, 50 μm, pixels) and non-negligible dark (integration) noise. Much larger format devices have been built for defense purposes, but are not yet available for the civilian market. They are well adapted to wavefront sensing for adaptive optics systems with their typical sub-millisecond integration times (see Section 9.4), but not yet – if ever – for hour-long integrations as the main science detector for spectrographic instruments. In the same vein, large format CCDs, which, owing to internal electron multiplication, reach full photon-counting capability, have been developed by an e2v-University of Montréal collaboration [14].

Finally, all 2D detectors – including photographic plates – currently feature plane light-sensitive surfaces. This does not mean that there is any strong technical difficulty in developing detectors with any other sensing surface curvature, just that zero curvature is by default *the* commercial standard. This is a real limitation, as spectrographic optics designers would love getting concave detectors when opting for refractive cameras and, conversely, convex detectors to match reflective optics. Developing 2D detectors is however an extremely expensive endeavor (hundreds of million USD/Euros), and there is little hope to ever get off-the-shelf curved detectors, except with the kind of massive financial investment that ground-based astronomy, and even space-based astronomy, could hardly afford.

1.6 Optics and Coatings

1.6.1 Introduction to Optics

As discussed above, the main optical path for the various spectrographic modes involves two basic subsystems, namely, a collimator to image either a full 2D field or a 1D field (an input slit) at infinity, and a camera to image back the field or the spectra on a 2D detector. Cameras and collimators can be based on lenses (dioptric systems), or mirrors (catadioptric systems), or a combination of both. The spectral range to be covered is usually one octave at most, unless multiple cameras/gratings are used in parallel.

It would be nice to use off-the-shelf optical subsystems, gaining very much in project cost and timeline, but that is not generally possible while retaining high

light transmission and excellent image quality. In particular, most spectrographic cameras necessarily feature an entrance pupil some 5–10 cm *before* their first lens, in order to insert their transmission grating; this is a very significant constraint on the camera optical system, and all off-the-shelf cameras are designed instead with the entrance pupil well inside their optical body. Main astronomical instrument optical systems are thus usually expensive One Off prototypes that are in-house designed and then contracted to industrial optical companies. Optical cost can then easily be in hundreds of thousand USD/Euros, with the total development time around 2 years. In this chapter, we will look at a few basic principles on how these complex optical systems are designed, fabricated, and tested.

1.6.2 Optical Computation

Optical computation of the various optical subsystems is the first development step. This endeavor remains much of an art, even if the market offers (generally at a cost, one of the very few exceptions in the mid-2010s being WinLens 3D basic for Windows) remarkably efficient optical design programs that are extensively used to optimize delivered image quality, taking into account the many constraints inserted by the designer. Besides elementary design constraints such as spectral range, entrance pupil position (significantly before the camera first lens for grating and Fabry–Pérot based spectrographs), and the collimator and camera focal lengths and aperture ratios, there are many other less obvious ones, for example: (i) near telecentricity,[4] that is, with the output pupil as seen by the detector located near infinity, when precise measurement of object locations or spectral lines positions is required; (ii) use of cheap glasses when budget is limited; (iii) reasonable glass lengths; (iv) small excursions of the final image position with temperature when the instrument is not kept at constant temperature; (v) achievable optical tolerances, that is, reasonable required precision on optical component parameters (refractive index, thickness, radius of curvature, tilt and centering, inter-lens/mirror separations); (vi) no exposed glass surface too close from an image of the field (as any dust particles would then create an artifact on the final image); (vii) no harmful parasitic images, with particularly no glass/mirror center of curvature close to the conjugate of the detector plane, as any bright point in the field would then get a small halo around it, the so-called Narcissus effect from the eponymous Greek demigod; (viii) no radioactive glasses or coatings, especially if close to the detector, in order to avoid extra detector noise, and so on.

One major difficulty in getting high-performance dioptric systems is in canceling their inherent chromatic aberrations: this requires at least two kinds of transmitting materials with different dispersions (relative variation of refraction index over the working wavelength range). This is relatively easy for 1D long-slit spectrographs since a large part of the chromatic effects can be easily offset by tilting slightly the detector. When working in the main part of the optical spectrum, say between 450 and 950 nm, cheap optical glasses can then be used (as an example, see the MUSE instrument case in Section 5.4.2). For the extreme

4 See Figure 1.11 for a visual appraisal of non-telecentricity effects.

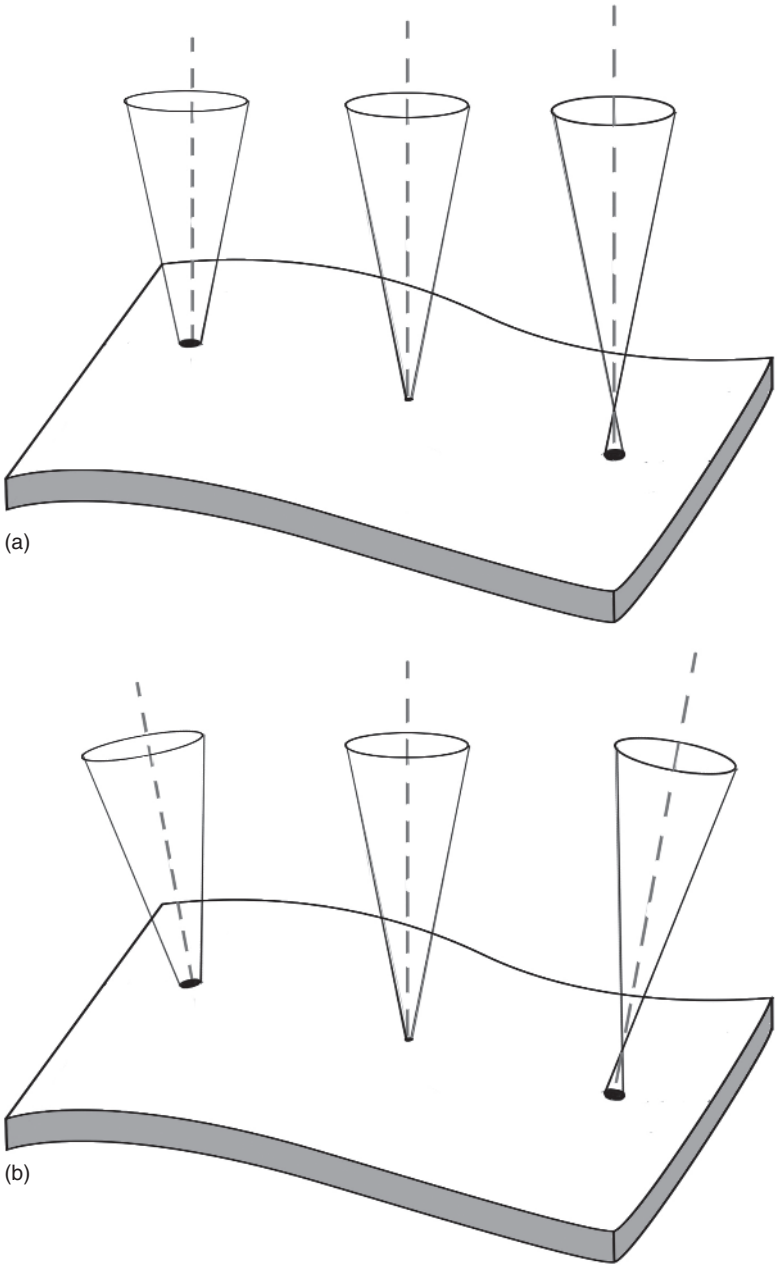


Figure 1.11 Illustration of the deleterious effect of non-telecentricity. (a) Three telecentric beams (with parallel optical axes) fall on a detector. Non-perfect flatness of the detector degrades the images, but does not move their centers of gravity with respect to each other. (b) The same, but for non-telecentric beams (optical axes not parallel). There is a similar image degradation, but now their centers of gravity are displaced with respect to each other. This leads to significant measuring errors, typically a few micrometers for, say, 10–20 μm flatness deviation. (Credit Colombine Majou.)

blue and near-ultraviolet, as well as the near-infrared domains, standard optical glasses, except the low-index low-dispersion fused silica, are no more transparent and more expensive and fragile crystalline glasses are used instead. One special difficulty for IR lenses is that the whole optical train is then usually cooled to cryogenic temperatures (say 77 °K for the near IR): knowledge of refractive indexes at these temperatures is quite sketchy, which makes optical optimization difficult. Also a remotely controlled cryogenic motorized system for accurate focusing on the detector surface is then required on such instruments.

A much better chromatic correction, called apochromatism, is needed for 2D field (multislit or slitless) spectrographs, for which the above tilting trick cannot work by design. In the optical domain this usually requires using fluoride glasses that are more expensive, difficult to polish, and easily cracked during the antireflection coating process: this results in steep cost escalation (say ×4) and long delays, easily 2 years for getting a full optical subsystem.

Mirrors have the huge advantage of no chromatic aberration at all, and on top give less geometrical aberrations than single lenses of equivalent converging/diverging power and aperture ratio. Their big disadvantage is geometrical in nature: light bounces back from mirrors and in most cases, especially for 2D fields, the output light beam is entangled with the input beam. Nevertheless, one popular solution for high-aperture cameras is the Schmidt design with a spherical mirror and the detector cum cryostat located at the mirror focal plane (see Figure 1.12). A different approach has been chosen for the JWST near-infrared multislit spectrograph NIRSpec: given its enormous spectral range, 0.6–5 μm, a train of three strongly aspherical off-axis mirrors (see Figure 1.13) is used for

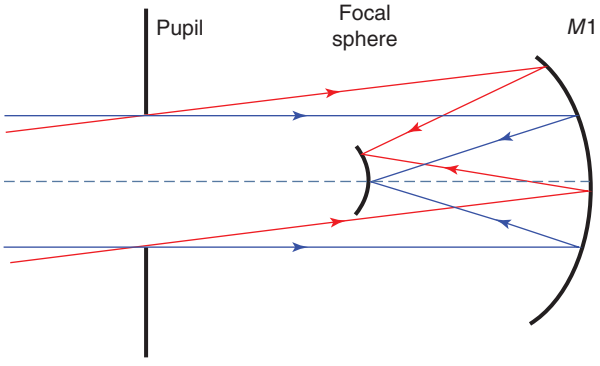


Figure 1.12 Illustration of the Schmidt mounting, in essence a spherical mirror with the entrance pupil located at its center of curvature. Two parallel light beams at two different inclinations are shown. Owing to the system's full rotational invariance, all input parallel beams are imaged with the same (small) aberration irrespective of their 3D inclination, a trick first discovered by the philosopher (and lens maker) B. Spinoza in 1600 and implemented by B. Schmidt in the 1930s. This field-invariant aberration can, for example, be canceled by adding an aspheric window on the pupil; its correction effect then varies over the field, but only as a cosine function, which in most cases is good enough. Again, because of rotational invariance, the images are located on a spherical segment with its center of curvature on the pupil. A field flattener (a thick convergent lens possibly doubling as the detector entrance window) can also be placed just before the focus.

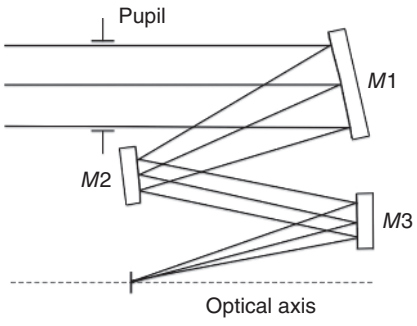


Figure 1.13 The so-called three mirror anastigmat (TMA) is a centered optical system (i.e., with a common optical axis) made of three highly aspheric mirrors, the shapes and positions of which are tuned to give extremely good images on a flat focal plane over a wide field of view. In real life, to avoid 100% beam obscuration by M_2, only off-axis cuts of the three mirrors are used. The TMA can be used as a camera with light reflected successively by M_1, M_2, and M_3, or as a collimator when used in reverse.

both the collimator and the camera. As can be seen on the figure, the off-axis part neatly solves the disentangling beam problem, at the (huge) cost of large diamond-machined monolithic optomechanical systems.

One integral part of the optical computation effort is to derive the manufacturing tolerances compatible with the required optical quality, so that the manufacturer can plan and test accordingly. Like for Goldilocks and the three bears, it is essential to design and fabricate the optics with the right tolerances, no more, no less: too loose ones than needed would give poor quality images and too tight ones, overly expensive or even unfeasible optics.

1.6.3 Optical Fabrication

Optical fabrication spans a huge range of technical procedures, adapted to the many different substrates (glass, crystal, metal, plastic, ceramic), sizes (from about 0.1 mm to 8.4 m diameter in the astrophysical domain), shapes (spherical, mildly or hugely aspherical), and required surface qualities (from a few λ for auxiliary lenses to $\lambda/4$ for precision lenses to $\lambda/100$ for interferometric components, where λ is the shorter wavelength at which the component is used).

For mirrors and lenses alike, the first step is usually to grind the substrate to get an approximate shape or even an accurate one, but with still rough surfaces at the micrometer size level. The next step is to lap/polish the surface to get a local smooth finish at close to the nanometer level, or even an extra-smooth one, for , for Fabry–Pérot or Michelson plates, or when the optics must image faint structures near an intensely bright one (the so-called coronagraphic grade optics).

Here are a few tidbits connected to optical surfacing:

- One might think that the most important manufacturing requirement is to get the exact theoretical shapes of the lenses/mirrors. Actually, this is not generally the most difficult part, as shape tolerances are often as "large" as a few micrometers, or even more, especially on large (meter-size) optics. What is more difficult is to avoid introducing significant slope deviations at all scales from the full diameter of the optical piece to about half a wavelength. Penalty for not reaching the required smoothness level is a significant fraction of scattered light, especially when working at short wavelengths. This means significant light losses and problematic artifacts around bright objects in the field. Slope requirements are especially tighter for the coronagraphic grade optics required to observe very faint sources near highly bright ones.

- Classical optical figuring uses a statistical process that automatically generates smooth spherical surfaces, including plane ones: the glass piece or blank, ground to the global shape, is put in contact with a same-size matching tool, with a mixture of abrasives and water in between. The blank and the tool rotate and oscillate with respect to each other in a pseudorandom manner. Over hours while using increasingly finer abrasives, this process automatically generates two spherical surfaces of opposite curvatures, since only two such spherical elements can remain in full contact for any orientation and lateral displacement. To produce plane surfaces, two blanks are lapped over a roughly plane tool, and the two blanks are also lapped against each other. This statistical process can be altered to get instead aspheric surfaces, for example, by stressing the optical surface during and/or after polishing, but these are slow, expensive processes, especially when one wants to avoid significant scattered light (say, no more than 1%). This is too bad because even only a few aspheric surfaces often allow the design of performant optical systems with much less lenses than their all-spherical variants.
- Molding techniques produce large quantities of dirt-cheap spherical or aspheric optics of quite reasonable optical quality. Molds are expensive to fabricate (say typically ~ 50 kEuros for a 10 cm diameter component), but can then produce tens of thousands of identical components at very low added cost. Owing to significant post-molding shrinking of the plastic materials, this process is however not well suited for production of high precision optical components. Molding is thus generally used for mass production of relatively low-tech optical components, for example, microlens arrays, cameras for smart phones, and so on. One nice feature is that a single mold can directly produce a full subcomponent, for example, a mirror with its mounting cell and reference alignment points, saving significant fabrication, integration, and adjustment time.
- Diamond-turning is a process of direct mechanical machining of precision optics using a computer-controlled lathe equipped with a diamond-tipped tool. Diamond-turning is used to manufacture spherical or aspheric lenses or mirrors alike from a number of crystals, metals, and plastics. Note that it is only since the beginning of the 2010s that it has been possible to get high-quality low-scatter diamond-turned optics good enough for the optical domain. One nice feature of this technique is that it is relatively easy to produce (at a cost) a monolithic all-mirror piece, for example, all aluminum or copper, incorporating, for example, three off-axis aspheric mirrors located at their precise theoretical locations within one micrometer or so. Such a built-in subsystem permits to evade having to perform tricky high-precision mechanical adjustments, and besides cannot subsequently become misaligned. Moreover, such subsystems can be put directly in a cryogenic environment; metal (isotropic) shrinkage will slightly alter the optics prescription in a homothetic way, but with the optics still aligned and focused. This does not work for lenses though, since optical train changes are then dominated by temperature shift of the refraction indexes of glasses.
- Optical testing during the fabrication process is essential for high-quality components and actually the only way to converge to the desired shape and surface

finish. This can easily be a full subproject in itself, with the study and fabrication of often huge testing devices. The usual rule for optical manufacturing is "if you can test it good enough and fast enough, you can get it good enough and fast enough." And the corollary, "if you cannot test it right, you will never get it right."

- For major subsystems (collimators, cameras, etc.) with exacting mounting tolerances, it is almost always better to get (at a cost of course) the optical components fully integrated inside their mechanical body from the industrial manufacturer. The client must nevertheless plan for independent end-to-end testing before closing the contract: image quality testing is generally quite easy in the optical domain, but takes much longer – usually days – for NIR optics working at cryogenic temperature. Checking light throughput is always difficult, and even more in the UV and NIR domains.

1.6.4 Anti-Reflection Coatings

When a beam of light crosses from a dielectric medium of index of refraction n_1 (e.g., air whose n is very close to 1) to another dielectric of index n_2 (e.g., an optical glass with index roughly in the 1.5–1.75 range), part of the light is transmitted (refracted) and the remainder is reflected. For optics made of transmitting elements – lenses, prisms, transmission gratings, beam splitters, Fabry–Pérot plates, and so on – the transmitted part is the useful one, while the reflected part is essentially a big nuisance, reducing overall light efficiency and possibly bouncing back from the other optical surfaces to finally create unwanted artifacts on the detector. The intensity of the reflected (R) and transmitted (T) components is given by the Fresnel equations as given in Section 1.1.1. In particular, the relative loss for light at normal incidence on a surface of refractive index n_2 immersed in air is $R = (n_2 - 1)^2/(n_2 + 1)^2$. This is about 4% loss for standard optical glasses, and is much higher with more exotic materials such as diamond at all wavelengths, or silicon, zinc sulfide, and zinc selenide in the near-IR range.

For astronomical optics, antireflection (AR) coatings are normally applied to the surface of all transmitting elements, including the front surface of transmission gratings and detectors. The simplest theoretical interference AR coating would consist of a single quarter-wave layer (meaning that its optical thickness ne is equal to $\lambda_c/4$) of a transparent material whose refractive index is the square root of that of the lens. This gives zero reflectance at λ_c for normal incidence light (of whatever polarization) and typically less than 2% reflectance over one octave in wavelength and for incidence angles less than 15°. In the optical and UV domains, but not the NIR, this is somewhat theoretical though as there are no two transparent and durable materials with such a wide refraction index contrast. Cheap AR coatings for cameras and prescription glasses are usually overcoated with a quarter wave layer of MgF_2 whose refractive index is 1.386 at 500 nm, versus typically 1.5–1.65 for cheap glasses.

On the opposite side, a perfect AR coating would consist of a material whose refractive index would continuously vary from the bulk material value on the

inside to the air index ($n = 1$) on the outside. In electrical terms, this would give perfect impedance matching between the surrounding air and the lens material, and hence no light loss at all. Incredibly enough, this can be rather closely achieved by spin-coating lens surfaces with variable porosity silicon prepared through a sol–gel process: see Cleveland Crystals Inc. for commercial availability up to 300 mm diameter, in particular for coating highly fragile crystals since the process is very gentle. Another process to cover a wide wavelength band, still at the research laboratory stage, is to print periodic nanostructures on the surface, mimicking biostructures on moth's eyes. With top-structured pyramids of about 1 μm size, better than 1% reflection has been obtained over 0.5–2.5 μm, or 2.3 octaves.

Most coatings for astronomical purposes consist instead of a sandwich of dozens of thin multilayers made of alternating high and low refractive index materials. Layer thicknesses of, for example, MgF_2 and ZnSe thin films are tailored to produce destructive interference in the beams reflected from the glass–air interfaces, and reciprocally constructive interference in the corresponding transmitted beams. In practice, one can get better than 1% reflectance over one octave wavelength and for incidence angles less than 15°. The layers are deposited one by one in a vacuum chamber. Whenever possible, it is better to get hard coatings that can be easily dusted off and cleaned. The hardening part involves baking the coated lenses at high temperature though and only soft coatings can be applied to fragile materials such as crystals and fluoride glasses. It is usually possible to get such complex vacuum-deposited coatings from commercial firms over about half a meter in diameter lenses (or mirrors).

1.6.5 High Reflectivity Coatings

Mirrors present a somewhat different set of challenges, whether working in the NIR or in the optical to near-UV domain.

For the NIR domain, beyond about 1 μm, an extremely durable very high reflectivity ($\geq 99\%$) can be readily obtained by overcoating the glass or metal mirror with a layer of vacuum-deposited gold. From Kirchhoff's law, this also means that the thermal emissivity of a such a mirror is below 1%, meaning that even at room temperature its thermal IR emission will usually be negligible.

For the optical domain, a protected silver coating gives good reflectivity (90–98%) in the 0.5–1 μm range, but does not cover the blue and near-UV domains. To do that, multilayer dielectric coatings are needed and can be obtained from a few industrial firms, very much like for AR coatings. Extremely high reflectivity $\geq 99.99\%$ can even be attained, but only for a single wavelength, a single incidence angle, and a single polarization to boot. For the usual astronomical requirements of 1 octave wavelength range, reasonable incidence angles, and unpolarized light, very good reflectivity $\sim 99\%$ is achievable, if by stacking up to ~ 100 coating layers.

One difficult case is the coating of the large telescope glass mirrors. They usually work from the atmospheric UV cutoff (0.31 μm) to the atmospheric IR cutoff at 24 μm, a huge spectral range for which there is no good solution, only tradeoffs. For almost all telescopes, the lesser bad choice is a thin (~ 100 nm) vacuum

deposited aluminum layer, spontaneously overcoated by a very thin Al_2O_3 layer as the vacuum chamber is opened to outside air. During the first few weeks of operation, this gives a respectable >92% reflectivity over the whole range (except for an 86% dip around 0.85 μm). Quite unfortunately, aluminum coatings do age though and, even with mirror cleaning every month or so to remove dust particles, reflectivity in the visible range drops to maybe 80% within 18 months, and recoating, usually with a custom plant at the telescope premises, needs to be performed. Alternatively, the Gemini-South 8-m diameter telescope mirror is currently overcoated with a custom-protected silver coating: it cannot be used below 0.4 μm, but is more durable and gives better and more stable reflectivity than even fresh aluminum for all wavelengths beyond 0.45 μm, also provided it is regularly cleaned.

1.6.6 Conclusions

By way of summary, here is the typical, if somewhat convoluted, way in which an optical subsystem for a major astronomical instrument is developed over possibly a 3–5 year time span:

- The instrument's main optical train is defined by the user, with all fundamental parameters (field, focal length, wavelength range, pupils, disperser, detector), and special requirements clearly set up.
- Detailed optical computation, including optical and mechanical tolerances, is performed by an optical engineer, with some iterations back and forth to the previous step and to the mechanical engineer in charge of the instrument design. A detailed specification and requirement document is then sent out for competitive tendering.
- Optical fabrication is contracted out to an optical firm, with some iterations back and forth to the previous step, for example, to refine glasses index of refraction to that of already available blanks, the radiuses of curvature to those of tools already available, etc. Acceptance tests both at the manufacturer premises and in-house are clearly spelled out. For ultra-high precision optics, it is not uncommon (at a cost though) that the most critical lens is produced first, and its actual parameters accurately measured and used for a next iteration of the whole optical system design.
- Coatings are usually subcontracted by the optical firm. Their specification and progress must be closely followed, because of the high potential for bad performance – for many possible reasons such as improper lens or mirror cleaning before deposition, deposition on the wrong surface, and even use of radioactive materials (!) that would saturate the detector. Besides, vacuum deposition of hard layers is not a gentle process, and lens/mirror cracking or permanent surface distortion is quite common, especially when exotic glasses are used. Any such event can easily delay the project by a year.
- Progress meetings/reports are regularly performed as contractually agreed upon.
- Subsystems delivery and acceptance are performed as contractually agreed upon.

1.7 Mechanics, Cryogenics and Electronics

1.7.1 Mechanical Design

Mechanical design of an astronomical instrument is a complex venture. The primary role of mechanics is to house the optical components of the instrument with accuracies/stabilities that can range from 1 mm to 1 nm (a factor of 1 million range!), depending on the component functions. It must allow for easy, precise, and stable adjustments, at first for the initial instrument assembly, but also during its whole lifetime. It almost always provides for the motion/exchange of key optical components (filters, gratings, scanning interferometric system, etc.), again with highly variable requirements in terms of accuracy and stability. Note that accuracy and stability are two separate issues; for example, exchanging one plane reflective grating with another for, say, modifying the spectral resolution calls for a modest accuracy (a 1 mm centering error for a 15 cm diameter grating would not make any sizeable harm), but requires an extremely good stability during a whole 1-h exposure, usually within a few micrometers. It is important to analyze quantitatively all these requirements before starting the mechanical design of the instrument.

Proper housing of optical components means putting them firmly and accurately in place in their holders, yet without exerting any strong mechanical constraint that would distort their shape, or even break the glass. One important mounting concept here is that of mechanical degrees of freedom: any optomechanical component has six degrees of freedom that fully define its position in space and must be all constrained to get a highly stable and repetitive positioning, possibly down to a fraction of a micron; see Figure 1.14 for an archetypal example using three spheres in contact with three grooves. In the laboratory, one can rely on gravity to ensure that components stay on their contact points (except during a strong earthquake!). For telescope-mounted instruments, one adds springs exerting forces orthogonal to the glass surface and directed toward the support points: any mismatch here would create constraints that would distort or even break the glass.

For less demanding applications, say at the 10 μm repetitiveness level, one can slightly overconstrain the system, for example, with extended soft contact points. For even less precise requirements, one can opt out entirely of the kinematic mounting business and overdefine the component positions, for example, insert a thin elastic rubber band between a lens and its housing: this is easier to install and much more gentle for a glass component, but also less stable and much less repetitive. For more on the fascinating issue of kinematic mountings, you may look at the nice University of Arizona tutorial at http://fp.optics.arizona.edu/.

Mechanical development also comprises housing of electronics systems (instrument and detector control), fluids cabling (electric power, data stream, cooling fluid, dry gas lines), and various calibration systems. Add that mechanical design complexity is more than often underappreciated and systematically "slaved" to the more glamorous optical design effort to boot, it is no wonder that this is historically **the** main source of frustration and delays in (astronomical) instruments development.

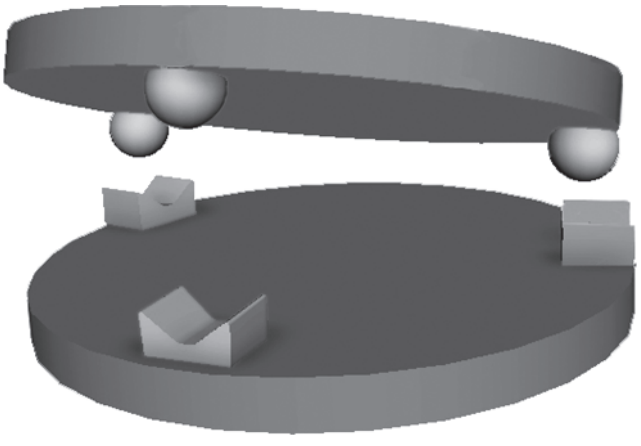


Figure 1.14 This classical kinematic mounting features three optically polished sapphire spheres glued at 120° to the underside of a component and three right-angle hardened ground steel grooves at 120° on top of its mounting plate. This gives the required six contact points, ensuring incredibly stable and repetitive positioning. The upper component can be removed and then put back (gently) in place, and without any adjustment repositions itself within a fraction of a micron. Note that this requires only very lax absolute accuracy (say only at the millimeter level) for the relative positions of the spheres and grooves. (Reproduced with permission of Colombine Majou.)

One significant difficulty is that most astronomical instruments call for extreme stability, yet are more than often moving a lot with respect to gravity with a typical angular speed of 15° per hour as the telescope tracks the sky: that is a lot when a few micron stability per hour is required, for instance, between the optics output and the detector. There are quite a number of ways for a telescope to feed its instruments (Figure 1.15): Prime Focus and Cassegrain instruments are actually moving in two dimensions; Nasmyth instruments are rotating orthogonally to gravity at the cost of one more mirror; folded Nasmyth are rotating along gravity (hence with no differential flexures) at the cost of two; finally coudé instruments are fully stable at a minimum additional cost of three mirrors (or alternatively tens of meters of optical fibers). It is usually a good idea to opt for a nonmoving instrument when stability requirements are stringent and the field of view is small enough to make it possible. Note that for moderately large fields (say 1′ diameter for a 10-m telescope), a so-called field derotator (e.g., a three-mirror combination, see Figure 1.16) can be inserted in the optical beam in order to feed a fixed-orientation instrument on a Nasmyth platform: the instrument still rotates slowly as the alt-azimuth telescope tracks an object on the sky, but only along gravity.

Still, many instruments are actually operated under big gravity changes and must carefully be designed to get negligible flexures, and, even more important, correctly built to avoid mechanical instabilities. It is not uncommon to find that a given instrument, which according to finite elements analysis should not flex

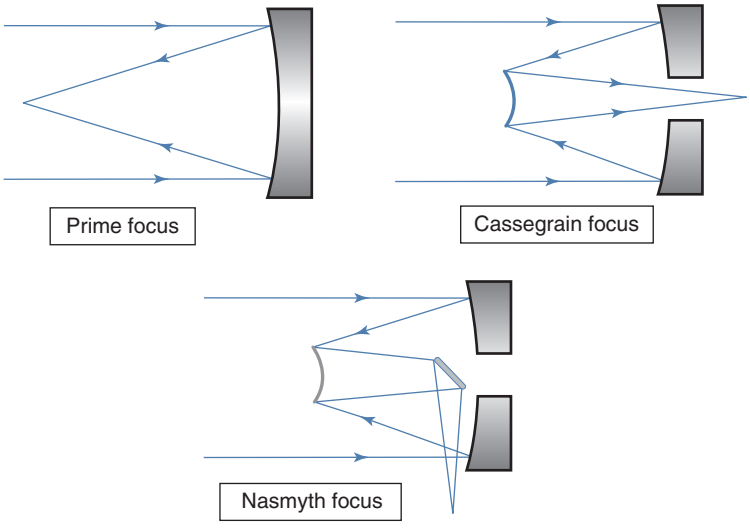


Figure 1.15 The three main telescope foci are shown, namely prime focus, Cassegrain focus, and Nasmyth focus. Additional mirrors are needed for the folded Nasmyth and coudé foci. As the telescope tracks during the night along two orthogonal axis, much like a warship turret, instruments at prime focus and Cassegrain focus move along, and on top usually rotate to cancel field rotation. At Nasmyth focus, owing to the rotating tertiary mirror, light is sent to a horizontal rotating platform when the instrument sits; field rotation has still to be canceled, though.

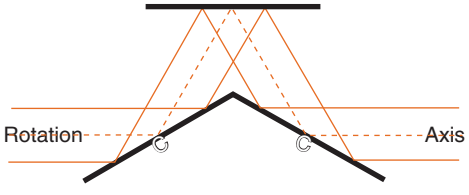


Figure 1.16 This is a schematic view of the classical three-mirror derotator in the case of a parallel beam input. It works also with a convergent beam, for example, with an image of the field on mirror #2. Field rotation is nulled by counterrotating the derotator around its horizontal axis. For small enough light beams, a prism with three internal reflections can be used instead.

by more than a few micrometer, actually internally moves by millimeters, due to, for example, an improperly tightened nut: this is a rather trivial example, but a number of "completed" instruments have never been operated because of unmanageable flexures. In some cases, flexure requirements are just too harsh to be attained with purely passive means; it is then necessary to incorporate active flexure compensating systems, despite the added complexity: as an example, no DVD player would work at its required submicron accuracy level without its many internal control loops.

1.7.2 Alignments

In any instrument design, and not only for astronomical ones, building and documenting an alignment/adjustment strategy is essential. Here are a few cardinal rules:

- Like for any crucial element of a project, if you fail to plan, you plan to fail.
- Start on it as soon as you are at the instrument concept level, certainly not as an afterthought at the end of instrument design. An alignment impossibility is arguably one of the most likely hidden traps that might spell doom for your project right from its start.
- Define carefully the instrument elements that are to be adjusted and how you are going to do it. If you do not put enough degrees of freedom, the instrument will never be aligned and so will never work. If you put more adjustments than needed, you might ultimately succeed, but that will take more efforts and might cause significant delays.
- Evaluate correctly the various adjustment ranges/accuracies required. If you set them too small/too loose, the instrument will never be adjusted correctly. If they are overly large/tight, the instrument might ultimately be adjusted, but with an impact on cost and timeline.
- Design and build adjustment devices that are repetitive, highly stable, and most preferably equipped with digital or analogic encoders. Manual adjustments are much cheaper to develop, but more than often could not be accessed safely, at least on large telescopes: fully motorized and encoded adjustment systems are then required. All that design and implementation effort might cause project overcosting and delays in the short term, but chance is that it is going to be recouped many times later on.
- When performing instrument adjustments, do not hesitate to be dumb and lazy and proceed empirically: see, for instance, Exercise 6 for a rather generic "blind" adjustment scheme that avoids figuring out the precise metrology of the adjustment scheme.

1.7.3 Cryogenics

Since the demise of the photographic plate, every astronomical instrument features at least one cryogenic system in order to operate its digital detector at proper temperature, around 150 °K for CCDs and 75 °K for NIR arrays. This requires integrating the detector and its proximity electronics in a cryostat (in essence a magnified thermos bottle) under good vacuum and developing a cooling system, generally either a liquid nitrogen bath or a cryocooler. In the latter case, much care is needed to avoid the vibrations from the cryocooler operation propagating inside the instrument. In the CCD case, the window cryostat is usually the interface between the cooled and uncooled parts of the instrument. To avoid frost formation on the outside face of the window, one can, for example, maintain a gentle nitrogen flow in front of the window. All in all, this requires a significant number of cables, pipes, and regulating mechanisms.

Near-IR spectrometers installed at normal telescope sites (meaning neither in space nor in Antarctica) must be entirely cooled when working above ~1.65 μm.

Looking backward along the light path, this must extend up to the entire focal plane of the instrument: the reason is that the disperser is actually "looking back" over a full hemisphere to any thermal radiation from the instrument mechanics and sending a good fraction of it along the main optical path, ultimately up to the detector. Any detector pixel thus sees thermal radiation over the full spectral range, but science light only over one resolved spectral element. NIR instruments are thus installed inside big cryostats, with usually a strong thermal gradient between the instrument entrance at say 150 °K, which is enough to get negligible thermal radiation up to 2.4 μm, and the detector support at 75 °K, about the best operating temperature for the detector array. Apart from the additional cost, this makes initial adjustment and integration, as well as subsequent repairs, excruciatingly slow: a week cycle forth and back to cryogenic operation for a few hours of repair on an open cryostat at room temperature is a frustrating but common occurrence. Also, any motorized motion inside the cryostat is difficult and expensive to develop, as the very few cryocompatible motors tend to exhibit vanishingly small torques and small lifetimes when operated at such low temperatures. It is also not uncommon that a failing motion at cryogenic temperature reworks spontaneously when the cryostat is still closed but already back to less frigid conditions. This makes repairs even more problematic, such as for proverbial car's faults that never occur at dealers' premises.

1.7.4 Electronics and Control System

Like modern cars, most astronomical instruments (including of course 3D-spectrometers) incorporate many motors and encoders and cannot be operated without fully automated control systems. Given that the time to, for example, reconfigure hundreds of fibers in the focal plane of a multiobject instrument to address a different sky field, exchange a grating to modify the spectral resolution, exchange a filter to modify the wavelength range, and so on, is time lost for observing, it is vital to develop optimized automatic sequences in order to perform these functions in parallel. Note that all this requires a lot of cabling, which must be done most professionally. It is not uncommon that analysis of the failures history of an operating instrument shows that more than half are connected to cabling problems, especially for instruments that are regularly taken in/out of their telescope focal locations.

Detector environments too usually require a number of controlled functions, such as a light shutter for all CCDs (to be closed during frame readout) or an internal focusing mechanism for most NIR instruments (because of the huge temperature difference between an opened and closed cryostat). A highly specialized electronic controller is also needed to set up the detector voltages, start/end the exposures and read out the detector pixels. This is often done with in-house built electronic racks; in some cases the required functions are (mostly) provided by an off-the-shelf integrated circuit developed by the detector provider.

Given the number and sophistication of controlled functions in any modern astronomical instrument, operation is now always fully computer controlled and closely integrated with the operation of the telescope. To save precious telescope time, whenever possible the various steps needed are done in parallel,

for example, disperser and filter exchange while the telescope moves toward the next target. User's friendly feedback is continuously sent to the observer, usually through a graphical user interface. In the few cases where the observer must actually provide a real-time input, for example, precisely set a narrow slit on the science target, a "super" user's friendly environment is provided. For instance, just clicking on two locations on a sky image taken with the instrument will automatically move the telescope and set its field rotator to place a slit or an integral field unit at the required sky position.

1.8 Management, Timeline, and Cost

As can be gathered from the technical complexity of 3D spectrometers, successful development of such full-scale instruments requires heavy management investment over a long timeline – up to a decade from first concept to start of routine operation – and carries a substantial global cost, easily in the tens of million Euros/USD.

Just to give a flavor of what is generally needed for a successful endeavor, here are a few pragmatic "rules" for the many project stages (see Figure 1.17):

- At start, there is a Concept, a Principle Investigator (P.I.) eager to transform his/her goal ("a goal is a dream with a deadline," Napoleon Hill 1925, The Law of Success) into a Project to be carried out over a 10-year period or so, and at least one Institution eager to get the instrument and ready to support a team of competent people led by the P.I. and find/provide adequate funding. Very often, this very first phase takes actually many years of preliminary studies and intense lobbying before the Project gets its first green light and moves to the definition phase, usually with a flashy acronym attached.
- In every case, at least most subsystems, or even the whole instrument, will be built by high-tech industrial firms. Always remember that they are most likely to know better than you how your specifications can be achieved at minimum cost and/or timeline by their technology: impose only really needed specs, and never how they are going to be met, making industry part of the solution, rather than of the problem. Remember also that tradeoffs are unavoidable, as

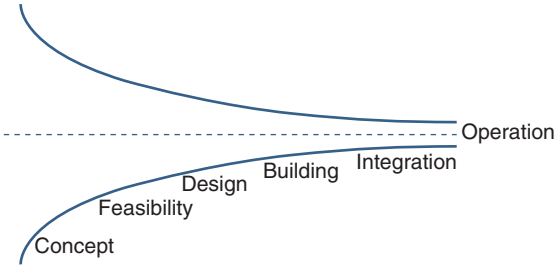


Figure 1.17 Project Funnel. This small cartoon illustrates how starting from a broad concept, any instrumental project becomes more and more tightly defined as it moves through successive stages toward start of operation. Along the way, uncertainties, in terms of cost, timeline, and/or performance drastically decrease, that is for a successful project.

per the (already optimistic) adage "Performance, Cost, Timeline, choose two." Competitive tendering (when at all possible) is a must, as a factor of 4 overall range in the industrial offers for the same optomechanical component is not uncommon. Beware, while you might have got a very cost-effective deal in the tendering process, changing later the specs at the building phase means reopening the whole cost issue, a recipe for heavy cost increases: this is a widely valid warning, not limited to astronomy or even high-tech developments, but which applies to any big project, as attested by many high-profile horror stories of public building developments.

- Any substantial Project goes through a number of phases, namely: Feasibility Study; Preliminary Design; Final Design; Fabrication; Assembly, Integration and Testing (AIT); Telescope Installation and Commissioning. Each step carries the Project to a well-defined level, including detailed planning of the next step, and ends with a Review, preferably led by external consultants. Projects can be deeply modified or even canceled following any Review by whatever organization the builders are working for, in particular, should meeting reasonable performance, cost, and/or timeline appears highly problematic.
- The basic aim of Feasibility Study is to show that the project can be achieved at minimum risk with the required performance at an affordable cost and within a reasonable timeline. This is for a substantial part "just" a paper effort, but with additional in-house and/or external technological developments to prove the validity and cost of the key concepts of the instrument. Preliminary and Final Design are in principle just design phases as their names imply, again producing a lot of (electronic) papers, but are often coupled with procurement in parallel of long-lead subsystems, such as the main optical components. These two phases lead to the Fabrication phase, which includes careful components/subsystems validation, both at the manufacturers' premises and subsequently in the AIT hall.
- AIT is the next crucial step, with the T (Testing) generally the longest and most expensive part. In particular, it usually requires special premises. That may include an integration hall with proper environmental parameters (temperature, hygrometry, dust level, vibration level, etc.) and various equipments (handling tools, vacuum pumps, gas containers, cryogenics, electric power, etc.), a metrology laboratory for subsystem test and acceptance, a clean room for detector integration, and any required custom-made test equipment: this can be a full subproject by itself that needs to be planned and even possibly fabricated well in advance as an integral part of the whole Project. In the 1980s, the then fashionable time- and cost-saving strategy of careful testing only the subcomponents as a way to avoid the higher burden of whole instrument testing resulted in the one billion dollars Hubble Space Telescope near-fatal disaster: a single error in the optical testing of the secondary mirror, compounded by NASA's adamant refusal for any full system testing, resulted in an orbiting telescope delivering strongly aberrated images. This was eventually fixed by adding optical correctors in front of each on-board instrument, but this now proverbial story remains as a reminder to come back to the harsh but much more sensible way of full instrument characterization prior to shipping to a distant mountain and even more to outer space.

- For a really big project there will be many specific team positions: Project Investigator; Project Manager; Instrument Scientist; Control Manager; Mechanical, Optical, Electronics and Software Leaders, and so on. For a smaller one, the same functions are still needed but are concatenated to be filled by fewer people. For both small and large projects, advanced work planning (often on a yearly basis), with at least monthly progress assessments, is a must. Keeping such large teams eager and enthusiastic over a decade or more, while facing unavoidable problems and setbacks, is by no means a small management challenge.
- Delivering proper documentation is a major part of any project, and a vital tool during the instrument operating phase. Yet, it is too often seen as a chore, to be (badly) done after instrument building has been achieved. *Au contraire*, it should be started early, preferably already at the feasibility study level: Early drafts of, for example, the Users manual, the Alignment, Control Software, and the AIT documents are actually a great way to catch well ahead of time intrinsic problems that might later kill the Project as such, or at least cause big delays and/or cost overruns.
- Providing, documenting, and maintaining a near real-time pipeline that extracts and visualizes the 3D data in physical units (intensity versus wavelength at each sky position) at most a few minutes after observation, is no small feat, but essential to evaluate the validity of typically 1-hour long observations. Together with a thorough off-line pipeline for proper data reduction delivering science-ready products, equally maintained during the whole life of the instrument, requires a big manpower investment over decades. This is also one of the best ways to 'sell' the instrument to its potential users, the ultimate touchstone for instrument success.
- Any project has risks. You can strive to minimize them, but you cannot eliminate them altogether. On one side, you can develop your 10-year project using only well-seasoned risk-free technologies right from the start, and chance is that when completed it will be fully uncompetitive. On the other side of the fence, you could redefine your project for any technological advance and/or new science drivers that happen along the line, and chance is that the instrument will never be completed. Staying at the optimum risk level that maximizes the expected scientific value of the instrument is arguably the best winning strategy, even if much easier said than done.

1.9 Conclusion

As can be gathered from this chapter, developing a state of the art (3D) instrument is a long, expensive, complex, and risky high-tech endeavor. It requires in particular building a large competent team in all relevant technical domains, selecting the best high-tech vendors and learning from them, and during the long development phases striking a delicate balance between rigorous long-term planning and creativity.

This long-term effort is not limited solely to the instrument design, building, and installation phases, but extends as well during its whole operating life, either

still by the original building team or by the instrument host observatory, or as a combination of both. It is in particular quite common to rejuvenate an ageing instrument, for instance, by implementing a new state-of-the-art detector or changing its spectral domain and/or spectral resolution. On the other hand, as over time an instrument becomes decidedly uncompetitive, it is usually better to reject aggressive and futile therapy, and build instead a worthy successor. In any case, only a cradle to grave investment can give an instrumental facility the chance to achieve its full observing potential over a reasonable length of time.

*** Exercise 1 Trying to beat etendue conservation #1

You have been tasked with detecting an all-sky night emission line with a ground flux of 20 ph.cm^{-2} s^{-1} sr^{-1}. Back in the early 1960s, your best detector is an 12% quantum efficiency (QE), 7 mm diameter, photomultiplier, accepting light from a 30° half-angle conic beam. Detector r.m.s. noise is 12 counts per second. The line is selected with a 38% peak transmission narrow-band interference filter. Since the detector QE varies over its sensitive area, for better signal stability, you image the pupil on the detector, with telecentric beams (i.e., sky image at infinity). A reminder: the solid angle of a cone of half-angle θ is $\pi \sin^2\theta$.

1. Shooting the detector directly at the sky, with a baffle somewhere to avoid input light out of the detector acceptance cone, what count number per second do you get? Hint: First compute the detector etendue.
2. To improve the situation, you put a telescope in front of your instrument in order to collect more light. What is the optimum telescope diameter (if any)?
3. Undeterred, you put a tapered cone in front of the photomultiplier with a d mm input diameter and of course a 7 mm diameter output. What is the best d (if any)? What happens to the light rays?
4. As a last resort, you now insert a tapered paraboloid with its focus at the center of the sensitive surface of the detector. With perfect concentration of all light rays parallel to the optical axis at the center of the detector, will you finally collect more light?

Answer of exercise 1

1. Detector maximum etendue: $(\pi/4) \times (0.7)^2 = 0.385$ cm^2 sr. Photons to detector counts efficiency: $0.38 \times 0.12 = 0.0456$. Detected flux: $20 \times 0.385 \times 0.0456 = 0.35$ counts per second. With the 12 counts per second r.m.s. detector noise, it looks like a short integration, ~ 300 s would lead to say a $5\,\sigma$ detection. However, at the time, photomultipliers had highly unstable noise properties, and a signal of about 3 counts per second was already at the detection limit, irrespective of integration time.
2. Light is collected on a much larger pupil area, but with a smaller solid angle on the sky. Because of etendue conservation, the detected flux is the same, actually smaller because of the not perfect light transmission of the telescope.

3. More of the same, again because of etendue conservation. One might wonder why a larger beam etendue than permitted by etendue conservation injected in the tapered fiber does not strike the detector. A careful look shows that these extra rays, after a few reflections on the cone wall, just come back along the cone, ultimately up to the sky.
4. Again no way, for the same reason. Yes, the whole beam parallel to the optical axis entirely strikes the detector. However, with zero etendue, it carries zero energy. Note that one of the authors (GM) has been indeed tasked at the time to find a way to detect such a source, and went through these steps one by one, including showing with a messy computation[5] that the paraboloid off-axis aberrations indeed prevent breaking the etendue limit.

******* Exercise 2 Trying to beat etendue conservation #2**

Let us take a hypothetical 10″ diameter ionized gas cloud at a galaxy center with a uniform brightness narrow emission line at 656 nm. The gas is rotating as a solid body, with constant integrated radial velocity along the sky plane projection of its rotation axis (the minor axis) and a linear radial velocity gradient $G = 20$ km s^{-1} $(″)^{-1}$ along the projected orthogonal axis (the major axis) from one edge to the other. You are using a long-slit spectrograph observing facility with the following parameters: telescope diameter $D = 3.6$ m, grating diameter $d = 150$ mm, and blaze angle ϕ to be derived.

1. Find the grating blaze angle for which a wide, 10″ width, slit parallel to the cloud minor axis (thus collecting the whole cloud light) gives nevertheless a narrow spectral line on the detector.
2. Assuming 100% optics transmission, this spectral line is an order of magnitude brighter at the detector location than on the sky. Now, here is the tough question: this is a clear violation of the sacrosanct second principle of thermodynamics, right?

Answer of exercise 2

1. Spectral resolution $\mathfrak{R} = \lambda/\delta\lambda$ for a 1″ wide slit is equal to $10^5 d\ (\tan\phi)/D$. $\mathfrak{R}$ is also equal to c/G', where G' is the radial velocity gradient across the 1″ slit. With $G' = -G$, the emission line peak beams are on top of each other and thus fall on the same detector pixels. This requires $\tan\phi = 10^{-5} cD/dG = 3.6$, or a steep but feasible 75° blaze angle high-order echelle grating.
2. Well, yes and no. The second law as expressed so far –never decreasing entropy in a closed system – is indeed spectacularly violated here. However, the fully correct extended 2^{nd} law – never decreasing entropy plus information in a closed system – is not. It is the known a priori information (location, orientation, and magnitude of the object radial velocity gradient) that is used to achieve this seemingly impossible feat. Note that a similar experiment had actually been done successfully in the 1950s (De Vaucouleurs G., private discussion).

5 You are of course most welcome to repeat it.

** Exercise 3 Prism etendue

The goal is to illustrate the small linear etendue of the prisms in the direction of dispersion. A spectrograph at the focus of a $D = 4$ m telescope is using a $d = 75$ mm, 60° apex angle A, BaF10 prism at minimum deviation. Slit width projected on the sky: $\alpha = 1''$ ($\sim 5.10^{-6}$ rad); camera aperture ratio $\Omega = 1/1.5$. Glass index of refraction n given by $n = B + C/\lambda^2$, with λ is in µm, $B = 1.67$; $C = 0.00743$. Central wavelength $\lambda_c = 0.55$ µm.

1. Compute the instrument spectral resolution at central wavelength. Hint: use the prism spectrograph cooking book insert.
2. Compute the slit width w projected on the detector.

Answer of exercise 3

1. Index $n = 1.695$ at 0.55 µm. From prism spectrograph insert, spectral resolution is $\mathfrak{R} = (K\Delta)d/D\alpha$. From the prism data: $K = 1.883$ and $\Delta = 2C/\lambda^2 = 0.049$. Finally $\mathfrak{R} = 346$, just a small value at the lower limit of spectrographic resolution.
2. Etendue conservation gives $w\,\Omega = D\,\alpha$, hence $w = 30$ µm, or typically 2.4 CCD pixels.

** Exercise 4 Grating etendue

The goal is to compare the linear etendue of gratings versus prisms in the direction of dispersion, with a similar instrumental setting as in Exercise 3. A spectrograph at the focus of a $D = 4$ m telescope is using a $d = 75$ mm diameter transmission grating (first order Littrow mounting) with a blaze angle $\phi = 30°$. Slit width projected on the sky: $\alpha = 1''$ ($\sim 5.10^{-6}$ rad); camera aperture ratio $\Omega = 1/1.5$. Central wavelength $\lambda_c = 0.55$ µm.

1. Compute the spectral resolution at central wavelength. Hint: use the grating spectrograph cooking book insert. Set up the grating ruling (number of grooves per millimeter).
2. Compute the slit width w projected on the detector.

Answer of exercise 4

1. From the grating spectrograph insert, spectral resolution is $\mathfrak{R} = 2\,(\tan\phi)\;d/D\alpha$. From the grating data: $2\tan\phi = 1.155$, giving $\mathfrak{R} = 4330$, a comfortable value, 12 times larger than with the equivalent prism-based instrument. Using $2\sin\phi = a\lambda_c$ gives $a = 1082$ grooves per millimeter.
2. Etendue conservation gives $w\,\Omega = D\,\alpha$, hence $w = 30$ µm, or typically 2.4 CCDpixels, of course the same as with the equivalent prism spectrograph.

*** Exercise 5 Grism Rotation Invariance

We consider a generic zero deviation grism, made of a blaze angle ϕ (in air at central wavelength λ_c) VPHG sandwiched between two identical apex angle ϕ

prisms of refractive index n_c. Because of their zero mean deviation, different grisms can be exchanged in the instrument, with, for example, different wavelength ranges and/or wavelength dispersion, with the spectra automatically centered on the detector. Given the use of remotely controlled exchange mechanisms, it is difficult to avoid small rotation angle uncertainties ϵ of the grisms in the dispersion plane, the consequences of which are evaluated here.

1. Find the emergent angle $(\phi + \epsilon')$ at central wavelength λ_c for an incident angle $(\phi + \epsilon)$. Derive the zero deviation departure $(\epsilon' + \epsilon)$. Hint: evaluate approximatively ϵ' as a second degree polynomial in ϵ.
2. Compute the corresponding spectral shift $\Delta\lambda$ versus rotation error ϵ. Find the relationship between the grism rotation error ϵ and the resulting spectral shift 'resolution' $R = \lambda_c/\Delta\lambda$. Does that ring a bell?
3. Taking the same spectrograph as in Exercise 4, what rotation angle error ϵ will shift the central wavelength by one slit width? How does it compare with the slit angular width as seen from the grating.

Answer of exercise 5

1. From the classical grating law, $\sin(\phi + \epsilon) + \sin(\phi + \epsilon') = 2\sin\phi$. Taking $\epsilon' = k_1\ \epsilon + k_2\ \epsilon^2$ with the well-known approximations $\sin x \sim x$ and $\cos x \sim (1 - x^2/2)$, one gets $k_1 = -1$ and $k_2 = \tan\phi$. The error angle $\epsilon + \epsilon'$ is equal to $\epsilon^2\ \tan\phi$.
2. Using again the classical grating law, but now looking at its λ dependence, we find $\epsilon = \sqrt{2/R}$. And, yes, the wavelength shift versus grism tilt angle (in the dispersion plane) and the wavelength shift versus Fabry–Pérot tilt angle (in any plane) obey similar laws.
3. From Exercise 4, $R = 4330$, hence $\epsilon = 0.0215$ or 1.23° at the grism level. Slit width is 1″ on the sky, hence $1'' \times (4000/75)$, or 0.000267 rad at the grism level. Their ratio is ∼ 80, which can be seen as the grism invariance figure of merit.

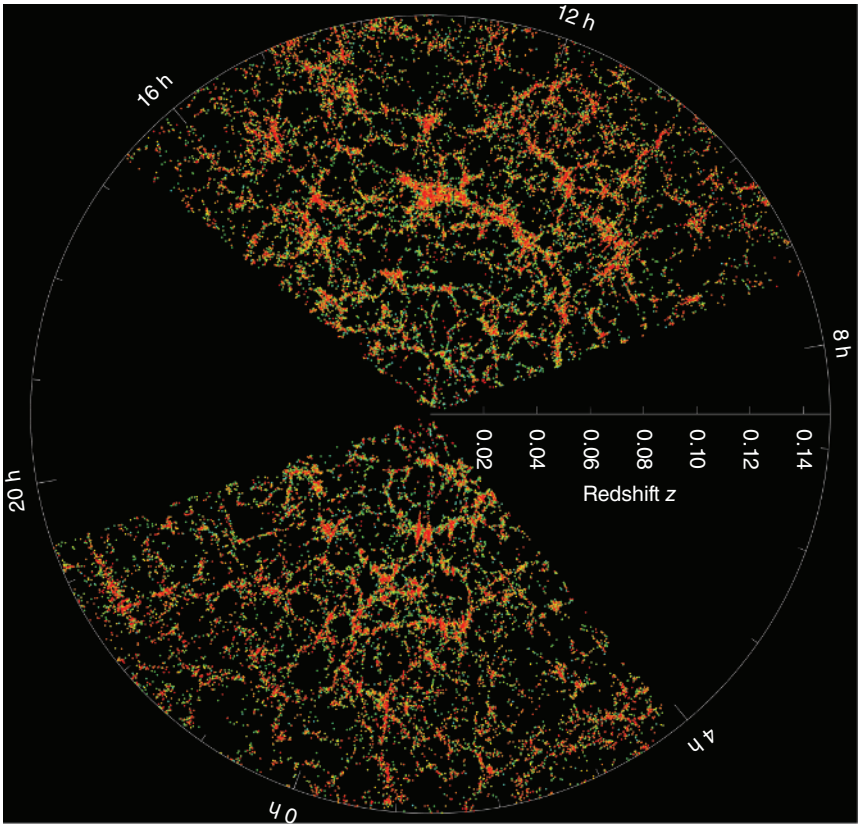
*** Exercise 6 Optical Adjustments for Dummies**

To adjust a mirror inclination inside your instrument, you use a point-light source centered on its input field, the image of which on the 15 µm pixel CCD must be exactly centered, that is, at $x_c = 2048, y_c = 2048$ in pixel units. Initially, the image is instead at $x = 2274.4, y = 1852.7$. You are using two rotating linear screws with encoders. From your own calibration, one positive (clockwise) 360° turn of the screws moves the light source on the detector by $\Delta x_1 = 122.7, \Delta y_1 = 85.4$ for screw #1, $\Delta x_2 = 97.1, \Delta y_2 = -101.6$ for screw #2.

1. Assuming fully linear behavior, what are the screw rotations $\Delta\theta_1, \Delta\theta_2$ in degrees required to make the adjustment?
2. Applying the computed corrections, you now find the source image at $x' = 2046.1, y' = 2048.9$. What screw rotations do you apply now?
3. You should be now happily dead centered: why? What can you say about the relative actions of the two screws in length scales and orientations?

Answer of exercise 6

1. From the calibration: $\Delta x = (122.7/360)\ \Delta\theta_1 + (97.1/360)\ \Delta\theta_2$ and $\Delta y = (85.4/360)\ \Delta\theta_1 - (101.6/360)\ \Delta\theta_2$. Inverting the corresponding 2×2 matrix gives $\Delta\theta_1 = 1.7621\ \Delta x + 1.6840\ \Delta y$ and $\Delta\theta_2 = 1.4811\ \Delta x - 2.1280\ \Delta y$. With the required adjustments $\Delta x = -226.4$ pixels and $\Delta y = 195.3$ pixels, this gives $\Delta\theta_1 = -70.05°$ and $\Delta\theta_2 = -750.92°$.
2. Now $\Delta x = 1.9$ pixels and $\Delta y = -0.9$ pixels. This gives $\Delta\theta_1 = +1.75°$ and $\Delta\theta_2 = +4.84°$.
3. The first iteration missed the target at the 1% level. The second one should give a similar accuracy, that is, this time better than 0.1 pixel adjustment: this has been done just from an empirical calibration, without having to figure out the mirror actual motions, a thankless task. Normalizing the two $\Delta\theta$ relationships gives $\Delta\theta_1 = k_1\ (x \cos\alpha_1 + y \sin\alpha_1)$ and $\Delta\theta_2 = k_2\ (x \cos\alpha_2 + y \sin\alpha_2)$, with $k_1 = 2.437$, $\alpha_1 = 43.7°$, $k_2 = 2.593$, $\alpha_2 = -55.2°$. The two screws thus move the light beam by slightly different amounts, their scale ratio being $k_2/k_1 = 1.064$. Their two directions are also not exactly orthogonal, with an angle $\alpha_1 - \alpha_2 = 98.9°$.



The Sloan Digital Sky Survey has created the most detailed three-dimensional maps of the local Universe ever made, with deep multicolor images of one third of the sky, and spectra for more than three million astronomical objects. The maps show the distribution of galaxies of the local universe [121]. Each dot is a galaxy; the color bar shows the local density. These observations have been obtained with the multiobject fiber-based spectrograph of the SDSS 2.5 m telescope at Apache Point Observatory. ([120]. Reproduced with permission of Michael Blanton.)