Optical practicalities: étendue, interferometry, fringe localization

11

11.1 Introduction

In this chapter we discuss several related topics. First we consider the flux of energy, mostly in the context of light from non-laser sources. An important geometrical factor is the étendue. It is the étendue that we must maximize when trying to get as much optical energy as possible into apparatus such as a spectrometer or an optical fibre. Optical instruments differ in their ability to collect light, and their relative merits in this regard are assessed by comparing their étendues.

We put this idea to use by analysing a Michelson interferometer used as a Fourier-transform spectrometer, and we discover that it is far better at light-gathering than a conventional (grating) spectrometer.

A property of interferometers of the Michelson type (achieving interference by division of amplitude) is that they generate **localized** fringes. We take the opportunity to explain localization, where the fringes lie, and how critical is the focusing on these fringes.

There is an insightful relationship between étendue and the number of transverse modes occupied by the electromagnetic field. There are further links with coherence area, with thermodynamics (entropy), and the understanding as to why laser light is so special.

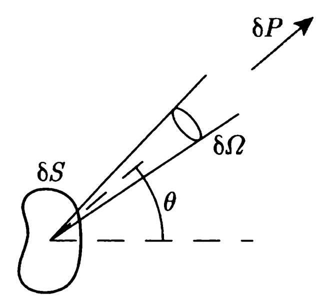

Fig. 11.1 Power δP is radiated from an element of surface area δS into an element of solid angle $\delta\Omega$.

11.2 Energy flow: étendue and radiance

Consider some area element δS that radiates optical power as shown in Fig. 11.1. The power δP that we can collect in a solid angle $\delta\Omega$ is proportional to $\delta\Omega$ and to $\delta S\cos\theta$ (the area projected along the line of sight). Given these dependences, we define a quantity **étendue** by[1]

$$\delta(\text{étendue}) \equiv n^2\,\delta S\,\delta\Omega\cos\theta. \tag{11.1}$$

The power δP collected is now given by[2]

$$\delta P = B \times n^2\,\delta S\,\delta\Omega\cos\theta. \tag{11.2}$$

If, as here, power δP is measured in watts, the coefficient B is called **radiance**;[3] if δP is measured in 'visual' units, weighted according to the

[1] The reason for including the factor n^2 (n is refractive index) will be made clear below.

[2] If B is independent of θ, the power δP is proportional to $\cos\theta$. This property is quite commonly encountered, and is called the Lambert cosine law.

[3] Old terminology dies hard, and B is still sometimes called *brightness*.

sensitivity to frequency of the human eye, then B is called **luminance**.

For a light source of finite area, radiating into a range of directions, eqn (11.2) can be integrated to give the total power P collected into apparatus with a defined étendue. In particular, if the source[4] has constant B (independent of locations within ΔS and directions within $\Delta\Omega$),

$$P = (\text{radiance } B) \times (\text{étendue}). \tag{11.3}$$

[4] Constant B means additionally that all parts of area ΔS radiate equally (apart from the $\cos\theta$ factor) into all parts of $\Delta\Omega$.

If we gather light radiated symmetrically into a cone of (non-small) semi-angle $\theta_{\max}$, we have (problem 11.5):

$$\text{étendue} = n^2\,\Delta S \int \cos\theta\,\mathrm{d}\Omega = \Delta S\,\pi(n\sin\theta_{\max})^2, \tag{11.4}$$

in which we encounter the **numerical aperture** defined by

$$\text{numerical aperture} \equiv (n\sin\theta_{\max}). \tag{11.5}$$

Although we have avoided approximating in the above, it often suffices to make small-angle approximations, and then eqn (11.1) integrates more simply to[5]

$$\text{small angles:} \qquad \text{étendue} \approx n^2\,\Delta S\,\Delta\Omega. \tag{11.6}$$

[5] Textbooks usually define étendue simply as $\Delta S\,\Delta\Omega$, with a small-angle approximation implied. The $\cos\theta$ factor of eqn (11.1) is, however, conventional in more careful treatments; see, e.g., Born and Wolf (1999), §§4.8.1 and 4.8.3. By contrast, the n^2 factor seems to be less usual, but its inclusion improves both the mathematics and the physics.

The presence of the numerical aperture in eqn (11.4), rather than $\sin\theta_{\max}$ on its own, is just one of the neat consequences of putting n^2 into definitions (11.1) and (11.2). Numerical aperture will be encountered again in §§12.5.1 and 12.6.

Equation (11.3) may be applied to the collection of light into an instrument such a spectrometer. In such a case, ΔS refers to the radiating area from which light is usefully collected,[6] and a similar understanding applies to the directions included within $\Delta\Omega$.

[6] Area ΔS may be the surface of a light source, or an image of a light source, or sometimes some other area, possibly one 'along the way' in an optical system. However, areas 'along the way' don't radiate uniformly into $\Delta\Omega$ save in special cases. See problem 11.1(2). As mentioned in sidenote 4 (this page) we restrict attention to areas ΔS that do radiate uniformly.

If a source such as a lamp has a non-uniform radiance from different parts of its area, or radiates unequally into different directions, we can, of course, attempt to allow for the non-uniformity by integrating eqn (11.2):

$$P = \int B n^2 \cos\theta\,\mathrm{d}S\,\mathrm{d}\Omega.$$

However, non-uniformity is likely to be a symptom of a design fault—or an inappropriate choice of area for ΔS—and should be removed rather than allowed for.

When using a spectrometer, we shall wish to obtain strong signals on a photographic film or an electrical photodetector. Optimizing the power collection is to be achieved by maximizing the étendue of the optical equipment, since (as we shall see) radiance B is not something we can usually do much about. This motivates much of the discussion in §§11.3–11.10.

11.3 Conservation of étendue and radiance

When the étendue of an optical system is evaluated, using a properly chosen area according to the understandings in sidenote 6 (this page), it remains unchanged as the light is transformed by lenses, mirrors, or other optical components, or passes from one medium to another:

$$\text{étendue is an invariant.} \tag{11.7}$$

A further conservation rule follows from eqn (11.7): if an image is formed of a source, using lenses or mirrors,

$$(\text{radiance of image of source}) \le (\text{radiance of original source}), \tag{11.8}$$

where equality holds when there is no loss of energy (**insertion loss**) in the imaging system. These two conservation rules are established in problem 11.1.

11.4 Longitudinal and transverse modes

A light beam can be decomposed into longitudinal and transverse modes, as was done in describing cavity modes in Chapter 8.

The étendue of a light beam is connected with the number of transverse modes that are populated by its photons. The relationship is:[7,8]

$$(\text{number of transverse modes occupied}) = \frac{\text{étendue}}{\lambda_{\text{vac}}^2}. \tag{11.9}$$

A derivation of eqn (11.9) is laid out in problem 11.3.

A wave occupying a single transverse mode has a photon flux, the number of photons per second crossing some plane, given by (problem 11.2)

$$(\text{photon flux}) = (\text{number of photons per mode}) \times \delta\nu, \tag{11.10}$$

where $\delta\nu$ is the range of frequencies occupied by the photons. The number of occupied longitudinal modes is related to $\delta\nu$, in a way that is presented in problem 11.2.

When light leaves an area ΔS into several longitudinal and transverse modes, the power $P(\nu)\,\delta\nu$ in frequency range $\delta\nu$ is easily seen to be

$$P(\nu)\,\delta\nu = (\text{number of photons per mode})\left(\frac{\text{étendue}}{\lambda_{\text{vac}}^2}\right) h\nu\,\delta\nu. \tag{11.11}$$

We take the opportunity to obtain the radiance of a source radiating into frequency range $\delta\nu$:

$$B(\nu)\,\delta\nu = \frac{P(\nu)\,\delta\nu}{\text{étendue}} = (\text{number of photons per mode})\,\frac{\nu^2}{c^2}\,h\nu\,\delta\nu. \tag{11.12}$$

Expressions (11.9), (11.11) and (11.12) are all to be doubled if two polarizations are excited.

[7]Equation (11.9) represents a purely geometrical relationship, so it applies to light of a single polarization. If we concern ourselves with unpolarized light, the number of occupied modes is doubled.

[8]A Gaussian beam occupies a single transverse mode. A direct check that it has étendue λ_{vac}^2 (with happy choices of definition for ΔS and $\Delta\Omega$) is made in problem 11.4.

11.5 Étendue and coherence area

Coherence area has been introduced in Chapter 9. The coherence area is related to the solid angle within which light arrives,[9] as is shown in Fig. 11.2: if angles are small, $\Delta S = \lambda^2/\Delta\Omega$.

Suppose that light is incident onto the area shown in Fig. 11.2, and arrives uniformly within solid angle $\Delta\Omega$. Then (small angles)

$$\begin{pmatrix}\text{étendue of light falling on}\\ \text{one coherence area } \Delta S\end{pmatrix} = n^2\Delta S\Delta\Omega = n^2\lambda^2 = \lambda_{\text{vac}}^2. \tag{11.13}$$

This result is independent of the illuminating geometry and of the refractive index of the medium; everything has cancelled out.

Comparing eqns (11.9) and (11.13), we see that whenever we seek to collect light falling within a single coherence area we are in fact aiming to receive no more than a single transverse mode. A simple interpretation of this idea is explored in problem 11.7.[10]

[9]Small-angle expressions are used here for simplicity. An equivalent result is correct for large angles, as mentioned below.

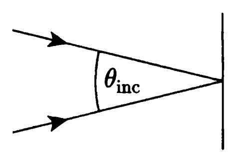

Fig. 11.2 This diagram shows a beam of light in cross section. It arrives at some plane within a range of angles θ_{inc} and in consequence the coherence area has width $\lambda/\theta_{\text{inc}}$. Something similar happens in directions at right-angles to the paper. Taking both directions together, the light arrives within solid angle $\Delta\Omega$ and yields a coherence area $\Delta S = \lambda^2/\Delta\Omega$.

[10]The derivation of expression (11.13) for coherence area has not taken detailed account of any particular geometry in the incident light beam; it is, therefore, less precise than the treatments in Chapters 9 and 10. However, there is an inevitable agreement in order of magnitude.

We have already seen an application of eqn (11.13). In the Hanbury Brown–Twiss experiment, it was necessary to observe fluctuations of intensity within a single coherence area. Problem 10.17 has shown that the experiments did precisely that.

Before we leave the present topic, we give a version of eqn (11.13) applicable to non-small angles. Let light be collected by a circular lens and focused into a cone with semi-angle θ_{max}, numerical aperture $(n \sin\theta_{\text{max}})$. The coherence area in the focus is a circular patch of area ΔS and radius r where, by the inverse of eqn (11.4),[11]

$$\Delta S = \frac{\text{étendue}}{\pi(n\sin\theta_{\text{max}})^2} = \frac{\lambda_{\text{vac}}^2}{\pi(n\sin\theta_{\text{max}})^2}, \qquad r = \frac{\lambda_{\text{vac}}}{\pi(n\sin\theta_{\text{max}})}. \quad (11.14)$$

Expressions (11.14) for ΔS and r also give us the dimensions of the area to which light can be focused if it occupies a single transverse mode.[12] If several transverse modes are excited (randomly, with nothing specially engineered), the achievable focus area will be increased by a factor equal to the number of occupied transverse modes. With this understanding, results (11.14) are of immediate relevance to the reading of a compact disc (problem 11.6) and the supply of power into an optical fibre (problem 14.2).[13]

11.6 Field modes and entropy

Consider a light beam that has some number of photons distributed among field modes, longitudinal and transverse. Like any other system of particles in quantum states, it must have entropy $k_B \ln W$, where W is the number of microstates in the macrostate and k_B is the Boltzmann constant. A chaotic (random, non-laser) light beam may occupy any of a large number of microstates that are macroscopically similar, so it has randomness and carries with it a non-zero entropy.[14]

The conservation laws for étendue and radiance stated in eqns (11.7) and (11.8) have a new interpretation: they tell us that a lens (or other optical component) doesn't change the number of field modes[15] occupied by photons; equivalently it doesn't alter the entropy.[16]

The prohibition on changing (in particular reducing) entropy-étendue applies with equal force to optical arrangements that don't form an image, e.g. where light is concentrated in a caustic.

By contrast with the above, a beam with only a single mode occupied can have no $k_B \ln W$ randomness, so it should not carry entropy.[17]

The idea that a (chaotic) light beam carries entropy justifies assumptions that were unexplained earlier in this book. It was assumed in Chapter 9 (particularly §9.15) that for random (non-laser, chaotic) light the only way to obtain a desired degree of coherence is to filter away all of the light that's doing things we don't want. This gloomy suspicion is now confirmed. Light in the 'wrong' field modes can't be rescued (somehow diverted into the 'right' modes) without diminishing the entropy and so violating the Second Law; it has to be discarded. In achieving

[11] A similar result can, of course, be obtained by using diffraction theory. The derivation given here is interesting for the rather different route that it follows. The reasoning is outlined in problem 11.5.

[12] This equality should 'feel' obvious. The van Cittert–Zernike theorem—which this is—should begin to feel obvious also ....

[13] Laser light is dangerous to eyes because it (usually) occupies a single transverse mode.

[14] As a light beam passes through some region of space, the photons that were present are swept out and replaced by others. So one microstate is succeeded by another, and what we observe over an ordinary timescale is a time average, equivalent to an ensemble average over microstates. All the conditions for $k_B \ln W$ to represent entropy are therefore met.

[15] *Aside*: There is a similarity to the Liouville theorem in classical particle mechanics (Mandl 1988, p. 208). A pretty discussion of the Liouville theorem is given by Penrose (1989), especially p. 235.

[16] I've made a bald statement for emphasis. More correctly: a lens can't reduce the number of occupied microstates while conserving the number of photons, because that would reduce W. Entropy can be increased if photons are scattered into additional directions. And the entropy of the light beam—though not that of the universe—can be reduced if photons are simply absorbed.

[17] We have $W = \prod_i \frac{(g_i + n_i - 1)!}{(g_i - 1)!\, n_i!} = 1$, whatever the value of n_i for the one state that's occupied.

Even in a single mode, photons can be 'loaded' into a variety of quantum states, see e.g., Loudon (2000), Chapter 5. Some randomness is therefore possible if the quantum state is changing. It seems best to treat such randomness as a quantum-mechanical phenomenon different in principle from entropy.

transverse coherence, we must throw away light from most of the source's area, or travelling in most directions, or both, until we have reduced the étendue to λ_{vac}^2. It's a high penalty. Likewise, to obtain a desired degree of longitudinal coherence (coherence time) we have to discard (by filtering) frequencies that are unwanted. Just how high these penalties are is a matter examined in problem 11.8.

11.7 Radiance of some optical sources

11.7.1 Radiance of a black body

The radiance $B(\nu)\,\delta\nu$ of a black body, for emission into frequency range $\delta\nu$, can be obtained at once from eqn (11.12):[18]

$$\text{radiance } B(\nu)\,\delta\nu = \frac{1}{e^{h\nu/k_B T} - 1}\,\frac{2\nu^2}{c^2}\,h\nu\,\delta\nu. \tag{11.15}$$

Both polarizations are included.[19]

In eqn (11.15), the Planck distribution $(e^{h\nu/k_B T} - 1)^{-1}$ is the value taken by the number of photons per mode when we have thermal equilibrium. Some representative numerical values are given in Table 11.1.[20]

[18]There is a 'Lambertian' dependence of the power radiated upon $\cos\theta$, which is not on display here because it has been included in the definition (11.2) of B.

[19]Radiance is here defined as in eqn (11.2), with a factor n^2. The fact that there is no n^2 on display in eqn (11.15)—it has cancelled—provides yet another reason why we favour defining things as in §11.2.

[20]These numerical values should be contrasted with the number of photons per mode in the output from a modest laser, as calculated in problem 10.20.

wavelength	$T = 2000\,\text{K}$	$T = 6000\,\text{K}$
435.8 nm	6.8×10^{-8}	0.41%
633 nm	1.2×10^{-5}	2.32%
780 nm	9.9×10^{-5}	4.85%
850 nm	2.1×10^{-4}	6.33%
1.3 μm	4.0×10^{-3}	18.8%

Table 11.1 The number of photons per mode for black-body radiation of two different temperatures.

11.7.2 Radiance of a gas-discharge lamp

The radiance of a black body represents a standard against which it is appropriate to compare the radiances of other light sources. In particular, we may ask whether a lamp emitting a line spectrum is or is not 'brighter' than a black body of comparable temperature—and what would be a 'comparable temperature'?

If we are to receive an intense line from a gas discharge, the discharge must be 'optically thick' at the middle frequencies of the spectral line. Otherwise, we could add more radiating gas 'behind' the discharge and we would be able to see extra light from that added gas through the original discharge. In an optically thick discharge, photons emitted in the middle of the discharge have a high chance of being absorbed by lower-state atoms before they can travel to the walls. There is frequent exchange of energy between atoms and photons, resulting in an approximation to local thermodynamic equilibrium within the bulk of the discharge; atoms acquire a Boltzmann distribution (and photons a Planck distribution) with a temperature T_d. This temperature is called the 'distribution-over-states' temperature.

At the middle of a spectral line, the considerations just given mean that the intensity we receive approximates to that of a black body at temperature T_d. For frequencies away from line centre, the atoms radiate less intensely, and they also absorb less intensely, so the mean free path for photons between absorptions is greater than at line centre. A point is reached where that mean free path is greater than the dimensions of the discharge, and the discharge then becomes 'optically thin'. The spectral line's emitted power falls with further frequency shift, following a profile that would have applied to the entire spectral line if the discharge had been thin at all frequencies.

As a rough rule of thumb, the distribution-over-states temperature for an intense discharge is about 6000 K; so a spectral line is at best about as bright (radiance, photons per mode) as sunlight of the same frequency.[21] A very clear and useful account of discharge physics is given by Wharmby (1997).

[21] Discharge physics is more complicated than this 'potted' account suggests. In particular, there is usually a region near the walls of the discharge where the gas is cooler. This cooler gas has its own T_d, lower than that for the bulk of the discharge. Photons near line centre come to equilibrium at this lower T_d on their way out while light a little away from line centre (outside the Doppler width for the cooler gas) does not. We have a case of 'self-reversal'. Such effects are described by Wharmby (1997).

11.7.3 Radiance of a light-emitting diode (LED)

Passage of a current through a suitably constructed diode results in a region of semiconductor where there is a surplus of electrons at the bottom of the conduction band and, likewise, a surplus of empty states at the top of the valence band.[22] Electrons in the conduction band come to thermal equilibrium with phonons and acquire a thermal distribution with a temperature equal to the lattice temperature (roughly room temperature). Relaxation to this distribution is fast, timescale about 10^{-11} s. The electrons then fall to empty states in the valence band with a radiative lifetime of order 1–10 ns. Once in the valence band they again thermalize (timescale 10^{-12} s) so that the states they entered at the top of the valence band remain empty.[23] These thermal effects mean that the photons emitted cover an energy range of roughly $2kT$.

[22] For a description of the basics of the device we refer the reader, e.g., to Fox (2001).

[23] The semiconductor is, therefore, optically thin, reabsorbing little of the emitted light. The timescales are given here because they are considerably shorter than those applying in a typical gas discharge, and result in a somewhat different way of thinking about what is happening.

The semiconductor almost always has a 'direct' band gap, so that the emission or absorption of photons is not accompanied by creation or absorption of phonons.

LEDs have a variety of constructions according to the purpose for which they are made. However, a diode intended for feeding an optical fibre might have an emitting area of 50 μm square radiating 1 mW at a wavelength of 850 nm into a numerical aperture of 1. Such an LED emits light with 3×10^{-2} photons per mode. We are again finding radiances roughly comparable to that for sunlight.

11.8 Étendue and interferometers

Any interferometer interferes light waves, obtained by dividing a beam of incident light into two (or more), either by division of wavefront or by division of amplitude. Consider first the case of division of wavefront, for which the paradigm is the Young slits. It's necessary to make the two divided beams coherent with each other, at least in one direction (the direction across the slits in a Young-slits case). To make discussion simple, imagine first that the beams must be coherent with each other in both transverse directions. In that case the light must be filtered, by narrowing its width and its range of directions, until it occupies only a single transverse mode. This filtering is known (problem 11.8) to result in only a small usable energy flow (for a non-laser source).

The penalty just introduced has arisen because the light beam has been made coherent in both transverse directions. We might expect that a Young-slits experiment, in which coherence is needed in only one direction, would be less disadvantageous: elongating the source from a pinhole to a slit as in Fig. 3.10 permits an increased flow of energy. This idea is explored in problem 11.9. The surprising result is that observation by eye is made more comfortable, but there is little improvement in visual brightness of the fringe pattern.

By contrast, an interferometer exploiting division of amplitude (the Michelson and its relatives) sets no requirement on transverse coherence. There is no need to perform any filtering down to one or a few transverse modes.[24] The bright fringes can be as bright as light received directly from the source without the interposition of the interferometer (if we ignore insertion loss).

Suppose we are to design an interferometer to make an optical measurement, perhaps to measure the refractive index of a gas. We could adapt a Young-slits apparatus, and we would end up with something resembling a Rayleigh refractometer. We could adapt a Michelson interferometer and end up with something like a Jamin refractometer. Of these two approaches, the second is far more practical, in terms of the optical power available in the fringes.[25]

The generalization is obvious:

- If you have any choice in the matter, design your interferometer to use division of amplitude.

[24] There is, of course, a requirement: that the frequency range of the light be small enough that the coherence length encompasses the intended path difference. However, this requirement is usually met by accepting the use of a limited range of path differences, rather than by filtering the light.

[25] This discussion is centred on the idea that we are going to use a gas-discharge lamp as light source. A laser gives high intensity in a single transverse mode, and a Rayleigh refractometer illuminated by a laser would work well. However, there is still no positive advantage in pursuing the wavefront-division route, so we would be wise to choose the greater versatility of the Jamin—which is also a slightly simpler instrument.

It might be objected that the 'textbook' description of the Rayleigh refractometer incorporates some elegant ideas, such as a set of fiducial fringes to assist measurement, which are not part of a 'textbook' Jamin. This, however, is to miss the point. Apparatus design is not set in stone, and there is nothing to prevent our hybridizing the designs to exploit the best features of both—or indeed of others. Excellent descriptions of the Rayleigh and Jamin designs may be found in Born and Wolf (1999).

11.9 Étendue and spectrometers

We consider here the light gathering by an optical apparatus such as a spectrometer: the larger we can make the étendue, the more energy we can collect and use.

In a spectroscopic instrument there is a trade-off between étendue and resolution. In the case of a grating spectrometer (problem 11.10), widening the entrance slit increases the étendue (by increasing area ΔS) but degrades the resolution, and conversely. A similar trade-off applies to other instruments, though for reasons that may not be quite so obvious.

Étendue further provides us with a means of comparing the merits of one instrument with another. We can ask: which is likely to be better for examining a weak source, a grating spectrometer or a Fourier-transform (Michelson-type) spectrometer? The answer (problems 11.14 and 11.15) may be surprising. A Michelson interferometer or a Fabry–Perot is a **whole-fringe instrument** possessing the properties (problem 11.15):

$$(\text{solid angle of acceptance}) \times (\text{chromatic resolving power}) = 2\pi \qquad (11.16)$$

and

$$(\text{étendue}) \times (\text{chromatic resolving power}) = 2\pi \times (\text{area of a mirror}). \qquad (11.17)$$

These properties represent a standard of comparison for the light gathering of a spectroscopic instrument. A Fourier-transform spectrometer is in line with this standard, while (it turns out) a comparable grating spectrometer falls short of it by a factor of order 200 (problem 11.14). This is the context for the ingenious measures described in §4.9.6: the grating spectrometer is coming from a long way behind, and there is correspondingly a great deal that might be gained from improving it.

11.10 A design study: a Fourier-transform spectrometer

The apparatus is a Michelson interferometer as described in §10.7. Our design study of it is pursued via problems 11.11–11.16, which discuss the light gathering (étendue) of the instrument, the resolution that can be achieved, and the trade-off between the two. The whole-fringe property of the interferometer, mentioned in §11.9, is derived in problem 11.15.

We have another agendum also in providing problems 11.11–11.16. There seems usually to be little time in a physics degree course for teaching the principles by which a piece of hardware (optical or otherwise) is designed. The author regards this as regrettable. It happens that the Fourier-transform spectrometer provides a rather good case study, where one can start from a blank sheet of paper and the relevant physics, and end up with a fully practical design for a piece of equipment.

11.11 Fringe localization

Any interferometer that exploits interference by division of amplitude (as in a Michelson interferometer) generates fringes that are **localized**. This may make the setting-up of the interferometer seem a less straightforward business than that for a division-of-wavefront arrangement such

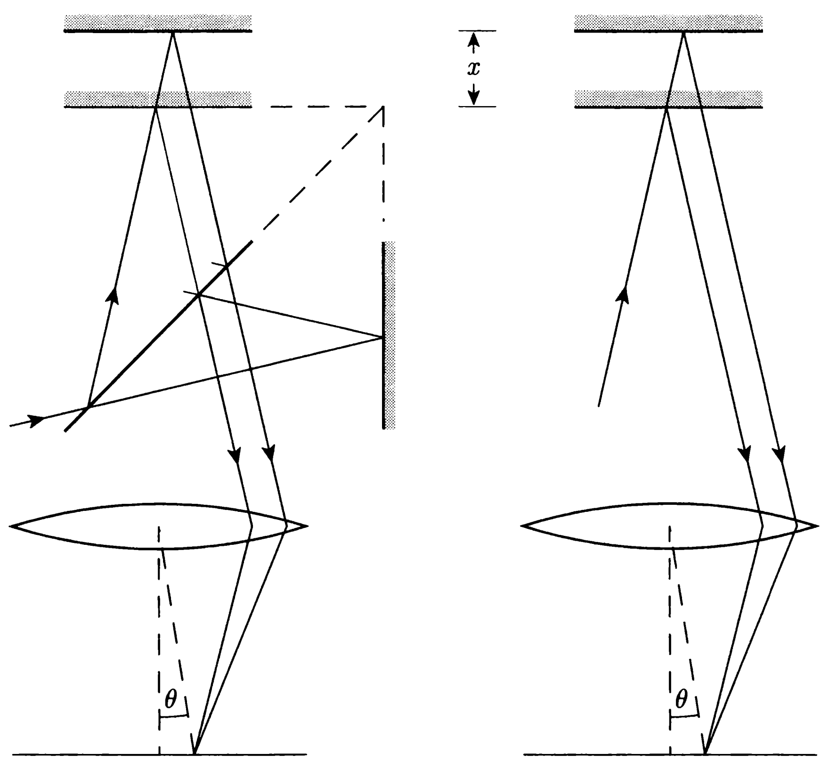

Fig. 11.3 Ray paths through a Michelson interferometer, compared with similar paths taken by light that is reflected at the two surfaces of a parallel-sided slab. The right-hand mirror of the Michelson, and the light rays to and from it, are shown mirror-imaged by the beam splitter.

To make the diagrams comparable, the 'slab' has been represented as two reflecting—or part-reflecting—surfaces with no refraction at the boundaries. The intention is to show that the interferometer is geometrically equivalent to the slab, without reproducing all the details of a real slab.

as Young's slits. However, the reasoning of §11.8 shows that there is a big advantage in energy gathering to be gained from using division of amplitude: localization, and the dealing with it, is a small price to pay. But, of course, we do have to understand localization: what it means; where the fringes are formed; and how critical is the focusing on them.

We shall restrict attention here to the case where a Michelson interferometer has its mirrors 'parallel'. Representative optical paths through the system are shown in Fig. 11.3. The same diagram shows rays reflected from a parallel-sided slab, to indicate that the ray geometry in the interferometer is the same as that for the simpler case of the slab.

For light following the paths shown, the optical path difference is $n\,2x\cos\theta$, where θ is the angle made by the ray paths *inside the slab*, x is the separation of the reflecting surfaces and n is the refractive index of the medium between the reflectors. The interference condition[26] is

$$\text{condition for dark fringe:} \qquad n\,2x\cos\theta = (p+\tfrac{1}{2})\lambda_{\text{vac}}. \qquad (11.18)$$

Given that the reflecting surfaces are parallel, x is being held constant. We also consider n and λ_{vac} to be constant. Then the phase difference between the two interfered light beams is controlled by $\cos\theta$ only.[27] To achieve good fringes, we must ensure that light ending up at a given place all has the same value of $\cos\theta$. What we need is shown in Fig. 1.3: a lens focused for infinity gathers all light travelling in a single direction and concentrates it in a single place.[28]

The analysis just given may seem rather abstract. It also gives us no idea of tolerances: the focal plane of the lens is the best place to look for fringes, but how critical is it to look just there? To investigate, we proceed by means of a series of diagrams. Figure 11.4 shows what happens if a pair of reflecting surfaces (which we'll call a slab) is illuminated by light from a monochromatic point source at S. Light reflected from the

[26] *Comment*: If we had a glass slab in air, there would be a phase change of π at one reflection or the other, and $(p+\frac{1}{2})$ would be replaced by p in eqn (11.18). However, our drawing of a slab is given to simplify the geometry while analysing the workings of a Michelson interferometer. The mirrors of the Michelson are likely to be identical to each other, so there should be no phase change of π here. (The physics of the beam splitter is another matter ....)

[27] This accounts for the description of these fringes as **fringes of equal inclination**.

[28] It may help to state that the interferometer with mirrors 'parallel' acts as an *angular filter*, with energy transmission dependent upon angle θ. We display this behaviour by following the interferometer by an *angular selector*, so we separate out light that has been filtered in the different ways.

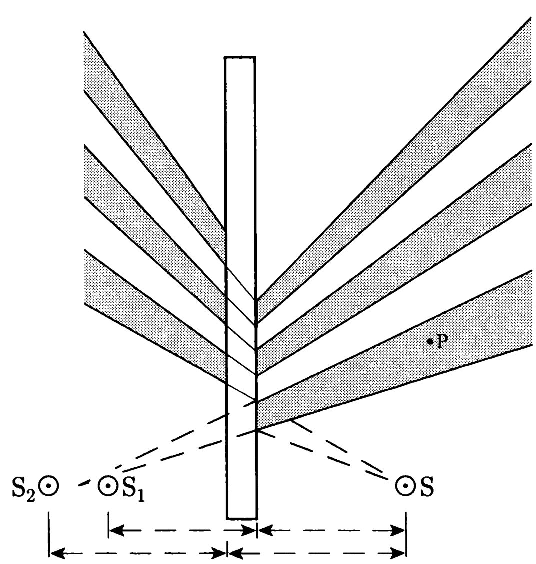

Fig. 11.4 A point source S illuminates a slab whose surfaces are part-reflecting, part-transmitting. Light travelling towards top right appears to have come from the two ray-optics images S_1, S_2 of the source. Equal-length arrows confirm the correctness of the image positions. Reflected light undergoes two-beam interference, with its energy redistributed into $\cos^2$ fringes; the grey shading indicates crudely where the intensity is less than average.

A slab acts as a low-grade Fabry–Perot: multiple reflections produce a succession of images of source S, of which only the brightest two are shown. At least one image to the right of the slab is needed to account for fringes in the light travelling towards top left.

The system has rotational symmetry about line S_2S.

right-hand surface appears to diverge from an image S_1 of the source, as far to the left of the reflector as the source is to the right of it. Light reflected from the other surface appears to diverge[29] from S_2. We have two (virtual-image) coherent sources S_1 and S_2 radiating spherical waves that overlap in the space to the right of the slab. When amplitudes are added at a point such as P, the result is light or dark, depending upon the value of the path difference $S_2P - S_1P$, which in turn depends upon angle θ according to eqn (11.18).[30] The shaded regions of the diagram indicate places where there is destructive interference and there is a beam of 'dark'. The fringes are, of course, $\cos^2$ fringes, so the intensity varies smoothly; the abrupt changes of shading are used in the figure to make the discussion that follows more stark.

The interference shown in Fig. 11.4 must, of course, conserve energy.[31] Therefore, where 'dark' travels one way away from the slab, there must be 'bright' travelling the other way. Lines on the diagram indicate how the fringes on the two sides of the slab 'fit between each other'.

The arrangement shown in Fig. 11.4 yields fringes wherever there are two beams of light to be added together: the fringes are *non-localized.* This setup resembles a Young-slits arrangement, so far as the localization of the fringes is concerned. We shall see that this has happened because the light originated from a *point* source.

'Non-localized' may seem a strange term, since the fringes do have

[29] This is not quite correct for a real slab because distances are modified by the slab's refractive index.

[30] Equation (11.18) is approximate here because the two waves have slightly different values of θ.

[31] In a Michelson interferometer the mirrors do not transmit, so there are no fringes equivalent to those going to the left in Fig. 11.4. Energy must, of course, still be conserved, but the surplus or deficit appears elsewhere; see problem 11.19.

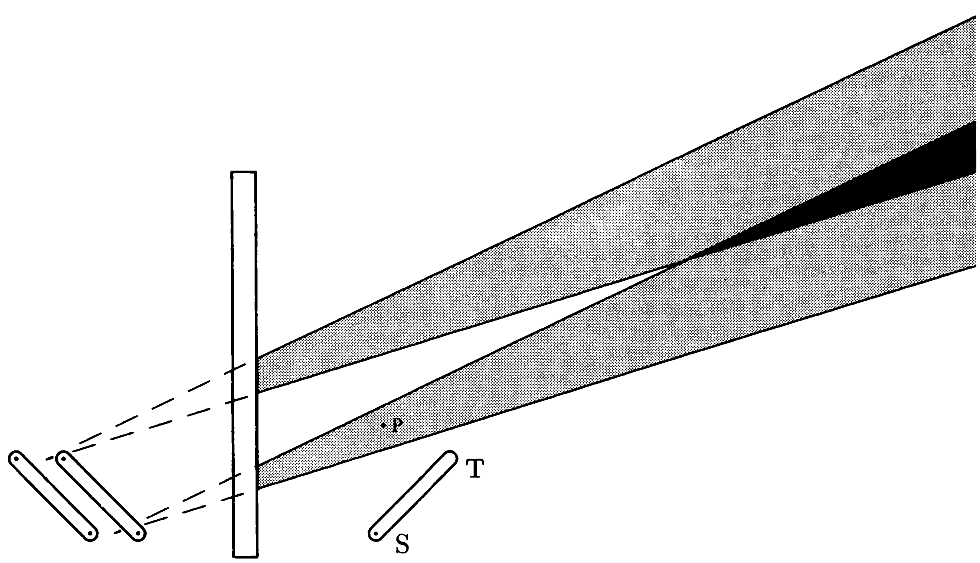

Fig. 11.5 The setup shown here is an elaboration of that shown in Fig. 11.4. The light source is now no longer a point, but is some luminous object ST. Source ST has been drawn in a nothing-special place and at a nothing-special orientation, in order to make it clear that the discussion is general. Light from S produces a set of fringes that radiate outwards from the images of S. Light from T produces a similar set of fringes, but displaced because they come from the images of T. Other similar fringes originate from places in between. A point such as P receives 'dark' from S but 'bright' from T; there can be no complete darkness there. But at large distances from the slab the fringes from S and T overlap with a displacement that becomes negligible compared with the width of a fringe: good fringes are seen, as is suggested by the black area.

locations; they do not exist over the whole of space. This is not the point. As we shall see shortly, there can be arrangements where interference happens, yet fringes are not seen: there is an additional condition to be met. 'Localized' fringes are those subject to such an additional condition; 'non-localized' fringes are those that are not.

Figure 11.5 moves us towards more practical conditions by showing what happens when light comes from an extended source. Light originating from point S produces a set of fringes, like those of Fig. 11.4, radiating out from the bottom end of the source–image pair. Light originating from point T does the same, but with everything shifted upwards and to the left. Light from points in between does something intermediate. Point P, to the right of the slab and fairly close to it, receives 'dark' from S but 'bright' from T, and little sign of any organized fringes can be seen at places similarly close to the slab.[32] However, we may see that the fringes overlap more and more as we go further away from the slab, and eventually they become so much wider than the separation of their sources that the overlap is total. The fringes are 'localized at infinity', in that they exist at sufficiently large distances from the slab—just such an 'additional' condition for seeing fringes as is claimed above.

The final step in our reasoning is made in Fig. 11.6. Rather than go to a large distance from the slab, we 'bring infinity closer' by using a lens. The 'beams of dark' are brought together so that their overlap is total in the lens's focal plane: the fringes are localized at that focal plane.[33]

Given that we have been made to concentrate attention on the lens's focal plane, we may view the lens and its focal plane as resembling a camera lens and film arranged to be 'focused for infinity'. This gives us another way of understanding 'localized at infinity'.

The reader will be able to imagine what happens to Fig. 11.6 if the source is made longer or shorter. If it is made shorter, the extreme

[32] It is possible to get into something of a tangle with irrelevancies. The reasoning given here proceeds by assuming that all points on the source radiate incoherently. That is: we assume that light from S never interferes destructively with light from T to yield 'dark' in the bright-expected areas. Would things be different if there were some significant degree of transverse coherence along the length of the source? It is for this reason that I have concentrated on where destructive interference sends 'dark'. If points S and T both send 'dark' to a given place, then it doesn't matter whether we should be adding amplitudes or intensities: the result is zero either way. What happens in the 'bright' areas is a separate issue; but that issue would have been with us in the absence of the slab ....

[33] To reconcile the jargon: The fringes are localized at the focal plane in the lens's image space; they are localized at infinity in its object space.

The reader is encouraged to confirm that the ray-optics rules of Fig. 1.3 have been rigorously applied in the preparation of Fig 11.6.

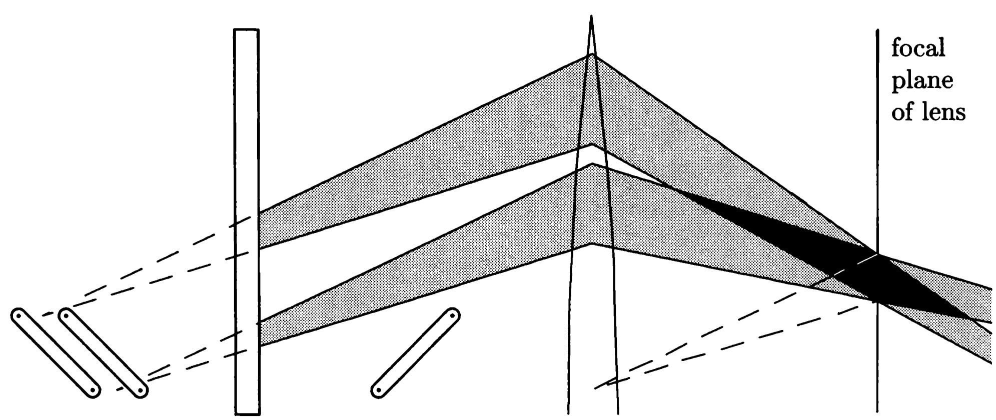

Fig. 11.6 The fringes of Fig. 11.5 are passed through a lens. Light from both ends of the source, and from points in between, is now focused so that all dark fringes come together at the lens's focal plane; and similarly all bright fringes come together (at other places) there. The best place to look for fringes is the focal plane. At the same time, the black diamond shape gives an idea of how far we can be from the focal plane and still see fringes of some sort; that is, it shows us the depth of focus.

beams of 'dark' get closer together, and the dark diamond of overlap gets longer. In the limiting case when the source is made a mere point, we are back to the non-localized fringes of Fig. 11.4. Conversely, as the source is made wider, the extreme beams of 'dark' get further apart; the diamond of overlap gets shorter and the focusing becomes more critical.

Given fixed values for slab thickness x, wavelength λ and refractive index n, the condition for a bright or dark fringe consists of a condition on θ only. Therefore, the fringes have axial symmetry about a direction normal to the reflecting surfaces (and passing through the centre of the lens if we use a lens). If we use an optical arrangement like that of Figs 5.2 or 5.8 or 11.6, the fringes are concentric circles centred on the lens axis. The reasoning of the present section fills gaps in the explanation of some diagrams that have appeared earlier in this book.

There is one other case of fringe localization that should be discussed: the case where the two reflecting surfaces form a thin wedge. The investigation of this case is left to the reader: problem 11.20.

Problems

Problem 11.1 (a) The conservation law for étendue
Figure 11.7 shows a ray-optics image formed when light is refracted at a curved surface separating two media; there is a similarity to Fig. 1.4. In the first instance, ΔS_1 is the area of some luminous source which radiates equally into all parts of solid angle $\Delta\Omega_1$. The image formed may be real or virtual, and is drawn real for simplicity.

(1) Use ray optics to show (small angles)[34] that

$$\frac{h_2}{h_1} = \frac{v\,n_1}{u\,n_2}, \qquad \frac{\Delta S_2}{\Delta S_1} = \frac{v^2\,n_1^2}{u^2\,n_2^2}, \qquad \frac{\Delta\Omega_2}{\Delta\Omega_1} = \frac{u^2}{v^2},$$

and hence

$$(\text{étendue})_2 = n_2^2\,\Delta S_2\,\Delta\Omega_2 = n_1^2\,\Delta S_1\,\Delta\Omega_1 = (\text{étendue})_1. \qquad (11.19)$$

[34] Small angles are used here for simplicity. However, the connection linking étendue to field modes and to entropy makes it clear that a large-angle generalization must exist.

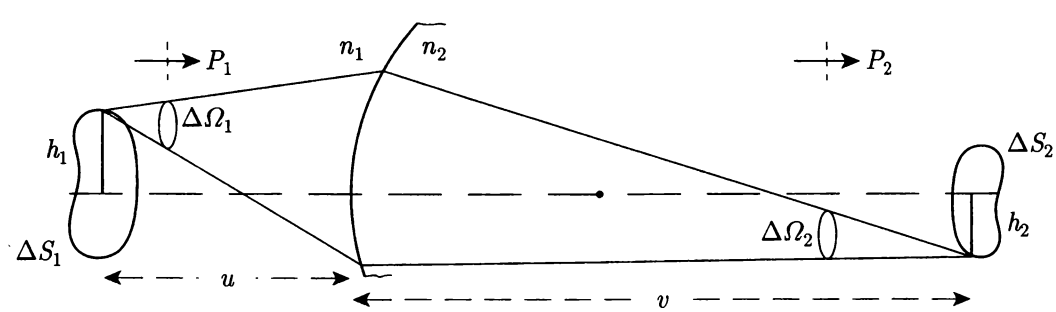

Fig. 11.7 The geometry for problem 11.1. Area element ΔS_1 is imaged to area ΔS_2. The solid angle collected is $\Delta\Omega_1$, and this results in an exit beam occupying solid angle $\Delta\Omega_2$.

(2) Show that the same value of étendue is obtained by using

$$n_2^2 \times \begin{pmatrix}\text{area of beam at}\\ \text{curved surface}\end{pmatrix} \times \begin{pmatrix}\text{solid angle into which}\\ \text{light leaves that surface}\end{pmatrix},$$

and that this works because there is no correlation between location and ray direction. Show that no other area 'along the way' through the optical system has such a lack of correlation, so étendue integrates to contain a simple area–solid-angle product only at the refracting surface or at an object or image.[35]

(3) In the diagram, P_1 and P_2 are the optical powers crossing the planes indicated. At best the system transmits all the power that's incident, and in practice it may lose a little.[36] Use the conservation of étendue to derive the 'conservation law' for radiance:

$$\frac{\text{radiance of image}}{\text{radiance of object}} = \frac{P_2}{P_1} \le 1, \qquad (11.20)$$

or, in words:

- an image formed by an optical system has (at best) a radiance equal to that of the original object.

(4) Argue that the conservation of étendue applies to the passage of light through any sequence of surfaces, and so applies (for example) to the case where a light beam is transformed by a glass lens in air.

Problem 11.2 (b) Longitudinal modes and their occupation
Consider a one-dimensional wave. It might be a Gaussian beam occupying a single transverse mode.[37] Let it occupy a very large length L, which is a 'quantization length', a length that we introduce simply to produce a countable number of longitudinal modes.[38]

(1) Apply periodic boundary conditions, and show that the number of longitudinal modes in frequency range $\delta\nu$ is $(L/v)\delta\nu$, where v is the speed of light (distinguished from c in case some medium is present). This is the number of (longitudinal) modes travelling in just one of the two possible directions.[39]

(2) Let a one-dimensional wavetrain have frequency range $\delta\nu$. Show that the number of photons per second that pass any fixed point within L, travelling in just one of the two possible directions, is[40]

$$(\text{photon flux}) = (\text{number } p \text{ of photons per mode}) \times \delta\nu. \qquad (11.21)$$

(3) Why were 'periodic boundary conditions' desirable in part (1)?

(4) Let each of the photons of part (2) be represented as a wave packet having duration τ_r. Show that the total length of these wave packets within L is $p(L/v)\delta\nu\, v\tau_r$, where p is the number of photons per mode. Show further that

$$\frac{\text{total length of wave packets in wavetrain}}{\text{length of wavetrain}} = p \times \tau_r\delta\nu. \qquad (11.22)$$

Interpret this as the average number of wave packets that overlap (if it's more than 1), or as the fraction of the time that is 'occupied' by light.

[35] Equivalently, at a field stop or an aperture stop.

[36] Some energy may be absorbed, or reflected from the interface. We may say that the refracting surface introduces an 'insertion loss'.

[37] We might equally well think of waves travelling along an electrical transmission line. Reasoning similar to that here is used in obtaining the thermal noise power travelling along such a line, and hence the thermal noise (Johnson noise, one-dimensional black-body radiation) radiated into the line by a resistor. See, e.g., Robinson (1974), §4.1.

[38] Compare with the large volume V that we introduce when deriving the density of states in statistical mechanics.

[39] Longitudinal and transverse modes are being distinguished in the same way as they were in Chapter 8. In all of this problem we consider a single polarization only.

[40] This result is easy to remember if we take an unconventional viewpoint and think that $\delta\nu$ modes pass per second, each occupied by p photons.

(5) Apply eqn (11.22) to a natural-broadened wavetrain, as modelled in §9.12 and problem 10.5: a sequence of identical wave packets each of duration τ_r. Show that eqn (11.22) simplifies for this case to

$$\text{fraction of wavetrain `occupied'} \approx p. \tag{11.23}$$

When applied to light of realistic intensity, this shows that photon wave packets are 'sparse', with only about 2% of the wavetrain occupied.[41]

[41] This numerical value is taken from Table 11.1. It is the basis of a number of arguments about intensity fluctuations in Chapter 10, especially in problem 10.21.

Problem 11.3 (b) Étendue and occupied transverse modes
Consider radiation occupying frequency interval $\delta\nu$, travelling in directions within solid angle $\delta\Omega$ and impinging on area ΔS at angle θ to the normal. We take $\delta\Omega = \sin\theta\,\delta\theta\,\delta\phi$ to encompass a small range of directions, so the light forms a near-collimated beam.

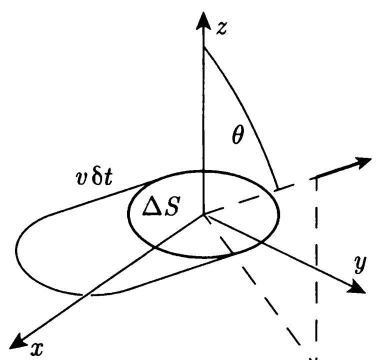

Fig. 11.8 Light arrives at area ΔS from within a cylinder making angle θ with the z-axis and of slant height $v\,\delta t$. All photons within the cylinder and travelling in direction θ meet or cross area ΔS in time δt.

(1) We know from statistical mechanics that the number of modes per unit volume within frequency range $\delta\nu$ is $(4\pi\nu^2\,\delta\nu)/c^3$, where the usual factor 2 for polarizations is omitted as it will be dealt with separately. Argue that the expression must be changed to $(4\pi\nu^2\,\delta\nu)/v^3$ if the radiation travels in a medium in which the speed of light is $v = c/n$, where n is the refractive index.

(2) The expressions given in part (1) apply to radiation travelling in all directions. Argue that if radiation travels within solid-angle range $\delta\Omega$ it must occupy a fraction $\delta\Omega/4\pi$ of these modes, so the number of modes (per volume) it occupies within $\delta\nu$ and $\delta\Omega$ is $(\nu^2\,\delta\nu/v^3)\delta\Omega$.

(3) Next adapt a standard discussion in gas kinetic theory to work out the number of photons that impinge on area ΔS within time δt from within the solid angle $\delta\Omega$. Figure 11.8 may help. Those photons that impinge within time δt are those that occupy a cylinder whose base area is ΔS and slant height is $v\,\delta t$, making an angle θ to the normal to the surface. The volume of this cylinder is $\Delta S\, v\delta t\cos\theta$. Within this are p photons per mode in $(\nu^2\,\delta\nu/v^3)\delta\Omega$ modes per unit volume. Put this together, and show that the number of photons arriving per unit time is

$$\text{photon flux} = p \times \delta\nu \times \left(\frac{n^2\,\delta\Omega\,\Delta S\,\cos\theta}{\lambda_{\text{vac}}^2}\right). \tag{11.24}$$

(4) The light beam occupies a small range of directions, so it can be described as composed of longitudinal and transverse modes, after the fashion[42] of the cavity modes in Chapter 8. Using eqn (11.21), identify $\delta\nu$ in eqn (11.24) as representing the number of longitudinal modes contributing to the photon flux.

[42] Any convenient set of eigenfunctions can be used to describe the transverse structure, not necessarily the Gauss–Hermite functions of eqn (8.1).

(5) Use the results of parts (3) and (4) to show that

$$(\text{number of transverse modes}) = \frac{n^2\,\delta\Omega\Delta S\cos\theta}{\lambda_{\text{vac}}^2} = \frac{\text{étendue}}{\lambda_{\text{vac}}^2}, \tag{11.25}$$

in agreement with eqn (11.9).

(6) Use eqn (11.24) to confirm eqns (11.11) and (11.12).[43]

[43] The present problem has concerned itself with the arrival of photons onto some surface element ΔS. The radiance of a black-body surface is obtained by appeal to the 'principle of detailed balance', requiring that what is radiated balances what is received.

Problem 11.4 (b) Étendue of a Gaussian beam

(1) Define the area ΔS of a Gaussian beam at its waist as a circle of radius w_0 where w_0 is the waist spot size defined in eqn (7.12). Take $\Delta\Omega$ as the (similarly defined) solid angle into which the beam diffracts in its far field. Show that with these definitions the étendue of a Gaussian beam comes out to be exactly λ_{vac}^2.

(2) What is wrong with the following argument? Take the same laser beam as in part (1) but take ΔS as the area of the beam at distance b (the confocal parameter) from the waist. Then ΔS is double what we had before, $\Delta\Omega$ is unchanged, so the étendue works out at $2\lambda_{\text{vac}}^2$. And an area ΔS farther from the beam waist would give an even larger value.

(3) [Harder] Generalize the discussion to the case of a Gauss–Hermite beam with transverse-mode eigenvalues l, m.

Problem 11.5 (a) Étendue and numerical aperture

(1) Confirm eqn (11.4).

(2) Combine eqns (11.4) and (11.9) to derive both parts of eqn (11.14).

Problem 11.6 (a) Reading a CD

The information track of a CD is read[44] by focusing a laser onto it through the body of the plastic disc. Show directly that refraction at the air–plastic interface has no effect on the numerical aperture, the étendue of the light beam, or the diffraction pattern formed at the focus (do not make a small-angle approximation).[45]

Does the refractive index affect the depth of focus?

[44] A fuller description of the optical system is given in §16.4.

[45] Thus, there is no advantage to be gained (and no disadvantage to be fought) by choosing any particular value for the refractive index of the plastic.

Problem 11.7 (a) Coherence area and transverse modes

A light beam consists of a mixture of the $(0,0)$ and $(1,0)$ transverse modes of eqn (8.1), with coefficients varying randomly with time. Show that the relative phases at locations $(x,y) = (-w,0)$ and $(x,y) = (w,0)$ are made unpredictable by the mixture.[46]

[46] This model shows very directly that a beam containing more than one transverse mode extends over more than one coherence area. Conversely, we see how it comes about that transverse coherence is associated with having only one transverse mode, as claimed in §11.5.

Problem 11.8 (a) The number of photons per mode

In black-body radiation, the number of photons per mode is given by the Planck distribution $(e^{h\nu/k_B T} - 1)^{-1}$. 'Mode' means that both longitudinal and transverse characteristics are specified, as is the polarization.

(1) Show that the number of photons per mode, for black-body radiation, is greatest when the frequency is low and the temperature is high.

(2) Calculate the number of photons per mode in black-body radiation for the cases given in Table 11.1 and check the values given there.

(3) Find the power that can be extracted from a black-body lamp, at wavelength 633 nm, into étendue λ_{vac}^2 and in a frequency range of 1 GHz. Consider temperatures of (a) 2000 K and (b) 6000 K. Express the results in units of photons per second and in watts.

[*My answers*: (a) 2.3×10^4 photons s^{-1}, equivalent to 7.3×10^{-15} W; (b) 4.6×10^7 photons s^{-1}; 1.5×10^{-11} W. If polarized light is required these figures must be divided by 2.]

Comment: These figures show why laser light is so spectacularly more intense than black-body light. Even a modest laser, say a He–Ne radiating 1 mW into several longitudinal modes spanning 1 GHz, is brighter by a factor between 10^8 and 10^{12}.

Comment: The conclusions here give a more quantitative view of the discussion of laser coherence in §9.15.1.

(4) Discuss whether a laser source is essential, or whether something cheaper will do, for:

(a) the 780 nm radiation source used in reading a CD in a domestic CD player

(b) the 1.3 μm radiation source used for sending information along a single-transverse-mode optical fibre carrying 2 Gbit s^{-1}

(c) the 850 nm radiation used for sending information along a 'multimode' glass fibre which supports 1622 transverse modes[47]

(d) the radiation used in a 'laser printer' giving a resolution on the paper of 600 dots per inch

(e) for exposing a hologram whose area is 100 mm by 100 mm.

[47]This number of transverse modes is obtained, for a particular set of fibre characteristics, in problem 14.1.

Problem 11.9 (a) Étendue required for the Young slits

Investigate a Young-slits experiment in which the fringes are observed by eye. We want the fringes to be bright enough to see comfortably. For this it isn't the total optical power that's of interest, but the power per unit area on the observer's retina. Consider the cases

(1) where the source is so small that the illumination is transversely coherent along the length of the Young slits as well as across their width

(2) where the source is elongated into a slit after the fashion of Fig. 3.10, giving fringes in the usual pattern of stripes.[48]

[48]*Answer*: Surprisingly, case (2) offers hardly any increase of visual brightness, though it's more comfortable to observe.

Problem 11.10 (a) Energy throughput for a grating monochromator

Let a per-frequency radiance $B(\nu)$ be defined as in §11.7, so that a light source (not a laser) is described (small angles) by

$$\begin{pmatrix}\text{power radiated in frequency range } \delta\nu \\ \text{from area } \Delta S \text{ into solid angle } \Delta\Omega\end{pmatrix} = B(\nu)\,\delta\nu\,(n^2 \Delta S\, \Delta\Omega).$$

Imagine that we send its light into a grating monochromator. The power gathered by the instrument is[49]

[49]The spectrometer is in air so we drop the n^2 factor.

$$\begin{pmatrix}\text{radiance of image of source} \\ \text{formed on entrance slit}\end{pmatrix} \times \begin{pmatrix}\text{area of ent-} \\ \text{rance slit}\end{pmatrix} \times \begin{pmatrix}\text{solid angle} \\ \text{collected}\end{pmatrix}.$$

The radiance of the image at the entrance slit, formed by a condenser lens, is at best (eqn 11.8) equal to the radiance $B(\nu)\,\delta\nu$ of the light source, so the power we can gather (ignoring insertion loss) is

$$P = B(\nu)\,\delta\nu \times (\text{area of entrance slit}) \times (\text{solid angle collected}).$$

Comment: In practical laboratories, students are instructed to 'fill the entrance slit with light, and fill the grating with light'. The reason for this injunction is now clear: you lose out if the area sending light into the monochromator is less than the full area of the entrance slit; and you lose out if light isn't sent into the full available solid angle.[50]

[50] A similar point is made in §4.9.2 and problem 4.13, but the significance of the product (area) × (solid angle) should now be clearer.

We'll now apply these ideas to a monochromator whose grating is square of side Nd, used with angle of incidence α; we are using the same configuration as in problem 4.6. The collimator lens's focal length is f.

(1) Show that light leaving the instrument's entrance slit is collected within the solid angle $\Delta\Omega = (Nd/f)^2 \cos\alpha$.

(2) Let the entrance slit have width $(\Delta y)_{\text{slit}}$. Imagine that $(\Delta y)_{\text{slit}}$ is increased from zero; it begins to degrade the instrument's resolution when $(\Delta y)_{\text{slit}}$ reaches about $f\lambda/(Nd\cos\alpha)$. Compare with problem 4.6(6) and example 4.1 part (7) and eqn (4.10). To make the entrance slit much narrower than $f\lambda/(Nd\cos\alpha)$ just wastes energy that could be admitted to the monochromator and used. So we'll probably choose to make $(\Delta y)_{\text{slit}}$ about equal to $f\lambda/(Nd\cos\alpha)$. We might even make the slit considerably wider, if we don't mind degrading resolution in struggling for all the energy we can get. Either way, we'll have

$$(\Delta y)_{\text{slit}} \approx (\mathrm{d}y/\mathrm{d}\lambda)\,\delta\lambda, \tag{11.26}$$

where $\mathrm{d}y/\mathrm{d}\lambda = pf/(d\cos\alpha)$ is the grating dispersion (referred to the *entrance* plane), p is the order of the spectrum, and $\delta\lambda$ is the resolution we agree to accept.

(3) Let the length of the entrance slit be h. Now assemble together the factors introduced above to show that

$P = B(\nu)$	per-frequency-interval radiance of source
$\times\, (\delta\nu)^2/\nu$	resolution demanded at frequency ν
$\times\, (Nd/f)^2$	$1/(f\text{-number})^2$; shape of apparatus
$\times\, hf$	(linear dimension)^2; larger apparatus permits larger slit area
$\times\, \sin\alpha$	($\approx p\lambda/d$); maximize this for the best energy transmission. (11.27)

The first two factors can't be controlled to any great extent. It's a good idea to have a small f-number, but we soon reach a limit set by aberrations in lenses or mirrors. After this, optimizing lies in doing the best we can with the last two factors. If all else fails we can increase the size of the entire spectrometer, but that has to be a last resort. The one thing we can optimize is $\sin\alpha$, making sure it isn't too far below 1. Since $\sin\alpha - \sin\theta = p\lambda/d$ and we are likely to aim to have $\theta \approx 0$, we have $\sin\alpha \approx \lambda(p/d)$. It follows that we must choose (p/d) with some care. There are reasons (unconnected with étendue, the possibility of overlapping orders) why it is best to choose p small, ideally $p = 1$. So we are pressed to choose *the most finely ruled grating that will work.*[51]

[51] 'That will work' is an allusion to the grating equation $d\sin\alpha \approx p\lambda$. Nothing will work if we choose values that make $p\lambda/d \approx \sin\alpha > 1$, either by making d too small or p too large. Additionally, a grating is usually required to record a *range* of wavelengths, and our choice of α must result in a range of θs that we can live with. Opportunities for dramatic improvement in the power P collected are not usually available unless we've been doing something silly.

(4) Why does the analysis given here need further elaboration before it can be applied to a *spectrograph*, in which the spectrum is photographed?

Comment: We compare eqn (11.27) with corresponding expressions for other instruments in problem 11.14.

Problem 11.11 (b) Resolution of Fourier-transform spectrometer[52]
An outline diagram of a Fourier-transform spectrometer is given in Fig. 10.1, and the important dimensions of it are identified in Fig. 11.9. Light reaching the detector is that which has passed through a small circular hole of diameter w at the focus of a lens of focal length f. The movable mirror on the right introduces a path difference $2x$ between the interferometer arms.

(1) Why is the limiting aperture located in front of the detector, rather than at the input?[53]

(2) Suppose that x is scanned from 0 to $x_{\max}$. Show that the resolution, in terms of wavenumber $\bar{\nu} = 1/\lambda_{\text{vacuum}}$, can be expressed as an instrumental width $\Delta\bar{\nu}$, where

$$\Delta\bar{\nu} \approx 1/(4x_{\max}), \tag{11.28}$$

within a factor 2 or so.[54] Check this against the general statement about resolution in eqn (4.6).

(3) The resolution of a spectroscopic instrument is best described by defining an *instrumental line profile*. This is the apparent frequency spectrum of a monochromatic wave when examined with the given instrument.[55] Show that a pure sinewave of angular frequency ω_0 gives an apparent frequency distribution (instrumental line profile), after Fourier computation, of[56]

$$L(\omega - \omega_0) \propto \frac{\sin(\omega - \omega_0)T}{(\omega - \omega_0)T}, \qquad \text{where} \qquad T = 2x_{\max}/c. \tag{11.29}$$

[52]This problem, and the following seven, may be compared with a rather similar treatment in Thorne (1988), especially §7.9.

[53]An aperture is shown at the input, but it is optically a little wider than the exit hole (meaning it is wider than the real image of the exit hole formed 'backwards' by the mirrors and lenses), and it is there only to minimize the admission of stray light.

[54]*Hint*: Figures 10.2 and 10.3 give a number of examples upon which a discussion may be based if desired.

[55]An instrumental line profile has been encountered previously in problems 4.8 and 5.8. Instrumental width has been encountered, for the case of a grating spectrometer in eqn (4.8), and for the case of a Fabry–Perot in eqn (5.10).

[56]*Hint*: This result is approximate, and assumes that $\omega_0 T \gg 1$.

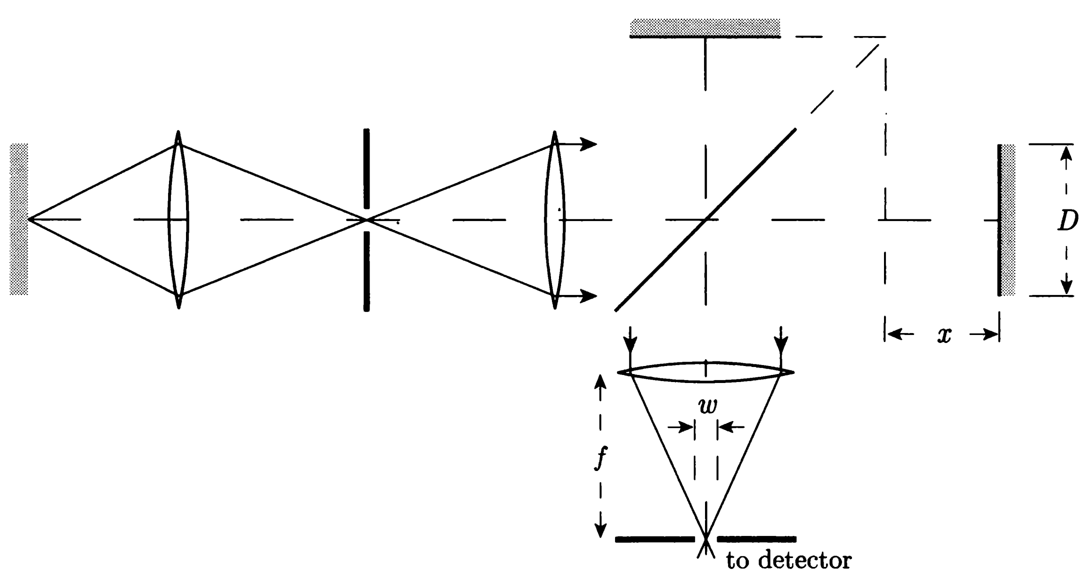

Fig. 11.9 A Michelson interferometer set up for use as a Fourier-transform spectrometer. The right-hand mirror is positioned to introduce an optical path difference of $2x$ between the two interferometer arms, distance x being adjustable over as large a range as may be necessary. Light passing through the interferometer has its range of directions limited by a small circular hole of diameter w in front of the detector.

(4) Draw sketch graphs of the functions $W(\tau)$ and $P(\omega)$ for this case and add them to the collection in §10.7.

(5) A spectral line with power spectrum $P(\omega)$ is analysed with the Fourier transform spectrometer. Consider each individual angular frequency component ω' received from the source, and work out what it contributes to the output. Show that the computed profile is

$$P_{\text{measured}}(\omega) \propto \int_{-\infty}^{\infty} P(\omega')\, L(\omega - \omega')\, \mathrm{d}\omega', \tag{11.30}$$

which is the convolution of $P(\omega)$ with the instrumental line profile.

(6) Obtain the result of part (5) by a different route. Since we stop the scan at x_{max},

$$W_{\text{measured}}(\tau) = W_{\text{source}}(\tau) \times H(\tau), \tag{11.31}$$

where $H(\tau)$ is a 'top hat' function

$$H(\tau) = \begin{cases} 1 & \text{for } -2x_{\text{max}}/c < \tau < 2x_{\text{max}}/c \\ 0 & \text{otherwise.} \end{cases} \tag{11.32}$$

Now use the convolution theorem (backwards).

(7) It is useful to define an arbitrary criterion for the limit of resolution. Since $L(\omega - \omega_0)$ has the same form as the amplitude (not intensity, note) in a single-slit diffraction pattern, it makes sense to adapt the Rayleigh criterion. Show that with such a choice,[57]

$$\Delta\bar{\nu} = 1/(4x_{\text{max}}). \tag{11.33}$$

[57] This might look 'too good' in relation to eqn (4.6). Remember that $W(\tau)$ is an even function of τ, so by measuring from 0 to x_{max} we are in fact obtaining knowledge of $W(\tau)$ from $-x_{\text{max}}$ to x_{max}. This property has been relevant from the beginning of this question, in particular in setting the limits on $H(\tau)$ in eqn (11.32).

Problem 11.12 (c) Something is wrong!

The $L(\omega - \omega_0)$ of eqn (11.29) is the result of measuring the power spectrum $P(\omega)$ for a special case. It is negative for some values of $(\omega - \omega_0)$. But power can't be negative, so something must have gone wrong. What?

Problem 11.13 (b) Improving the instrumental line profile

The $L(\omega - \omega_0)$ calculated in problem 11.11 has obvious inconvenient properties:

- it is an oscillating function of $(\omega - \omega_0)T$, so it can yield negative values of $P_{\text{measured}}(\omega)$, which must be physical nonsense
- its positive excursions can create small peaks in $P_{\text{measured}}(\omega)$ that might be misinterpreted as weak spectral lines
- by falling slowly away from line centre, it could hide a quite distant, weaker, line.

(1) Show that the most negative value of $L(\omega - \omega_0)$ is 22% of its most positive value, so the oscillations of $L(\omega - \omega_0)$ are quite serious.

(2) Suggest a way in which these artefacts could be reduced or eliminated.[58]

[58] *Hint*: Think of similar problems. Problem 2.17. Apodizing.

Problem 11.14 (b) Light gathering: Fourier-transform and grating spectrometers compared

(1) Show that if the beam splitter in a Michelson interferometer divides the amplitude into the two arms unequally, the intensity falling on the detector is reduced, but the fringe visibility is unaffected.[59]

(2) Show that, if the resolution is not to be degraded from the value found in problem 11.11, the exit pinhole of the Fourier-transform instrument must have diameter w, where[60]

$$\frac{w}{f} \ll \left(\frac{2\lambda}{x_{\max}}\right)^{1/2} \tag{11.34}$$

More realistically, the resolution may be limited both by the finite travel $x_{\max}$ of the movable mirror, and by the finite diameter w of the exit pinhole. We should construct an 'error budget' so that these two limitations result in a combined performance lying within a design requirement. A reasonable start is to allow the resolution to be degraded about equally by the two contributions. Then

$$\frac{w}{f} = \left(\frac{2\lambda}{x_{\max}}\right)^{1/2}, \tag{11.35}$$

and the overall resolution limit is changed to $\Delta\bar{\nu} = 1/(2x_{\max})$, twice the value obtained previously.

(3) Let the interferometer mirrors be square with side D. The étendue can be worked out by using the area $\pi w^2/4$ of the exit pinhole, taken with the solid angle D^2/f^2 within which light arrives at the pinhole. Show that the étendue is $\pi D^2\lambda/(2x_{\max}) = \pi D^2\lambda\Delta\bar{\nu}$. Notice that w and f have cancelled out, so there is nothing we can do to optimize the power by choosing favourable dimensions for the pinhole or the focal lengths of lenses.[61] The result obtained here should 'feel right', given the invariance of étendue.

(4) Let the source have radiance B. Show that the power reaching the detector in a bright fringe is (small angles)

$$\text{peak power at detector} = \pi B D^2\lambda\Delta\bar{\nu}. \tag{11.36}$$

(5) Now make a comparison with corresponding quantities for a grating monochromator. Assume that the grating is used with incidence angle α, the same conditions as were assumed in problem 11.10. To make the instruments roughly comparable, we'll take the grating to be square with side $Nd = D$, and the entrance slit width will be set to give the same resolution $\Delta\bar{\nu}$ as is achieved with the Fourier-transform instrument.[62]

Use eqn (11.26) to show that the spectrometer's entrance slit has width $\Delta y = (pf_s/d\cos\alpha)\Delta\lambda = f_s\lambda\tan\alpha\,\Delta\bar{\nu}$, where f_s is the focal length of the monochromator's collimating lens. Show that the étendue is

$$\text{étendue of grating monochromator} = (Nd)^2(h\lambda/f_s)\sin\alpha\,\Delta\bar{\nu}. \tag{11.37}$$

Show that the power reaching the monochromator's detector is

$$\text{power reaching detector} = B(\nu)\,\delta\nu\,(Nd)^2(h\lambda/f_s)\sin\alpha\,\Delta\bar{\nu}. \tag{11.38}$$

[59] *Comment*: There are implications here for the effort we need to put into the quality of our equipment: the reflectivity of the beam splitter is not critical; spend money and effort elsewhere.

[60] For the precise numerical factor on the right see problem 11.16.

[61] *Comment*: This should agree with previous experience in problem 4.13, where all attempts at optimizing a condenser lens in front of a spectrometer cancelled out in a similar way.

[62] It is assumed that $4x_{\max} \le Nd\sin\alpha$. If this condition is not met, the instruments are not comparable because the interferometer achieves a resolution outside the capability of the grating.

(6) Assemble these results to show that the ratio

$$R \equiv \frac{\text{detector power in interferometer}}{\text{detector power in monochromator}} = \frac{B_{\text{total}}}{B(\nu)\,\delta\nu}\left(\pi\frac{f_{\text{s}}}{h}\frac{1}{\sin\alpha}\right). \tag{11.39}$$

(7) Show that the quantity in large brackets, $R_1 = \pi f_{\text{s}}/(h\sin\alpha)$, is the ratio of the two instruments' étendues. For a monochromator, the entrance slit length h cannot be made very large because of aberrations in the focusing optics (lenses or mirrors), so that $f_{\text{s}}/h \sim 40$. We can, therefore, estimate the geometrical factor, the ratio of étendues R_1, as about 200 in favour of the interferometer.

(8) Consider now the other factor in eqn (11.39). The numerator B_{total} is the radiance of the source totalled over the whole range of frequencies that is being accepted by the interferometer. The denominator is the radiance for frequencies that lie within the range $\delta\nu$ transmitted by the grating instrument. If the radiation were 'white', we should have $\delta\nu = c\Delta\bar{\nu}$ and

$$\frac{B_{\text{total}}}{B(\nu)\,\delta\nu} \approx \frac{\text{frequency range accepted by interferometer}}{\text{resolvable frequency difference}}, \tag{11.40}$$

which can easily be of order 10^4. Even if the radiation is far from white, say a line spectrum containing 100 lines, the ratio $B_{\text{total}}/B(\nu)\,\delta\nu$ will still be 100 or so. Altogether then,

$$R \approx 200 \times \frac{B_{\text{total}}}{B(\nu)\,\delta\nu}, \tag{11.41}$$

which is likely to be in the range 10^4 to 10^7.

Comment: This problem shows that the interferometer has a large superiority in energy gathering, composed of two factors

- a factor of order 200, of purely geometric origin (larger étendue), which permits the interferometer to accept more light energy for a given resolution
- a factor because the interferometer 'looks at all of the spectrum at once'.

Of course, it is not yet obvious what the accuracy will be after our computer has used the data to calculate a Fourier transform: we wonder how errors propagate through the calculation .... Nevertheless, we do win a great deal. The factor from étendue gives an unqualified advantage to the interferometer. The second factor requires more careful analysis, and may or may not favour the interferometer depending on conditions. There is a nice discussion in Thorne (1988), §7.9.

Problem 11.15 (b) The Fourier-transform spectrometer as a whole-fringe instrument

Problem 11.14 raises an obvious general question: what étendue is theoretically possible in a spectrographic instrument with a given resolution? Here we investigate.

(1) Show that the Fourier-transform spectrometer, with the design decisions of problem 11.14, has the properties:

$$\text{(solid angle within which light is accepted)} \times \text{(chromatic resolving power)} = \pi \qquad (11.42a)$$

$$\text{(étendue)} \times \text{(chromatic resolving power)} = \pi \times \text{(area of a mirror)}. \qquad (11.42b)$$

Here the solid angle referred to is $\pi w^2/(4f^2)$ within the notation of problem 11.14; and the chromatic resolving power is $\lambda/\Delta\lambda$ as in eqn (4.5).

(2) Explain in words why étendue has to be traded against resolution.

(3) Investigate whether a Fabry–Perot has a similar property.

Comment: Equation (11.42a) is similar to the property that defines a **whole-fringe instrument**:[63]

$$\text{(solid angle)} \times \text{(chromatic resolving power)} = 2\pi. \qquad (11.43)$$

For discussion of this concept, see Jacquinot (1960). Jacquinot identifies a whole-fringe instrument as representing a useful standard of comparison so far as the gathering of light energy is concerned. We learn that a Fourier-transform spectrometer conforms to this standard.

Conversely, we learn that a grating spectrometer is *not* a whole-fringe instrument, and that it has by comparison a remarkably unfavourable étendue.

[63]Equation (11.42a) seems to fall short of this by a factor 2. This is merely because we have allowed the finite pinhole diameter and the finite mirror travel both to degrade the resolution. Show this. The whole-fringe property of eqn (11.43) applies if all of the limitation on resolution comes from the pinhole. The same factor 2 accounts for a difference between eqns (11.42b) and (11.17).

(4) Look at §8.10. The confocal Fabry–Perot accepts an even larger solid-angle range than does a normal Fabry–Perot. So does a whole-fringe instrument represent an unsurpassable optimum?

(5) Come to think of it, we could invent an improvement to a Fourier-transform spectrometer or a Fabry–Perot. Where the ring-fringes lie in front of the detector we could place an aperture that transmits several rings,[64] rather than just the central spot. Since all bright rings have the same area, we would gain a factor equal to the number of fringes transmitted. Why is this recourse rarely attempted?

[64]Not a large hole, which would wreck the resolution. A set of transparent annuli with black between them, rather like a Fresnel zone plate, arranged to match the sizes of the rings in the fringe pattern.

Problem 11.16 (c) Fourier-transform spectrometer: diameter of exit pinhole

A Fourier-transform spectrometer delivers light to its detector through a pinhole of diameter w. This diameter is now non-negligible.

(1) Show that a monochromatic input gives an intensity at the detector of $I(\tau)$, where $\tau = 2x/c$ and

$$W(\tau) = I(\tau) - \tfrac{1}{2}I(0) \propto \cos\{2kx\{1 - w^2/(16f^2)\}\} \times \frac{\sin\{kxw^2/(8f^2)\}}{kxw^2/(8f^2)}. \qquad (11.44)$$

Interpret the two factors here.

(2) Find the instrumental line profile that results from the finite size of the source aperture. Show that it is a top-hat function with full width

$$\Delta\bar{\nu} = \frac{1}{\lambda}\frac{w^2}{8f^2}. \qquad (11.45)$$

(3) Show that the $\Delta\bar{\nu}$ just calculated is the least spacing between resolvable spectral lines, if the resolution is limited entirely by the finite size of w.

(4) Problem 11.11(7) has shown that when the resolution is limited entirely by the finite travel of the moving mirror, the resolution is $\Delta\bar{\nu} = 1/(4x_{\text{max}})$. Show that, if we agree to make both limits on the resolution equally damaging, then we must choose w/f to have the value $\sqrt{2\lambda/x_{\text{max}}}$ that was claimed in eqn (11.35) of problem 11.14(2).

Problem 11.17 (b) The rate of scanning and digital sampling
A Fourier-transform spectrometer works by scanning one mirror between $x = 0$ and $x = x_{\text{max}}$. Discuss the allowable speed of the scanning and its relation to the response time of the 'slow' detector ('slow' defined as in §9.10).

During a scan, the intensity $I(x)$ reaching the detector is to be sampled at discrete values of x, digitized and fed to a computer for the Fourier analysis. What[65] is the greatest allowed separation Δx between samples?

[65] *Hint*: The idea needed is contained the **Nyquist sampling theorem**, which is well known in this kind of digital electronics. For example, it sets a lower limit on the sampling rate that must be used during the recording of sound that will end up impressed on a CD.

Problem 11.18 (a) Mechanical considerations
Consider a Fourier-transform spectrometer used at a wavelength of order 500 nm. Let the mirror diameter D be 50 mm.

(1) Estimate the tolerance on angular orientation of the movable mirror.

(2) Estimate the precision with which x must be known during its scan.

(3) Repeat these estimations for a wavelength of order 10 μm.

(4) Use your values to comment on the desirability of a Fourier-transform spectrometer as an instrument for the visible and for the infrared.[66]

[66] This merely scratches the surface. Problem 11.18 shows that a Fourier-transform instrument is easy to engineer for infrared use. But the really important thing about the infrared is how much we need the superior energy-gathering of the Fourier-transform technique. Reasons:

- Radiative lifetimes are long (roughly $\propto \omega^{-3}$) so spontaneous emission is infrequent, and spectroscopy must be done in absorption.
- Continuum sources for absorption spectroscopy are weak ($\propto \omega^2$ according to Rayleigh–Jeans).
- Detectors are notoriously insensitive once the photon energy $\hbar\omega$ gets too small to activate a quantum detection process.

Problem 11.19 (a) Energy conservation in slab interference

(1) Consider the equivalent of Fig. 11.4 for a Michelson interferometer. Light travelling to the right exhibits bright and dark fringes. In the case of the slab, energy conservation was assured by a second set of fringes, with light and dark interchanged, travelling to the left. But in the Michelson interferometer the mirrors are opaque and the 'transmitted' fringes are absent. This is drawn attention to in sidenote 31 on p. 258. Yet energy must still be conserved: what does not go into a dark fringe must go somewhere else. Where?

(2) A slab is illuminated as in Fig. 11.4, but its left-hand surface is perfectly reflecting. Discuss the conservation of energy in the resulting interference pattern.

Problem 11.20 (b) Localization of 'wedge' fringes
When light is reflected from two surfaces that are close together and whose spacing varies, we have 'fringes of equal thickness'. They are localized *at* the thin layer. Construct an explanation of this localization.