3 The Abbe theory of microscope image formation

Introduction—the interference of waves

The Abbe theory of image formation in the microscope involves the phenomena of diffraction and interference. Geometrical optics makes use of the concept of rays of light. If a point source of light (more accurately, a coherent illumination system) illuminates an opaque screen with a circular hole, an observation screen placed behind the hole will show a bright circular image of the hole. As the hole decreases in size, the image decreases in size, and in this manner one might hope to isolate a single light ray.

However, as the hole's diameter reaches a dimension smaller than about 0.1 mm, the image assumes a complex pattern, and appears to fan out from the hole, a process called **diffraction**: the apparent bending of rays at the edge of an aperture. The image is no longer a single spot, but a central spot surrounded by a series of dark and light fringes as in **1.1**, due to the **interference** of the diffracted rays. If white light is used, the fringes are coloured. Physical optics, describing light in terms of waves, attempts to explain the two processes.

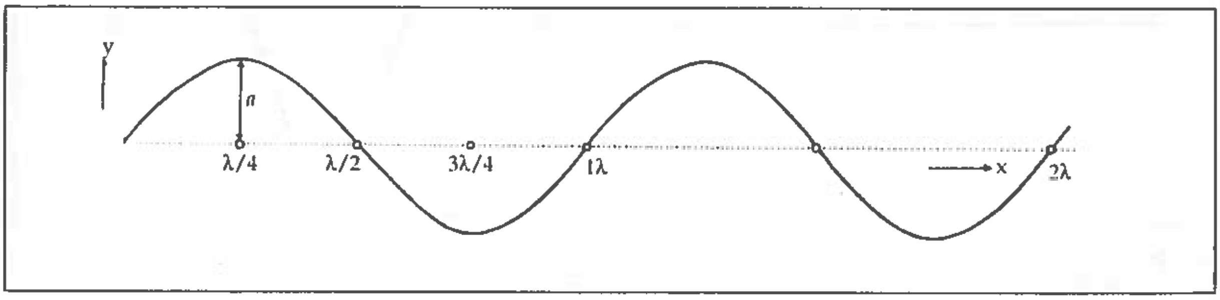
A light wave may be represented as a sine wave (**3.1**) for which the classical mathematical concepts of simple harmonic motion apply, describing the displacement of a point in terms of time (t), as well as space (x and y).

$$y = a \sin\frac{2\pi}{\lambda}(x - vt)$$

where v is the velocity of the wave.

The **phase** relationships between two particles, and between two waves, are also subject to mathematical treatment. For the immediate purposes of this discussion, the more important concepts relate to: λ, the **wavelength**, which the eye perceives as colour; a, the **amplitude**, or maximum displacement of a particle (the intensity is proportional to a^2); and the **phase difference** between two or more waves, the difference usually being defined in terms of the wavelength of the waves.

There is particular interest in an understanding of the superposition (combining) of coherent waves, especially the recombining of two waves which originate from the same wave source. Young's principle of superposition (1802) states that the resultant displacement of any one point



3.1

3.1 The sine wave (at zero time).

is the sum of the displacements due to each wave separately. The modification of intensity obtained by the superposition of two or more beams of light is called interference; *constructive interference* if the resultant intensity is greater than each of the separate intensities, *destructive interference* if less. The two simplest cases of interference are shown in **3.2**.

The figure suggests that if two waves from a single light source are recombined after traversing two different optical paths, darkness (zero intensity) will result if the path difference is $\lambda/2$, $3\lambda/2$ or $(m + ½)\lambda$, where m is a whole number; and maximum intensity will result if the path difference is $m\lambda$.

Both constructive and destructive interference occur in the same system. The effects, seen as intensity differences, are due to a redistribution of energy, and the principle of conservation of energy is not violated.

Young's experiment

Two slits illuminated by coherent light (**3.3**) act as secondary sources of light waves. All points on the screen receive waves from S_1 and S_2. The figure shows the distance S_1P_1 to be shorter than S_2P_1; interference of the waves will occur at P_1.

If the path difference S_2A is equal to 1λ, through constructive interference a bright image of the light source, S, is formed on the screen at P_1. It can be shown that bright images will be formed whenever

$$y = m\lambda\,\frac{x}{d}$$

and dark bands will occur whenever

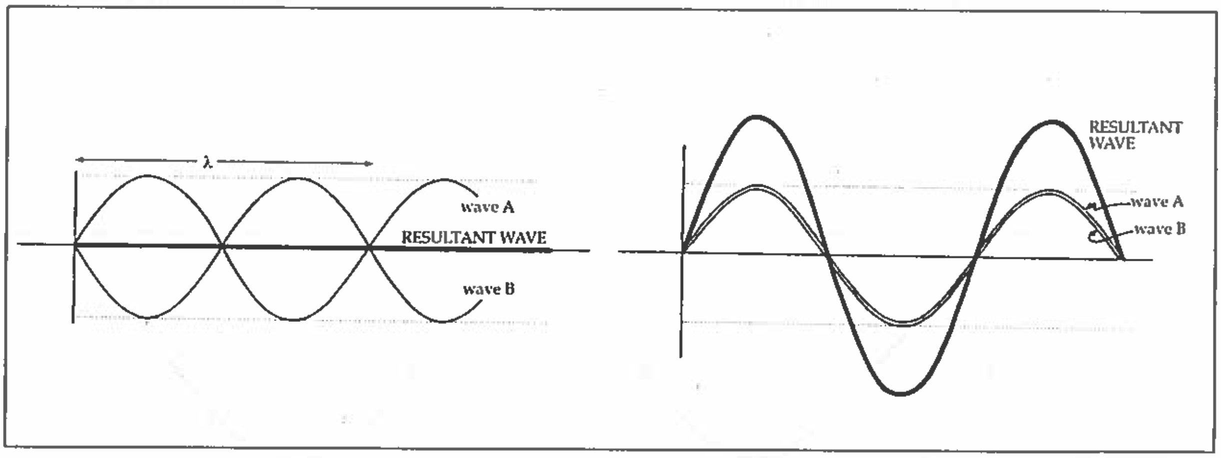
$$y = (m + ½)\,\lambda\frac{x}{d}$$

where y is the distance of the band from the axis, m is a whole number, λ is the wavelength of the illumination, x is the distance from slit to screen, and d is the distance between the centres of the slits.

The zero order maximum is formed when $m = 0$.
The first order maxima are formed when $m = \pm 1$.
The second order maxima are formed when $m = \pm 2$.

For the first order maxima, where $m=1$ in the equation $y = m\,\lambda x/d$, the following relationships

3.2



3.2 Interference: two waves of identical wavelength, showing (left) destructive interference where the phase difference between them is $\lambda/2$, $3\,\lambda/2$, $5\,\lambda/2$, etc.; and (right) constructive interference where the phase difference is $1\,\lambda$, $2\,\lambda$, $3\,\lambda$, etc.

can be seen to apply:

- As d decreases, y increases, i.e. as the slits come closer together, the interference maxima are further displaced from the axis.
- As λ decreases, y decreases, i.e. the shorter the wavelength, the closer to the axis are the maxima. If white light is used, interference spectra will form, with the blue fringe innermost.

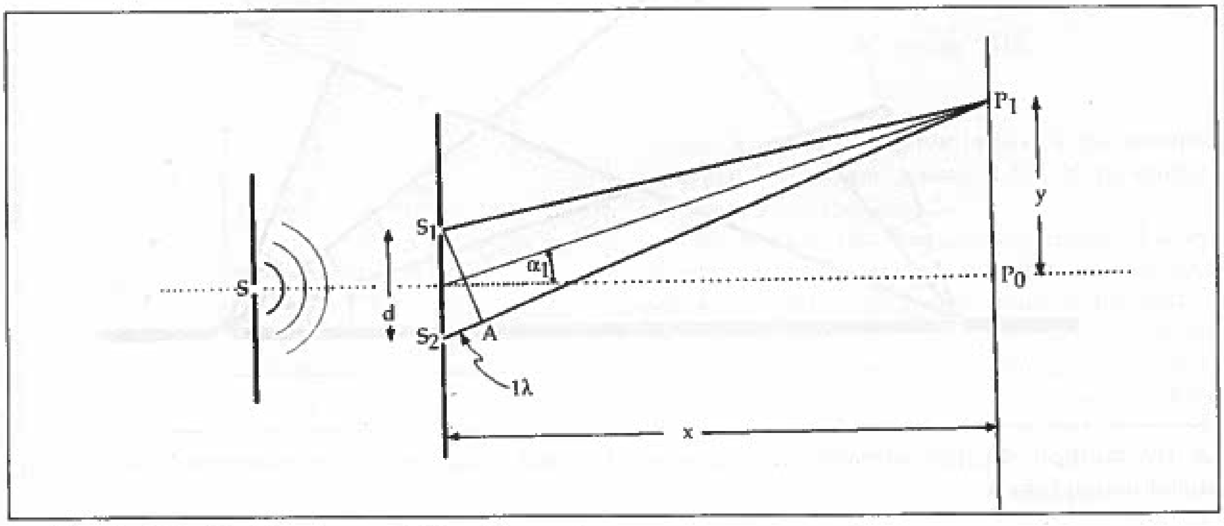
For simple demonstrations of the phenomena described above, it is common to employ multiple, rather than double, slits. The positions of the interference maxima are identical to those formed by a double slit, but the greater intensity of the maxima due to multiple slits makes for easier laboratory experimentation.

Exercise 3.1 and **Exercise 3.2**, in which cheap replica gratings are used, permit ready display and identification of the maxima. It will be noticed that there is a marked decrease in the intensity of the maxima as they are formed further from the axis. As with the precise location of the successive maxima, their intensities (energies) can be rigorously defined mathematically. Each set of maxima produced is unique to the multiple slits which produced them. If the orientation of the slits, or the distance between successive slits, is changed then the maxima are changed. The usual mental image that we have of the grating labelled '600 lines/mm' is simply a set of fine lines, 600 of them in each millimetre; but an equally valid description of the grating is a set of interference maxima with a precise orientation and a unique and precise displacement. Mathematically, the slits and their interference maxima are said to be *transforms* of each other.

The interference maxima form the (Fraunhofer) diffraction pattern of the object and, in mathematical terms, the pattern is the Fourier transform of that object. See Lipson and Lipson, 1981.

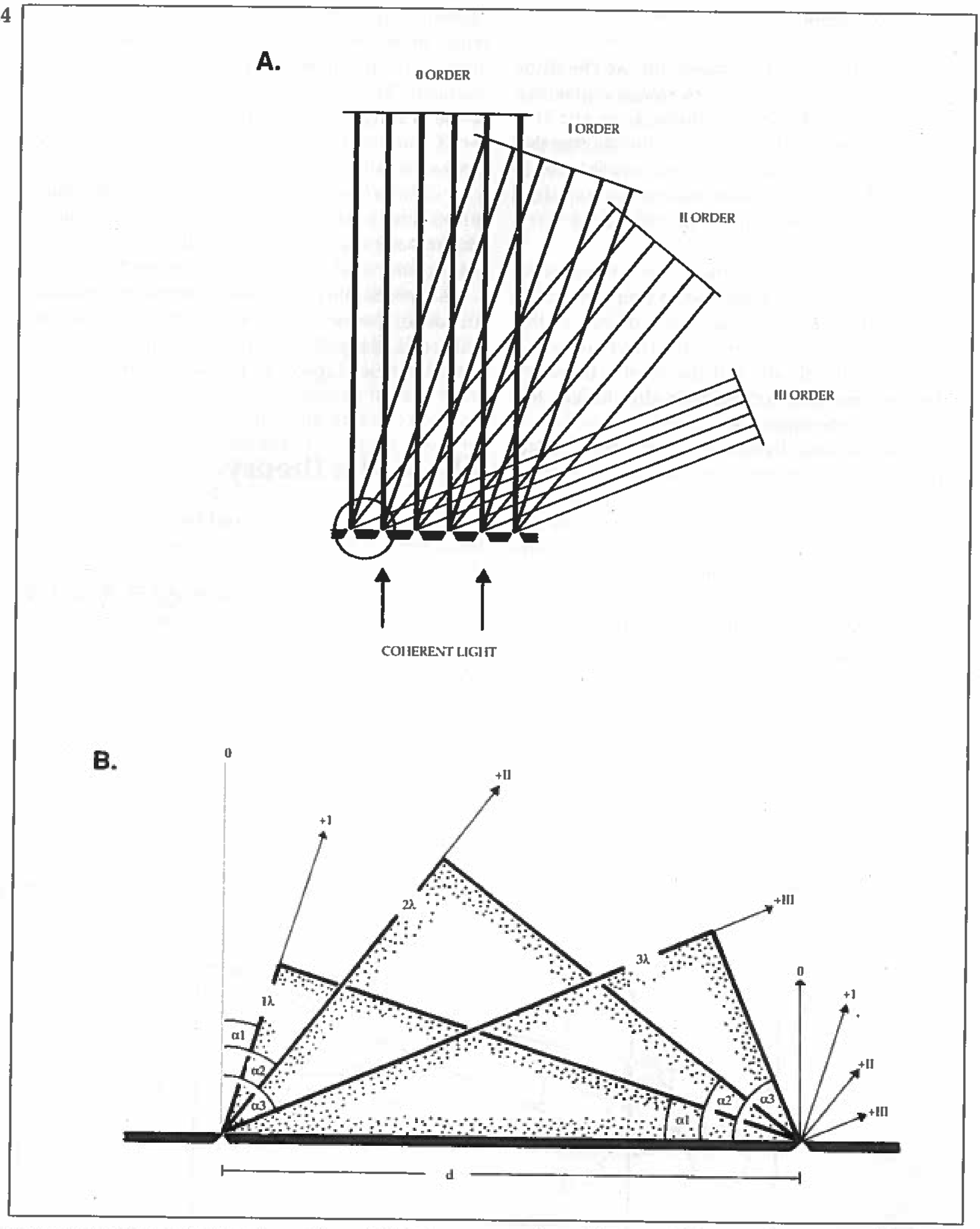
The Abbe theory

Multiple-slit diffraction and interference is the basis for the Abbe theory of image formation in the microscope. Abbe gave an explanation to the formation of the microscope image by the use of periodic structures, such as gratings, as object specimens.



3.3 The double slit, showing one first order maximum.

3.4



3.4 The multiple slit. The formation of interference maxima. Diagram **B** is an enlargement of the area circled in diagram A.

The use of periodic objects

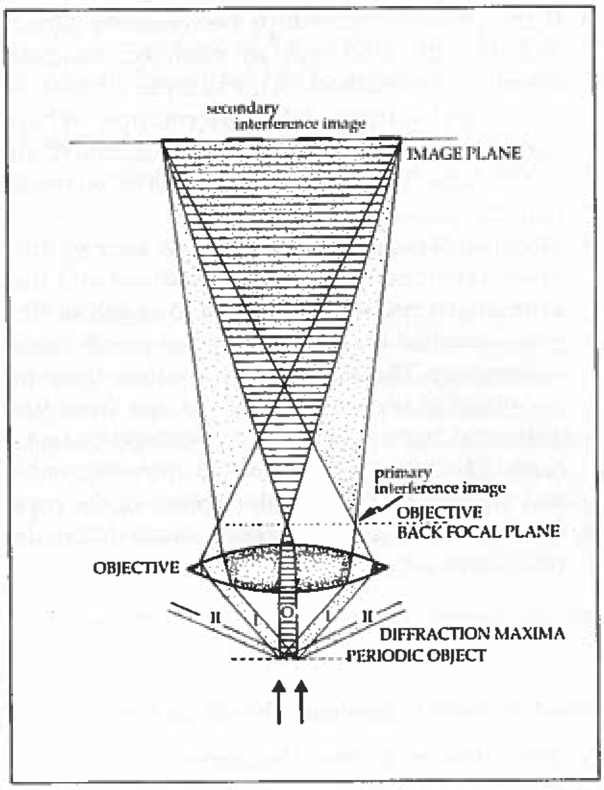
A grating, consisting of alternating dark and light bands of very fine width, acts as a multiple slit. Illuminated by parallel rays of light, the grating gives rise to diffraction and interference, so that interference maxima are formed in positions where the path difference between rays from adjacent slits is an integral number of wavelengths.

For the maxima, $\sin\alpha = m\frac{\lambda}{d}$

($m = 0, \pm1, \pm2, \pm3$, etc.; λ= wavelength. See 3.4 for definition of α and d.)

In the microscope, the objective lens collects these rays and brings the interference maxima to a focus in the objective back focal plane (3.5). This focused image is described as the **primary interference** image. The grating image is formed by **secondary interference** of these diffraction maxima. The secondary interference image is found in the intermediate image plane, in or near the eyepiece, from which plane the image normally undergoes further magnification, with eventual imaging on the retina of the eye.

3.5



3.5 The Abbe theory of image formation.

Summary

Abbe theory is briefly summarised as one which postulates diffraction by the specimen to give a primary interference image, and microscope image formation by the secondary interference of the primary interference phenomena. Secondary interference occurs, and is required, between the zero order (direct or undiffracted) rays and the higher order (diffracted) rays. If secondary interference fails to occur, image formation (a resolved image) of the grating will not occur.

The Abbe formula

The most obvious way in which secondary interference will be prevented is if d, the grating or slit width, becomes smaller and smaller; so small that the first order maxima (which are further displaced from the axis — as d decreases, y increases) fail to be accepted by the lens. The theory postulates that such a system will fail to image, or more strictly fail to resolve, the grating.

When an objective just accepts the zero and the two first order maxima as required by this theory,

$m = 1$ in the equation $\quad \sin\alpha = m\frac{\lambda}{d}$

Therefore,

$$d = \frac{\lambda}{\sin\alpha} = \frac{\lambda}{n\sin\alpha} = \frac{\lambda}{\mathrm{NA}}$$

where n is the refractive index of the medium between grating and lens and NA is the numerical aperture of the lens.

This defines the limit of resolution of a lens illuminated with parallel rays, that is, a low aperture illumination system (see page 10 for MRD of electron microscopes). If light is directed obliquely into the grating, the grating will still be resolved if only two maxima, instead of three, are accepted by the lens, say 0 and +1. In this instance the resolving power is doubled, and the limit of resolving power of a high aperture lens system can be defined as:

$$d = \frac{\lambda}{2 \times \text{NA}}$$

which is the equation used earlier in this work, and generalised to MRD = $k\frac{\lambda}{\text{NA}}$.

The practical work will seek answers to the following questions: Will any two maxima permit object structure to be resolved? **(Exercise 3.8.)** Does the zero order maximum have to be involved? **(Exercise 3.8.)** Must the maxima selected be successive maxima? **(Exercise 3.9.)**

Some consequences of Abbe theory

The theory indicates that it is secondary interference between the direct and the diffracted rays which produces the microscope image. Image formation of regular, periodic structures would now appear to have an acceptable explanation, and simple demonstrations can illustrate Abbe theory. But regular periodic structures are seldom encountered in routine microscopy. Does diffraction occur with non-periodic specimens? **(Exercise 3.3.)**

Where does diffraction occur in the microscope? A reasonable answer is everywhere, but there are certain planes, such as the objective back focal plane, where the special, focused, diffraction image is formed. Conjugate with the objective b.f.p. is the exit pupil (also called the Ramsden disc or Ramsden circle) of the eyepiece. Is the diffraction image observable in that plane? **(Exercise 2.12.)**

The diffraction image is claimed to be an optical transform of the specimen. Is there any information available concerning the specimen which is better interpreted from the diffraction image? **(Exercise 3.4.)** Since lens systems are not available for X-rays, it is only the diffraction pattern which is examined in the techniques of crystal X-ray diffraction. Orientation and spacing of atomic lattices and molecular planes are deduced from the diffraction images, although there is no 'picture' of the crystal lattice.

- If only the direct rays reach the image plane (the diffracted rays are not accepted by the objective) image formation (resolution) will not occur. It is incorrect to say that there is no image; there IS an image, but it is an unresolved image of the specimen. See **1.6 (Exercises 3.6, 3.7 and 3.8.)**
- If only the diffracted rays reach the image plane, there can be a reversal of image intensities. This is the basis for darkfield microscopy (although reflection from the specimen also plays a major role) in which a central opaque stop, or very oblique lighting, prevents the zero order rays from reaching the objective. Secondary interference between diffraction maxima permits image formation (resolution). **(Exercise 3.8.)**
- Contrast effects can be produced by modification of the zero order, undiffracted rays, for example, by closing of the condenser iris diaphragm. The reason for this increase in image contrast as the condenser (aperture) iris diaphragm is closed, is yet to be explained.
- If the phase relationships between the direct and the diffracted rays are changed, contrast effects are obtained. An example of this is extra-focal setting of the microscope. When very poor contrast specimens are examined, an out-of-focus image actually has better contrast than the perfectly focused system.
- More sophisticated techniques for altering the phase relationships between the direct and the diffracted rays have been the foundation for phase contrast and some forms of interference microscopy. The techniques have their basis in the physical separation of the direct from the diffracted rays, even with non-periodic structures. The two sets of rays can then be modified separately, so that interference of the rays in the image plane produces some desirable effect such as contrast enhancement.

Further reading

Barer, R. 1968. *Lecture Notes on the Use of the Microscope.* Blackwell Scientific Publications, Oxford. pp. 63-65.

Spencer, M. 1982. *Fundamentals of Light Microscopy.* 93 pp. Cambridge University Press, Cambridge.

Zieler, H.W. 1973. *The Optical Performance of the Light Microscope,* part 2. 110 pp. Microscope Publications Ltd., Chicago.