Chapter 2

Introduction to the Theory of Fourier Imaging

In order to appreciate the fundamental limitations of the resolution and image formation properties of optical microscopes, it is necessary to begin by discussing the foundations of diffraction theory. We shall then carry on to discuss the most important component of an optical imaging system—the lens—and finally combine our findings in a consideration of the role of Fourier analysis in the theory of coherent and incoherent image formation.

2.1 Kirchhoff diffraction and the Fresnel approximation

We take as our starting point the Kirchhoff diffraction formula, which is the mathematical interpretation of Huygens' principle. For paraxial optics the electromagnetic field may be expressed as a scalar field amplitude. If we are concerned only with radiation of angular frequency $\omega$, this may be written as

$$U = \text{Re} \{ U e^{i\omega t} \},$$

(2.1)

where $U$ is the complex amplitude and $\text{Re} \{ \}$ denotes the real part. Kirchhoff's diffraction formula [2.1] gives the amplitude in the plane $x_2, y_2$ in terms of the distribution in the plane $x_1, y_1$ (Fig. 2.1) as

$$U_2(x_2, y_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{j\lambda R} U_1(x_1, y_1) \exp(-j\lambda R) \, dx_1 \, dy_1,$$

(2.2)

where $k$ is the wavenumber, given by $k = 2\pi/\lambda$, and $\lambda$ is the wavelength. This expression assumes that $U_1$ is slowly varying compared to the wavelength, and that both $U_1$ and $U_2$ are only appreciable in a region around the optic axis which is small compared to the axial distance $z$. According to Huygens' principle each element of the wavefront $U_1$ may be considered to give rise to a spherical wave with strength proportional to $U_1$. The double integral then represents a summation over all elements of the wavefront.

![Fig. 2.1. Diffraction geometry.](image)

If we impose a more rigid condition on the maximum values of $x_1, y_1$ and $x_2, y_2$, we may replace the distance $R$ in the denominator by $z$ and expand the $R$ in the exponent by the binomial theorem to give the Fresnel approximation

$$U_2(x_2, y_2) = \frac{\exp(-j\lambda z)}{j\lambda z} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} U_1(x_1, y_1) \exp \left[ -\frac{jk}{2z} [(x_2 - x_1)^2 + (y_2 - y_1)^2] \right] \, dx_1 \, dy_1.$$

(2.3)

It should be noted that the assumption that $U_1$ is slowly varying is necessary for paraxial optics, as a quickly varying amplitude will result in diffraction through large angles.

2.2 The Fraunhofer approximation

If $z$ is large compared to the maximum of $x_1$ and $y_1$, the Fresnel approximation to the diffraction integral may be used. However, if the more stringent condition that

$$z \gg \frac{1}{2} k (x_1^2 + y_1^2),$$

(2.4)

is also satisfied, we may make the further approximation of neglecting the terms involving $x_1^2$ and $y_1^2$, to give the Fraunhofer approximation

$$U_2(x_2, y_2) = \frac{\exp(-j\lambda z)}{j\lambda z} \exp \left[ -\frac{jk}{2z} (x_2^2 + y_2^2) \right] \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} U_1(x_1, y_1) \exp \left( \frac{jk}{z} (x_1 x_2 + y_1 y_2) \right) \, dx_1 \, dy_1.$$
The condition (2.4) may be written in terms of the Fresnel number \( N \) defined as
\[
N = \pi(x_1^2 + y_1^2)_{\text{max}}/\lambda z,
\] (2.6)
so that if \( N \ll 1 \) the Fraunhofer form of the diffraction integral may be used.

We now introduce the two-dimensional Fourier transform \( \hat{U}(m, n) \) of \( U(x, y) \) which we define as
\[
\hat{U}(m, n) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} U(x, y) \exp(-2\pi i(mx + ny)) \, dx \, dy;
\] (2.7)
the inverse relationship is
\[
U(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \hat{U}(m, n) \exp(-2\pi i(mx + ny)) \, dm \, dn.
\] (2.8)

Then we can write (2.5) in the form
\[
U_2(x_2, y_2) \exp\left(\frac{jk}{2z}(x_2^2 + y_2^2)\right) = \exp\left(\frac{-jkz}{j\lambda z}\right) \hat{U}_1(x_2/\lambda z, y_2/\lambda z),
\] (2.9)
The exponential term on the left-hand side of equation (2.9) shows that a Fourier transform relationship between the original and diffracted field is satisfied on a spherical surface (to the paraxial approximation) centred on the axial point of the \( x_1, y_1 \) plane.

We shall now give three examples of Fraunhofer diffraction which we shall find important later.

2.2.1 The rectangular aperture

We define the rectangular function \( \text{rect}(x) \) by [2.1]
\[
\text{rect} x = 1, \quad |x| < \frac{1}{2},
\]
\[
= 0, \quad |x| > \frac{1}{2}.
\] (2.10)
The Fourier transform of \( \text{rect}(x/a) \) is a sinc \((am)\), the sinc function being defined as
\[
sinc \frac{\zeta}{\pi} = \sin (\pi \zeta)/(\pi \zeta).
\] (2.11)

If an aperture in an opaque screen is illuminated with a plane wave, then, providing the dimensions of the aperture are large compared to the wavelength (as they must be for our paraxial approximation), the amplitude after the aperture is simply the product of the illuminating amplitude and the transmission of the aperture. This amounts to saying that the presence of the aperture does not affect the distribution in amplitude before the aperture, which corresponds to the Kirchhoff boundary conditions. Thus if a rectangular aperture is illuminated with a uniform plane wave, the amplitude \( U_1 \) is given by a rect function. The diffracted amplitude \( U_2 \) for a rectangular aperture of dimensions \( a \) and \( b \) may thus be written
\[
U_2(x_2, y_2) \exp\left(\frac{jk}{2z}(x_2^2 + y_2^2)\right) = \frac{\exp\left(-\frac{jkz}{j\lambda z}\right) ab \sin\left(\frac{a x_2}{\lambda z}\right) \sin\left(\frac{b y_2}{\lambda z}\right)}{\zeta^2}.
\] (2.12)

Defining the angles \( \theta, \phi \) as \( \sin^{-1}(x_2/z), \sin^{-1}(y_2/z) \), we may write for the diffracted intensity, which is merely the modulus squared of the diffracted amplitude,
\[
I_2(\theta, \phi) = \left(\frac{ab}{\lambda z}\right)^2 \sin^2\left(\frac{a \sin \theta}{\lambda z}\right) \sin^2\left(\frac{b \sin \phi}{\lambda z}\right).
\] (2.13)

2.2.2 The circular aperture

If a function of two coordinates \( x, y \) is radially symmetric such that it is a function of \( r \) only then its two-dimensional Fourier transform is also radially symmetric and may be written as a Fourier–Bessel (or Hankel) transform
\[
\hat{U}(\rho) = \int_{0}^{\infty} U(r) J_n(2\pi r \rho)2\pi r \, dr,
\] (2.14)
where \( J_n \) is a Bessel function of the first kind of order \( n \). If the original amplitude is that of an evenly illuminated circular aperture radius \( a \), written as \( \text{circ}(r/a) \), where the circ function is defined
\[
\text{circ} (r) = 1, \quad r < 1,
\]
\[
= 0, \quad r > 1,
\] (2.15)
the diffracted amplitude is given by
\[
U_2(r_2) \exp\left(\frac{jk r_2^2}{2z}\right) = \frac{\exp\left(-\frac{jkz}{j\lambda z}\right) 2J_n(2\pi r_2 a / \lambda z)}{2\pi r_2 a / \lambda z}.
\] (2.16)

Here we have made use of the integral
\[
\int_{0}^{\infty} J_n(2\pi r \rho)2\pi r \, dr = \frac{J_{n+1}(2\pi \rho)}{2\pi \rho}.
\] (2.17)
2.2.3 The annular aperture

For an evenly illuminated annular aperture, outer and inner radii $a$ and $ya$ respectively, the amplitude is given by (using (2.16))

$$I(z) = \left(\frac{\pi a^2}{\lambda z}\right)^2 \left[\frac{2J_1(2\pi a \sin \theta/\lambda)}{2\pi a \sin \theta/\lambda}\right]^2, \quad \text{(2.18)}$$

where $\sin \theta$ is $r_3/z$.

2.2.3 The annular aperture

For an evenly illuminated annular aperture, outer and inner radii $a$ and $ya$ respectively, the amplitude is given by (using (2.16))

$$U(r) = \frac{\exp(-jkd)}{\exp(-jkz)} J_1(2\pi a \sin \theta/\lambda) \left[\frac{2J_1(2\pi a \sin \theta/\lambda)}{2\pi a \sin \theta/\lambda}\right], \quad \text{(2.19)}$$

In the limiting case of a thin annulus given by $(1 - \gamma) = \epsilon$, with $\epsilon$ small, the amplitude is

$$U(r) = \frac{2\pi a \sin \theta/\lambda}{2\pi a \sin \theta/\lambda} \delta(r_1 - a) J_1(2\pi a \sin \theta/\lambda) 2\pi a r_1 dr_1, \quad \text{(2.20)}$$

where $\delta$ is the Dirac delta function, giving

$$U(r) = \frac{\exp(-jkz)}{\exp(-jkz)} J_1(2\pi a \sin \theta/\lambda) 2\pi a r_1 dr_1. \quad \text{(2.21)}$$

2.3 The thin lens

We shall now consider the effects of a thin lens, that is a lens which is so thin that the rays do not experience a significant displacement upon traversing it. The lens produces a phase change proportional to the optical path, given by the line integral of the refractive index along the ray. Restricting our analysis to the quadratic approximation, we now consider a perfect thin lens, by which we mean a lens producing the quadratic phase change required to collimate a spherical wave diverging from a point at distance $f$ from the lens into a plane wave. By considering equation (2.9), we see that the phase change produced by the lens must be such as to multiply the amplitude by the factor

$$t(x, y) = \exp\left(-\frac{jk}{2f}\right)(x^2 + y^2). \quad \text{(2.22)}$$

It may be shown that a thin lens constructed from a dielectric medium with two spherical surfaces gives the required phase variations [2.1].

A real lens also has a finite physical size. The effect of this may be taken into account by introducing the pupil function of the lens $P(x, y)$, which is unity within the pupil and zero outside. In general, the pupil function can be a varying complex function of position to account for absorption in the lens, reflection at the surfaces, or variations which are purposely introduced.

Let us now illuminate the lens with a plane wave of unit strength. The amplitude just after the lens may be written

$$U_1(x_1, y_1) = P(x_1, y_1) \exp\left(-\frac{jk}{2f}\right)(x_1^2 + y_1^2). \quad \text{(2.23)}$$

The exponential factor represents a spherical wave convergent on the point at distance $f$ beyond the lens. The amplitude in the plane at distance $z$ away may be calculated using equation (2.3), to give

$$U_2(x_2, y_2) = \exp\left(-\frac{jkz}{2f}\right) \int \int P(x_1, y_1) \exp\left(-\frac{jk}{2f}\right)(x_1^2 + y_1^2) \times \exp\left(-\frac{jk}{2f}\right)(x_2^2 + y_2^2) \quad \text{(2.24)}$$

Multiplying out the squared brackets, the terms in $x_1^2, y_1^2$ will be seen to cancel if $z$ is equal to $f$. Then

$$U_2(x_2, y_2) = \exp\left(-\frac{jkf}{2f}\right) \exp\left(-\frac{jk}{2f}\right)(x_2^2 + y_2^2) \times \int \int P(x_1, y_1) \exp\left(-\frac{jk}{2f}\right)(x_1^2 + y_1^2) dx_1 dy_1. \quad \text{(2.25)}$$

The integral is the Fourier transform of the pupil function, which we denote by $h(x_2, y_2)$. Thus we can write for the intensity in the focal plane of the lens

$$I_3(x_2, y_2) = |h(x_2, y_2)|^2/\pi f^2. \quad \text{(2.26)}$$

We are often concerned with radially symmetric pupils. Then the two-dimensional Fourier transform may be written as a Fourier–Bessel transform and in the focal plane

$$U_2(r_2) = \exp\left(-\frac{jkf}{2f}\right) \exp\left(-\frac{jkr_1^2}{2f}\right) \int_0^\infty P(r_1) J_0\left(\frac{2\pi r_1 r_2}{\lambda f}\right) 2\pi r_1 dr_1. \quad \text{(2.27)}$$
If the lens has a circular pupil radius \( a \), then the amplitude is identical to that written in equation (2.16) for Fraunhofer diffraction by a circular aperture. It is convenient to introduce a normalised optical coordinate \( v \), defined by

\[
v = kr_2 \sin \alpha \approx 2\pi r_2 a/jf,
\]

where \( \sin \alpha \) is the numerical aperture of the lens (assuming imaging in air). Then we can write for the focal amplitude

\[
U_2(v) = -jn \exp(-jkf) \exp\left(-\frac{jkv^2}{4N}\right) J_0\left(\frac{2J_1(v)}{v}\right),
\]

where \( N \) is the Fresnel number given by

\[
N = \frac{\pi a^2}{j\lambda}.
\]

If \( N \) is large then the condition

\[
N \gg \frac{v^2}{4}
\]

ensures that the quadratic phase variation in equation (2.29) is negligible for reasonable values of \( v \). The intensity is thus proportional to \( |J_0(v)|^2 \), which is plotted in Fig. 2.2.

For a pupil function of the form of a thin annulus of fractional thickness \( \varepsilon \), equation (2.21) gives for the amplitude in the focal plane

\[
U_2(v) = -2jN \varepsilon \exp(-jkf) \exp\left(-\frac{jkv^2}{4N}\right) J_0(v).
\]

The intensity is now proportional to \( J_0^2(v) \), which is also plotted in Fig. 2.2. The central spot is now narrower, but the strength of the outer rings is seen to be increased.

### 2.4 The effect of defocus

We have considered the amplitude in the focal plane of a lens illuminated by a plane wave, and now discuss the amplitude in a plane a distance \( \delta z \) from the focal plane. The squared terms in \( x_1, y_1 \) in equation (2.24) no longer cancel, and we can thus write for the radially symmetric case

\[
U_2(r_z) = \exp\left(-\frac{jkr_2^2}{j\delta z}\right) \exp\left(\frac{jnr_2^2}{\lambda\delta z}\right) \times \int_0^\infty P(r_1) \exp\left[\frac{jkr_2^2}{z} - \frac{1}{z}\right] J_0\left(\frac{2\pi r_1 r_z}{\lambda z}\right) 2\pi r_1 dr_1.
\]

We can regard the integral as the Fourier–Bessel transform of the product of the pupil function and the complex exponential which is present as a result of the defocus. This product may be thought of as a generalised pupil function, which is complex to account for the aberration of the wavefront at the pupil.

For a circular pupil we can write, introducing \( \rho = r_1/a \),

\[
U_2(v) = -jn \exp(-jkf) \exp\left(-\frac{jv^2}{4N}\right) \times \int_0^1 2 \exp\left[-\frac{jkr_2^2a^2}{z} - \frac{1}{z}\right] J_0(v) \rho \, d\rho.
\]

The wavefront aberration rises to its greatest value at the edge of the pupil where it is equal to \( \frac{1}{4}a^2(1/f - 1/z) \) wavelengths. If we define the normalised optical coordinate \( u \) by

\[
u = ka^2\left(\frac{1}{f} - \frac{1}{z}\right),
\]
the amplitude is

\[ U_2(u, v) = -jN \exp(-jkz) \exp\left(-\frac{ju^2}{4N}\right) \]
\[ \times \int_0^1 2 \exp\left(\frac{1}{2}j\mu \rho^2\right) J_0(\nu \rho) \rho \, d\rho. \quad (2.36) \]

If \( z = f + \delta z \) with \( \delta z \) small

\[ u \approx k \delta z a^2 f^2 \approx 4k \delta z \sin^2 (\alpha/2) \quad (2.37) \]

and \( u \) is linearly related to the distance from the focal plane. Then the maximum wavefront aberration is \( 2 \delta z \sin^2 (\alpha/2) \).

Along the optic axis we obtain for the amplitude

\[ U_2(u, 0) = -jN \exp(-jku) \exp\left(-\frac{k}{4}u^2\right) \]  
\[ \frac{\sin (u/4)}{u/4}. \quad (2.38) \]

Or for the intensity

\[ I(u, 0) = N^2 \left[ \frac{\sin (u/4)}{u/4} \right]^2. \quad (2.39) \]

In general the intensity may be written

\[ I(u, v) = N^2[C^2(u, v) + S^2(u, v)]. \quad (2.40) \]

where \( C(u, v) \) and \( S(u, v) \) are defined as \[2.2\]

\[
C(u, v) = \left\{ \begin{array}{l}
\frac{1}{2} \cos\left(\frac{1}{2}j\mu \rho^2\right) J_0(\nu \rho) \rho \, d\rho, \\
0
\end{array} \right. \]
\[ S(u, v) = \left\{ \begin{array}{l}
\frac{1}{2} \sin\left(\frac{1}{2}j\mu \rho^2\right) J_0(\nu \rho) \rho \, d\rho, \\
0
\end{array} \right. \quad (2.41) \]

These integrals may be evaluated numerically or expressed in terms of Lommel functions. The behaviour of the intensity in the focal region is illustrated in Fig. 2.3, which shows contours of constant intensity, normalised to unity at the focal point. The lines \( u = v \) correspond to the shadow edge given by geometrical optics for the paraxial case.

If the pupil function is a thin annulus, the integral in equation (2.32) may be evaluated directly, to give for the amplitude

\[ U_2(u, v) = -2jNz \exp(-jkz) \exp\left(-\frac{j\pi^2}{4N}\right) \exp\left(\frac{1}{2}j\mu \right) J_0(\nu \rho). \quad (2.42) \]
The important feature here is that the intensity variation with distance from the optic axis is independent of the value of \( u \) within the range of the Fresnel approximation, that is the depth of focus is exceedingly large. As the beam propagates the radiation diffracts away from the axis, but power is simultaneously diffracted inwards from the strong outer rings. A beam with intensity distribution given by \( J_0(t) \) will propagate without spreading as a result of this dynamic equilibrium: it is a mode of free space.

The imaging properties of lenses and mirrors with annular aperture have been the subject of considerable interest since the work of Airy [2.3] in 1841. In the annular lens the central peak is sharpened but at the expense of increasing the strength of the outer bright rings (Fig. 2.2). The intensity in the focal plane and along the optic axis for an annulus of finite width has been calculated by Steward [2.4, 2.5] who also showed [2.5] that the intensity distribution along the optic axis is stretched out relative to that of a circular lens, that is, the depth of field is increased. This increased depth of field, however, is unfortunately not useful for examining extended objects in the conventional microscope [2.6], as the increase in brightness in the outer diffraction rings results in a loss of contrast, and an \( n \)-fold increase in focal depth involves an \( n \)-fold loss of light. Because a laser is used as a light source in scanning microscopy this latter point, however, is not a serious drawback in this case.

2.5 Coherent imaging

Let us now assume we have a transparency which is sufficiently thin that it may be described completely by a complex amplitude transmittance \( t(x, y) \), of which the variations in modulus represent the variations in absorption in traversing the transparency, whereas the variations in phase account for the optical path travelled. If this is illuminated with an axial plane wave of unit strength the amplitude after the transparency is similarly \( t(x, y) \).

If the transparency is placed at a distance \( d_1 \) in front of a lens (Fig. 2.4), the amplitude \( U_2(x_2, y_2) \) immediately behind the lens is found by applying the Fresnel diffraction formula and multiplying by the pupil function and phase factor for the lens, to give

\[
U_2(x_2, y_2) = P(x_2, y_2) \exp \left( \frac{jk}{2d_1} \right) (x_2^2 + y_2^2)^{1/2} \frac{\exp (-jkd_1)}{jkd_1} \times \int \int \int t(x_1, y_1) \exp \left( \frac{-jk}{2d_1} \right) ((x_2 - x_1)^2 + (y_2 - y_1)^2) \, dx_1 \, dy_1.
\]

(2.43)

The amplitude \( U_3(x_3, y_3) \) in a plane at a distance \( d_2 \) behind the lens is then given by a further application of equation (2.3)

\[
U_3(x_3, y_3) = -\exp \left( \frac{jk(d_1 + d_2)}{\lambda^2 d_1 d_2} \right) \int \int \int P(x_2, y_2) t(x_1, y_1)
\times \exp \left( \frac{-jk}{2d_1} \right) ((x_2 - x_1)^2 + (y_2 - y_1)^2)
\times \exp \left( \frac{-jk}{2d_2} \right) ((x_3 - x_2)^2 + (y_3 - y_2)^2)
\times \exp \left( \frac{jk}{2d_2} \right) (x_3^2 + y_3^2) \, dx_1 \, dy_1 \, dx_2 \, dy_2.
\]

(2.44)

If the condition known as the lens law

\[
\frac{1}{d_1} + \frac{1}{d_2} = \frac{1}{\lambda f}
\]

(2.46)
is satisfied, and furthermore with
\[ d_2 = Md_1, \] (2.47)
we obtain
\[ U_3(x_3, y_3) = -\frac{\exp \left\{ -jkd_1(1 + M) \right\}}{\lambda^2 Md_1^2} \]
\[ \times \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} P(x_2, y_2) \, t(x_1, y_1) \exp \left\{ -\frac{jk}{2d_1} (x_1^2 + y_1^2) \right\} \]
\[ \times \exp \left\{ -\frac{jk}{2Md_1} (x_3^2 + y_3^2) \right\} \]
\[ \times \exp \left\{ \frac{jk}{d_1} \left( x_2 \left( x_1 + \frac{x_3}{M} \right) + y_2 \left( y_1 + \frac{y_3}{M} \right) \right) \right\} \, dx_1 \, dy_1 \, dx_2 \, dy_2. \] (2.48)

Performing the integral in \( x_2, y_2 \) we have
\[ U_3(x_3, y_3) = -\frac{\exp \left\{ -jkd_1(1 + M) \right\}}{\lambda^2 Md_1^2} \]
\[ \times \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} t(x_1, y_1) \exp \left\{ -\frac{jk}{2d_1} (x_1^2 + y_1^2) \right\} \]
\[ \times \exp \left\{ \frac{jk}{d_1} \left( x_2 \left( x_1 + \frac{x_3}{M} \right) + y_2 \left( y_1 + \frac{y_3}{M} \right) \right) \right\} \, dx_1 \, dy_1. \] (2.49)

where
\[ h(x, y) = \int_{-\infty}^{+\infty} P(x_2, y_2) \exp \left\{ \frac{jk}{d_1} (x_2 x + y_2 y) \right\} \, dx_2 \, dy_2 \] (2.50)
is the Fourier transform of the pupil function, as introduced in equation (2.26). Suppose now that our object consists of a single bright point in an opaque background, so that
\[ t(x_1, y_1) = \delta(x_1) \, \delta(y_1). \] (2.51)

Then the amplitude is a constant times \( h(x_3/M, y_3/M) \), and the latter is called the amplitude point spread function or impulse response of the optical system. The distance \( x_3/M \) represents a distance in the object plane, and \( M \) is the linear magnification of the image. The intensity is given by the modulus squared of \( h(x_3/M, y_3/M) \), again multiplied by a constant, and this is known as the intensity point spread function. For a circular pupil the intensity is (following equations 2.16, 2.29)
\[ I(v) = \left| \frac{2J_1(v)}{v} \right|^2, \] (2.52)
where the normalised coordinate \( v \) is given by
\[ v = \frac{2\pi r a}{\lambda d_1} \] (2.53)
and we have normalised the intensity to unity on the optic axis. Equation (2.52) represents the Airy disc, as shown in Fig. 2.2.

If the lens law (2.46) is not satisfied then for small departures from the focal plane the amplitude point spread function is as given by the previous section on the effects of defocus. For a circular pupil we have
\[ h(u, v) = C(u, v) + jS(u, v), \] (2.54)
where we have normalised to unity for \( u = v = 0 \), and \( u \) is given by
\[ u = ka^2 \left( \frac{1}{d_i} - \frac{1}{d_1} - \frac{1}{d_2} \right). \] (2.55)

Introducing \( x' = x_1 + x_3/M, y' = y_1 + y_3/M \) in equation (2.49) we have
\[ U_3(x_3, y_3) = -\frac{\exp \left\{ -jkd_1(1 + M) \right\}}{\lambda^2 Md_1^2} \]
\[ \times \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} t(x_1, y_1) \exp \left\{ -\frac{jk}{2Md_1} (x_3^2 + y_3^2) \right\} \]
\[ \times \exp \left\{ -\frac{j}{d_1} \left( x_2 \left( x_1 + \frac{x_3}{M} \right) + y_2 \left( y_1 + \frac{y_3}{M} \right) \right) \right\} \]
\[ \times \exp \left\{ \frac{jk}{Md_1} \left( x_2 \left( x_1 + \frac{x_3}{M} \right) + y_2 \left( y_1 + \frac{y_3}{M} \right) \right) \right\} \, dx_1 \, dy_1 \] (2.56)

For an imaging system of reasonable quality the spread function falls off quickly, so that \( x', y' \) are small. The exponential terms in \( x'^2, y'^2 \) and \( x'x_3, y'y_3 \) can therefore be replaced by unity to give
\[ U_3(x_3, y_3) = -\frac{\exp \left\{ -jkd_1(1 + M) \right\}}{\lambda^2 Md_1^2} \]
\[ \times \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} t(x_1, y_1) h(x_3/M, y_3/M) \, dx_1 \, dy_1. \] (2.57)
The integral is the convolution of the object transmittance with the point spread function, the M's resulting in a magnification $M$ in the image, and the positive sign in the argument of the spread function corresponding to an inverted image. Of the two complex exponential terms in (2.57) the first is a constant phase term which therefore does not affect the image.

The second phase factor in (2.57) represents a spherical phase variation which may also be neglected if we are concerned with small changes in $x_3$, $y_3$. In most of the following, however, we are interested in optical systems where the optical beam travels along the axis, in which case there is no phase variation to worry about.

The intensity is clearly given by

$$I_3(x_3, y_3) = \frac{1}{\lambda^2 MD^2} \left| \int_{-\infty}^{+\infty} t(x_1, y_1) h(x_1 + x_3 / M, y_1 + y_3 / M) \, dx_1 \, dy_1 \right|^2. \quad (2.58)$$

### 2.6 Imaging of line structures in coherent systems

In the previous section we derived a general expression for the image in a coherent imaging system, and also considered the image of the single point object. An important class of objects are those in which the transmittance is a function of one direction only, let us say $t(x_1)$. Using equation (2.57) the image is, disregarding the phase-terms,

$$U_3(x_3, y_3) = \frac{1}{\lambda^2 MD^2} \int_{-\infty}^{+\infty} t(x_1) h\left(x_1 + \frac{x_3}{M}, y_1 + \frac{y_3}{M}\right) dx_1 \, dy_1. \quad (2.59)$$

Considering the integral $y_1$, we see that

$$\int_{-\infty}^{+\infty} h\left(x_1 + \frac{x_3}{M}, y_1 + \frac{y_3}{M}\right) \, dy_1 = \int_{-\infty}^{+\infty} h(x_1) \, dy_1, \quad (2.60)$$

with the result that, as we might expect, the image is independent of the coordinate $y_3$. The integral may be written in terms of the pupil function,

using equation (2.50) to give

$$\int_{-\infty}^{+\infty} h(x_1, y_1) \, dy_1 = \int_{-\infty}^{+\infty} P(x_2, y_2) \exp\left(\frac{jk}{d_1} (x_1 x_2 + y_1 y_2)\right) \, dx_2 \, dy_2 \, dy_1$$

$$= \int_{-\infty}^{+\infty} P(x_2, 0) \exp\left(\frac{jkx_1 x_2}{d_1}\right) \, dx_2 = g(x_1). \quad (2.61)$$

The image may thus be written

$$U_3(x_3, y_3) = \frac{1}{\lambda^2 MD^2} \int_{-\infty}^{+\infty} t(x_1) g\left(x_1 + \frac{x_3}{M}\right) \, dx_1, \quad (2.63)$$

which is the convolution of the object transmittance and $g(x_1)$. The quantity $g(x_1)$ is called the line spread function, and is the amplitude image of a bright line.

If the pupil is radially symmetric, the line spread function is given by equation (2.62) as the one-dimensional Fourier transform of $P(r_2)$, as compared with the point spread function, which is the Fourier-Bessel, or two-dimensional Fourier, transform of $P(r_2)$. For a lens with $|x| \leq a$

$$g(x_1) = \int_{-\infty}^{+\infty} \exp\left(\frac{2\pi j x_1 x_2}{\lambda d_1}\right) \, dx_2$$

$$= 2a \left(\frac{\sin v}{v}\right) \quad (2.64)$$

where (equation 2.28)

$$v = k x_1 a / d_1. \quad (2.65)$$

This is plotted in Fig. 2.5, where it is normalised to unity at the origin. This result may also be derived by direct integration of the point spread function as in (2.61). For a thin annular pupil, on the other hand, we have

$$g(x_1) = \int_{-\infty}^{+\infty} \{\delta(x_2 - a) + \delta(x_2 + a)\} \exp\left(\frac{2\pi j x_1 x_2}{\lambda d_1}\right) \, dx_2$$

$$= 2 \cos v, \quad (2.66)$$

which again is shown normalised in Fig. 2.5. The side-lobes are now as strong...
as the main lobe, and hence imaging of such an extended object with a thin annular pupil is completely useless. The problem is that in the spread function of the annular pupil, the intensity in successive outer rings does not decay to zero.

Another object of great importance is the step object, which has a transmittance defined by

\[ t(x_1) = 0, \quad x_1 < 0, \]
\[ = 1, \quad x_1 > 0. \]  

(2.68)

The image may be calculated for a circular pupil using (2.63) to give

\[ U_3(x_3) \propto \int_{-\infty}^{\infty} \left( \frac{\sin v}{v} \right) \, dv, \]  

(2.69)

\[ \propto \frac{1}{2} - \frac{1}{\pi} \text{Si}(v), \]  

(2.70)

where Si is a sine integral and (2.70) is normalised to unity for large negative v. The intensity is plotted in Fig. 2.6, showing that the image exhibits fringes. For a thin annular pupil, the image integrates to a constant.

2.7 The coherent transfer function

An alternative approach to imaging is to consider the object in terms of its spatial frequencies. A periodic object may be resolved into a Fourier series, whereas for a non-periodic object we must use a Fourier transform. The Fourier transform of the object transmittance is

\[ T(m, n) = \iint \! t(x, y) \exp \left( \frac{2\pi j(mx + ny)}{\lambda} \right) \, dx \, dy \]  

(2.71)

FIG. 2.5. The coherent line spread function for both a circular and an annular lens.

where m, n are spatial frequencies in the x, y directions respectively, so that 1/m, 1/n are the corresponding spatial wavelengths. We are thus considering the object as a superposition of gratings. The inverse transform relation gives us

\[ t(x, y) = \iint \! T(m, n) \exp \left( \frac{-2\pi j(mx + ny)}{\lambda} \right) \, dm \, dn \]  

(2.72)

and substituting this in equation (2.57) we have

\[ U_3(x_3, y_3) = \frac{1}{\lambda^2 M \Delta z^2} \iint \iint \! T(m, n) \exp \left( \frac{x_1 + x_3}{M} + \frac{y_1 + y_3}{M} \right) \times \exp \left( \frac{-2\pi j(mx_1 + ny_1)}{\lambda} \right) \, dm \, dn \, dx_1 \, dy_1. \]  

(2.73)
Performing the integrals in \( x_1, y_1 \) using the inverse of equation (2.50) we obtain

\[
U_3(x_3, y_3) = \frac{1}{\lambda^2 M d_i^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} T(m, n) c(m, n) \times \exp \frac{2\pi j}{M} (mx_3 + ny_3) \, dm \, dn. \tag{2.74}
\]

The image may thus be found by resolving the object transmittance into its Fourier spectrum, multiplying by a coherent transfer function, which gives the strength of the various Fourier components in the image, and then inverse transforming to give the image amplitude. We may thus write

\[
U_3(x_3, y_3) = \frac{1}{\lambda^2 M d_i^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} T(m, n) c(m, n) \times \exp \frac{2\pi j}{M} (mx_3 + ny_3) \, dm \, dn \tag{2.75}
\]

where the coherent transfer function \( c(m, n) \) is given by

\[
c(m, n) = P(m \tilde{d}_i, n \tilde{d}_i). \tag{2.76}
\]

The positive exponent in (2.75) indicates that the image is inverted. For the case of a circular pupil we may further write

\[
c(m, n) = P(\tilde{m} a, \tilde{n} a) \tag{2.77}
\]

where \( \tilde{m}, \tilde{n} \) are normalised (dimensionless) spatial frequencies and \( a \) is the lens radius.

If we now consider a line structure for which \( n = 0 \), then the transfer function is

\[
c(m, 0) = \begin{cases} 1, & |\tilde{m}| < 1, \\ 0, & |\tilde{m}| > 1. \end{cases} \tag{2.78}
\]

This spatial frequency cut-off corresponds to a spatial wavelength in the object of \( \lambda/2\pi \).

Let us consider an object which may be described by

\[
t(x) = 1 + b \cos 2\pi vx, \tag{2.79}
\]

that is, it consists of only one pair of spatial frequency, one positive and one negative. The Fourier transform of the object is

\[
T(m, n) = \left[ \delta(m) + \frac{b}{2} \delta(m - v) + \frac{b}{2} \delta(m + v) \right] \delta(n). \tag{2.80}
\]

If the transfer function is even, we have for the image amplitude

\[
U(x_3) = c(0) + b c(v) \cos 2\pi vx, \tag{2.81}
\]

and for the intensity

\[
I(x_3) = |c(0)|^2 + \frac{1}{2} |b|^2 |c(v)|^2 + 2 \Re \{ c^*(0)c(v)b \} \cos 2\pi vx + \frac{1}{2} |b|^2 |c(v)|^2 \cos 4\pi vx. \tag{2.82}
\]

where \( \Re \{ \} \) denotes the real part and \( * \) denotes the complex conjugate. If the modulus of \( b \) is small such that we can neglect \( |b|^2 \) this becomes

\[
I(x_3) = |c(0)|^2 + 2 \Re \{ c^*(0)c(v)b \} \cos 2\pi vx, \tag{2.83}
\]

which is a linear image of the object amplitude transmittance. If the transfer function is real, as it is for an aberration free pupil, and \( b \) is purely imaginary, there is no image.

For an object which has a cosinusoidal variation in absorption coefficient, refractive index or thickness (or height in a reflection specimen) we can write

\[
t(x) = \exp \{ b \cos 2\pi vx \}. \tag{2.84}
\]

In this case there are an infinite series of spatial frequencies present, but if \( |b| \) is small we can expand as a power series

\[
t(x) = 1 + b \cos 2\pi vx + \frac{1}{2} b^2 \cos^2 2\pi vx + \ldots \tag{2.85}
\]

If we can neglect terms in higher order than \( b \), equation (2.85) reduces to (2.79) and the image is given by equation (2.83). If the term of order \( b \) in equation (2.83) is zero, however, then we must include terms of order \( b^2 \) in equation (2.85), in which case equation (2.82) is no longer valid.

### 2.8 The angular spectrum

The amplitude of a plane wave of unit strength at a point \( r \) is given by

\[
U(r) = \exp -j(k \cdot r) \tag{2.86}
\]

where \( k \) is the wave vector such that if the direction cosines of the direction of propagation are \( \alpha, \beta, \gamma \), this may be written

\[
U(x, y, z) = \exp \frac{-2\pi j}{\lambda} (ax + by + cz), \tag{2.87}
\]
where \(\alpha, \beta, \) and \(\gamma\) are related by
\[
\alpha^2 + \beta^2 + \gamma^2 = 1. \tag{2.88}
\]

In the plane \(z = 0\), the plane wave is thus
\[
U(x, y, 0) = \exp \left(\frac{-2\pi j}{\lambda} (\alpha x + \beta y)\right). \tag{2.89}
\]

If we illuminate an object with transmission \(t(x, y)\) we have, in terms of the Fourier Spectrum of the object (equation 2.72),
\[
t(x, y) = \int_{-\infty}^{+\infty} T(m, n) \exp \left(-2\pi j (mx + ny)\right) dm \, dn. \tag{2.90}
\]

Comparing equations (2.89) and (2.90), we see that we can think of the amplitude immediately behind the object as being made up of many plane waves travelling in directions
\[
\alpha = m\ell, \quad \beta = n\ell, \tag{2.91}
\]
where the strength of the particular plane wave is
\[
T(m, n) = T(\alpha/\ell, \beta/\ell). \tag{2.92}
\]

The limits of the integral in equation (2.90) require that \(\alpha, \beta\) be allowed to vary in the range \(-\infty\) to \(+\infty\). Our assumptions of paraxial optics assume that \(\alpha\) and \(\beta\) are small, and this condition is satisfied if the object transmittance is slowly varying relative to the wavelength. Otherwise waves with \(\alpha\) and \(\beta\) greater than unity are produced. These evanescent waves decay quickly with \(z\), as by equation (2.88) \(\gamma\) is complex. In our case the lens, assumed to be of small aperture, collects only waves for small \(\alpha\) and \(\beta\), and the presence of these other components need not concern us.

We now see a physical picture for the transfer function of an imaging system, for if a spectral component with spatial frequency \(m\) in the object results in a wave propagating at an angle \(\theta\) to the optic axis we have from equation (2.90)
\[
\theta \approx m\ell, \tag{2.93}
\]
and the transfer function will cut off (Fig. 2.4) when
\[
\theta \approx \frac{\pi}{\lambda d_1} \approx m\ell. \tag{2.94}
\]

In the plane \(z = 0\), the plane wave is thus
\[
U(x, y, 0) = \exp \left(\frac{-2\pi j}{\lambda} (\alpha x + \beta y)\right). \tag{2.95}
\]

where \(a\) is the radius of the lens pupil. This is the basis of the Abbe theory of microscope imaging.

2.9 Incoherent imaging

In the previous sections of this chapter we have been considering coherent imaging systems. These systems produced images which were linear in amplitude, in the sense that the amplitude image of each point in the object transparency added to give the final amplitude image. The intensity image is given by the modulus square. We now consider the other extreme of incoherent imaging, which is linear in intensity such that the intensities of individual point images add. Such an object might be formed if it is self-luminous, if it emits light such that there is no phase coherence between the different points. Alternatively a transparency may be illuminated incoherently, as we shall discuss in the next chapter.

As the intensities in the images of the individual points add, we have for an object of amplitude transmittance \(t(x_1, y_1)\)

\[
I(x_3, y_3) = \frac{1}{\lambda^4 M^2 d_1^4} \int_{-\infty}^{+\infty} |\hat{h}(x_1 + x_3, y_1 + y_3)T(x_1, y_1)|^2 \, dx_1 \, dy_1 \tag{2.96}
\]

that is, the convolution of the intensity transmittance of the object and the intensity point spread function.

For a single point situated at \(x_1 = x, y_1 = y\), the image intensity is given by

\[
I(x_3, y_3) = \frac{1}{\lambda^4 M^2 d_1^4} \int_{-\infty}^{+\infty} |\hat{h}(x_1 + x_3, y_1 + y_3)\delta(x_1 - x, y_1 - y)| \, dx_1 \, dy_1 \tag{2.97}
\]

which leads to precisely the same result as equation (2.58) for the coherent case.

If we again consider line structures such that the transmittance is a function of one direction only, we can obtain from equation (2.95), by analogy to equation (2.62), an incoherent line spread function \(g'(x_1)\), such that

\[
g'(x) = \int_{-\infty}^{+\infty} |\hat{h}(x_1, y_1)| \, dy_1. \tag{2.98}
\]
In general it is not possible to express this integral simply, in terms of the pupil function of the lens, as in the coherent case. However, by substituting the appropriate point spread function into equation (2.95), we obtain \( g'(x) \) by direct integration. For a circular lens we have

\[
g'(v) = \int_{-\infty}^{\infty} T(m, n) \exp(-2\pi j mx + ny) \, dm \, dn
\]

which may be written as

\[
g'(v) = \frac{3\pi}{v^2} H_1(2v)
\]

where \( H_1 \) is a first-order Struve function and where we have normalised to unity at the origin. The incoherent line spread function is illustrated in Fig. 2.7.

We may also consider the image of the straight edge (equation 2.68) which may be calculated from equation (2.100), for a system using circular lenses, as

\[
I(x_3, y_3) = \frac{1}{\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} T(m, n) C(m, n) \exp(-2\pi j mx_3 + ny_3) \, dm \, dn
\]

with

\[
C(m, n) = P(m, n^2, n^2) \otimes P^*(m, n^2, n^2)
\]

where \( H_0 \) is a zero order Struve function, and where we have used Struve's integral [2.7, p. 497] and normalised (2.102) to unity at large negative values of \( v \). This is plotted in Fig. 2.6 and we see that the fringing which characterised the coherent response is absent. It is also important to note that the apparent position of the edge is different in the two cases. If we assume (arbitrarily) that the edge occurs at the position of the half-intensity response, we would introduce a slight error in the coherent case. We now finally consider incoherent imaging in terms of spatial frequencies. Following section 2.7 we introduce the object intensity spectrum \( \tau(m, n) \), such that

\[
|\tau(x, y)| = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \tau(m, n) \exp(-2\pi j(mx + ny)) \, dm \, dn
\]
The convolution of two circles is given by the area in common when the centre of one is displaced. For line structures the transfer function may be evaluated analytically to give

\[
C(m) = \frac{2}{\pi} \left[ \cos^{-1} \frac{m}{2} - \frac{m}{2} \left(1 - \left(\frac{m}{2}\right)^2\right)^{1/2} \right].
\]

We may also consider the images of weak objects, as in section 2.7. The remarks made there apply equally here, but with the proviso that \(C(m, n)\) is now always real, even in the presence of aberrations, and thus only the real part of \(b\) is ever imaged.

![Graph of the incoherent transfer function.](image)

It would not be correct to give the impression that an incoherent system is always to be preferred. For example, we have just seen above that no phase information will ever be imaged. We should also be rather careful in comparing the cut-offs of the two transfer functions, as they are not strictly comparable: one is concerned with amplitudes and the other with intensities. However it is important to understand the differences between these two imaging modes. They can be regarded as two limits in microscope imaging which, as we shall see in the next chapter, is in general neither strictly coherent nor incoherent, but rather partially coherent.

References


Chapter 3

Image Formation in Scanning Microscopes

We have already demonstrated, by physical arguments, the equivalence of the conventional microscope and the Type 1 scanning microscope and suggested the arrangement of the Type 2 or confocal scanning microscope which was predicted to have far superior imaging properties to conventional microscopes.

We now put these assertions on a more rigorous basis by using the Fourier imaging approach introduced in the previous chapter to compare the various imaging configurations.

3.1 Imaging with the STEM configuration

We begin our discussion of imaging in practical microscope systems by considering the arrangement of Fig. 3.1. Here we have one lens of pupil function \(P_1(\xi_1, \eta_1)\) which focuses light onto the scanning object of amplitude transmittance \((x_0, y_0)\); the transmitted radiation is then collected by a large area detector which has an amplitude detection sensitivity of \(P_2(\xi_2, \eta_2)\). This is the same as the electron optical layout employed in the scanning transmission electron microscope and we should remember that although we are primarily concerned with light the remarks apply equally to electrons provided we assume appropriate functions for \(P_1, P_2\) etc.

We can write the field just after passing through the object as

\[
U_1(x_0, y_0; x_t, y_t) = h_1(x_0, y_0)(x_0 - x_t, y_0 - y_t),
\]

where \((x_t, y_t)\) represents the scan position and \(h_1\) is the amplitude point function.
interference term depends on \( g_2(2x_0, 0) \), that is on the size of \( P_2 \). If we consider our two limiting cases we have for a vanishingly small detector

\[
I(x, 0) = \left| h_1(x + x_0, 0) + h_1(x - x_0, 0) \right|^2,
\]

(3.12)

where the amplitude images add together and imaging is therefore coherent. Conversely for a large area detector

\[
I(x, 0) = \left| h_1(x + x_0, 0) \right|^2 + \left| h_1(x - x_0, 0) \right|^2,
\]

(3.13)

and the intensity images add together, as one would expect for incoherent imaging. Generalising, we may thus say that altering the size of the detector alters the coherence of the system.

The question as to how close together the two points may come before they are said to be no longer resolved is not easy to answer as various criteria have been proposed giving different results. The two most widely used criteria are the Sparrow criterion which is concerned with the rate of change of the slope of the image at the midpoint and the Rayleigh criterion which somewhat arbitrarily states that the two points will be just resolved when the intensity at the midpoint is 0.735 times that at the points. The Rayleigh criterion was introduced for incoherent imaging with a circular aberration-free pupil, in which it corresponds to the condition that the first zero of the image of one point coincides with the position of the central peak of the image of the second point object. We will discuss two-point imaging a little further by supposing that we have two circular aberration-free pupils such that

\[
P_1(r) = 1, \quad r < a_1,
\]

(3.14)

\[
P_2(r) = 1, \quad r < a_2.
\]

(3.15)

We further introduce a parameter \( s \), defined by

\[
s = \frac{a_2}{a_1}.
\]

(3.16)

The value \( s = 0 \) corresponds to coherent imaging and \( s \to \infty \) to incoherent imaging. Using these definitions we can immediately write from equation (3.6)

\[
g_2(r) = \frac{2J_1(sr)}{sr},
\]

(3.17)

where \( v \) is the normalised optical coordinate of equation (2.53). Equation (3.11) may now be rewritten as

\[
I(v, 0) = \left( \frac{2J_1(v + v_d)}{v + v_d} \right)^2 + \left( \frac{2J_1(v - v_d)}{v - v_d} \right)^2
\]

\[
+ 2 \left( \frac{2J_1(2sv_d)}{2sv_d} \right) \left( \frac{2J_1(v + v_d)}{v + v_d} \right) \left( \frac{2J_1(v - v_d)}{v - v_d} \right).
\]

(3.18)

Thus we see that the image depends on the size of the detector relative to that of the objective rather than the absolute size of the detector. The parameter \( s \) is called the coherence parameter. It is interesting to note that whenever \( 2sv_d \) is a root of \( J_1(2sv_d) = 0 \) the product term is absent and the image is the same as would have been obtained if the object had been incoherently illuminated.

In particular for equal pupils \( (s = 1) \) this will be the case when \( 2v_d \) is a non-zero root of \( J_1(2v_d) = 0 \) which means practically that the geometrical images of the pinholes are separated by a distance equal to the radius of any dark ring of the Airy pattern of the objective. Thus if the two points are separated by a distance such that the Rayleigh criterion is satisfied for an incoherent system, the Rayleigh criterion is also satisfied for a system with equal objective and detector pupils. So the two-point resolution as given by the Rayleigh criterion is identical in these two systems.

We can discuss the effect of the coherence parameter \( s \) on the two-point resolution by introducing a function

\[
L(s) = 2v_d,
\]

(3.19)

which is the distance in optical coordinates between two point objects such that the Rayleigh criterion

\[
\frac{I(0, 0)}{I(v_d, 0)} = 0.735
\]

(3.20)

is satisfied.

This function is plotted in Fig. 3.2 and we can see that the separation for the points to be just resolved for equal pupils or very large detector is 0.61 optical units. The best resolving power is obtained with \( s \sim 1.5 \). We have included in Fig. 3.2 the curve for a microscope employing a full circular

![FIG. 3.2. Two point resolution in a Type 1 scanning microscope.](image-url)
objective lens and an infinitely narrow annular detector. For the particular case of a two-point object the limiting resolution is improved by employing such a collector [3.2].

**3.2 The partially coherent Type 1 scanning microscope**

Although the arrangement of Fig. 3.1 is employed in the scanning transmission electron microscope it is usual in scanning optical microscopy to employ a second, collector lens to gather the radiation which has passed through the object and focus it onto the detector. Figure 3.3 shows two possible configurations in which the detector collects all the light which is incident on \( P_2 \) and so they both have exactly the same imaging properties as each other and as the STEM configuration. In Chapter 1 we discussed how the Type 1 scanning microscope is equivalent to the conventional microscope. In the scanning microscope of Fig. 3.3(a) the radiation is focused onto the detector, and as it is analogous to the critical illumination system in conventional microscopy it may be termed critical detection. In Fig. 3.3(b) on the other hand the detector is placed in the back focal plane of the collector lens and we may call this Köhler detection, again by analogy with Köhler illumination in conventional microscopy. The Köhler system relies on the response of the detector being uniform across the whole area and so the preferred approach is the critical detection arrangement of Fig. 3.3(a).

This is in contrast to conventional microscopy where, as a complete object field must be imaged, Köhler illumination is preferred to give uniform illumination over the field.

The imaging of the Type 1 scanning microscope is still described by equations (3.5) and (3.6) but now \( P_2(\xi_2, \eta_2) \) represents the pupil function of the collector lens if we assume that the detector has a uniform response. We now proceed to discuss the imaging in terms of spatial frequencies. This concept was introduced in section (2.7) where we represented a non-periodic object in terms of its Fourier transform or spectrum. Thus we can write (equation 2.72)

\[
I(x, y) = \iint T(m, n) \exp \{-2\pi j(mx + ny)\} \, dm \, dn \tag{3.21}
\]

and for the complex conjugate

\[
t^*(x, y) = \iint T^*(p, q) \exp \{2\pi j(px + qy)\} \, dp \, dq \tag{3.22}
\]

where \( m, p \) are spatial frequencies in the \( x \) direction and \( n, q \) are similarly spatial frequencies in the \( y \) direction. We have to introduce the spatial frequencies \( p, q \), which are dummy variables which disappear upon integration of (3.22), in order to be able to write the product of (3.21) and (3.22) with the integral signs at the front.

Substituting equations (3.21) and (3.22) into (3.5) we obtain

\[
I(x, y) = \iiint \int \int h_1(x_0, y_0) h_1^*(x_0, y_0) T(m, n) T^*(p, q) \times g_2(x_0 - x, y_0 - y) \times \exp \{-2\pi j[(m - p)x - (n - q)y - p(x_0 - x) + n(y_0 - y)]\} \times dm \, dn \, dp \, dq \, dx_0 \, dy_0 \, dx' \, dy'. \tag{3.23}
\]

This may be written as

\[
I(x, y) = \iiint \int \int C(m, n; p, q) T(m, n) T^*(p, q) \times \exp \{2\pi j[(m - p)x + (n - q)y]\} \, dm \, dn \, dp \, dq. \tag{3.24}
\]
with

$$C(m, n; p, q) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h_1(x_0, y_0) h_1^*(x_0, y_0) g_2(x_0 - x_0, y_0 - y_0) \times \exp \left\{ -2\pi j \left[ m x_0 - p x_0' + n y_0 - q y_0' \right] \right\} dx_0 dy_0 dx_0' dy_0'$$

which using equations (3.6) and (3.2) may be recast as

$$C(m, n; p, q) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |P_2(\xi_2, \eta_2)|^2 P_1(\xi_2 + m \lambda d, \eta_2 + n \lambda d) \times P_1^*(\xi_2 + p \lambda d, \eta_2 + q \lambda d) d\xi_2 d\eta_2.$$  

(3.25)

This represents a very important result as we have been able to express the intensity variation in the image of an arbitrary specimen by equation (3.24) in which $C(m, n; p, q)$, the partially coherent transfer function (sometimes also called the transmission cross coefficient), is a function only of the optical system and not the object. This is the real power of this approach whereby we can introduce an imaging function which is common to all objects. We can see from equation (3.24) that "perfect" imaging is obtained if the transfer function $C(m, n; p, q)$ is always unity; this is not possible in practice and the aim in microscope design is to make this function as smooth and great in extent as possible. It should be noted, however, that a "perfect" image does now show up phase variations, so that "perfect" imaging may not even be desirable in practice.

In order to fix our ideas concerning the transfer function method let us now consider a line structure which has detail in the $x$-direction only. The transfer function may now be contracted to

$$C(m; p) = C(m, 0; p, 0)$$

(3.27)

and $C(m; p)$ is the transfer function which gives the magnitude of the spatial frequency component $(m - p)$ in the intensity image.

We further consider the case where the microscope has aberration-free circular pupils of the form of equations (3.14) and (3.15). We may graphically calculate the transfer function as the area of overlap of the two circles representing $P_1$ centred on $(-m \lambda d, 0)$ and $(-p \lambda d, 0)$, which also falls within the circle $P_2$. This is illustrated in Fig. 3.4. There are two limiting cases of interest as the diameter of $P_2$ is varied. When $P_2$ is large we see that $C(m; p)$ becomes a function of $(m - p)$ only. This is in fact what one expects for incoherent imaging, and imaging is indeed incoherent for this limiting case as discussed in section (2.9). On the other hand as $P_2$ becomes vanishingly small we find that

$$C(m; p) = P_1(m \lambda d) P_1^*(p \lambda d)$$

(3.29)

$$= c(m) c^*(p),$$

(3.30)

and thus the image may be written as

$$I(x_i) = \int_{-\infty}^{+\infty} c(m) T(m) \exp \left\{ 2\pi j m x_i \right\} dm,$$

(3.31)

which of course corresponds to the coherent imaging of section (2.7). Again the positive exponent corresponds to an inverted image.

A practically important case is where the two pupils are of equal size. Under these circumstances the imaging is partially coherent and the $C(m; p)$ function somewhat more complicated. It is shown in Fig. 3.5(b) in $(m; p)$ space where it is seen to exhibit an "hexagonal cut-off. It should be remembered that although $m$ and $p$ are plotted here in orthogonal directions they represent two spatial frequencies in the same direction. The symmetry of
the surface is also shown, $A(\tilde{v})$ being the radial variation of the convolution of two circles which may be written

$$A(\tilde{v}) = \frac{2}{\pi} \left[ \cos^{-1} \left( \frac{\tilde{v}}{2} \right) - \frac{\tilde{v}}{2} \right] \left\{ 1 - \left( \frac{\tilde{v}}{2} \right)^{1/2} \right\}, \quad |\tilde{v}| < 2. \quad (3.32)$$

Here $\tilde{v}$ is the normalised spatial frequency given by $v\pi/a$, where $a$ is the radius of the pupil, so that the cut off is at $\tilde{v} = 2$.

We also show, for comparison, the transfer functions for the coherent and incoherent microscopes in Fig. 3.5. It is clear that Fig. 3.5(b) represents a transition between the two extremes. The $C(m; p)$ surfaces are shown in Fig. 3.6.

The Type 1 scanning microscope behaves in an identical way to the analogous conventional microscope. Imaging in partially coherent conventional microscopes was analysed by Hopkins [3.3] using the theory of partial coherence. The spread function $g_2$ of equation (3.18) may be recognised as nothing other than the degree of coherence [3.4] which may be derived from the van Cittert-Zernike theorem [3.5, p. 510].

3.3 The confocal scanning microscope

3.3.1 Introduction

As explained in section 1, the Type 1 scanning microscope has imaging properties identical to those of conventional non-scanning microscopes. The Type 2, or confocal scanning microscope, on the other hand has completely different imaging properties. The confocal microscope is formed by placing a point detector in the detector plane of Fig. 3.3(a). We can write the field in the detector plane $(x_2, y_2)$ as the convolution of the amplitude in the object plane with the point spread function of the collector lens

$$U(x_2, y_2; x_s, y_s) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h_1(x_0, y_0) h_2 \left( \frac{x_2}{M} - x_0, \frac{y_2}{M} - y_0 \right) dx_0 dy_0.$$

However if we employ a point detector at $x_2 = y_2 = 0$ the detected intensity is

$$I(x_i, y_i) = \left| \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h_1(x_0, y_0) h_2(-x_0, -y_0) dx_0 dy_0 \right|^2 \quad (3.34)$$

which may be written for even spread functions

$$I(x_i, y_i) = |h_1 h_2 \otimes t|^2, \quad (3.35)$$

that is the microscope behaves as a coherent microscope with an effective point spread function given by the product of those for the two lenses.

We have thus found that the combination of point detector and scanning converts a convolution of the form $(h_1 \otimes h_2) \otimes t$ (which is obtained for a Type 1 scanning microscope (equation 3.16)) into one of the form $(h_1 h_2) \otimes t$.

Another way of considering the confocal microscope is to calculate first the amplitude $U$ in terms of $U_2$,

$$U(x_2, y_2) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} U_2(\xi_2, \eta_2) \exp \frac{jk}{Md} (\xi_2 x_2 + \eta_2 y_2) d\xi_2 d\eta_2 \quad (3.36)$$

so that if $x_2 = y_2 = 0$

$$I(0, 0) = \left| \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} U_2(\xi_2, \eta_2) d\xi_2 d\eta_2 \right|^2 \quad (3.37)$$
that is the effect of the point detector is to integrate the amplitude over the pupil \( P_2 \). This compares with the Type 1 scanning microscope where the detector integrates the intensity over the pupil \( P_2 \). Substituting for \( U_3 \) from (3.3) we obtain for the confocal case

\[
I(x, y) = \left| \iiint h_1(x_0, y_0)k(x_0 - x, y_0 - y)P_2(\xi_2, \eta_2) \right|^2 \exp \left( \frac{jk}{\delta} (x_0 \xi_2 + y_0 \eta_2) \right) dx_0 dy_0 d\xi_2 d\eta_2 \]

and using (3.2) to introduce \( h_2 \) we reproduce (3.34).

The point detector integrates amplitude over the pupil \( P_2 \); it therefore has the same effect as the amplitude-sensitive detector used in acoustic microscopy [3.6]. The scanning acoustic microscope is a confocal microscope and exhibits many of the properties of confocal microscopes.

### 3.3.2 Image formation in confocal microscopes

If the two lenses in a confocal microscope are circular and of equal numerical aperture the image of a point object is (from 3.35)

\[
I(v) = \left( \frac{2J_1(v)}{v} \right)^4, \quad (3.39)
\]

which is shown in Fig. 3.7, the central peak being sharpened up by 27% relative to the image in a conventional microscope (at half the peak intensity). The sidelobes are also drastically reduced, so there is thus a marked reduction in the presence of artefacts in confocal images.

If we calculate the image of two closely spaced point objects we find that when the Rayleigh criterion is satisfied the points are separated by a normalised distance \( 2d_0 = 0.56 \). This is 32% closer than in a conventional coherent microscope and 8% closer than in a conventional microscope with equal lens pupils. The relative values are illustrated in Fig. 3.2. The fact that the sidelobes are weaker in confocal microscopy suggests that it should be possible to use an annular lens in a confocal microscope. With one circular and one annular lens of equal radii the intensity is given by

\[
I(v) = \left( \frac{2J_0(v)J_1(v)}{v} \right)^2, \quad (3.40)
\]

so that the central peak is now even narrower (40% narrower compared with a conventional microscope) and the sidelobes are extremely weak as the maxima of the spread function of the annulus coincide with the zeros of that of the circular lens. The two-point resolution is now 28% better than a conventional microscope with equal lens pupils.

For large values of \( v \) the intensity in the image of a single point in a conventional microscope falls off as \( v^{-3} \). In a confocal microscope with two circular lenses it falls off as \( v^{-4} \), whereas with one annular lens it falls off as \( v^{-4} \). With two narrow annuli however it falls off as \( v^{-2} \), that is the power in successive sidelobes only falls off as \( v^{-1} \) and total normalised power does not converge, so that such an arrangement is clearly unusable.

![Fig. 3.7. The image of a single point object.](image)

We may now turn our attention to the Fourier imaging and substituting equation (3.21) into (3.34) are able to write

\[
I(x, y) = \left| \iiint c(m, n)T(m, n) \exp 2\pi i(mx + ny) dm dn \right|^2 \quad (3.41)
\]

with

\[
c(m, n) = P_1(m\delta, n\delta) \otimes P_2(m\delta, n\delta), \quad (3.42)
\]

where \( c(m, n) \) is a coherent transfer function. For two circular pupils of equal radii the coherent transfer function is identical to the incoherent transfer function for an incoherent system (equation 2.106).
If we again restrict ourselves to considering the images of line structures the function of interest is $C(m; p)$ given by

$$C(m; p) = c(m)c^*(p)$$

(3.43)

for the confocal microscope.

Figures 3.8(a) and 3.9(a) show the form of this transfer function for a confocal microscope with equal pupils. For a given pair of spatial frequency moduli the response is higher if they have the same sign (difference frequencies) than if of opposite sign (sum frequencies). The $m-p$ axis is shown in Figs 3.8(a) and 3.8(b): the greater the distance along the axis the higher the spatial frequency in the image. For the confocal microscope the response for the sum frequencies is improved, but for the difference frequencies is reduced as compared to the conventional microscope, Fig. 3.5(b). This accounts for the fact that the imaging in confocal microscopy is generally improved even though the coherent transfer function in the confocal microscope is identical to the incoherent transfer function for a conventional incoherent microscope. This coherent transfer function is compared with that for a conventional coherent microscopy in Fig. 3.10, the cut-off frequency being twice as great. The transfer function also falls off gradually and thus we do not expect excessive fringing to be present in the image of a straight edge. Also shown is the transfer function for a confocal microscope with one annular pupil, illustrating that the response for higher spatial frequencies is improved. This transfer function is given by the radial variation of the convolution of a circle with an annulus, given by

$$A(\tilde{v}) = \frac{2}{\pi} \cos^{-1}\left(\frac{\tilde{v}}{2}\right), \quad \tilde{v} < 2.$$ (3.44)

This should be compared with equation (3.32) for the convolution of two circles.

The confocal microscope with one annular pupil may be compared with that with two circular pupils by studying the region of $m, p$ space within

FIG. 3.8. Contours of constant $C(m; p)$ showing lines of normalised spatial frequency $(m - p)$ for (a) circular lenses and (b) one annular and one circular lens in confocal microscopes.

FIG. 3.9. The $C(m; p)$ surface for a confocal scanning microscope with (a) circular lenses and (b) with one annular lens and one circular lens.

FIG. 3.10. The coherent transfer function for various microscope geometries.
which the transfer function is greater than one half, as illustrated by shading in Fig. 3.8.

If we now consider the imaging of a weak object of the form of equation (2.84), that is
\[ t(x) = \exp \left( b \cos 2\pi vx \right) \] (3.45)
with \( b \) small so that terms in \( b^2 \) may be neglected, we find that by substituting in equation (3.24) the image is given by
\[ I(x) = C(0; 0) + 2 \Re \{ bC(v; 0) \} \cos 2\pi vx, \] (3.46)
that is it depends only on \( C(v; 0) \). Imaging of weak objects in conventional and confocal microscopes with circular pupils is thus identical. However, if aberrations are present this is no longer the case and the confocal microscope may behave very differently.

### 3.4 Aberrations in scanning microscopes

#### 3.4.1 Introduction

Until now we have discussed the imaging properties of various microscopes assuming perfect lenses. In practice the lenses will not be perfect and it is important to know how seriously lens aberrations or defocus affect the optical performance of the microscope. The most common practical arrangement of scanning optical microscope that has been constructed involves mechanically scanning the object across a stationary spot. This has the advantage that the optical system is axial and so the lenses need only strictly be corrected for axial aberrations, although in practice high-quality microscope objectives should be used if specially corrected lenses are not available.

In this section we therefore restrict our attention to the axial aberrations and consider the effects of defocus and primary spherical aberration. Chromatic aberration is not considered specifically as usually only a single laser wavelength is used. These effects are studied not only to examine the degradation of the imaging but also because in certain circumstances they serve to introduce an imaginary part to the transfer function and advantage could be taken of this to crudely image phase detail. We can see from equations (3.45) and (3.46) that if \( b \) is complex the intensity may be written as
\[ I(x) = 1 + 2(b, C_r - b, C_i) \cos 2\pi vx, \] (3.47)
where we have taken \( C(0; 0) \) to be unity and set
\[ b = b_r + jb_i \] (3.48)
and
\[ C(v; 0) = C_r + jC_i. \] (3.49)

#### 3.4.2 Defocus in scanning microscopes

We begin by considering the conventional or scanning microscope of Type 1; we see from equation (3.28) that for equal sized pupils the function \( C(m; 0) \) is wholly real. Furthermore, as it is given by the convolution of a function with its complex conjugate, the aberrations of the collector lens \( P_2 \) are unimportant, this being equivalent to the well-known result in conventional microscopy that the aberrations of the condenser do not affect the imaging.

We will restrict ourselves to one-dimensional pupils for ease of analysis and introduce aberrated pupil functions, following equations (2.35), (2.37),
\[ P_1(x) = \exp \left\{ \frac{1}{2} \mu_1 \left( \frac{x}{a_1} \right)^2 \right\}, \quad |x| < a_1, \] (3.50)
\[ P_2(x) = \exp \left\{ \frac{1}{2} \mu_2 \left( \frac{x}{a_2} \right)^2 \right\}, \quad |x| < a_2, \]
where
\[ \mu_1 = 4k \delta s_1 \sin^2 \left( a_1/2 \right), \]
\[ \mu_2 = 4k \delta s_2 \sin^3 \left( a_2/2 \right). \] (3.51)

The wholly real transfer function for the conventional microscope is shown in Fig. 3.11 for varying degrees of defocus. The similarity of this result with

![Fig. 3.11. The transfer function $C(m; 0)$ for a conventional microscope with varying degrees of defocus.](image-url)
that obtained by Hopkins [3.7] for full circular pupils justifies our one-
dimensional model. An imaginary part may be introduced into the transfer
function by stopping down the collector lens. It is found, however [3.8], that
this lens must be stopped down considerably before phase imaging becomes
appreciable which, with the associated reduction in spatial frequency cut-off,
is the major reason why this method of obtaining phase contrast is not widely
used.

The properties of the confocal microscope, however, are very different. We
recall that (equations 3.42, 3.43)

$$C(m; 0) = \{P_1(\tilde{m}\alpha) \otimes P_2(\tilde{m}\alpha)\} \{P^\dagger(0) \otimes P^\dagger(0)\},$$

(3.52)

$\tilde{m}$ again representing the normalised spatial frequency.

If the defocus of the pupils is equal and opposite, such that $P_1 = P^\dagger_2$ the
transfer function for weak objects is always real. Furthermore it is of exactly
the same form as for the conventional microscope in Fig. 3.11. In a
transmission instrument this corresponds to the case when the object is moved
along the axis relative to the stationary lenses. For an object of uniform
optical thickness, therefore, if the lenses are correctly spaced, no phase
imaging will result from any depth within the object. Conversely if the pupils
are different, for example one circular and one annular, then phase imaging
will result from the defocused parts.

We now consider the case of equal defocus, $P_1 = P_2$, which corresponds to
a displacement of both lenses such that the object remains midway between
them. This is always the case in a reflection microscope and is the result of
moving either the lens or the object: an imaginary part is introduced for the
defocused parts of an object with variations in height. Equal defocus is also
introduced if a transmission object has varying optical thickness which alters

**Fig. 3.12.** The real and imaginary parts of the coherent transfer function of a
transmission mode confocal microscope with lens defocus, or a reflection mode
confocal microscope.

**Fig. 3.13.** The real and imaginary parts of the coherent transfer function of a
confocal microscope with one annular pupil with varying degrees of defocus.
the effective spacing of the lenses. The transfer function is shown in Fig. 3.12 and we see that the imaginary part experiences a sign change which effectively rules it out as a practical method of phase contrast imaging. Care should therefore be taken when focusing by maximising the contrast if the object has comparatively strong phase variations. The strength of the transfer function at zero and low spatial frequencies is reduced with defocus by the spreading of the radiation in the detector plane. This property is associated with depth discrimination, which is discussed in section 3.7.

If we use an infinitely thin annular lens in a confocal microscope defocus of either object or the lens has the same result due to the large depth of focus of an annular lens. The transfer function is again complex (Fig. 3.13), a very small wavelength aberration ($\sim \lambda/2\pi$) being necessary to produce a substantial imaginary part to the transfer function. This is a major disadvantage of using an annular lens in a confocal microscope as parts of a thick object away from the focal plane are imaged with poor fidelity.

**3.4.3 Defocus and primary spherical aberration in scanning microscopy**

We now consider pupil function of the form

$$P(\xi) = \exp j\phi(\xi), \quad |\xi| < 1,$$

(3.53)

where the wave distortion function is given by

$$\phi(\xi) = A[(\xi/a)^n + B(\xi/a)^2].$$

(3.54)

The coefficient $A$ measures the total amount of fourth-power error on the wavefronts in the clear aperture, whereas the coefficient $B$ specifies the focal setting, $B = 0$ corresponding to the paraxial setting, $B = -2$ to the marginal one while $B = -1$ indicates a "mid-focus" setting. For this last condition the wave deviations vanish at the centre and edge of the clear aperture (Fig. 3.14).

**FIG. 3.14.** The wave distortion function in the presence of spherical aberration and defocus.

**FIG. 3.15.** The transfer function $C(m; \theta)$ for a Type 1 scanning microscope with equal pupils for $A = 1$.

**FIG. 3.16.** The transfer function $C(m; \theta)$ for a Type 1 scanning microscope with equal pupils for $A = 2$. 
It should be noted that, unlike the case of pure defocus, the system is not symmetrical in focal setting and that by making $B$ negative the effects of the spherical aberration may be to a certain extent be alleviated. We now consider the effect of this pupil function on the performance of various microscope types.

For the conventional microscope, $C(m; 0)$ is wholly real and independent of the aberrations of the second lens. This has been evaluated numerically and is plotted for various focal settings with degrees of spherical aberration corresponding to $A = 1$ and 2 in Figs 3.15 and 3.16 again for a one-dimensional model. The effect of the spherical aberration is drastically to reduce the mid- to higher spatial frequency components, but the effect is almost focused out at $B = -1$, the mid-focus setting. Positive values of $B$, on the other hand, reduce the performance and with higher degrees of spherical aberration and defocus phase reversal occurs for some spatial frequencies.

**Note:**

FIG. 3.17. The real and imaginary parts of the transfer function $C(m; 0)$ for a confocal scanning microscope with equal lenses, $A = 1$. Defoci of both lenses are equal.

FIG. 3.18. The real and imaginary parts of the transfer function $C(m; 0)$ for a confocal scanning optical microscope with equal lenses, $A = 2$. Defoci of both lenses are equal.
We now turn to confocal scanning microscopes [3.9]. Here the aberrations of both lenses are important and the transfer function is in general complex. It is reasonable to assume that if two equal lenses are used that they would both suffer from similar degrees of spherical aberration and we have assumed that the coefficients $A$ for the lenses are equal. We have plotted the transfer function for two special cases. The first is when the lenses have equal degrees of defocus (corresponding to a change in the separation of the lenses). It is seen (Figs 3.17 and 3.18) that again the mid-focus setting almost cancels out the effect of spherical aberration.

The second special case is that of equal and opposite defocus of the two lenses (corresponding to a movement of the specimen relative to the fixed lenses). The curves (Figs 3.19 and 3.20) indicate that defocus does not improve the high spatial frequency response and does not decrease the magnitude of the imaginary part. The curves for positive and negative defocus are identical because of the commutative property of the convolution operation.

The results of using an annular objective in a confocal microscope [3.10] are shown in Figs 3.21 and 3.22. These are equally applicable to conventional microscopes with an annular condenser. There is again no "aberrated" contribution from the annulus and so the curves apply equally to reflection microscopy. We can see that for $A = 1$ and 2 the mid-focus setting, $B = -1$ has almost cancelled out the effect of spherical aberration. However at higher values of $A$ it is more difficult to focus out these effects and so a confused image consisting of both amplitude and phase information would result.

3.4.4 Discussion

In the conventional or Type 1 scanning microscope with two equal pupils the transfer function is purely real and although spherical aberration results in a degradation of the spatial frequency response the effect may be reduced by appropriate defocusing.
In a confocal microscope, there is an imaginary part introduced. This may be reduced by adjustment of the lens spacing but not by movement of the object relative to the lenses. If the object transmittance exhibits both phase and amplitude variations interpretation of the micrographs may prove difficult. Thus, in general, for a confocal microscope to operate properly it is important that the lenses should be well corrected for the laser wavelength used, but, on the other hand, if the object is mechanically scanned off-axis aberrations are of course, unimportant. If the object is such that it exhibits only very small amplitude variations in transmittance the spherical aberration could result in useful phase imaging without this being associated with the detrimental reduction in spatial frequency bandwidth introduced by stopping down the collector lens.

The scanning acoustic microscope has imaging properties similar to those of the confocal optical microscope and the imaging is characterised by a coherent transfer function identical to the \( C(m; 0) \) function of the confocal microscope. A typical figure for the maximum path error in an acoustic microscope might be \( \lambda/4 \). A value of \( A \) of unity corresponds to a maximum path error of \( \lambda/6.3 \) and thus the spherical aberration in the acoustic microscope could well give rise to phase imaging.

We should note that although in some cases the effect of aberrations on the real part may be small care should be taken in the interpretation of micrographs especially if the phase delay in the object is appreciable.

Fig. 3.21. Transfer function \( C(m; 0) \) for a conventional microscope with annular condenser or confocal microscope with one annular lens in the presence of spherical aberration, \( A = 1 \).

Fig. 3.22. Transfer function \( C(m; 0) \) for a conventional microscope with annular condenser or confocal microscope with one annular lens in the presence of spherical aberration, \( A = 2 \).
3.5 The image of a straight edge

In sections 2.6 and 2.9 we calculated the image of a straight edge in coherent or incoherent imaging systems by finding the response to a δ-line and integrating. This method is not the most convenient for partially coherent systems. Instead we use the object spectrum, and in order that the integrals reduce to summations choose to consider the response to a bar pattern object and arrange for its spatial frequency to be sufficiently small that neighbouring edges do not affect each other [3.11]. We consider an object of the form (Fig. 3.23)

\[ t(x) = M - \frac{2A}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^n}{2n-1} \cos (2n-1)2\pi vx, \]  

(3.55)

which when Fourier transformed and substituted in equation (3.24) gives

\[ I(x) = M^2 - \frac{4AM}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^n}{2n-1} C\{(2n-1)v; 0\} \cos (2n-1)\theta \]

\[ + \frac{4A^2}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^n}{2n-1} \left[ C\{(2n-1)v, (2r-1)v\} \cos \{(2n-1)\theta - (2r-1)\theta\} \right. \]

\[ \left. + C\{(2n-1)v, -(2r-1)v\} \cos \{(2n-1)\theta + (2r-1)\theta\} \right], \]

(3.56)

where

\[ \theta = 2\pi vx, \]

(3.57)

and we have used the identities

\[ C(m; p) = C(-m; -p), \]

(3.58)

\[ C(m; 0) = C(0; m). \]

(3.59)

For the conventional microscope, the transfer function of which is shown in Fig. 3.24, we may write

\[ I(x) = M^2 - \frac{4AM}{\pi} S_1 + \frac{2A^2}{\pi^2} S_2 + \frac{2A^2}{\pi^2} S_3 + \frac{4A^2}{\pi^2} S_4 + \frac{4A^2}{\pi^2} S_5, \]  

(3.60)

where the Ss represent sums over portions of the \( C(m; p) \) surface as shown in Fig. 3.2 and are given by

\[ S_1 = \sum_{n=1}^{\infty} \frac{(-1)^n}{(2n-1)} C\{(2n-1)v\} \cos (2n-1)\theta, \]

\[ S_2 = \sum_{n=1}^{\infty} \frac{C\{(2n-1)v\}}{(2n-1)^2}, \]

\[ S_3 = \sum_{n=1}^{\infty} \frac{C\{(2n-1)v\}}{(2n-1)^2} \cos \{2(2n-1)\theta\}, \]

\[ S_4 = \sum_{n=1}^{\infty} \sum_{r=n+1}^{\infty} \frac{(-1)^{r-1} C\{(2r-1)v\}}{(2r-1)(2n-1)} \cos \{(2n-1)\theta - (2r-1)\theta\}, \]

\[ S_5 = \sum_{n=1}^{\infty} \sum_{r=n+1}^{\infty} \frac{(-1)^{r-1} C\{(2r-1)v, (2r-1)v\}}{(2r-1)(2n-1)} \cos \{(2n-1)\theta + (2r-1)\theta\}, \]

(3.61)

where \( C(\cdot) \) is either the convolution of two circular pupils for which equation (3.32) holds, or the convolution of a circle with a thin annulus (equation 3.44). The straight edge responses for conventional microscopes with two circular pupils and with an annular condenser and circular objective are

---

**Fig. 3.23.** The periodic object used in the calculation of the image of a straight edge object.

**Fig. 3.24.** The area of summation in \((m, p)\) space.
shown in Fig. 3.25 where the normalised distance \( v \) is given by equation (2.28). It should be noted that the image in the microscope with an annular condenser is relatively poor. Turning now to the confocal microscope we obtain the much simpler result

\[
I(x) = \left[ M - \frac{2A}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^n C_n (2n - 1)v}{(2n - 1)} \cos \left( (2n - 1)\theta \right) \right]^2. \quad (3.62)
\]

The relevant responses are plotted in Fig. 3.25 and we see that the confocal microscope with two circular lenses gives a better image than a conventional microscope but that a confocal microscope with one annular and one full circular lens gives the best response of all.

In a confocal microscope the use of two equal pupils results in an amplitude point spread function which is never negative and hence the straight edge response under these conditions does not exhibit fringing even when the object is defocused. There are thus advantages in using two equal pupils in a confocal microscope and we conclude this section by examining this case.

The impulse response for a confocal microscope with circular lenses has very weak side lobes. We know that the use of annular lenses results in a sharper central peak and also an increase in the strength of the side lobes. It is interesting to examine which of these two effects is dominant on the straight edge response for small central obscurations.

In a confocal microscope the use of two equal pupils results in a much simpler result. The impulse response for a confocal microscope with circular lenses has very weak side lobes. We know that the use of annular lenses results in a sharper central peak and also an increase in the strength of the side lobes. It is interesting to examine which of these two effects is dominant on the straight edge response for small central obscurations.

Figure 3.26 shows the results for various values of \( \gamma \), the ratio of the inner to the outer radii of the annuli [3.10]. Unfortunately, we can see that any obscuration of the aperture degrades the performance. However the increase in depth of focus may warrant the use of a small obstruction. As the side lobes with circular lenses are so small it still seems likely that there is some apodisation which would result in an improved straight edge response.

### 3.6 The image of a phase edge

We have just discussed the image of a strong amplitude object and so we now move on to discuss the image of a strong phase object such as a phase edge.
where the phase change is abrupt and not small. The image of such an object is of considerable importance in cell sizing and counting in biology and linewidth measurement in integrated circuit technology, where the edge of the cell or the metallisation may be thought of as the phase step.

We consider an object whose amplitude transmittance is alternatively \( \exp j \phi_1 \) and \( \exp j \phi_2 \), which may be written

\[
\psi(x) = \exp j \phi_1 \left[ \frac{1 + \exp(j \Delta \phi)}{2} - \frac{2}{\pi} (1 - \exp(j \phi)) \sum_{n=1}^{\infty} (-1)^{n} \cos(2n - 1) \phi \right]
\]

where \( \Delta \phi = \phi_2 - \phi_1 \) and \( \phi \) is as defined in equation (3.57).

(3.63)

Following the methods of the previous section we are able to write for the partially coherent conventional microscope [3.12]

\[
I(x) = \left( \frac{1 + \cos \Delta \phi}{2} \right) C(0; 0) - \frac{4 \sin \Delta \phi}{\pi} \text{Im} \{ S_1 \} - \frac{4}{\pi^2} (1 - \cos \Delta \phi) \left[ S_2 + S_3 + 2(S_4 + S_5) \right]
\]

where the \( S_s \) are given by equation (3.61).

FIG. 3.27. The intensity in the image of a phase edge in a conventional scanning microscope.

FIG. 3.28. The intensity in the image of a phase edge in a confocal scanning microscope.
The expressions are again much simpler for the confocal microscope and may be written

\[ I(x) = \left( \frac{1 + \exp j \Delta \phi}{2} \right) C(0) \left[ 2(1 - \exp j \Delta \phi) S_k \right]^2. \]  

(3.65)

The responses are plotted in Fig. 3.27 and 3.28. We see that the phase edge manifests itself as a central dip in intensity which in all cases gives greater contrast for the confocal microscope. The extreme case of a 180° phase step gives a limiting central intensity of 0.333 in the conventional microscope compared with zero in the confocal instrument. The central intensity in the confocal case may be written as

\[ I(0) = \cos^2 \left( \frac{\Delta \phi}{2} \right) \]  

(3.66)

which becomes smaller as \( \Delta \phi \) increases as indicated in Fig. 3.28 reaching zero for \( \Delta \phi = \pi \).

3.7 Depth discrimination in scanning microscopes

A large depth of field is often a desirable property in a microscope, such as when observing rough surfaces. Equally, when observing thick biological specimens in transmission it is often useful to limit the depth of field to avoid confusion in interpretation of the micrographs.

If we observe the image of a single point in a conventional or Type 1 scanning microscope as the object is taken out of the focal plane we find that the image broadens and that the axial intensity decreases. Either of these quantities may be taken as an indicator of depth of field.

The intensity variation near the focus of a lens may be written as

\[ I(u, v) = |h_1(u, v)|^2 \]  

(3.67)

where \( h_1 \) is the impulse response of the lens and \( u, v \) the optical coordinates (equations 2.28 and 2.37), a constant multiplier being neglected. For the case when the lens has a circular pupil, Fig. 2.3 shows contours of equal intensity in the \( (u, v) \) plane. These results also show the intensity variation in the image plane \( (r \) is now proportional to the distance in the image plane) of a point object placed a normalised distance \( u \) from the focal plane of the lens. For a given object position the axial intensity in the image is given by

\[ I(u, 0) = \left( \frac{\sin u/4}{u/4} \right)^2. \]  

(3.68)

In a scanning microscope the intensity in the image depends on the properties of both the objective lens and the collector lens. Assuming that the impulse response of the latter is \( h_2 \), the intensity is given by (equation 3.35)

\[ I(u, v) = |h_1(u, v)h_2(u, v)|^2. \]  

(3.69)

If both lenses are equal the intensity is simply the square of that in the Type 1 microscope. For two circular pupils the contours of Fig. 2.3 are still valid except that the value of the intensity of the contour is squared. The axial intensity varies as

\[ I(u, 0) = \left( \frac{\sin u/4}{u/4} \right)^2, \]  

(3.70)

which shows that if depth of field is defined in terms of the fall-off in maximum intensity for a point image, the depth of field for the confocal microscope is reduced relative to that of a conventional microscope, the difference not being very great.

There are, however, other associated properties which may be of practical importance. For instance we may investigate the variation in the integrated intensity for the image of our point source, which is a measure of the total power in the image [3.13]. This tells us how our microscope discriminates against parts of the object not in the focal plane. We define the integrated intensity as

\[ I_m(u) = 2\pi \int_0^\infty I(u, v) \, dv. \]  

(3.71)

For the Type 1 microscope we know from Parseval’s theorem that this is equal to the integral of the modulus squared of the effective (that is the defocused) pupil function, and since the effect of the defocus is merely to introduce a phase factor which disappears when the modulus is taken we come to the conclusion that the integrated intensity does not fall off in the Type 1 microscope. We may argue this equally from conservation of energy. Since our function \( I(u, v) \) for the Type 1 microscope is the same as the intensity near the focus of a lens the integrated intensity is proportional to the power crossing any plane perpendicular to the optic axis, and this must be constant.

There is thus no discrimination of this kind in the Type 1 or conventional microscopes, but this is not as bad as it seems as the defocused image eventually becomes a constant background which is rejected by the observer, although of course it does reduce contrast.
Turning now to the confocal microscope we have, using the expression for the intensity in the focal region of a single lens (equation 2.40),

\[ I_{\text{int}}(u) = 2\pi \int_0^\infty \left( C^2(u, v) + S^2(u, v) \right) J_0^2(v) \, dv. \]  

(3.72)

This integral is plotted in Fig. 3.29 where we have normalised it to unity in the focal plane. The integrated intensity falls off monotonically, reaching the half power point at a distance from the focal plane of 0.70\(\lambda\) for a numerical aperture of unity. For large values of \(u\) the integrated intensity falls off according to an inverse square law, as may be shown by considering the geometrical optics approximation.

We have already discussed the use of annular lenses in scanning microscopes earlier in this chapter and it is therefore of interest to examine the effect of using one annular lens on the depth of field of the microscope. The variation in the intensity in the image of the single point is now exactly as for a conventional microscope (equation 3.5) as the impulse response of the annular lens does not vary along the axis. It should be noted that we have not said that the variation in the breadth of the point image is as in the conventional microscope: these are difficult to compare as the intensity variations differ in shape. The integrated intensity may again be computed, in this case from

\[ I_{\text{int}}(u) = 2\pi \int_0^\infty \left( C^2(u, v) + S^2(u, v) \right) J_0^2(v) \, dv, \]  

(3.73)

and this is also shown in Fig. 3.29 again normalised to unity at the focal plane. The curve decreases monotonically, but the discrimination is not now so great as for the microscope with circular pupils. The integrated intensity has fallen to one half at a distance of 1.05\(\lambda\) for a numerical aperture of unity.

We have seen that using one annular lens in a confocal microscope reduces the discrimination against objects outside the focal plane. If two annular lenses are used we would expect this discrimination to be further reduced. In the limiting case as the cross-section of the annulus is very small the integrated intensity becomes constant as the impulse response does not then vary along the axis. This is consistent with the claim of increased depth of field in a confocal microscope with two annular lenses [3.14].

3.8 Contrast mechanisms in confocal microscopy

In sections 3.1 to 3.6 we considered imaging of a thin object whose amplitude and phase effect on the transmitted (or reflected) beam is completely characterised by a complex function of position \(t(x, y)\). Let us further consider two particular objects of this type. A linear variation in phase, as would be produced by a wedge of dielectric viewed in transmission or a sloping surface in reflection, has a single spatial frequency in its spectrum so that the image consists of a constant intensity the strength of which is given by \(C(m, m)\). This imaging of phase gradients results from the fact that the refracted (or reflected) beam tends to miss the collector lens. The transfer function \(C(m, m)\) falls off quicker with increasing slope in the confocal microscope than in a conventional microscope with equal pupils so that contrast resulting from variations in this phase gradient are imaged more strongly in the confocal system. The second example is a surface of small cosinusoidal height variations (equation 2.84). If \(b\) is the amplitude of the oscillations there is no image at the spatial frequency of the oscillations in an aberration free system. The image variations at twice this spatial frequency are of strength \(b^2\) and are also proportional to the value of \(C(m; -m) - C(2m; 0)\). For an incoherent conventional microscope this vanishes as might be expected, but it also vanishes for conventional microscopes with equal pupils. For confocal microscopes it does not vanish and hence an image is formed. There are thus some differences in the contrast produced in confocal and conventional microscope images.
However the most important differences result from the depth discrimination properties described in section 3.7. If a reflection object has height variations of sufficient magnitude to result in a change in the spread function then the signal in a confocal microscope will vary accordingly. Imaging of this type is very difficult to analyse because the spread function is not spatially invariant. However if the height is only slowly varying diffraction by the object may be neglected and the image signal in a confocal microscope will result entirely from the depth discrimination. Consider an object consisting of a perfect reflector in a reflection microscope. The Fourier transform of the reflectance is simply $\delta(m)\delta(n)$ and consequently the signal is just equation (3.24) $C(0; 0)$. For a conventional microscope we can see from equation 3.26 that the signal is independent of focus position. But for a confocal microscope we have (equation 3.42)

$$I = \left| \int \int P_1(x, y)P_2(-x, -y) \, dx \, dy \right|^2$$

or for two circular pupils

$$I(u) = \left| \int \exp j\mu u^2 \rho \, d\rho \right|^2 = \left( \frac{\sin u/2}{u/2} \right)^2.$$  

So if the object is displaced from the focal plane in either direction the signal falls. It should be noticed that $C(0; 0)$ is the intercept on the defocused transfer function (Fig. 3.12), but it should be remembered that this figure is for a one-dimensional model.

If the object is mounted with its normal slightly away from the optic axis, some parts will appear out of focus in a conventional or Type 1 image, but in a confocal microscope the image is modulated by equation (3.76) so that only the part of the object in focus is imaged [3.15] as shown in Figure 5.2.

At high numerical apertures only a very small height variation is necessary to produce a substantial depth discrimination effect so that this is an important source of image contrast. For higher spatial frequencies such that diffraction at the object may not be neglected there will be a combination of depth discrimination and diffraction imaging. In particular there will be an interaction caused by the complex defocused transfer function resulting in phase imaging of height variations.

Similar effects occur in transmission microscopy, a slab of dielectric of varying optical thickness in the object plane producing a variation in signal resulting from the change in effective separation of the lenses.

3.9 Scanning microscopes with partially coherent effective source and detector

We should finally briefly mention that all the microscope systems we have discussed so far are merely special cases of a more generalised scanning microscope, that is, one with a finite source and a finite detector. This system has been analysed in great detail by Sheppard and Wilson [3.16] but we will do no more here than discuss a few special cases.

We consider the optical system of Fig. 3.30 and restrict ourselves to a one-dimensional analysis for simplicity, the two-dimensional result being an obvious extension. Using either the concept of mutual coherence [3.4] or the methods of Chapter 2 we may write the image intensity as

$$I(x_0) = \int \int \int S(x_1)h_1\left( x_0 + \frac{x_1}{M} \right)$$

$$\times h^*\left( x_0 + \frac{x_1}{M} \right)I(x_0 - x_1)^*I(x_0 - x_2)^*D(x_2) \, dx_1 \, dx_0 \, dx_0 \, dx_2$$

where $S$ and $D$ are the source and detector intensity sensitivities respectively.
We may now as a simple example consider the image of a point-like object. Then

\[
I(x_i) = \left\{ \int_{-\infty}^{\infty} S(x_1) \left| h_1 \left( x_1 + \frac{x_1}{M} \right) \right|^2 dx_1 \right\} 
\times \left\{ \int_{-\infty}^{\infty} D(x_2) \left| h_2 \left( x_2 + \frac{x_2}{M} \right) \right|^2 dx_2 \right\}
= \{S(Mx_i) \otimes |h_1(x_i)|^2\} \{D(Mx_i) \otimes |h_2(x_i)|^2\}. \tag{3.79}
\]

The image is clearly sharpest when both source and detector are points (the confocal case) and is degraded when either or both have finite size. As both become large the imaging becomes poor.

It is now possible to introduce the Fourier transform of \( \mathcal{F} \) and obtain a general expression for the transfer function \( C(m; p) \). However this is beyond the scope of our present intention; the conclusion of such an analysis being that in general the confocal arrangement possesses superior imaging properties.

### 3.10 The limitations of scalar diffraction theory

We have based our analyses so far on the application of the Kirchhoff diffraction formula in the paraxial limit. This approach necessarily involves certain simplifications and approximations apart from the use of the Kirchhoff boundary conditions. The paraxial condition essentially requires approximately \( \tan \theta \approx \theta \) but this is clearly not appropriate if we wish to apply the results to microscopes involving high numerical aperture objectives. Further we have assumed light to be a scalar phenomenon, i.e. only the scalar amplitude of one transverse component of either the electric or magnetic field has been considered, it being assumed that any other components of interest can be treated independently in a similar fashion. This entirely neglects the fact that the various components are coupled to each other through Maxwell’s equations and cannot strictly be treated independently.

To emphasise this we follow Hopkins [3.17] and consider a convergent spherical wavefront with the electric vector parallel to one principal section (Fig. 3.31). We would expect the electric vectors from \( A \) and \( B \) to combine vectorially at the focus to give a field less than their algebraic “scalar” sum. Conversely disturbances from the equivalent point of the meridian perpendicular to the plane of the diagram would be expected merely to add directly.

### Fig. 3.31

The vector nature of the electric fields in the focal plane of a convergent spherical wavefront.

The important differences are as follows. The distribution of time-averaged electric energy density (which is the quantity which would be detected by, for example, a photographic emulsion) is not radially symmetric: the resolving power for measurements in the azimuth at right angles to the electric vector of the incident wave is increased relative to that of the paraxial theory, whereas in the plane containing the incident electric vector the resolving power is reduced. The minima of these distributions on one of the principal azimuths in the focal plane and along the optic axis are no longer zeros. The total time-averaged energy (electric plus magnetic) however is radially symmetric, but again the minima are not zeros. The electric field is not in general in the same direction as the incident electric field; it has both a cross and a longitudinal component. The magnitude of the time-averaged Poynting vector is also radially symmetric, the power flow forming closed eddies in the region of the focal plane.

These results are very important in certain applications, for example the analysis of telescope performance, although they are not, however, sufficient per se for microscope imaging, for in a microscope the object is illuminated by one system of high aperture and observed using another. It is quite surprising therefore that very little has been written on microscope imaging.

As an example Sheppard and Wilson [3.22] have calculated the image of one and two bright points in an opaque background for aberration-free
microscope systems at high numerical apertures. They find that the image of a single point object remains radially symmetric even if the light source is plane polarised. For the conventional microscope the overall effects are quite small but the aperture of the condenser affects the image of the single point as well as determining the degree of coherence in the imaging process. The confocal microscope is found to give a slightly broader image than that predicted by paraxial theory and the depth of the minima are reduced. This is in agreement with the practical observations of Brakenhoff et al. [3.23].

The general conclusion of this discussion is that further work on the electromagnetic aspects of image formation would be immensely valuable but is likely to be very difficult. The fortunate fact is that for all its assumptions the scalar theory is able to predict to a surprising accuracy the effects that are observed in practice.

References


Chapter 4
Imaging Modes of the Scanning Microscope

4.1 General imaging considerations

It will be useful before we move on to discuss practical imaging schemes to return to our general partially coherent microscope of Chapter 3 and ask what form the transfer function should take in order that the variations in image intensity represent the required variation of object property. We recall that for an object which varies only in the x-direction, the image intensity may be written as

$$I(x) = \int_{-\infty}^{+\infty} C(m; p) T(m) T^*(p) \exp 2\pi j(m - p)x \, dm \, dp$$

(4.1)

from which we see that

$$C(m; p) = C^*(p; m)$$

(4.2)

as the intensity must be a real quantity.

We now choose to ignore diffraction effects and consider an object whose amplitude transmittance may be written as $$t(x) = a(x) \exp j\phi(x)$$. If

$$C(m; p) = 1$$

(4.3)

then

$$I(x) = |t|^2$$

(4.4)

which is often referred to as a perfect amplitude image. On the other hand if

$$C(m; p) = mp$$

(4.5)
We now move on to include the effects of diffraction, and to this end consider in detail the image of an object of the form
\[ t(x) = \exp b \cos 2\pi vx \] (4.13)
where \( b \) is, in general, complex. The case when \( b \) is small is representative of many biological specimens, and under these circumstances the image intensity may be written
\[ I(x) = C(0; 0) + 2 \text{Re}\{bC(v; 0)\} \cos 2\pi vx \\
+ \frac{1}{2} bb^*C(v; v) + \frac{1}{2} \text{Re}\{bb^*C(v; -v)\exp(4\pi jvx)\} \] (4.14)
where we have assumed the pupil functions to be symmetric.

\[ I(x) = Ic(0) + bb^*C(v; v) \] (4.15)
where \( C(m; p) = c(m)c^*(p) \).

If \( |b| \) is not small, the image intensity is related in a complicated fashion to the real and imaginary parts of \( b \). If we now specialise to the confocal rather than the conventional microscope, we have
\[ C(m; p) = c(m)c^*(p) \] (4.16)
which is the modulus square of the amplitude image of the object.

We notice here again that although the conventional and confocal microscopes have \( C(m; p) \) functions which display the same symmetries, the quality of the final image depends crucially on the precise form of the transfer function.
Returning to the general case if \(|b|\) is small, equation (4.14) reduces to
\[
I(x) = C(0; 0) + 2 \Re \{b C(v; 0)\} \cos 2\pi vx \tag{4.17}
\]
and the imaging depends only on the properties of the weak object transfer function \(C(v; 0)\). If the system is in focus, then only the real part of \(b\) is imaged. If on the other hand it is defocused, we have, as in Chapter 3, a complex \(C(v; 0)\).

Introducing
\[
C(v; 0) = C_r + jC_i
\]
and
\[
b = b_r + jb_i \tag{4.18}
\]
we have
\[
I(x) = C(0; 0) + 2(b_r C_r - b_i C_i) \cos 2\pi vx \tag{4.19}
\]
and phase information is present in the image. Even if the phase variation in the object is small, a purely imaginary weak contrast transfer function \(C(v; 0)\) results in an image of the phase information only.

If we impose the condition that \(b\) must be small and expand equation (4.1), we can show that for a differential image the condition [4.1]
\[
C(v; 0) = -C(-v; 0) \tag{4.20}
\]
must be satisfied, whereas for a standard non-differential image
\[
C(v; 0) = C(-v; 0) \tag{4.21}
\]
as in the derivation of equation (4.14). In addition we have conditions for pure amplitude contrast
\[
C(v; 0) = C^*(-v; 0) \tag{4.22}
\]
and for pure phase contrast
\[
C(v; 0) = -C^*(-v; 0). \tag{4.23}
\]

We notice that these conditions obey the symmetry requirements of Fig. 4.1. A further requirement follows from equations (4.1) and (4.20)-(4.23), that \(C(0; 0)\) must be zero for either pure differential contrast or pure phase contrast.

4.2 The dark-field and Zernike microscope arrangements

4.2.1 The dark-field microscope

This is a simple method of examining low contrast specimens, and consists of obscuring the passage of light directly from the source to the image plane. Thus only light which reaches the image is that which has been diffracted or scattered by detail in the object, and it is only this detail which is observed.

The conventional dark-field microscope employs an annular condenser and a full objective. A similar arrangement may also be used in the scanning microscope. For the sake of simplicity, we idealise the pupil functions such that the condenser (collector) pupil is unity over a narrow annular region and zero elsewhere, and the objective pupil function is unity over a circular region of radius equal to the inner radius of the annular condenser pupil, such that when the two pupils are placed on top of each other there is no common area. This ensures that \(C(0; 0)\) and \(C(m; 0)\) are zero for both Type 1 and confocal microscopes.

The \(C(m; p)\) surfaces, which were obtained from a geometrical interpretation of equations (3.28) and (3.42) are shown in Fig. 4.2(a) and (b). We can see that for the conventional dark-field microscope \(C(v; -v)\) is zero, and so the image of a weak object (equations 4.13 and 4.14) is spatially constant. Thus single spatial frequency objects are not imaged. This has been noted previously by Burge and Dainty [4.2] who consider similar objects, but consisting of two spatial frequencies, and deduce that the image contains difference but not sum frequencies in addition to constant terms.

The confocal dark-field microscope behaves very differently as \(C(v; v)\) is non-zero [4.1]. Thus single spatial frequencies are imaged [4.3], and the image intensity can now be written as
\[
I(x) = |bc(v)\cos 2\pi vx|^2 \tag{4.24}
\]
Thus we see that the dark-field method renders detail visible in terms of its diffraction pattern, and so is useful in examining weak objects. We can also see from equation (4.24) and the dark-field form of equation (4.14) that phase detail is also present in the image. This, then, is also a slightly more satisfactory way of imaging weak phase objects than the crude defocus technique of the previous chapter. It may be regarded as a limiting case of the more sophisticated Zernike phase contrast system, which modifies the bright-field image such that phase detail is imaged, rather than relying entirely on a dark-field image.
We have thus far been concerned with the images of weak phase objects. We conclude this section by discussing a strong phase object, the phase step. The images may be calculated by the methods of section 3.6 [4.4]. Figure 4.3 indicates that a phase edge is imaged in a conventional microscope by a local rise in intensity at the edge, the rise being greater the larger the phase step. We also note that far from the edge the intensity has fallen to a value of \( \frac{i}{2}(1 - \cos \Delta \phi) \), which is to be expected, as our choice of an infinite annulus means that we are averaging over the entire field of view.

The confocal image is shown in Fig. 4.4. The most striking feature is that the intensity falls to zero at the edge regardless of the value of the phase step. This is a consequence of the coherent imaging of the confocal microscope, in which we essentially form an image of the amplitude transmittance of the object minus its mean value, and this function is zero at the edge. This, of course, is not the case for the partially coherent microscope, and as a result the image is indicated by a local rise in intensity.

For the sake of completeness, we should also mention that a very simple dark-field confocal microscope may be constructed by placing the limiting pinhole over the first dark ring in the Airy disc in the detector plane [4.5]. However, although this arrangement produces dark-field conditions, the transfer function is not the same as that obtained by conventional dark-field microscopy. Hence the two schemes produce slightly different images of the same object.

4.2.2 The Zernike phase contrast method

This was the first practical method of converting the phase differences suffered by the light passing through a phase object into observable differences in amplitudes [4.6]. The method consists of using an annular condenser and an annular phase ring in the objective. It is possible to increase the sensitivity of the method slightly by using a slightly absorbing phase ring [4.7].

Let us assume that the phase ring has a transmittance \( c \exp(j\pi/2) \). We idealise the pupil functions such that the collector pupil is unity over a narrow annular region and zero elsewhere, and the objective pupil is \( c \exp(j\pi/2) \) over a narrow annular region of equal diameter to that of the condenser: unity within the circular region inside the annulus and zero outside. The transfer function \( C(m; p) \) is given for the conventional and Type 1 scanning microscope in Chapter 3 as the weighted area in common between the condenser pupil, the objective pupil displaced a distance \( m \) in spatial frequency space and the complex conjugate of the objective pupil displaced a distance \( p \). We can break up the transfer function into three parts, the first being the area in common between the condenser pupil and the circular parts.
of the objective pupils, the second between the condenser pupil and the annular parts of the objective pupils, and the third part between the condenser pupil and the circular part of one objective pupil and the annular part of the other. The first part is real and its form is shown in Fig. 4.5(a). For \( m = 0 \) or \( p = 0 \) the transfer function is zero, but for small values of \( m \) and \( p \), if they are both of the same sign, it quickly rises to a value corresponding to half the area of the annulus. Thereafter, for small \( p \), the transfer function falls off as \( \cos^{-1}(m/2a) \). If \( m \) and \( p \) are of opposite sign, the transfer function is zero. The second part of the transfer function is also real (Fig. 4.5(b)). The three annuli intersect only if two annuli are coincident, that is, along the \( m = 0 \), \( p = 0 \), or \( m = p \) lines. The value at \( m = 0 \), \( p = 0 \) is \( e^{-2\pi} \) times the area of the annulus, this being normalised to unity in the diagrams. The third part of the transfer function is imaginary. In Fig. 4.5(c) we have shown the contribution from the convolution of a circle and an annulus for \( m = 0 \) and \( p = 0 \). There is also a very small non-zero value of the transfer function for other values of \( m \) and \( p \) if the width of the annulus is non-zero, but we have neglected this for the sake of simplicity.

We consider imaging of a weak object of the form given by equation (4.13) such that equation (4.19) applies. The zero spatial frequency component in the image is taken to have a relative magnitude of unity, as given by Fig. 4.5(b). The real part of the cosinusoidal component is imaged with a strength given by the values of \( C(m;0) \) in Fig. 4.5(b). Similarly, the imaginary component is imaged with a strength given by the value of \( C(m;0) \) in Fig. 4.5(c). The properties of the image of a weak object are as follows:

i. The spatial frequency cut-off is the same as for a conventional microscope with full illumination, but the spatial frequency response for phase information is better in the Zernike arrangement than for a conventional amplitude contrast microscope.

ii. As Born and Wolf state, the intensity distribution produced in the image is directly proportional to the phase changes introduced by the object. This follows from equation (4.17), with the condition that the phase change is small.

iii. The contrast of the phase information is enhanced by a factor \( 1/c \) relative to the constant term.

iv. Weak amplitude information is imaged poorly, with fringes and low contrast.

The Zernike method suffers from a number of disadvantages, although it is a very convenient way of showing up phase information. The ideal geometry of the phase plate depends on the particular form of the specimen, there being four variables to consider: the diameter and thickness of the annulus, and the transmission and phase delay of the phase plate. De and Mondal [4.8] have discussed the effect of varying the thickness of the annulus with constant outer diameter. As might be expected, the cut-off frequency increases as the annulus width decreases. Mondal and Slansky [4.9] have considered what happens as the diameter of the annulus is varied with constant thickness. The cut-off frequency is greater for larger diameter annuli, but contrast is reduced.

For the confocal microscope, the transfer function is again made up of three parts [4.1]. If we represent the annular collector pupil by \( P_{2a} \) and the objective \( P_{1} \) by \( P_{1f} + jP_{1a} \), where \( P_{1f} \) represents the full part and \( P_{1a} \) the annular part, we can write

\[
C(m; p) = \{ P_{2a}(m) \otimes P_{1f}(m) \} \{ P_{2a}(p) \otimes P_{1a}(p) + jP_{2a}(m) \otimes P_{1a}(m) \}
\]

\[
\times \{ P_{2a}(p) \otimes P_{1a}(p) \} + j \{ P_{2a}(m) \otimes P_{1a}(m) \} \{ P_{2a}(p) \otimes P_{1f}(p) \}
\]

\[
- \{ P_{2a}(m) \otimes P_{1a}(m) \} \{ P_{2a}(p) \otimes P_{1f}(p) \}.
\] (4.25)

The first two terms here representing the real part are shown in Fig. 4.6(a) and (b). We note that these pupil functions ensure that \( C(0;0) \), \( C(m;0) \), \( C(p;0) \), etc. are all zero in Fig. 4.6(b). The imaginary part is also made up of
two components, each of opposite sign. If we assume that the annulus is narrow, the imaginary part is of greatest magnitude near to the \( m \) and \( p \) axes, and this is what we have shown, for simplicity, in Fig. 4.6(c).

The image is again composed of a constant background, a dark-field image and the phase contrast image. For a weak object, the image is precisely the same as a Type 1 or conventional Zernike arrangement. For a strong object, there are differences, but they become smaller as the thickness of the annulus becomes smaller, except for the superimposed dark-field image, which is always superior in the confocal microscope.

Again we conclude by discussing the image of a phase edge. The images are shown in Figs 4.7 and 4.8, and certain differences are noticeable between the conventional and the confocal images, in particular the high degree of fringing on the conventional image. This may be explained by considering the form of the image indicated by equation (3.64). The image we have here may be thought of as being a superposition of a phase given by the first two terms and a dark-field image given by the last term. Thus the extra fringing in the conventional image is a consequence of the form of the conventional dark-field image, which has a central maximum rather than the minimum of the confocal response. Nonetheless, the Zernike method is the first to give a steplike image. The intensity far from the edge is given by \((1 \pm \sin \Delta \phi)\), from which we can confirm that for low values of \( \Delta \phi \), constant phase changes are not imaged well, and that such an edge would only be visible in terms of the

FIG. 4.6. The transfer function \( C(m; p) \) for Zernike phase contrast imaging in a confocal scanning microscope.

FIG. 4.7. The intensity in the image of a phase edge in a Type 1 Zernike phase contrast microscope.
fringing. We also note that, unlike the dark-field case, the method is sensitive
to the direction of change of phase, i.e. the sign of $\Delta\phi$ is important. It is
interesting that in the limiting case of $\Delta\phi = \pi$ (where the average value of the
amplitude is zero), the image is exactly the same as would be obtained in a
dark-field microscope.

![Image](image.png)

Fig. 4.8. The intensity in the image of a phase edge in a confocal Zernike phase
contrast microscope.

4.2.3 The halo effect in dark-field and Zernike phase contrast microscopy

Our discussion so far has been concerned exclusively with the images of weak
phase objects. While such objects are representative of a great many
biological specimens, we must now turn our attention to strong objects, and
the phase edge in particular. This is not a pure abstraction: it may, for
example, reasonably model the edge of a biological cell.

The dark-field and Zernike schemes are two examples of a whole range of
imaging techniques which rely on the presence of a spatial filter in the Fourier
transform plane to remove or modify the zero spatial frequency component
of the image [4.8]. All these methods have the disadvantage that the image of
a sharp discontinuity exhibits a distinct halo, which may extend for many
resolution elements [4.9]. The presence of this fringing is inherent in the
method, as absolute phase is not imaged. Thus a phase edge, for example,
would not be visible without fringing. We might think that the fringing is
carried mainly by the annular region of the pupil, as might be expected for
imaging with an annular lens, but in fact this is not the principal mechanism.

The haloes are produced irrespective of the degree of coherence of the
imaging system. For the sake of clarity, then, we restrict ourselves to a
coherent system, although this means that the results will generally not be
applicable near the edge [4.10].

A straight-edge object is defined to have an amplitude transmittance

\[ t(x', y') = \begin{cases} 1; & x' > 0 \\ 0; & x' < 0; \ \forall y' \end{cases} \]

for which

\[ T(m, n) = \frac{1}{2} \left\{ \delta(m) + \frac{1}{jm} \right\} \]

and thus the image, which is only a function of $x$, becomes

\[ I(x) = \frac{1}{4} \left| c(0) + \frac{1}{jm} \left( \frac{C(m)}{m} \exp(2\pi jmx) \right) dm \right|^2. \]  

In a conventional coherent microscope, the coherent transfer function, $C(m)$,
is given by the pupil function of the lens which for a Zernike system may be
written as

\[ C(m) = d \exp(j\pi/2) \begin{cases} 0 < |m| < a \\ 1 & a < |m| < b \end{cases} \]

where $b$ is related to the outer radius of the lens and $a$ to the radius of the
central phase disc, which has a transmittance $d$. By setting $d = 0$, we
reduce to the case of the dark-field microscope [4.11]. We therefore have

\[ I(x) = \frac{1}{4} \left| dj + \frac{2}{\pi} (dj - 1) \text{Si}(2\pi ax) + \text{Si}(2\pi bx) \right|^2. \]

where $\text{Si}$ is the sine integral.

The image now consists of a fine structure of fringes superimposed on a
slowly varying background, the scale of the fine structure being determined
by the outer radius of the lens pupil. The limiting case as $b$ becomes much
larger than $a$ describes the slowly varying background as

\[ I(x) = \left( 1 - \frac{2}{\pi} \text{Si}(2\pi ax) \right)^2 ; \ x \neq 0. \]
The halo is therefore the result of diffraction by the finite size of the central disc. The extreme case of \( a \) tending to zero results in a halo that does not decay within the extent of the image, which then becomes a central dip on a uniform background. If we now move on to consider a phase edge which may be defined as

\[
I(x', y') = \exp j\phi_1; \quad x' > 0 \\
= \exp j\phi_2; \quad x' < 0
\]  

(4.32)

we can combine the previous results to obtain an expression for the halo, in this case (when \( b \gg a \)) as

\[
I(x) = \cos \left( \frac{\Delta\phi}{2} + \frac{1}{\pi d} \sin \left( \frac{\Delta\phi}{2} \right) \left[ (dj - 1) \sin (2\pi ax) + \frac{\pi}{2} \right] \right)^2. \quad (4.33)
\]

The essential features are sketched in Fig. 4.9, where the fringes have been omitted for clarity.

For the partially coherent Zernike arrangement employing an annular source and phase ring, the fringes are almost absent, and they could be further reduced by apodisation of the phase ring. The resolution is principally determined by the sum of the outer aperture of the annular source and the aperture of the objective. The width of the halo is determined by the minimum spatial frequency transmitted without change of phase, that is, by the difference between the outer aperture of the annular source and the inner aperture of the phase ring (or vice versa). In practice, we should aim for the width of the halo to be either as wide or as narrow as possible. If it is very wide, the image gives an intensity change proportional to the change of phase superimposed onto the halo, whereas if it is very narrow, we obtain imaging somewhat similar to that in dark-field. If we choose to make the halo as wide as possible, this width is limited by three major factors:

i. The annular source must be of sufficient area to provide a suitable level of illumination. This is particularly important in conventional microscopes.

ii. The thickness of the annular phase ring must be sufficiently large compared to the annular source to allow easy alignment.

iii. There is a fundamental limit set by the requirement that the thickness of the annuli must be large compared to the wavelength.

In practice, the dimensions are usually such that the halo does not fill the entire field of view.

In phase contrast or dark-field, we are imaging a change relative to some average background amplitude. The effects which we have discussed are a result of the fact that the averaging is performed over some finite region of the object. The principles are the same for the case of edge enhancement using a split pupil in which one half experiences a phase change relative to the other, as the transition region must be of non-zero width and cannot be aligned exactly with the zero spatial frequency position. The Schlieren method, in which one half of the pupil is obscured, behaves similarly, as in practice it is impossible to obscure the zero spatial frequency component without also obscuring some of the lower spatial frequencies.

### 4.3 Interference microscopy

The basic criterion for a source in a scanning optical microscope is that it should be sufficiently bright to give a reasonable signal to produce a good picture. Hence a laser is now usually used. A further advantage of using a laser is that the coherence length of the beam makes the construction of an interference microscope relatively easy, as the optical path lengths in the two arms do not have to be made equal within such close tolerances.

We will begin by discussing the general scheme of Fig. 4.10, in which light transmitted through two dissimilar parallel optical paths illuminates two photodiodes which are arranged to give signals corresponding to the sum and difference of the amplitudes of the two beams \([4,1]\). The arrangement is analogous to the Mach–Zehnder interferometer. Thus, each detector essentially gives

\[
I_x = |O \pm R|^2 \quad (4.34)
\]

where \( O \) and \( R \) refer to the object and reference beam respectively. If we now
electronically subtract these two signals, we are left with an interference term of the form

\[ I_+ - I_- \sim \text{Re} (\text{OR}^*) \]  
(4.35)

Thus by careful choice of reference beam, we can image either the real or the imaginary part of the object amplitude transmittance.

Thus by careful choice of reference beam, we can image either the real or the imaginary part of the object amplitude transmittance.

\[ I(x) = 4 \text{Re} \left[ w^* \int_{-\infty}^{\infty} h_1^*(x_0/\lambda d) h_1^*(x_0 + x_2/\lambda d) dx_0 \right] \]

whereas for a confocal scanning microscope

\[ I(x) = 4 \text{Re} \left[ \left\{ \int_{-\infty}^{\infty} h_1(x_0/\lambda d) t_1(x - x_0) h_1^*(x_0 + x_2/\lambda d) dx_0 \right\} \right] \]

(4.37)

The expression for the Type 1 microscope is quite complicated. For the confocal microscope, however, the first integral is a constant giving the modulus and phase of the reference beam, while the second is the convolution of \( h_1 h_1^* \) with the object transmittance. If the lenses in the first optical path are aberration-free such that \( h_1^* \) and \( h_1 \) are real, we can image the real or imaginary part of the object transmittance by choosing \( w \) such that the expression in square brackets is either real or imaginary. This is true even with strong contrast objects, as is the fact that the image is a linear function of the amplitude transmittance of the object.

It is of great practical importance that the image depends only upon the amplitude of the reference beam at the detector pinhole. Thus, the shape of the reference beam is immaterial.

In practice, we might arrange for the elements in the two paths to be identical, in which case the equations become much simpler.

If we now introduce the Fourier transforms of the object transmittances, we can write the image in the form

\[ I_\pm (x) = \int_{-\infty}^{\infty} C(m; p) [T_1(m) \pm T_2(m)] \{ T_1^*(p) \pm T_1^*(p) \} \exp 2\pi i(m - p) x \ dm \ dp \]  
(4.39)
where $C(m; p)$ is the transfer function for a single optical path. If the second object has constant transmittance $w$, we obtain

$$I_2(x) = \int \int C(m; p)T_1(m)T_2(p) \exp \left( 2\pi j (m - p)x \right) \, dm \, dp + |w|^2$$

$$\pm 2 \Re w^* \int C(m; 0)T_1(m) \exp \left( 2\pi jm\cdot x \right) \, dm.$$

We note that the important interference term is modified by the transfer function $C(m; 0)$. This is the case for any source and detector distribution. Reference [4.1] shows that this reduces to the corresponding transfer function for the non-interference case. However, as we know that $C(m; 0)$ is identical for Type 1 and 2 microscopes with circular aberration-free pupils, so these systems should behave identically. In the Type 1 case it is important that the lenses in the two paths are identical and that the two beams are accurately aligned over the detector surface. With the Type 2 microscope, on the other hand, the amplitude and phase of the reference beam at the point detector are the only important properties, and the alignment of the system is much less critical.

We recall from Chapter 3 that the use of one full lens and one annular lens results in the enhancement of the $C(m; 0)$ function at high spatial frequencies. Such a combination has obvious advantages. If we choose to employ such a system, the relative size of the annulus is important. If the annular lens is larger than the full lens, such as in dark-field, and if this combination is used in the reference path, there will be, for constant $t_2$, no reference beam. On the other hand, if it is used only in the object arm, we can build a dark-field interference microscope which, unlike conventional dark-field instruments, can image single spatial frequencies.

If we now return to our phase step object, we see that we may produce an image which depends on the absolute value of $\Phi_1$ and $\Phi_2$ rather than just their difference. In particular the real part of the phase step will be imaged as a step of height $\cos \Phi_1 - \cos \Phi_2$ and the imaginary part as a step of height $\sin \Phi_1 - \sin \Phi_2$. Figure 4.11 shows a typical image when $\Phi_1 = 0$ and $\Phi_2 = \pi/3$ for a microscope with two equal circular pupils. The response may also be improved slightly by introducing an annular lens into both paths, as this serves to enhance the higher spatial frequencies.

It is possible to extend the schemes we have discussed above to include two reference beams and two dummy objects. By carefully selecting the dummy objects, we can arrange to image the real and imaginary parts of the object simultaneously. These electrical signals could then be processed, and we could, for example, display an image in which intensity changes in direct proportion to phase changes in the object.

A practical arrangement for a reflection confocal interference microscope is shown in Fig. 4.12, in which the radiation is focused into a plane parallel beam before passing through the first beam splitter [4.12]. It is apparent...
from equations (4.34) and (4.35) that the sum signal \( I_s \) consists of the normal confocal image \( \langle 0 \rangle \); superimposed on a constant \( |R|^2 \), while the difference signal, \( I_d \), leaves a pure interference image. Because this signal can be bidirectional, an electronic offset is added to it, producing, for display purposes, a signal which is always positive-going.

Figure 4.13 shows the image of an area of a TEM grid formed just from the signal \( I_s \), showing a series of fringes representing variations in surface height superimposed on the confocal image. Local deformations of the fringes correspond to surface detail. Figure 4.14 shows a comparison confocal image of the same region. A deep scratch is clearly visible in both images.

The images which have been shown so far have been produced using a single detector. The effects of using the two-detector system will now be described. The interference image, \( I_s \), from the second detector is similar to \( I_s \), except that the fringes are displaced by half a period, so that the signal \( I_s \) produced by addition should exhibit no fringes. In practice small optimisations of the relative gains and pinhole alignments are needed to achieve the optimum results shown in Fig. 4.15 which is, as predicted, similar to the confocal images of Fig. 4.14. Figure 4.16 then records the image from the difference signal \( I_d \), with no further adjustments made to the system, and shows the fringe pattern without the superimposed confocal image. Even though as equation (4.35) predicts, the brightness of the fringes is modulated by the reflectivity of the object, their positions are now much more easily observed, allowing the surface topography to be deduced more precisely.

The arrangements discussed so far use a reference beam external to the object, but it is also possible to use a reference beam which passes through the specimen as well [4.1]. For example, in a differential interference microscope the same optics are used to focus light simultaneously on two adjacent points, the beam being split into two and recombined using birefringent elements. Let us assume that the objects as seen by the displaced beams are

\[
\begin{align*}
t_1(x) &= t(x - \Delta) \\
t_2(x) &= t(x + \Delta)
\end{align*}
\]  (4.40)

which, by applying the shift theorem gives for the Fourier transforms gives

\[
\begin{align*}
T_1(m) &= T(m) \exp(-2\pi j m \Delta) \\
T_2(m) &= T(m) \exp(2\pi j m \Delta)
\end{align*}
\]  (4.41)

Substituting in equation (4.39), and assuming that the signals are added with a phase difference \( 2\phi \), we have

\[
I(x) = \int_{-\infty}^{+\infty} C(m; p) T(m) T^*(p) \cos(2\pi m \Delta - \phi) \times \cos(2\pi p - \phi) \exp(2\pi j (m - p) x) \ dm \ dp
\]  (4.42)

where we have neglected a constant phase factor. If \( \Delta \) is very small, the case when \( \phi = 0 \) reduces to the conventional image of \( t \), whereas when \( \phi = \pi/2 \) we may write

\[
I(x) = \int_{-\infty}^{+\infty} C(m; p) T(m) T^*(p) 4\pi^2 \Delta^2 m p \exp(2\pi j (m - p)x) \ dm \ dp
\]  (4.43)
This may be written in the form of an effective transfer function

\[ I(x) = \int \int C_{\text{eff}}(m; p) T(m) T^*(p) \exp \left( 2\pi i (m - p) x \right) \text{dm dp} \]  

(4.44)

where

\[ C_{\text{eff}}(m; p) = 4\pi^2 \Delta^2 mp C(m; p). \]  

(4.45)

This effective transfer function satisfies the conditions of equations (4.20), (4.23) and Fig. 4.1 for differential phase contrast imaging of weak objects. On the other hand, it also has the property that \( C_{\text{eff}}(m; 0) = 0 \), so that weak objects are not imaged at all. The imaging of weak objects may be improved by setting

\[ \phi = \frac{\pi}{2} - \delta \]  

(4.46)

which modifies the effective transfer function to

\[ C_{\text{eff}}(m; p) = C(m; p)(2\pi \Delta m + \delta)(2\pi \Delta p + \delta) \]  

(4.47)

The image of an object of the form

\[ t(x) = 1 + jb_1 \cos 2\pi vx \]  

(4.48)

is given by

\[ I(x) = 1 - 2b_1 \left( \frac{2\pi \Delta v}{\delta} \right) \sin 2\pi vx + \frac{1}{2} b_1^2 C(v; v) \left\{ 1 + \left( \frac{2\pi \Delta v}{\delta} \right)^2 \right\} \]

\[ + \frac{1}{2} b_1^2 C(v; -v) \left\{ 1 - \left( \frac{2\pi \Delta v}{\delta} \right)^2 \right\} \cos 4\pi vx. \]  

(4.49)

For imaging over a wide range of spatial frequencies we may take

\[ \delta = 2\pi \Delta v. \]  

(4.50)

The weaker the object modulation, the greater is the range of spatial frequencies that is imaged linearly. It is therefore important to provide some means for the fine tuning of the relative phase of the two beams in the differential interference microscope.

As a second example of a system with a reference beam which passes through the object, we should mention the method of axial separation, in which the reference beam is focused at a point on the axis outside the specimen. Because this reference beam is defocused as it traverses the object, we may assume that the effective specimen seen by the reference beam is of unity magnitude. The image is then given by equation (4.39), and phase information may be obtained by suitable choice of the relative phase of the two signals.

### 4.4 Differential microscopy

Although a great deal of useful information may be obtained by studying a microscope image in which the contrast depends on the variation of either the amplitude (absorption) or phase of the amplitude transmittance of the object, it is often desirable to form an image which indicates how these quantities vary spatially across the object. This differential imaging is extremely useful, for example, in visualising the shape of an object or enhancing the contrast of point features.

There are many practical schemes of obtaining this contrast, and they may be divided into two main types. In the first, the differentiation is achieved by differentially detecting the radiation which has probed the object, whilst in the second a differential probe is employed.

#### 4.4.1 Differentiation in the detector plane

We begin by describing a powerful scheme proposed initially by Dekkers and de Lang [4.13] to obtain phase information in the scanning transmission electron microscope without having to resort to the cruder methods involving the defocus and spherical aberration of the electron lenses. The basic principle may be explained by considering a weak object of the form of equation (4.13) in the STEM configuration of Fig. 4.17, such that the field just after the object may be written as

\[ U(x, x_0) = (1 + b \cos 2\pi v(x_0 - x)) h(x_0) \]  

(4.51)

![Fig. 4.17. The STEM configuration.](image)
where $x$ denotes the scan and $h$ is the impulse response of the lens. The field at the detector is given by the Fourier transform of this expression, thus

$$U(x, \zeta) = \int_{-\infty}^{\infty} U(x, x_0) \exp \left(-\frac{2\pi j}{\lambda d} x_0 \zeta \right) \, dx_0$$  \hspace{1cm} (4.52)

which for a symmetrical pupil gives

$$U(\zeta, x) = P(\zeta + \lambda dv) \exp (-2\pi jv x) + b \frac{P(\zeta - \lambda dv)}{2} \exp (2\pi jv x)$$

which may be associated with three patches of light on the detector corresponding to an undeflected wave and two diffracted beams, one to the left and one to the right. This is depicted in Fig. 4.18, where the detector has also been included. As the object scans ($x$ varies) the intensity in the detector plane will also vary owing to the interference between the beam deflected to the left and the undeflected beam and the beam deflected to the right and the undeflected beam. This is indicated by the shaded area in Fig. 4.18. If we assume an unaberrated pupil function, the intensity in the left area of interference is proportional to

$$1 + \text{Re} \left\{ b \exp (-2\pi jv x) \right\}$$  \hspace{1cm} (4.54)

while that to the right is proportional to

$$1 + \text{Re} \left\{ b \exp (2\pi jv x) \right\}. \hspace{1cm} (4.55)$$

We can see from these two equations that by choosing $b$ to be imaginary, the modulation in the two areas of interference are opposite in phase. This is to be expected, since a phase object does not alter the total amount of power transmitted. However, with an amplitude object the intensity modulations in the two areas are in phase with each other.

Dekkers and de Lang took advantage of this different behaviour of phase and amplitude objects to divide their detector into two semicircular areas as shown in Fig. 4.18. If the signals from the two halves are added, amplitude information is imaged, whereas if they are subtracted, phase information is present in the image. The form of equations (4.23), (4.54) and (4.55) suggest that phase structure appears differentiated in the final image whereas amplitude structure, of course, is not differentiated.

We may analyse this system further by considering the transfer function which may be written as

$$C(m; \rho) = \int_{-\infty}^{+\infty} D(\xi, \eta) P(\lambda dm - \xi, \eta) P^*(\lambda dp - \zeta, \eta) \, d\xi \, d\eta$$  \hspace{1cm} (4.56)

where $D(\xi, \eta)$ represents the detector sensitivity.

The case when the two signals are added clearly results in the usual partially coherent imaging of the conventional microscope or the STEM. However, if we neglect aberrations, we see that when the signals are subtracted, $C(m; \rho)$ is a real, odd function with an hexagonal cut-off. The symmetry of Fig. 4.19(a) confirms that this arrangement gives differential phase contrast. The weak object transfer function $C(m; 0)$ may be evaluated in the case of a circular pupil together with a split detector of the same size as

$$C(m; 0) = \frac{2}{\pi} \left( \cos^{-1} m - \frac{1}{2} m (1 - m^2)^{1/2} \right) \hspace{1cm} \text{if} \hspace{1cm} m < \frac{1}{2}$$

$$= \frac{2}{\pi} \left( \cos^{-1} m - \frac{1}{2} m (1 - m^2)^{1/2} \right) \hspace{1cm} \text{if} \hspace{1cm} m > \frac{1}{2} \hspace{1cm} (4.57)$$

where $m = m d/2a$, and $a$ is the radius of both the detector and lens pupil. This is plotted in Fig. 4.19(b). We might also consider the effects of using different detectors and in particular the wedge detector, defined by

$$D(\xi) = \text{const} \circ \text{circ} \left( \frac{\xi}{a} \right). \hspace{1cm} (4.58)$$

Differential phase contrast results again if the signal from the two halves of the detector are subtracted. Although a slightly superior image is obtained [4.14], the simplicity of the split detector together with the high-quality conventional image that is simultaneously available during the scanning process makes this a very versatile arrangement.

In practice, as the difference signal is bidirectional, a constant voltage is usually added to it such that zero difference signal gives approximately half full screen brightness. Figure 4.20(a) shows the image of an integrated circuit.
formed by adding the signals from the detector halves. It is identical to a conventional image, the bright regions corresponding to metallisation which stands up above the silicon surface. Figure 4.20(b) shows the differential phase image formed by subtracting one of the two detected signals from the other. (Positive or negative contrast may be produced according to which signal is the subtrahend.) This image shows a pronounced effect of relief and considerable detail, particularly on the surface of the metallisation.

Fig. 4.19. (a) The symmetry of the $C(m, p)$ function indicating that differential phase contrast imaging results. (b) The weak object transfer function for a split detector (A) and a wedge detector (B).

Fig. 4.20. An integrated circuit viewed in reflection: (a) amplitude image, (b) differential phase image and (c) a similar region viewed in a conventional Zeiss microscope using Nomarski DIC.
image from just one half of the detector consists of a conventional image with differential phase contrast superimposed, again leading to an impression of relief. Figure 4.20(c) represents a comparison image obtained with a conventional microscope using the Nomarski differential contrast technique described in the last section. The same optics are used to focus light simultaneously on two adjacent points of the object, the beam being split into two and recombined using birefringent elements.

A detailed comparison of these two techniques has been carried out [4.15] and the main differences are as follows. In the Nomarski technique, the image is a complicated mixture of different contrast mechanisms, the relative strengths of which can be altered by adjusting the compensator, while the split detector produces pure differential phase contrast. The Nomarski method also results in an asymmetrical response to phase gradients, unlike the split detector method. Further the Nomarski technique is unsuitable for examining birefringent objects.

It is, however, usual in scanning optical microscopy to include a lens after the object to collect the transmitted light onto a photodetector. We might ask what kind of image would result from the use of a coded detector in this case. We recall from the discussion in Chapter 3 that the two detector distributions which prove useful are the split detectors, by analogy with the conventional microscope. The two-point detector would consist of two pinholes placed symmetrically about the optic axis of the microscope. The transfer function for a microscope with a point source may be written as

$$C(m; \rho) = \int_{-\infty}^{\infty} F_D\left(\frac{\xi_1 - \xi_1}{\lambda M d}\right) P_3(\xi_1) P_1(\lambda dp - \xi_1) \times P'_1(\lambda dm - \xi_1) d\xi_1 d\xi'_1 \quad (4.59)$$

where $F_D$ is the Fourier transform of the detector sensitivity function, which for a split detector, if we subtract the signals, is given by

$$F_D(v) \sim j \frac{1}{v} \quad (4.60)$$

and again if we subtract the signals from a two-point detector

$$F_D(v) \sim j \sin \beta v \quad (4.61)$$

where $\beta$ is a constant proportional to the spacing between the point detectors. If $\beta$ is small then

$$F_D(v) \sim jv. \quad (4.62)$$

In both cases $F_D$ is an imaginary odd function, and so, from equation (4.59) is $C(m; \rho)$. The symmetry is shown in Fig. 4.21, which indicates that this scheme results in differential amplitude contrast. We also note that the imaging is no longer coherent in the two point detector case.

![Fig. 4.21. The symmetry of the transfer function $C(m; \rho)$ indicating that differential amplitude contrast results.](image)

4.4.2 Differentiation by a differential probe

In this approach, the optical arrangement in front of the object is modified in order that the object may be probed with what might be termed a differential spot. We have already discussed one such system in the section on interferometry, where birefringent elements were used to produce spatially separate spots to probe the object, thus producing differential phase contrast. However, in this section we shall be more concerned with the effects of using lenses with coded pupil functions [4.14]. If the pupil function of the objective were an amplitude wedge, then we would probe the object with the differential of the Airy disc. In practice, it might be easier to use a split pupil, in which case the object is probed by

$$h = \frac{H_1(v)}{v} \quad (4.63)$$

in the differentiating direction, where $H_1$ is a first-order Struve function. Figure 4.22 illustrates these two forms of illumination.

We may now ask what kind of contrast would result from the use of such pupil functions. The easiest way to find out is to investigate the symmetry of the $C(m; \rho)$ function. We can see from equation (3.28) that for a conventional or Type 1 scanning microscope the transfer function is real and even.
We have already indicated in Chapter 3 that the performance of a confocal microscope may be improved by employing lenses with annular apertures. Thus we now calculate the transfer function $c(m)$ for various combinations of pupils. A combination of one split detector and one annular lens gives

$$c(m) = P_1 (\lambda dm) \otimes P_2 (\lambda dm)$$  \hspace{1cm} (4.64)$$

we find that $c(m)$ is a real odd function and that the cut-off in $(m; p)$ space is a square with the symmetry shown in Fig. 4.23 which results in what we have termed differential imaging. This is another instance of the power of coherent imaging in the confocal microscope: we have been able to incorporate the tools of coherent signal processing, but to a system having twice the spatial frequency bandwidth than is usual with such systems.

$$\begin{align*}
2c(m) &= -\cos^{-1} m - \cos^{-1} 2m; \quad m < \frac{1}{2} \\
2c(m) &= -\cos^{-1} m; \quad m > \frac{1}{2}
\end{align*}$$  \hspace{1cm} (4.65)$$

One full lens and one split annular lens gives

$$c(m) = \frac{2}{\pi} \cos^{-1} m \, sgn (m)$$  \hspace{1cm} (4.66)$$

while one full circular and one split circular lens has the transfer function of equation (4.57). These functions are plotted in Fig. 4.24. We may also calculate similar curves for the case of wedge pupils [4.14].

In confocal microscopy, the detected amplitude may be written

$$U = h_1 h_2 \otimes t$$  \hspace{1cm} (4.67)$$

and if one of the lenses has a wedge pupil function this becomes

$$U = h_1' h_2 \otimes t$$  \hspace{1cm} (4.68)$$

FIG. 4.22. The differential probe resulting from the use of a split pupil (A) and a wedge pupil (B).

FIG. 4.23. The transfer function symmetry indicating that differential imaging results.

FIG. 4.24. The transfer function for a combination of one split pupil with one full circular pupil (A); one split pupil with one annular pupil (B); one full circular pupil with one annular split pupil (C).
the prime denoting differentiation. However, for the special case of $h_1 = h_2 = h$, this becomes

$$U = \frac{1}{2} h^2 \otimes t'. \quad (4.69)$$

That is, we form an image of the differential of the amplitude transmittance of the object.

We finally consider the image of a single point. Figure 4.25 shows the amplitude image for the various practically important cases, all of which have been normalised such that they have the same slope at the origin. We recall that in the non-differential coherent microscope a combination of one annular and one full lens resulted in good imaging. The same is true here, but it is now important that the annular lens should be responsible for the differentiation. We have also plotted the ideal differential response with one annular and one full lens in Fig. 4.25, that is

$$U = \frac{\partial}{\partial x} \left( J_0(x) \frac{2J_1(x)}{x} \right). \quad (4.70)$$

Upon comparison, we see that it is reasonably similar to the wedge annulus case.

Again, we may perform similar calculations for split pupils which display similar trends, the preferred combination being one full and one split annular lens.

If we place an object within an optical resonator, we can construct a multiple beam interference microscope which is very sensitive to changes in optical path [4.1]. A schematic diagram of the method is shown in Fig. 4.26. We calculate the effect on the signal of introducing an object into the resonator, and assume that the loss in the resonator is made up of a reflector power loss $S_1$, a diffraction loss $S_2$, and that the power transmission factors of the input and output mirrors are $S_3$ and $S_4$ respectively. Let us assume that the system is at resonance with no object, so that the electric fields at the first mirror add in phase. If the incident field is $E_o$, the circulating field after reflection from the input mirror is $E$, and the signal field is $E_s$, we have, noting that the power loss per round trip is $2(S_1 + S_2) + S_3 + S_4$,

$$E_s \frac{1}{2} + E \{1 - (S_1 + S_2 + S_3/2 + S_4/2)\} = E \quad (4.71)$$

and

$$E_s = E \left\{1 - \frac{S_1 + S_2}{2}\right\}^{1/2}. \quad (4.72)$$

Taking

$$S_3 = S_4 = s' \text{ and } S_1 + S_2 = s \quad (4.73)$$

we have for the signal intensity, if $s$ is small,

$$I_s = \frac{I_o s'^2}{(s + s')^2}. \quad (4.74)$$

Let us now assume that we have an object of transmittance $t$ in the resonator. The fields still add in phase to give

$$E_o s \frac{1}{2} + E \{1 - (S_1 + S_2 + S_3/2 + S_4/2)\} t^2 = E \quad (4.75)$$
and
\[ E_v = E \left(1 - \frac{s_1 + s_2}{2}\right) s_1^{1/2} t \] (4.76)

whence, assuming \( s' = s \) and that \( s \) is small,
\[ I_v = \frac{4t^2 s^2 I_{so}}{\left(1 - (1 - 2s)^2\right)^{3/2}}. \] (4.77)

A typical figure of \( s \) obtainable with multilayer dielectric reflectors [4.16] of \( 2 \times 10^{-3} \) gives a signal which varies as shown in Fig. 4.27. The behaviour is highly nonlinear, but for an almost transparent object
\[ t = 1 - \delta \] (4.78)

where \( \delta \) is very small. We can write for the intensity
\[ I_v = \left(1 - 2\delta \left(1 + \frac{1}{2}\right)\right) I_{so} \] (4.79)

and so obtain a contrast enhancement of about 500 times. For a pure phase object
\[ t = \exp j\theta \] (4.80)

we obtain from equation (4.77)
\[ I_v = \frac{I_{so}}{1 + (1 - 2s)e^{-2} \sin^2 \theta}. \] (4.81)

This is the normal expression for the fringes in multiple beam interferometry [4.7]. Our typical value of \( s \) gives a fringe half-width that, with visible light, corresponds to a path length of the order of one nanometre.

So far we have not mentioned the spatial resolution that can be obtained with a resonant microscope. We can obtain some idea of this by calculating the field in the focal plane of a concentric resonator. Assuming that only the lowest-order mode is present, the higher-order modes being attenuated by higher diffraction losses, an approximate expression for the amplitude at the mirrors in Fig. 4.26 is
\[ \psi = \frac{v_{01} r}{a(1 + \beta(t - j)/M)} \] (4.82)

where \( v_{01} \) is the first zero of the zero-order Bessel function, \( a \) is the radius of the mirrors, \( \beta \) is given by
\[ \beta = -\frac{\zeta(\frac{1}{2})}{\sqrt{\pi}} = 0.824 \] (4.83)

where \( \zeta \) is Riemann's zeta function, and \( M \) is related to the Fresnel number \( N \) by
\[ \frac{M^2}{8\pi} = N = \frac{a^2}{2\lambda d}. \] (4.84)

In practice one would use a very high Fresnel number, for which equation (4.82) reduces to
\[ \psi = \frac{v_{01} r}{a}. \] (4.85)

In the focal plane we have an amplitude distribution given by
\[ \psi(v) = \int_0^\pi J_0(v_{01} r/a) J_0(v) v \, dv \] (4.86)

where \( v \) is the optical coordinate. This integral may be evaluated [4.17] to give
\[ I = \frac{J_2^2(v)}{\left(1 - (v/v_{01})^2\right)^2}. \] (4.87)

As \( v \to v_{01} \) both the numerator and denominator vanish and we obtain the
limit by differentiation of these quantities

\[ I_{\text{foc}(v_{01})} = \frac{v_{01}^2 J_2^2(v_{01})}{4} \]  \hspace{1cm} (4.88)

The focal plane intensity is shown in Fig. 4.28. It is somewhat wider than the corresponding Airy disc, but exhibits extremely weak side lobes.

This, then, gives us an estimate of the resolution obtainable. If the object structure is coarse compared with this distribution, the radiation in the resonator will not be disturbed. However, to actually calculate the modes of the resonator with an object present, the case of a microscope in which the radiation traverses the object twice is discussed in Chapter 6.

The resonant microscope has potential applications in studying small phase variations, and also small changes in height of a reflecting specimen. It is the coherence of the laser beam which allows the reflecting elements to be separated by such large distances. This geometry allows a large numerical aperture to be employed as the beam strikes the mirrors of the resonator normal to the coating.

4.6 Synthetic aperture imaging

It has long been known that by laterally displacing one of the lenses in a microscope, the imaging properties are altered. This leads to easier visualisation of certain features of certain classes of objects. In order to understand this effect we consider weak objects and the changes in the \( C(m; 0) \) function as the lens is moved. These functions are plotted in Fig. 4.29, and we see that in the final case, the transfer function passes spatial frequencies twice as high as in the in-line arrangement. This may be explained physically, as high spatial frequencies in the object result in large diffraction angles and as such will not be collected by the lens unless it is offset. The effect of this single sideband transfer function on the imaging is illustrated in Fig. 4.30, where we have redrawn the weak object transfer function as the sum of two functions. These may be recognised as resulting in differential phase contrast and dark-field images, but with the advantage of having twice the usual cut-off. Imaging of a similar form is obtained in the Schleiren system, where a knife edge is used to remove all the diffracted orders to one side of the zeroth, thereby resulting in a single sideband transfer function.

We might take advantage of the fact that the signal in a scanning microscope is obtained in an electrical form and can thus be processed before being displayed. If we consider the arrangement of Fig. 4.31 where we have one axial and two displaced lenses, we can combine the three signals to

---

**Fig. 4.28.** Focal plane intensity distribution for a concentric resonator (lowest-order mode).

**Fig. 4.29.** The effect on the transfer function of displacing an objective lens.
produce various pure contrast forms. These are illustrated in Fig. 4.32. Adding the outer two gives pure dark-field imaging, while subtraction results in pure differential phase contrast. Summing all three signals clearly results in a conventional image. The great advantage of this approach is that in all cases the resolution is twice as high as usual but there is no need to use very expensive, high numerical aperture lenses. Although we have limited our discussion to three lenses it would be entirely possible to extend the argument to include the use of a multi-element fly's eye lens.

We may now go on to consider the images of strong objects. Fig. 4.33 indicates the form and symmetry of the \( C(m; p) \) function when one of the outer signals is subtracted from the other, and confirms that in general differential phase contrast does result. However, recalling that \( C(m; p) \) is the

---

**Fig. 4.30.** A geometrical construction to illustrate that single sideband imaging results in differential phase contrast and dark-field imaging.

**Fig. 4.31.** Schematic diagram of a synthetic aperture imaging microscope.

**Fig. 4.32.** The transfer functions resulting from various combinations of \( I_1, I_2 \) and \( I_3 \). (a) Darkfield; (b) differential phase contrast; (c) conventional imaging.
4.7 Stereoscopic microscopy

In a conventional stereoscopic microscope [4.18] the object is viewed from two slightly different angles. The stereoscopic effect comes from a combination of mechanisms, the most important of which is the parallax effect. This results in a lateral separation of the left- and right-hand images of each object point, the magnitude of which depends on the axial distance from the focal plane. As a secondary mechanism, an inclined surface appears with different intensity in the two images, resulting in highlighting. The shape of the object also differs in the two images according to the position in the object field and associated with this is a variation in the lateral parallax shift over the field.

Stereoscopic microscopes have a limited resolution, as it is not possible to arrange for two high numerical aperture optical systems to be placed side by side. The maximum numerical aperture which may be achieved in principle without immersion using a single objective with the pupil split into two semicircles, one being used for the left-hand image and the other for the right, is 0.707, rather than unity in a non-stereoscope microscope. In practice, however, it would be difficult to approach this limit closely. If a single objective with a numerical aperture of \( \sin z \) is used in this way then the spatial frequency cut-off for each semicircular pupil is \( 2 \sin (\frac{\alpha}{2})/\lambda \) rather than \( 2 \sin \alpha/\lambda \) for the whole objective.

In the confocal microscope, however, the spatial frequency cut-off for a reflection mode system with oblique illumination [4.19] is given by the maximum angle to the optic axis subtended at the object, rather than the semi-angle subtended by the objective at the object as in a conventional microscope. Thus if this instrument is used to examine an object, with semicircular halves of the pupil being used to produce left- and right-hand images (Fig. 4.34), the spatial frequency cut-off in the resulting image is maintained at \( 2 \sin \alpha/\lambda \). Points of the object far away from the focal plane undergo a lateral shift in the stereoscopic image pairs. The object may be both illuminated and detected using semicircular pupils, or alternatively, illuminated and detected using one semicircular and one circular pupil or vice versa, the latter method resulting in a weaker stereoscopic effect. If two semicircular pupils are employed, dark-field imaging results, so that in practice the pupils are obscured by a segment slightly smaller than a semicircle. A compromise must be made between high-quality imaging and a noticeable stereoscopic effect.

Figure 4.35 shows a high magnification stereoscopic image pair produced using a reflection microscope [4.20] and a semiconductor device as object. The objective had a numerical aperture of 0.5. The pair may be viewed by conventional methods, the separation being adjusted so that all object heights appear to be a positive distance in front of the observer. Figure 4.36 shows corresponding images produced with a large area detector. This arrangement is optically similar to a conventional stereoscopic microscope, except that in scanning methods there is no variation in object shape or in frequency cut-off for each semicircular pupil is \( 2 \sin (\frac{\alpha}{2})/\lambda \) rather than \( 2 \sin \alpha/\lambda \) for the whole objective.
parallax over the object field, as each point is brought to the centre of the field
when imaged. In the resulting images, resolution is clearly much poorer than
in those produced by the confocal system.

The confocal stereoscopic pairs exhibit lateral shifts corresponding to the
axial distance from the focal plane. There is also a highlighting effect which
results from a sensitivity to surface slope as shown in Fig. 4.37, where $\varepsilon$ is the
fraction of the pupil radius obscured, adding to the stereoscopic sensation.
The intensity detected for a surface whose normal is at an angle $\theta$ to the optic
axis is given by

$$I(\theta) = c(\sin \theta) e^{*} \left( \frac{\sin \theta}{\sin \alpha} \right)$$

(4.90)

with

$$c(m) = 0; \quad m > 0$$

$$= \Lambda(m); \quad 1 > m > \varepsilon/2$$

$$= \Lambda(m) - \Lambda(1 - \varepsilon + 2m); \quad 0 < m < \varepsilon/2$$

$$= \Lambda(m) - \Lambda(1 - \varepsilon); \quad 0 > m > -(1 - \varepsilon)$$

$$= 0; \quad m < -(1 - \varepsilon).$$

(4.91)
The transfer function for weak object modulation is given by \( c(m)c^*(0) \), which may be resolved into an even part, resulting in amplitude imaging, and an odd part giving rise to differential phase contrast. These are shown in Fig. 4.38. The differential phase contrast gives opposite contrast in the two images, thus producing a confusing stereoscopic image. The method as described so far is therefore only applicable for objects in which contrast comes from variations in amplitude or surface slope. For a weak phase object, on the other hand, a stereoscopic pair may be formed by inverting the contrast of one of the pair.

References


Chapter 5

Applications of Depth Discrimination

5.1 Introduction

We have already seen in Chapter 3 that an inevitable consequence of the point detector of the confocal microscope is that the microscope images detail only from the parts of the object around the focal plane. This depth discrimination property is not a feature of conventional or Type 1 scanning microscopes. The effect may be explained physically by considering the light from outside the focal plane of the collector lens, which forms a defocused spot at the detector plane (Fig. 5.1). The use of a central point detector results in a much weaker signal and so provides discrimination in the image against detail outside the focal plane.

The effect has been illustrated by deliberately mounting a planar microcircuit in a scanning optical microscope with its normal at an angle to the optic axis [5.1]. Figure 5.2(a) shows the Type 1 microscope image: only one portion, running diagonally, is in focus. Figure 5.2(b) is the corresponding confocal image: here the discrimination against detail outside the focal plane is apparent. The areas which were out of focus in Fig. 5.1(a) have now been rejected. Furthermore the confocal image appears to be in focus throughout the width of the visible band, demonstrating that the depth discrimination effect is dominant over the depth of focus. This is important, as it means that any detail that is imaged efficiently in a confocal microscope will be in focus.

5.2 Imaging with extended depth of field

For many purposes it is preferable to use a microscope with a small depth of field. One such application is the examination of a thick slice of biological tissue, where the section of interest may be surrounded by some transparent...
supporting medium. The depth discrimination property of the confocal scanning microscope makes it ideally suited for such applications. On the other hand, a large depth of focus is often desirable; if, for example, we are interested in imaging a rough metallographic surface. An increased depth of field may be obtained in a conventional instrument, but at the price of poorer resolution. It is possible, however, to take advantage of the depth discrimination property and combine this with axial scanning of the object to produce a high resolution imaging instrument with vastly increased depth of field.

![Diagram of confocal microscope](image)

**Fig. 5.1.** Depth discrimination in the confocal microscope. When the object is in the focal plane (dashed lines) the transmitted light is focused on the pinhole, when the object is out of the focal plane, a defocused spot is formed at the pinhole, and the measured intensity is greatly reduced.

For example, if we had moved the specimen of Fig. 5.2 axially, we would have obtained a corresponding pair of images, but with a different portion of the object brought into focus [5.1]. Thus we can produce the image of a rough object such that all areas appear in focus by scanning the object in the axial (z-direction) with a amplitude sufficiently large for every part of its surface to pass through the focal plane. Needless to say, we must still scan in the x- and y-directions. The required frequency for the z-scan is determined from consideration of the horizontal (x-) and vertical (y-) resolution of the image: the z-scan frequency must be high enough to ensure that each picture point passes through focus at least once. If the optical resolution is such that there are \( N \) picture points per line of the raster and if the line scan time is \( T \) seconds, the axial scan frequency must be at least \( (N/2T) \) Hz to meet this requirement. (Two picture points can pass through focus in each cycle of axial scan, one as the specimen travels each way.)

**Fig. 5.2.** (a) Conventional scanning microscope image of tilted specimen: the parts of the object out of the focal plane are blurred. (b) Confocal scanning microscope image of tilted specimen: only the part of the specimen in the focal plane is imaged strongly. 0.5 numerical aperture objective was used with a He-Ne (0.6328 \( \mu \)m wavelength) laser. A 10-\( \mu \)m diameter pinhole was used in the confocal mode.

**Fig. 5.3.** (a) Conventional scanning microscope image of tilted specimen with axial scan. (b) Confocal scanning microscope image of tilted specimen with axial scan.
extension of more than two orders of magnitude has already been achieved [5.3]. An example of this technique at high resolution is shown in Fig. 5.4 where the hairs on an ant's leg, with two hairs projecting to the left, have been imaged. The axial distance between the tips of these hairs is 30 μm, and Fig. 5.4(a) is an extended focus image which shows both excellently resolved

![Image of ant's leg with hairs imaged](image)

**Fig. 5.4.** The hairs on an ant's leg. (a) Extended focus image with 0.85 numerical aperture objective and He-Ne laser light. (b) Conventional scanning image with reduced (1/32) numerical aperture.

along their full length, as well as much detail on the leg itself. Figure 5.4(b), in which the microscope has been focused on the tip of the projecting hair, shows an attempt to increase the depth of focus in conventional microscopy by using a very low numerical aperture lens, but it is clear that as a result the resolution has suffered dramatically. Even so, nothing approaching the depth of field of the extended focus image has been achieved.

### 5.3 Surface profiling with the confocal scanning microscope

In the reflection confocal microscope, in the intensity $I(u)$, in the image of a point object placed on the optic axis a normalised distance $u$ from the focal plane of the lens, is given by

$$I(u) = \left( \frac{\sin (u/2)}{u/2} \right)^2$$  \hspace{1cm} (5.1)

where $u$ is measured in optical coordinates.

As the object passes through focus, the image intensity shows a sharp maximum. This result can be exploited to measure the surface profile of an object along a line scan (the $x$-direction, say). As the specimen is scanned in the axial ($z$-) direction, the position of the stage in each cycle of the scan where the intensity maximum occurs will depend on the height of the object's surface. For a depression in the surface, the maximum occurs when the stage is nearer to the lens, and for a projection when the stage is further away.

A measure of the stage position at the occurrence of the maximum in each cycle therefore gives a measure of the surface height of the specimen. In order to fully exploit the resolution of the microscope in the $x$-direction, the $z$-scan frequency must be such that at least one cycle of $z$-scan occurs for each resolution element in the $x$-direction: the object height is then being sampled once per picture point. In practice, however, to ensure accurate operation of the measuring system, the $z$-scan rate should be made many times this value.

It might be thought that a digital signal processing system would be required to analyse this data, but in fact a simple analogue arrangement is found to be perfectly satisfactory [5.4]. The aim of the system is to detect the time position of the maximum, and at this instant to sample the $z$-scan waveform and hold this value until the next cycle. This output will then give a measure of the surface profile.

Figure 5.5 shows a block diagram of the system, and Fig. 5.6 the waveforms at various points in it. The $z$-scan frequency will be assumed to be sufficiently high that over one cycle the object moves a negligibly small distance in the $x$-direction. Over one cycle, therefore, the detector output will vary as shown in Fig. 5.6(b), passing through its maximum twice as the object scans through focus in each direction.

![Diagram of measurement system](image)

**Fig. 5.5.** Block diagram of the measurement system.
The detector signal is fed into a differentiator, whose output passes through zero at the signal maxima. However, there will be other turning points in the signal owing to the side lobes of the \(I(u)\) variation, and it is necessary to ensure that all turning points apart from the large central maximum are ignored. To this end only the maximum in the second half cycle is used for the actual measurement, the amplitude \(V_p\) of the first maximum being measured by a peak detector (which is reset to zero at the beginning of each cycle). The comparator then ensures that the trigger pulse to the sample-and-hold circuit is operative only when the signal level is greater than about 0.8\(V_p\), that is, for a short period enclosing the second maximum. The object height is hence measured once per cycle of the z-scan. The sample-and-hold circuit is followed by a low pass filter which limits the bandwidth to that calculated from the x-scan rate, the x-scan amplitude, and the spot size—any higher frequency components must be due to noise from the sampling system. The filter output drives the vertical channel of a pen recorder, its horizontal channel being fed from the x-scan waveform.

The technique's potential may also be demonstrated with the microcircuit test object (Fig. 5.7). The profile of a metal strip is presented in Fig. 5.8. The information is presented as a three-dimensional map of the surface in isometric coordinates, with the x- and y-directions at 30° to the horizontal and the z-direction (specimen surface height) vertically on the plot. This method gives a fairly effective subjective impression of the surface shape [5.5]. The raster comprised 112 lines, which was optimum for clarity of display; and did not cause any serious loss of resolution as long as the y-scan amplitude was not greater than about 50\(\mu\)m.

Figure 5.8 shows a profile across a metal conductor strip on the device. The shape of the strip is clearly defined and the difference between the surface texture of the metal and the surrounding semiconductor is apparent. The reflectivity of the metal and semiconductor is also quite different, but it does

![Image of waveforms in the measurement system](image)

**Fig. 5.6.** Waveforms in the measurement system.

![Image of area of microcircuit](image)

**Fig. 5.7.** Area of microcircuit. A, Metal; B, semiconductor.

![Image of profile of metal strip of microcircuit](image)

**Fig. 5.8.** Profile of metal strip of microcircuit.
not affect the height measurement system. The fringing at the edge of the step is thought to be an artefact; theoretical calculations suggest that the higher numerical aperture of the imaging lens, the less this effect is noticeable. Nevertheless, the technique is a useful non-contacting method of surface profilometry. Depth variations of the order of 0·1 μm have been clearly resolved [5.5].

5.4 The theory of image formation in extended field microscopy
In the previous chapters we have discussed the image formation properties of conventional and confocal microscopes where the object remains firmly in the focal plane. We have, however, seen in the previous sections of this chapter that high resolution images can be obtained by axial scanning in the confocal instrument. Now we shall analyse in detail the imaging properties of this kind of microscopy.

In Chapter 3 we examined the total power in the image of a single point object. We showed that it was independent of defocus in a conventional microscope, but fell off sharply in a confocal system (Fig. 3.29)—so strongly, in fact, that the integral of the power over all axial positions converges. For the extended focus method, then, we elect to integrate in \( u \) from \(-\infty\) to \(+\infty\).

The image of a point object in extended focus is thus

\[
I(v) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left[ C^2(u,v) + S^2(u,v) \right] \, du
\]

where \( C \) and \( S \) are defined in equation (2.41).

This is plotted in Fig. 5.9, normalised to unity on the axis, together with the point images for conventional and confocal systems. It is seen that the extended focus image, although very slightly broader than the confocal image, is still sharper than the conventional image. The outer rings are also weaker than in the conventional microscope, and in fact the intensity decreases monotonically with radius. If axial scanning and integration are used with a conventional microscope, the intensity for a point object may be written

\[
I(v) = 4 \int_{-\infty}^{\infty} \int_{0}^{1} P(p)P^*(p') \exp \left\{ \frac{1}{2} (\rho^2 - \rho'^2) \right\} J_0(\rho p) J_0(\rho' p') \, dp \, dp' \, du.
\]

Performing the integral in \( u \) we obtain

\[
I(v) = 4 \int_{0}^{1} \int_{0}^{1} P(p)P^*(p') J_0(\sqrt{v} p) J_0(\sqrt{v} p') \delta(p - p') \, dp \, dp'
\]

where \( \delta \) is the Dirac delta function. Making the substitutions

\[
p^2 = p \begin{cases} \rho^2 \end{cases}
p^2 = p' \begin{cases} \rho'^2 \end{cases}
\]

this can be written

\[
I(v) = \int_{0}^{1} \int_{0}^{1} P(p)P^*(p') J_0(\sqrt{v} p) J_0(\sqrt{v} p') \delta(p - p') \, dp \, dp'
\]

Somewhat surprisingly, this is independent of aberrations. For an unshaded circular pupil we obtain

\[
I(v) = 2 \int_{0}^{1} J_0^2(\sqrt{v} p) \, dp
\]

which may be written [5.6]

\[
I(v) = J_0^2(v) + J_0^2(v)
\]

with \( J_1 \) a Bessel function of order unity, which is shown in Fig. 5.9. The curve is broad, indicating a poor image, and the total power within a radius \( v \) diverges for large \( v \). This infinite incident power results from the integration over an infinite axial distance. In practice, of course, the axial scan and total power must be finite, but the resulting image is still poor if the axial scan is appreciable.

For the case of two closely spaced points, the Rayleigh criterion shows that in reflection, the extended focus case results in a small decrease in two-point resolution as compared to the incoherent conventional microscope, but that the resolution is superior by 7% in transmission [5.3].
We now turn our attention to general objects with spectrum \( T(m, n) \). We recall that for a coherent linear space invariant system with coherent transfer function \( c(m, n) \), the image intensity is given by the modulus square of the amplitude and may be written as

\[
I(x, y) = \int \int \int c(m, n) c^*(p, q) T(m, n) T^*(p, q)
\times \exp \left\{ -2\pi i \left[ (m - p)x + (n - q)y \right] \right\} \, dm \, dn \, dp \, dq. \tag{5.11}
\]

Restricting the analysis to line structures such that

\[
t(x, y) = t(x) \tag{5.14}
\]

only, equation (5.11) reduces to

\[
I(x) = \int \int C(m; p) T(m) T^*(p) \exp -2\pi i (m - p)x \, dm \, dp \tag{5.15}
\]

where

\[
C(m; p) = \int \int c(m, n) c^*(p, u) \, du \tag{5.16}
\]

for the extended focus case.

The coherent transfer function of a confocal system is given by the convolution of the defocused pupil functions of the illuminating and collecting lenses (equation 3.42). Thus for equal circular pupils in reflection mode, the transfer function is given by the integral

\[
c(m, u) = \int \int \exp \left( \frac{\pi i u^2}{\lambda} \right) \exp \left( \frac{\pi i u^2}{\lambda} \right) dS \tag{5.17}
\]

over the shaded \( S \) (Fig. 5.10).

Equation (5.11) holds for coherent systems. For a general partially coherent system we must write

\[
I(x, y) = \int \int \int C(m, n; p, q) T(m, n) T^*(p, q)
\times \exp \left\{ -2\pi i \left[ (m - p)x + (n - q)y \right] \right\} \, dm \, dn \, dp \, dq. \tag{5.12}
\]

and in general \( C(m, n; p, q) \) does not separate.

We obtain the extended focus confocal image by integrating equation (5.11) over all values of \( u \), remembering that \( c(m, n) \) is dependent on the degree of focus. Equation (5.12) thus holds for the extended focus case, in which for reflection

\[
C(m, n; p, q) = \int \int c(m, n, u) c^*(p, q, u) \, du. \tag{5.13}
\]

Fig. 5.10. Area of integration for the defocused transfer function of a confocal microscope.
Introducing a normalised spatial frequency

\[ \hat{m} = \frac{m \lambda}{2 \sin \alpha} \]  

(5.18)
yields

\[ \rho_1^2 = \rho^2 + \hat{m}^2 + 2 \rho \hat{m} \sin \theta, \]
\[ \rho_2^2 = \rho^2 + \hat{m}^2 - 2 \rho \hat{m} \sin \theta. \]  

(5.19)

Thus

\[ \frac{1}{2}(\rho_1^2 + \rho_2^2) = \rho^2 + \hat{m}^2 \]  

(5.20)
is independent of \( \theta \). Substituting in equation (5.17) and normalising so that

\[ c(m = 0, u = 0) = 1 \]  

(5.21)
one obtains

\[ c(m, u) = \exp \{ j \hat{m} \rho \} \left\{ \frac{1}{2} \int_{-\infty}^{\infty} \exp \{ j \hat{m} \rho \} \rho \, d\rho + \frac{4}{\pi} \int_{-1}^{1} \sin^{-1} \left( 1 - \frac{\hat{m}^2 - \rho^2}{2 \rho \hat{m}} \right) \exp \{ j \hat{m} \rho \} \rho \, d\rho \right\} \]  

(5.22)
of for the real and imaginary parts

\[ c_{re}(m, u) = 2 \int_{0}^{1} \cos u(\hat{m}^2 + \rho^2) \rho \, d\rho + \frac{4}{\pi} \int_{-1}^{1} \sin^{-1} \left( 1 - \frac{\hat{m}^2 - \rho^2}{2 \rho \hat{m}} \right) \cos u(\hat{m}^2 + \rho^2) \rho \, d\rho \]  

\[ c_{im}(m, u) = 2 \int_{0}^{1} \sin u(\hat{m}^2 + \rho^2) \rho \, d\rho + \frac{4}{\pi} \int_{-1}^{1} \sin^{-1} \left( 1 - \frac{\hat{m}^2 - \rho^2}{2 \rho \hat{m}} \right) \sin u(\hat{m}^2 + \rho^2) \rho \, d\rho \]  

(5.23)
in each case the first integral can be performed analytically. This coherent transfer function has been evaluated numerically, and its real and imaginary parts are shown in Fig. 5.11. For \( m = 0 \) equations (5.23) may be evaluated numerically, yielding

\[ \begin{align*}
    c_{re}(m = 0, u) &= \frac{\sin u}{u} \\
    c_{im}(m = 0, u) &= -\frac{1 - \cos u}{u}
\end{align*} \]  

(5.24)

The real part of \( c(v; 0) \) results in the imaging of amplitude variations, and the imaginary part of \( c(v; 0) \) in the imaging of phase structure. Thus for the...
confocal microscope in reflection we have
\[
C_{Re}(v; 0, u) = \left( \frac{\sin u}{u} \right) c_{Re}(v, u) + \left( \frac{1 - \cos u}{u} \right) c_{Im}(v, u)
\]
\[
C_{Im}(v; 0, u) = \left( \frac{\sin u}{u} \right) c_{Im}(v, u) - \left( \frac{1 - \cos u}{u} \right) c_{Re}(v, u)
\] (5.28)

The real and imaginary parts of this weak object transfer function are shown in Fig. 5.12. It is seen that the real part is well behaved for values of \( u \) less than about six, by which time the magnitude of the real part is small anyway. The imaginary part, on the other hand, has a magnitude as large as 0.1

around \( u = 2 \), which results in the image of phase variations, perhaps caused by specimen height variations. There are two completely different mechanisms whereby height variations may be imaged in confocal microscopy. In the first, the small variations in height are imaged by the transfer function, assumed constant over the corresponding part of the image. For example, if the transfer function is defocused by the presence of a constant displacement from the focal plane, small variations in height about this constant displacement are imaged by its imaginary part. In the second mechanism, the absolute height variations are imaged by the optical sectioning property. It is difficult to develop a theory which adequately incorporates both these effects, although it is apparent that the first mechanism results in contrast which is independent of numerical aperture, while the second becomes increasingly more important at high numerical apertures.

If we now return to consider the transfer function of the extended focus confocal microscope, we have from equation (5.16), for the weak object transfer function
\[
C(m; p) = \int_{-\infty}^{+\infty} \{ c_{Re}(m, u)c_{Re}(p, u) + c_{Im}(m, u)c_{Im}(p, u) \} \, du
\] (5.29)
that is, it is wholly real. Here we have made use of the relationship
\[
c_{Re}(m, -u) = c_{Re}(m, u),
\]
\[
c_{Im}(m, -u) = -c_{Im}(m, u)
\] (5.30)

Furthermore the relationships
\[
c_{Re}(-m, u) = c_{Re}(m, u)
\]
\[
c_{Im}(-m, u) = c_{Im}(m, u)
\] (5.31)
result in the symmetry conditions
\[
C(\pm m; \pm p) = C(\pm p; \pm m).
\] (5.32)

The weak object transfer function may be written
\[
C(m; 0) = \int_{-\infty}^{+\infty} c(m, u)c^*(0, u) \, du
\] (5.33)
where
\[
c^*(0, u) = \frac{j}{u} (\exp - j u - 1).
\] (5.34)
Using equation (5.22) and evaluating the integral in \( u \) first we have

\[
F(\hat{m}, \rho) = \frac{1}{\pi} \int_{-\infty}^{\infty} \left( \exp -ju - 1 \right) \exp ju(\hat{m}^2 + \rho^2) \, du
\]

\[
= \pi [\text{sgn} (\hat{m}^2 + \rho^2) - \text{sgn} (\hat{m}^2 + \rho^2 - 1)]
\]

where the sgn function is plus one for positive argument and minus one for negative argument, and the integral

\[
\int_{-\infty}^{0} \exp \frac{ja}{x} \, dx = \pi \text{sgn} a
\]

has been used. Thus

\[
C(m; 0) = 2 \int_{0}^{\frac{1}{1+m^2}} F(\hat{m}, \rho) \, d\rho + \frac{4}{\pi} \int_{\frac{1}{1+m^2}}^{1} \sin^{-1} \left( \frac{1 - \hat{m}^2 - \rho^2}{2\hat{m}\rho} \right) F(\hat{m}, \rho) \, d\rho
\]

\[
= C(m; 0, n = 0).
\]

The weak object transfer function for the extended focus microscope is thus purely real and identical to that for the confocal microscope (and also the conventional microscope with equal pupils).

The transfer function \( C(m; m) \) for the extended focus microscope is also purely real, and given by

\[
C(m, m) = \int_{-\infty}^{+\infty} \left[ c_{Re}(m, u) + c_{Im}(m, u) \right] \, du
\]

which is shown in Fig. 5.13. It is seen that this is a smooth, monotonically decreasing, well-behaved function which falls off more quickly than the conventional microscope with equal pupils, but more slowly than the confocal microscope.

The transfer function theory which has been presented here has been derived for objects which do not depart appreciably from one plane. In practice, if the object height is slowly varying such that the transfer function may be assumed spatially invariant over each patch of the object, the results may also be applied. In general, if the object height is allowed to vary substantially within the region where the point spread function is appreciable, the method breaks down as the transfer function is not spatially invariant; that is, it is not independent of the object. As we explained earlier, this restriction is more severe at higher numerical apertures.

**Fig. 5.13.** The transfer function \( C(m, n, u) \) for focused confocal, conventional and extended focus confocal imaging modes. This corresponds to the image of a perfect reflector at an angle \( \sin^{-1} (m/2) \) to the optic axis.

### 5.5 Conclusions

The introduction of axial scanning into a confocal microscope has resulted in a technique capable of dramatically extending the depth of field in optical microscopy. The resultant images bear some similarities to those produced in the scanning electron microscope, but do not exhibit the shadowing effects, as the illuminating and detection optics are coaxial. These images are also similar to confocal images in many respects, but do not, of course, exhibit contrast from variations in absolute phase. The confocal microscope requires that the object be scanned while retaining the axial position with high precision as otherwise intensity fluctuations appear in the image. In extended focus microscopy, however, this requirement is relaxed so that the method is less affected by external vibrations.

### References