PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

On the Assessment of Optical Images

P. B. Fellgett and E. H. Linfoot

Phil. Trans. R. Soc. Lond. A 1955 **247**, 369-407
doi: 10.1098/rsta.1955.0001

ON THE ASSESSMENT OF OPTICAL IMAGES

By P. B. FELLGETT and E. H. LINFOOT

The Observatories, University of Cambridge

(*Communicated by R. O. Redman, F.R.S.—Received* 29 *April* 1954)

In the formation of an optical image, each surface element of the object gives rise to a more or less blurred distribution in the image surface, of total brightness proportional to that of the object element. The image is the sum of these distributions in the appropriate sense: when the object is coherently lit, the image is built up by adding their complex amplitudes; when the object elements are regarded as incoherent it is the intensities which are added.

In both cases the image can be expressed as the convolution of the object with a spread function which characterizes the optical system. In systems for which the spread function does not change appreciably from one part of the field to another, the Fourier transform of the image is obtained to a sufficient approximation on multiplying the Fourier transform of the object with that of the spread function. More generally, this holds for any part of the field of a non-isoplanatic system over which the changes in the form of the spread function are small enough to be disregarded; we call such an area an 'isoplanatism-patch'. Working over such an area, an optical system can be regarded as a linear filter in which the Fourier components of the object reappear in the image multiplied by 'transmission factors'. These factors, first considered by Duffieux, depend on the aperture and aberrations of the system, and in §2 they are evaluated in terms of an ikonal function.

The qualities required of an optical image are so varied that an assessment valid over the whole range of practical applications seems out of the question. Two extreme cases are considered in the present paper. In the first of these it is assumed that the aim of an optical design is to produce an image which is directly similar to the object. This is appropriate when no process of image interpretation or reconstruction is envisaged. In the second case, the aim is to produce an image containing the greatest possible amount of information about the object, without regard to the complexity of the interpretation processes which may be needed to extract it.

For the first case, a criterion of image fidelity is proposed in §2·4 which gives a numerical measure of the resemblance of image to object in terms of the transmission factors of the optical system.

In the second case, assessment is based on the information content of the image in Shannon's sense. This depends not only on the transmission factors of the system but also on the statistical properties of the presumed object set and of the unpredictable fluctuations which necessarily disturb observation; the analysis is carried through in §3.

In §4 the assessment of optical images is discussed in terms of these two criteria.

CONTENTS

1. Introduction

The difficulties in formulating a satisfactory assessment theory of optical images are in large measure due to the wide range of purposes for which optical systems are used. It is only to be expected that the qualities which an image is required to have will differ, or will have different emphasis, in different applications; and a change in aberration balancing which improves an optical system for one purpose may well make it worse for another. It seems hardly reasonable, therefore, to expect to be able to define a unique measure of image quality; the most that one can hope to do is to devise measures which are each satisfactory over a wide range of applications. Such measures will be the more intelligible if they represent some simple physical property of the image; if they are to be susceptible to mathematical analysis they must be precisely and objectively defined; and if they are to be useful in the design of optical systems of practical interest, they must be of a simple analytical form.

The most general description of the purpose of an optical instrument is perhaps that it is to give information about the object. If the information is to be immediately available, this means that the image should resemble the object as closely as possible, so that it can be treated simply as a reproduction of the object. Any optical differences between them, other than change of scale, will then be regarded as image defects. A corresponding theory of image assessment will aim at a quantitative estimate of the similarity between object and image, and this estimate can be regarded as a measure of the amount of information that is explicitly displayed in the image.

Sometimes, however, it may be required to extract all the information that is implicit in the image without regard to the possible complexity of the interpretation process. Although such cases are less common, they are often important, since the requirement will tend to arise in connexion with photographs which it would be difficult or impossible to repeat. Photographs of a rare astronomical event are an example. Moreover, interpretative processes of a kind are inevitably introduced whenever, as is common in scientific work, photographs are measured objectively instead of being looked at (cf. Fellgett 1953, § 3). In these circumstances it is no longer a prime necessity that the image should resemble the object, and the quality of an image is more appropriately assessed by the amount of implicit information which it contains and which can be made explicit by a suitable interpretation process.

A particularly interesting example of the way in which improved performance may sometimes be obtained by dropping the requirement of similarity is provided by the 'holograms' proposed by Gabor (1949) for electron microscopy, in which the process of image interpretation, or 'reconstruction', is carried out by means of diffraction.

The amount of information implicit in an image evidently depends in an essential way on the accuracy with which the intensity distribution over the image can be measured. In practice, the accuracy of such measurements is limited by the presence of random irregularities in the image itself. In photography these arise mainly from the granularity of the emulsion; in astronomy there is also a contribution from atmospheric turbulence ('seeing'). Even in the absence of disturbances of this kind, there would always remain the fluctuations resulting from the quantum nature of the light itself. That this fundamental limit is of practical importance has been demonstrated by Rose (1948) and vividly illustrated by him in six pictures (Rose 1953) taken at different brightness levels which are evaluated in terms of the number of photons used by the light-sensitive receiver.

When similarity of image to object is not the overriding consideration, aberrations are only to be regarded as harmful in so far as they cause loss of information; the loss of similarity alone could be put right, as in Gabor's holograms, by a suitable reconstruction. Such reconstruction may increase the effects of inaccuracies in the measurement of the original image, and this interaction between interpretation processes and the errors of measurement can be regarded as the real source of the loss of information in an aberration-loaded image.

In the present paper, optical images are considered from the two points of view just described and corresponding assessment theories are developed.

Section 2 develops systematically the Fourier treatment of optical imaging. The full importance of this method, which was foreshadowed by Michelson 50 years ago, was first brought out by Duffieux (1946), who expressed the properties of an optical system by means of its transmission factors for the Fourier components of the object. As Elias, Grey & Robinson (1952), among others,† have pointed out, this approach provides a link between optical theory and the fruitful ideas and methods which have been developed during the last 30 years in connexion with linear wave filters. In § 2·4, a quantitative definition of image fidelity is introduced.

In §§ 3·11 to 3·13 an outline is given of the concepts of optical noise and information; and a more detailed analysis follows in §§ 3·2 and 3·3, leading to the calculation in § 3·233 of the information content of an optical image under prescribed conditions.

Section 4 discusses the two types of image assessment which correspond respectively to the fidelity evaluation introduced in § 2·4 and to the evaluation of information content made in § 3·233. As already explained, the first assessment is appropriate when the aim is to obtain, without any interpretation process, the greatest possible similarity between image and object; the second when information implicit in the image is to be extracted by reconstruction processes.

† A valuable paper by Blanc-Lapierre (1953) deals with the same topic; see also Gabor (1952).

2. Transmission factors and the fidelity of optical images

1. *Optical images and Fourier transforms*

Let (x, y) be scale-normalized co-ordinates in the object surface S of an optical system (see figure 1) and (ξ, η) scale-normalized Cartesian co-ordinates in the exit pupil. We can define a set of co-ordinate numbers (x', y') in the image surface S' of the system by first assigning to the Gauss image of the object point (x, y) the co-ordinate numbers $(x' = x, y' = y)$. It is convenient to choose the scale normalizations so that, near the optic axis, (x, y) agree with the (small) angular off-axis displacements in radians of the point (x, y) in the object surface, while (ξ, η) agree with the (small) angular off-axis distances of the point $Q = (\xi, \eta)$ in the exit pupil as seen from O'.

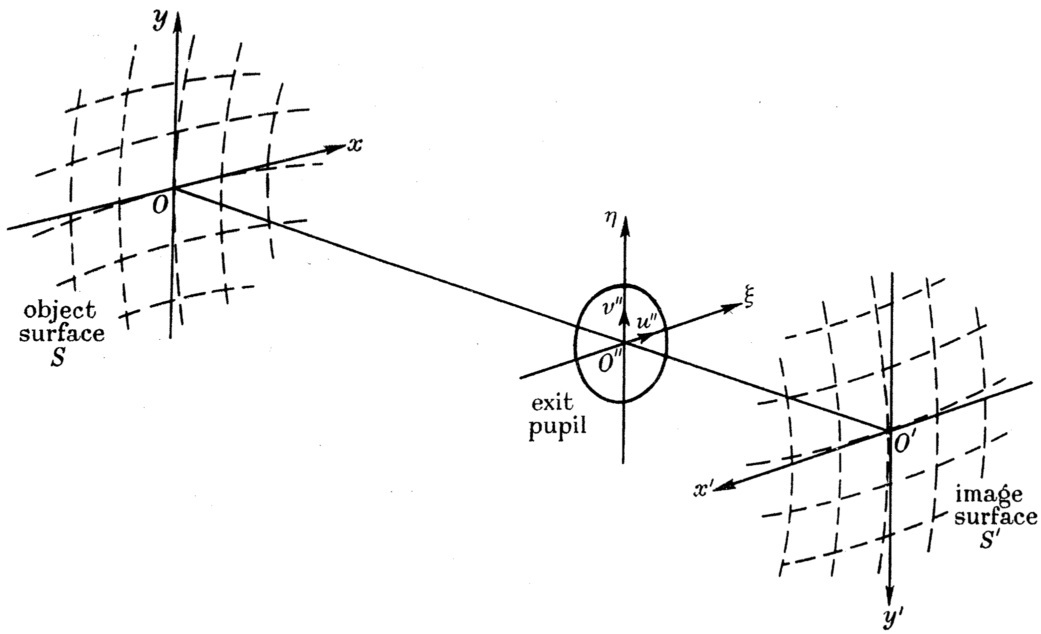


Figure 1

Let the total power of the radiation reaching the image surface from a surface element $dx\,dy$ situated at (x, y) in the object surface be $\sigma(x, y)\,dx\,dy$. Because of aberrations and diffraction, this energy is spread out in the image surface S' over a region surrounding the point (x, y) in this surface, and the intensity contribution $dI(x', y')$ received from the element $dx\,dy$ at an arbitrary point (x', y') of S' can be written in the form

$$dI(x', y') = \sigma(x, y)\, w(x', y'; x, y)\, dx\, dy, \tag{2·1}$$

where

$$\iint_{S'} w(x', y'; x, y)\, dS' = 1. \tag{2·2}$$

If the object is incoherently lit, or is self-luminous, the total intensity $I(x', y')$ at (x', y') in the image is the sum of the intensity contributions (2·1) from all the elements $dx\,dy$ of the object surface, and if $\sigma(x, y)$ be set equal to zero everywhere outside the working field, we can write

$$I(x', y') = \iint_{-\infty}^{\infty} \sigma(x, y)\, w(x', y'; x, y)\, dx\, dy. \tag{2·3}$$

In an actual system, the shape of the spread function $w(x', y'; x, y)$ varies only slowly as (x, y) explores the working field in S. More precisely, the working field can be divided up into patches, each large compared with the size of the finest detail which can be resolved, with the property that in each patch A we can write

$$w(x', y'; x, y) = w_A(x'-x, y'-y) \tag{2·4}$$

to a sufficient approximation. A patch with these properties is called an *isoplanatism-patch* of the system. A system is called *isoplanatic* when the whole of its working field is an isoplanatism-patch.

When the system is isoplanatic, equation (2·3) takes the form

$$I(x', y') = \iint_{-\infty}^{\infty} \sigma(x, y)\, w(x'-x, y'-y)\, \mathrm{d}x\,\mathrm{d}y. \tag{2·5}$$

If we define, by means of the equations

$$\epsilon(u, v) = \iint_{-\infty}^{\infty} \sigma(x, y)\, \mathrm{e}^{-2\pi\mathrm{i}(ux+vy)}\, \mathrm{d}x\,\mathrm{d}y, \tag{2·6}$$

$$\tau(u, v) = \iint_{-\infty}^{\infty} w(x, y)\, \mathrm{e}^{-2\pi\mathrm{i}(ux+vy)}\, \mathrm{d}x\,\mathrm{d}y, \tag{2·7}$$

the functions $\epsilon(u, v)$ and $\tau(u, v)$ of the new variables u, v, and apply to (2·5) the Fourier product theorem and the Fourier inversion theorem, we obtain at once

$$\iint_{-\infty}^{\infty} I(x, y)\, \mathrm{e}^{-2\pi\mathrm{i}(ux+vy)}\, \mathrm{d}x\,\mathrm{d}y = \epsilon(u, v)\, \tau(u, v), \tag{2·8}$$

$$I(x, y) = \iint_{-\infty}^{\infty} \epsilon(u, v)\, \tau(u, v)\, \mathrm{e}^{2\pi\mathrm{i}(ux+vy)}\, \mathrm{d}u\,\mathrm{d}v. \tag{2·9}$$

Here (u, v) are regarded simply as spatial frequencies and the equation

$$\sigma(x, y) = \iint_{-\infty}^{\infty} \epsilon(u, v)\, \mathrm{e}^{2\pi\mathrm{i}(ux+vy)}\, \mathrm{d}u\,\mathrm{d}v, \tag{2·10}$$

obtained from (2·6) by Fourier inversion, is a representation of $\sigma(x, y)$ in terms of its 'spatial spectrum', which is described by the spectral function $\epsilon(u, v)$. Equations (2·9) and (2·10) show that the 'Fourier element' $\epsilon(u, v)\, \mathrm{e}^{2\pi\mathrm{i}(ux+vy)}\mathrm{d}u\,\mathrm{d}v$ in the object appears in the image multiplied by the 'transmission factor' $\tau(u, v)$, where $\tau(u, v)$ is the Fourier transform of the spread function $w(x, y)$, so that, by (2·2), $\tau(0, 0) = 1$ and $|\tau(u, v)| \leqq 1$.

If we use $F[f]$ to denote the Fourier transform of f in the sense of (2·6) and $F^*[g]$ to denote the inverse Fourier transform of g in the sense of (2·10), the last five equations, together with the inverse of (2·7), can be written in the more compact form

$$\epsilon(u, v) = F[\sigma], \quad \sigma(x, y) = F^*[\epsilon], \tag{2·11}$$

$$\tau(u, v) = F[w], \quad w(x, y) = F^*[\tau]. \tag{2·12}$$

$$\epsilon(u, v)\, \tau(u, v) = F[I], \quad I(x, y) = F^*[\epsilon\tau], \tag{2·13}$$

while Fourier's inversion theorem can be written as the operator equation

$$FF^* = 1. \tag{2·14}$$

If the object is coherently lit, the analysis is changed only in that it is now the complex displacement contributions in the image surface which have to be added. Thus, if the complex function $\hat{E}(x, y)$ describes the phase-amplitude distribution in the coherent object and if $g(x', y'; x, y)$ describes the normalized complex displacement at (x', y') in the image surface due to a point source of unit strength and zero phase situated at the point (x, y) in the object surface, then the image is described by the complex distribution

$$\hat{D}(x', y') = \iint_{-\infty}^{\infty} \hat{E}(x, y)\, g(x', y'; x, y)\, \mathrm{d}x\, \mathrm{d}y \tag{2·15}$$

and the intensity distribution in the image is

$$\hat{I}(x', y') = |\,\hat{D}(x', y')\,|^2. \tag{2·16}$$

From the definitions of g and of w it follows† that

$$w(x', y'; x, y) = |\,g(x', y'; x, y)\,|^2. \tag{2·17}$$

When the system is isoplanatic over the area occupied by the object, $g(x', y'; x, y)$ takes the special form $g(x'-x, y'-y)$, (2·15) becomes

$$\hat{D}(x', y') = \iint_{-\infty}^{\infty} \hat{E}(x, y)\, g(x'-x, y'-y)\, \mathrm{d}x\, \mathrm{d}y, \tag{2·18}$$

and we obtain in place of (2·11) to (2·13) the equations

$$\hat{\epsilon}(u, v) = F[\hat{E}], \quad \hat{E}(x, y) = F^*[\hat{\epsilon}], \tag{2·19}$$

$$\hat{\tau}(u, v) = F[g], \quad g(x, y) = F^*[\hat{\tau}], \tag{2·20}$$

$$\hat{\epsilon}(u, v)\,\hat{\tau}(u, v) = F[\hat{D}], \quad \hat{D}(x, y) = F^*[\hat{\epsilon}\hat{\tau}], \tag{2·21}$$

while (2·17) becomes

$$w(x, y) = |\,g(x, y)\,|^2. \tag{2·22}$$

From the spectral representations

$$\hat{E}(x, y) = F^*[\hat{\epsilon}], \quad \hat{D}(x, y) = F^*[\hat{\epsilon}\hat{\tau}] \tag{2·23}$$

of the object and image given in (2·19) and (2·21), we see that $\hat{\tau}(u, v)$ is the cofactor with which a Fourier element $\hat{\epsilon}(u, v)\, \mathrm{e}^{2\pi \mathrm{i}(ux+vy)}\, \mathrm{d}u\, \mathrm{d}v$ of the coherent object reappears in the image; it is the transmission factor of the optical system for coherent objects (Duffieux 1946).

2·2. *Transmission factors and ikonal function*

The results obtained so far have been derived without assuming any specific mechanism of image propagation; they depend only on the assumptions that the radiation from each element dS of the object is spread out in the image surface according to a fixed distribution function, and that the contributions from different parts of the object are added as complex amplitudes in the coherent case, as energies in the incoherent case. They take on a new significance when combined with the equations which express, in the Huyghens approximation, the relations between the spread function $g(x', y'; x, y)$ and the ikonal function of the optical system.

† Not quite trivially, since an infinitesimal surface element of unit total power cannot be represented by a single wave train.

A point source of unit strength and zero phase, situated at (x, y) in the object surface, produces at (x', y') in the image surface a complex displacement proportional to

$$d_\lambda(x', y'; x, y) = \iint \exp\{-ike(\xi, \eta; x, y)\}\, e^{ik[\xi(x'-x)+\eta(y'-y)]}\, d\xi\, d\eta, \tag{2·24}$$

where λ denotes the wave-length, $k = 2\pi/\lambda$ and the integration is over the (ξ, η)-region defined by the exit pupil. $e(\xi, \eta; x, y)$ is an ikonal function of the system and

$$\exp\{-ike(\xi, \eta; x, y)\}$$

measures the complex displacement produced by the point source (x, y) in the object surface over that part of a spherical reference surface, centred at (x, y) in the image surface and with radius $O'O''$ (see figure 1), which approximately coincides with the exit pupil.† When the imaging is isoplanatic as (x, y) explores the object, $e(\xi, \eta; x, y)$ takes the special forms $e(\xi, \eta)$ in the coherent case, $e(\xi, \eta) + f(x, y)$ in the incoherent case, where $f(x, y)$ is a smooth function of x and y.‡

In the special case where the exit pupil is circular, it has often been found convenient to replace (ξ, η) by co-ordinates (u', v') renormalized to make the circle $u'^2 + v'^2 < 1$ represent the exit pupil and to express this ikonal e as a function of u', v' and the normalized off-axis angle Θ of the point (x, y) (cf. Linfoot 1955, § 3·1). In the present case, where the exit pupil is not restricted to be circular, or even to be a single area, it is more convenient to use a different renormalization, introducing the new co-ordinates u'', v'' in the exit pupil by means of the equations

$$\xi = \lambda u'', \quad \eta = \lambda v'', \tag{2·25}$$

where, as before, λ denotes the wave-length of the light.

Equation (2·24) then gives, with a suitable normalization of g,

$$g(x', y'; x, y) = |\mathscr{A}|^{-\frac{1}{2}} \iint_{-\infty}^{\infty} \mathscr{E}(u'', v''; x, y)\, e^{2\pi i[u''(x'-x)+v''(y'-y)]}\, du''\, dv'', \tag{2·26}$$

where

$$\begin{aligned} \mathscr{E}(u'', v''; x, y) &= \exp\{-ike(\lambda u'', \lambda v''; x, y)\} \quad (u'', v'') \subset \mathscr{A} \\ &= 0 \quad (u'', v'') \not\subset \mathscr{A}, \end{aligned} \tag{2·27}$$

and $\mathscr{A}$ is the (u'', v'')-region which represents the exit pupil.

In the coherent isoplanatic case $e(\xi, \eta; x, y) = e(\xi, \eta)$, and (2·26) takes the more special form

$$g(x', y'; x, y) = g(x' - x, y' - y), \tag{2·28}$$

where

$$g(x, y) = |\mathscr{A}|^{-\frac{1}{2}} \iint_{-\infty}^{\infty} \mathscr{E}(u'', v'')\, e^{2\pi i(u''x + v''y)}\, du''\, dv'' = |\mathscr{A}|^{-\frac{1}{2}} F^*[\mathscr{E}] \tag{2·29}$$

and

$$\left.\begin{aligned} \mathscr{E} = \mathscr{E}(u'', v'') = \mathscr{E}(u'', v''; x, y) &= \exp\{-ike(\lambda u'', \lambda v'')\} \quad (u'', v'') \subset \mathscr{A} \\ &= 0 \quad (u'', v'') \not\subset \mathscr{A}. \end{aligned}\right\} \tag{2·30}$$

In the incoherent isoplanatic case, $e(\xi, \eta; x, y) = e(\xi, \eta) + f(x, y)$ and (2·28) is replaced by

$$g(x', y'; x, y) = \exp\{-ikf(x, y)\}\, g(x' - x, y' - y). \tag{2·31}$$

† Variations in amplitude on this surface are taken care of by the imaginary part of $e(\xi, \eta; x, y)$. In the present paper we assume that these variations are negligible and that $e(\xi, \eta; x, y)$ is real.

‡ $f(x, y)$ represents a phase shift, constant over the whole reference sphere but different at different points of the image.

Applying Fourier's inversion theorem to (2·29), we obtain

$$\mathscr{E}(u,v) = |\mathscr{A}|^{\frac{1}{2}}F[g], \tag{2·32}$$

from which, on comparing it with (2·20), we see that the transmission function $\hat{\tau}(u,v)$ for extended coherent objects is, apart from a constant factor $|\mathscr{A}|^{-\frac{1}{2}}$, identical with $\mathscr{E}(u,v)$, and that u'', v'', which were introduced as normalized co-ordinates in the exit pupil, can be identified with the spatial frequencies u, v in the image (2·15).

This identification appears the more natural in view of the relation, familiar in elementary diffraction theory, between the angle of an obliquely diffracted beam and the spatial frequency in the structure of the surface from which it is diffracted. In the case where the image surface is part of a sphere centred at O'', it is easily seen† that radiation from the vicinity of the point $u''=u$, $v''=v$ in the exit pupil corresponds to a spatial frequency (u,v) in the image structure. That the same is true for any smooth receiving surface lying in the image layer of the optical system follows from the properties of the ikonal function $e(\xi,\eta;x,y)$, which automatically provides phase factors covering the generalization.

Putting the results together and writing generally $C[h]$ for the autocorrelation function $\iint_{-\infty}^{\infty} h(u',v')\,h^*(u'-u,v'-v)\,\mathrm{d}u'\,\mathrm{d}v'$, where h^* denotes the complex conjugate of h, we obtain for the imaging of a coherent isoplanatism patch the equations

$$\hat{E}(x,y) = F^*[\hat{\epsilon}], \tag{2·33}$$

$$\hat{D}(x,y) = F^*[\hat{\epsilon}\hat{\tau}] = |\mathscr{A}|^{-\frac{1}{2}}F^*[\hat{\epsilon}\mathscr{E}], \tag{2·34}$$

$$\hat{I}(x,y) = |\hat{D}(x,y)|^2 = |\mathscr{A}|^{-1}F^*[C[\hat{\epsilon}\mathscr{E}]], \tag{2·35}$$

and for the imaging of an incoherent isoplanatism patch the equations

$$\sigma(x,y) = F^*[\epsilon], \tag{2·36}$$

$$I(x,y) = F^*[\epsilon\tau] = |\mathscr{A}|^{-1}F^*[\epsilon C[\mathscr{E}]]. \tag{2·37}$$

In the first case the transmission factor

$$\begin{aligned}\hat{\tau} = \hat{\tau}(u,v) = |\mathscr{A}|^{-\frac{1}{2}}\mathscr{E}(u,v) &= |\mathscr{A}|^{-\frac{1}{2}}\exp\{-ike(\lambda u,\lambda v)\} \quad (u,v)\subset\mathscr{A}\\ &= 0 \quad (u,v)\not\subset\mathscr{A},\end{aligned} \tag{2·38}$$

where $e(\xi,\eta)$ is the ikonal function of the isoplanatism-patch; in the second case this ikonal function has the form $e(\xi,\eta)+f(x,y)$ and, by (2·12), (2·22) and the Wiener-Khintchine theorem, the transmission factor $\tau=\tau(u,v)$ satisfies the equation

$$\tau(u,v) = \frac{1}{|\mathscr{A}|}\iint_{-\infty}^{\infty}\mathscr{E}(u',v')\,\mathscr{E}^*(u'-u,v'-v)\,\mathrm{d}u'\,\mathrm{d}v' = |\mathscr{A}|^{-1}C[\mathscr{E}], \tag{2·39}$$

and from this it follows at once that $|\tau|\leqq\tau_0$ for all (u,v), where $\tau_0(u,v)$, defined by the equations

$$\tau_0 = C[\kappa_{\mathscr{A}}], \quad \kappa_{\mathscr{A}} = \kappa_{\mathscr{A}}(u,v) = \begin{cases}1 & (u,v)\subset\mathscr{A}\\ 0 & (u,v)\not\subset\mathscr{A}\end{cases}$$

is the transmission function of an ideal aberration-free system of aperture $\mathscr{A}$.

† Linfoot (1946), equation (2·14).

Because $e(\xi, \eta; x, y)$ is a slowly changing function of (x, y), the above analysis can be applied without essential change to the approximate discussion of problems of resolution and of information content in images by non-isoplanatic systems. In fact, to each point (x, y) of the working field of an optical system corresponds a neighbourhood $R_{x,y}$ called the isoplanatism-patch belonging to (x, y). $R_{x,y}$ is the set of points (x', y') near (x, y) at which $\mathscr{E}(u, v; x', y')$ may be taken as effectively equal to $\mathscr{E}(u, v; x, y)$ (or, in the incoherent case, to $\mathscr{E}(u, v; x, y)$ multiplied by a smooth function of x, y) for all values of (u, v). Because $e(\xi, \eta; x, y)$ is a slowly changing function of (x, y), $R_{x,y}$ is large compared with the finest detail in the image; its precise dimensions depend on the analytical interpretation given to the words 'effectively equal' in the previous sentence.

In each $R_{x,y}$ the image of an extended coherent (incoherent) object is essentially the convolution of its phase-amplitude (intensity) distribution with the phase-amplitude (intensity)

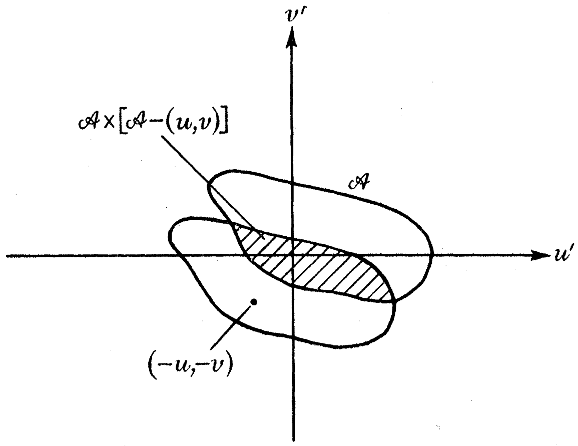


FIGURE 2. $\mathscr{A} \times [\mathscr{A} - (u, v)]$.

distribution in the image surface corresponding to a point object at (x, y). The latter is essentially the Fourier transform of the aperture-aberration function $\mathscr{E}(u, v; x, y)$ of the system in the first case, and of the (u, v)-autocorrelation function of $\mathscr{E}(u, v; x, y)$ in the second case.

2·3. *Frequency-boundedness of optical images*

From (2·34) and (2·38) it follows at once that the frequency (u, v) of every non-null Fourier element of $\hat{D}(x, y)$ lies in $\mathscr{A}$. This is expressed by saying that *the image of a coherent object* $\hat{E}(x, y)$ *is frequency-limited to* $\mathscr{A}$.

An analogous result holds for incoherent objects. We can write (2·39) in the more explicit form

$$\tau(u, v) = \frac{1}{|\mathscr{A}|} \iint_{\mathscr{A} \times [\mathscr{A} - (u, v)]} \exp\{-ik[e(\lambda u', \lambda v') - e(\lambda u' + \lambda u, \lambda v' + \lambda v)]\} \, du' \, dv', \quad (2·40)$$

where the domain of integration arises in the manner illustrated in figure 2. Evidently $\tau(u, v)$ can only differ from zero when this domain is non-null; that is, when (u, v) satisfies the condition

$$\mathscr{A} \times [\mathscr{A} - (u, v)] \neq 0. \quad (2·41)$$

Let $\mathscr{F} = \mathscr{F}(\mathscr{A})$ denote the set of values of (u, v) satisfying (2·41); then $\tau(u, v) = 0$ for $(u, v) \notin \mathscr{F}$ and (2·37) shows that *the image* $I(x, y)$ *of an incoherent object is frequency-limited to* $\mathscr{F}$.

If $\mathscr{A}$ is the circle $u^2+v^2 \leqq a^2$, then $\mathscr{F}$ is the circle $u^2+v^2 \leqq 4a^2$, and the image of an incoherent object may contain frequencies twice as rapid as those which can appear in the complex displacement image of a coherent object by the same system.

$\mathscr{A}$ and $\mathscr{F}$ are bounded regions in the (u, v)-plane; thus *the images of coherent and of incoherent objects are frequency-bounded.* It is not difficult to show that the same is true of the images of partially coherent objects.

2·4. *Fidelity of images*

An intuitively acceptable and analytically convenient measure of the extent to which the image resembles the object can be defined in terms of the root mean square (r.m.s.) distance between the functions which represent them.

Let $f_1(x, y)$, $f_2(x, y)$ be two (real or complex) functions whose total powers

$$P(f_1) = \iint_{-\infty}^{\infty} |f_1|^2 \,dx\,dy, \quad P(f_2) = \iint_{-\infty}^{\infty} |f_2|^2 \,dx\,dy \tag{2·42}$$

are both finite. Their r.m.s. distance $d(f_1, f_2) \geqslant 0$ is defined by the equation

$$d^2(f_1, f_2) = \iint_{-\infty}^{\infty} |f_1 - f_2|^2 \,dx\,dy. \tag{2·43}$$

Evidently $\quad P(f_1) = d^2(f_1, 0) \quad$ and $\quad P(f_2) = d^2(f_2, 0)$.

The fidelity defect of an optical image can now be defined as the normalized mean square distance between this image and the corresponding object. More precisely, if f_1 represents the object and f_2 the image, we define the fidelity defect of the image as the quotient

$$\frac{d^2(f_1, f_2)}{d^2(f_1, 0)} = \frac{\iint_{-\infty}^{\infty} |f_1 - f_2|^2 \,dx\,dy}{\iint_{-\infty}^{\infty} |f_1|^2 \,dx\,dy}. \tag{2·44}$$

In the case of an incoherent object $\sigma(x, y)$ and its image $I(x, y)$, the mean square distance

$$\begin{aligned} d^2(\sigma, I) &= \iint_{-\infty}^{\infty} (\sigma - I)^2 \,dx\,dy \\ &= \iint_{-\infty}^{\infty} |\tau\epsilon - \epsilon|^2 \,du\,dv, \end{aligned} \tag{2·45}$$

by (2·36), (2·37) and Parseval's theorem,

$$= \iint_{\mathscr{F}} |1-\tau|^2 |\epsilon|^2 \,du\,dv + \left(\iint_{-\infty}^{\infty} - \iint_{\mathscr{F}}\right) |\epsilon|^2 \,du\,dv, \tag{2·46}$$

while
$$d^2(\sigma, 0) = \iint_{-\infty}^{\infty} \sigma^2 \,dx\,dy = \iint_{-\infty}^{\infty} |\epsilon|^2 \,du\,dv. \tag{2·47}$$

Thus the fidelity defect of the image is

$$\left(1 - \frac{\iint_{\mathscr{F}} |\epsilon|^2 \,du\,dv}{\iint_{-\infty}^{\infty} |\epsilon|^2 \,du\,dv}\right) + \frac{\iint_{\mathscr{F}} |1-\tau|^2 |\epsilon|^2 \,du\,dv}{\iint_{-\infty}^{\infty} |\epsilon|^2 \,du\,dv}. \tag{2·48}$$

In the case of a coherent object $\hat{E}(x,y)$ and its image $\hat{D}(x,y)$, (2·48) is replaced by the expression

$$\left(1-\frac{\iint_{\mathscr{A}}|\hat{\epsilon}|^2\,\mathrm{d}u\,\mathrm{d}v}{\iint_{-\infty}^{\infty}|\hat{\epsilon}|^2\,\mathrm{d}u\,\mathrm{d}v}\right)+\frac{\iint_{\mathscr{A}}|1-\hat{\tau}|^2\,|\hat{\epsilon}|^2\,\mathrm{d}u\,\mathrm{d}v}{\iint_{-\infty}^{\infty}|\hat{\epsilon}|^2\,\mathrm{d}u\,\mathrm{d}v}. \tag{2·49}$$

In each of (2·48) and (2·49), the first term depends only on the aperture of the optical system, while the second term depends on both aperture and aberrations. In the coherent case (2·49), the second term vanishes when the aberrations are zero; and we can regard the two parts of (2·49) as the respective contributions to the fidelity defect from the frequency cut-off in the image through the finite aperture $\mathscr{A}$ and from the aberrations. In the incoherent case (2·48), $|\tau|<1$ for $(u,v)\neq(0,0)$ and the second term represents the contribution to the fidelity defect from the 'damping' of the higher object frequencies which is enhanced by aberrations, and accompanied by phase distortion when aberrations are present, but which occurs even in the absence of aberrations; while the first term represents the contribution from the frequency cut-off (this time to $\mathscr{F}$) in the image through the finite aperture $\mathscr{A}$.

It is of fundamental importance in this connexion that object frequencies outside a certain finite region ($\mathscr{A}$ or $\mathscr{F}$) in the (u,v)-plane do not appear at all in the image; two object functions σ_1, σ_2 whose Fourier transforms differ only outside this region have identical images. In fact, the image of an incoherent object σ only contains such information about σ as can be expressed in terms of the 'τ_0 cut-off' of σ, namely, the function

$$\sigma^{\tau_0}=F^*[\tau_0\epsilon];\quad \tau_0=C[\kappa_{\mathscr{A}}];\quad \kappa_{\mathscr{A}}=\kappa_{\mathscr{A}}(u,v)=\begin{cases}1 & (u,v)\subset\mathscr{A}\\ 0 & (u,v)\not\subset\mathscr{A};\end{cases} \tag{2·50}$$

that is to say, the image of σ through an ideal aberration-free system of the same aperture $\mathscr{A}$ as the actual system.

A result of a more special character which follows from (2·46) is that the mean-square distance between an incoherent object σ and its image can never fall below

$$\left(\iint_{-\infty}^{\infty}-\iint_{\mathscr{F}}\right)|\epsilon|^2\,\mathrm{d}u\,\mathrm{d}v.$$

3. Noise, aberrations and information

3·1. *Introductory*

The term 'noise' is already used in many fields of physics to denote those fluctuations which in the circumstances of a given experiment must be regarded as unpredictable in detail and therefore a bar to perfectly exact measurement. No apology seems necessary, therefore, for using it here to denote such fluctuations affecting an optical image.

Every optical image is affected by noise; as already pointed out, the quantum nature of light itself renders the production of noise-free images impossible in principle, and the practical limitations of the eye or photographic plate give a noise level that is substantially greater than this fundamental limit.

In this introduction, comprising §§ 3·11, 3·12 and 3·13, we set out briefly and without proofs the ideas and results underlying the main discussion, which begins with § 3·2.

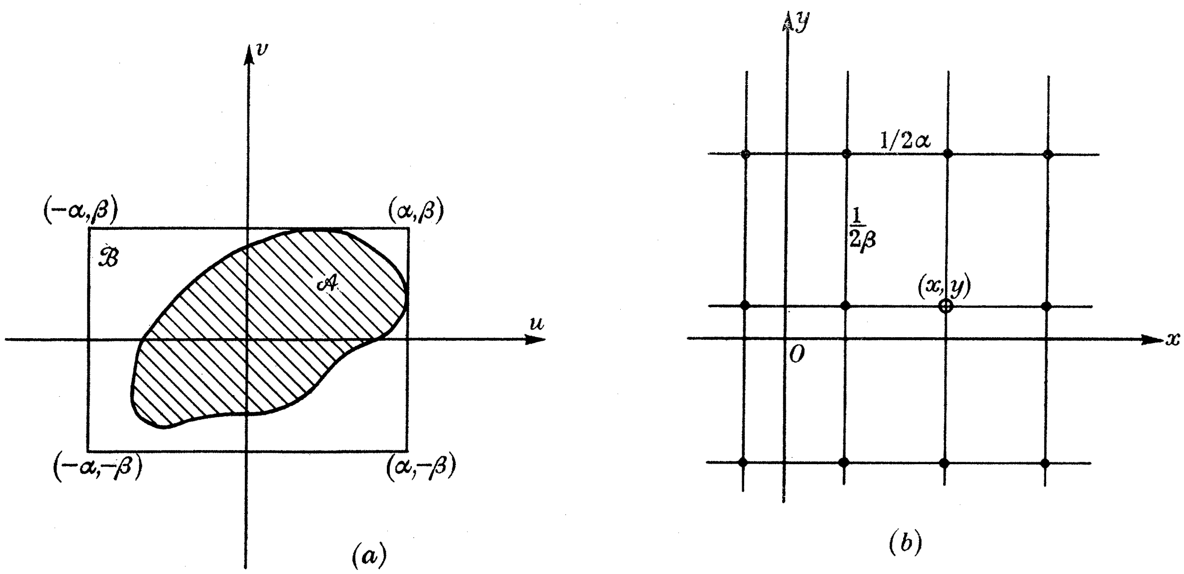
3·11. *Distinguishable and indistinguishable images*

Let the aperture $\mathscr{A}$ of an optical system be enclosed by a rectangle $\mathscr{B}$ with sides parallel to the (u, v)-axes and corner points $(u, v) = (\pm\alpha, \pm\beta)$, as shown in figure 3*a*. If a point object situated at (x, y) be imaged without aberrations through the rectangle $\mathscr{B}$, its diffraction pattern is defined by the (normalized) complex displacement

$$g_0(x', y'; x, y) = \operatorname{sinc}[2\alpha(x'-x)] \operatorname{sinc}[2\beta(y'-y)], \tag{3·1}$$

where sinc t stands for $\dfrac{\sin \pi t}{\pi t}$.† The corresponding intensity distribution is

$$w_0(x', y'; x, y) = \operatorname{sinc}^2[2\alpha(x'-x)] \operatorname{sinc}^2[2\beta(y'-y)]. \tag{3·2}$$



(*a*) The (u, v)-plane can be identified both with (u'', v'')-space in the exit pupil and with frequency-space in the object and image surfaces.

(*b*) The (x, y)-plane provides a map of both the object and image surfaces.

FIGURE 3

It will be seen that g_0, w_0 vanish at all the points

$$x' = x + \frac{r}{2\alpha}, \quad y' = y + \frac{s}{2\beta} \quad (r, s \text{ integers}) \tag{3·3}$$

for which $(r, s) \neq (0, 0)$, while for $(r, s) = (0, 0)$ they have the value 1. These points mark the corners of the lattice rectangles in figure 3*b*.

Now let $E(x, y)$ be the phase-amplitude distribution in a coherent object filling the region $A\,(-a \leqq x \leqq a,\ -b \leqq y \leqq b)$ in the object surface, and let $D(x, y)$ be its image through the aperture $\mathscr{B}$. We take 'sampling points'

$$P_{rs} = (x_r, y_s) = \left(\frac{r}{2\alpha}, \frac{s}{2\beta}\right) \tag{3·4}$$

in the image surface and denote by D_{rs} the value of $D(x, y)$ at P_{rs}.

† This notation is due to Woodward (1953).

By the two-dimensional form of Shannon's sampling theorem,† any function $f(x, y)$ which is frequency-limited to the rectangle $\mathscr{B}$ $(-\alpha \leqq u \leqq \alpha,\ -\beta \leqq v \leqq \beta)$ satisfies the identity

$$f(x,y) = \sum_{rs} f\left(\frac{r}{2\alpha}, \frac{s}{2\beta}\right) \operatorname{sinc}\left[2\alpha\left(x - \frac{r}{2\alpha}\right)\right] \operatorname{sinc}\left[2\beta\left(y - \frac{s}{2\beta}\right)\right]. \tag{3·5}$$

$D(x, y)$ is such a function; therefore by (3·1), (3·4) and (3·5)

$$D(x,y) = \sum_{rs} D_{rs} \operatorname{sinc}(2\alpha x - r) \operatorname{sinc}(2\beta y - s) \tag{3·6}$$

for all values of x, y. Interpreted physically, (3·6) expresses $D(x, y)$ as a sum of the image displacements corresponding to a rectangular lattice of independent point sources in the object surface.

When P_{rs} is outside A, D_{rs} is very small and consequently $D(x, y)$ is characterized to a good approximation by its values at the finite set of $|A|\,|\mathscr{B}|$ sampling points lying in the region A. It can be verified that the Fourier transform $\hat{r}\hat{\epsilon}$ of $D(x, y)$ is effectively frequency-limited to A in the sense that the 'overspill'

$$\left(\iint_{-\infty}^{\infty} - \iint_{A}\right) |F^*[\hat{r}\hat{\epsilon}]|^2 \, dx\, dy \tag{3·601}$$

is small. From this it results that $\hat{r}\hat{\epsilon}$ is sufficiently well characterized by the finite set of $|A|\,|\mathscr{B}|$ sampling points

$$Q_{pq} = (u_p, v_q) = \left(\frac{p}{2a}, \frac{q}{2b}\right) \ (p, q \text{ integers}) \tag{3·602}$$

which lie in the region $\mathscr{B}$ of the frequency plane, the approximation being of the same order as that in the representation of $D(x, y)$ by the function

$$D_{(A)}(x,y) = \sum_{P_{rs} \subset A} D_{rs} \operatorname{sinc}(2\alpha x - r) \operatorname{sinc}(2\beta y - s). \tag{3·603}$$

How small the expression (3·603) can be depends on the areas A, $\mathscr{B}$; more particularly, it tends to zero as the product $d_A d_{\mathscr{B}}$ of the diameters of the largest circular areas contained respectively in A and $\mathscr{B}$ tends to infinity, and this point is taken up later (§§ 3·22, 3·23).

Because the accuracy with which $D(x, y)$ can be measured is limited by noise, the statistical effect on $D(x, y)$ of replacing the sampling values D_{rs} outside A by zero is negligibly small if d_A is sufficiently large compared with the size of the resolution limit of the system.

Suppose now that $\{D(x, y)\}$ is a statistical set of functions, frequency-limited to $\mathscr{B}$ and effectively limited to A, which represents a set of images together with noise. We may call $D(x, y)$ and $D_{(A)}(x, y)$ *effectively indistinguishable* from each other when

$$d^2(D_{(A)}, D) = \iint_{-\infty}^{\infty} |D_{(A)} - D|^2 \, dx\, dy \ll \iint_{-\infty}^{\infty} \overline{|D - D_0|^2} = \overline{N},$$

where bars denote statistical means over the full set $\{D(x, y)\}$, D_0 denotes the function D as it would be without the noise and $\overline{N}$ is the statistical mean total noise power. The statistical mean

$$\overline{d^2(D_{(A)}, D)} = \overline{\iint_{-\infty}^{\infty} |D_{(A)} - D|^2 \, dx\, dy} = \iint_{-\infty}^{\infty} \overline{|D_{(A)} - D|^2} \, dx\, dy = \frac{1}{|\mathscr{B}|} \sum_{P_{rs} \not\subset A} \overline{|D_{rs}|^2}.$$

† Shannon (1948, theorem 13). A proof of the form used here is given incidentally in § 3·21 below.

Because D is effectively limited to the image patch A (in the sense of § 3·22 below), this sum is small compared with the statistical mean noise power $\overline{N}$ provided $d_A d_{\mathscr{B}}$ is sufficiently large compared with unity. In these circumstances we may replace the statistical set $\{D(x, y)\}$ by a new statistical set $\{D_{(A)}(x, y)\}$, obtained by grouping together those members of $\{D(x, y)\}$ which have identical sampling values D_{rs} in A and then replacing each group by the corresponding $D_{(A)}(x, y)$ with the appropriate probability weighting. The sampling points P_{rs} in A are $|A||\mathscr{B}|$ in number, where $|A| = 4ab$ and $|\mathscr{B}| = 4\alpha\beta$ are the areas of $A, \mathscr{B}$ in (x, y)-units and (u, v)-units respectively, and the final conclusion is that, provided $d_A d_{\mathscr{B}}$ is large enough, a set of $|A||\mathscr{B}|$ parameters D_{rs} with known joint probability distribution may replace the original set $\{D(x, y)\}$ for the purpose of calculating the effective number of distinguishable images contained in the set.

3·12. *Information in an optical image*

Shannon has developed a quantitative theory of information which is in strikingly close agreement with ordinary intuitive notions and has shown that only one mathematical definition of quantity of information is possible without violating the conditions which these notions impose. In the case where the effect of an observation is to single out one from a number N of distinguishable states, all equally probable *a priori*, the amount h of information derived from the observation is given by the equation

$$h = \log N. \tag{3·7}$$

Changing the base of the logarithm introduces a multiplicative factor; to take the logarithm to base 2 is equivalent to choosing as unit of information the binary unit or 'bit', which is the amount gained in an observation which results in the selection of one from two equally likely possibilities. For example, the answer to a 'yes or no' question provides one bit of information if (and only if) neither answer was more to be expected than the other.

Information is additive only when the result of a first observation leaves the expected result of a second unchanged; being told what one already knows adds little information. Thus, repetition of a measurement only adds information in so far as it reduces inaccuracy. In this case, the second measurement gives less information than the first, but if the first one is of very low accuracy the difference is small, because the first measurement hardly changes the expected result of the second.

Suppose now that, at each sampling point (3·4) in A, only m discrete values of D_{rs} can be distinguished. It will appear in § 3·13 that a situation essentially equivalent to this results from the limitation on accuracy of measurement which necessarily follows from the presence of noise. Then the number of distinguishable image states is

$$n = m^{|A||\mathscr{B}|},$$

and if all these states can be regarded as independent and equally probable, the quantity of information in the image is, on Shannon's definition,

$$H' = |A||\mathscr{B}| \log m. \tag{3·8}$$

For the images formed by the system with aperture $\mathscr{A}$, smaller than $\mathscr{B}$, the representation (3·6) is still valid, but it is no longer possible to regard the D_{rs} as completely independent;

there is now a certain correlation between the values of different D_{rs}, so that a knowledge of some of them produces an expectation of the values of the rest. We might expect this to reduce the amount of information in the image; it will appear later (see § 3·2) that the amount is in fact reduced to

$$H = |A|\,|\mathscr{A}| \log m. \tag{3·9}$$

Aberrations, which broaden the diffraction patterns of point sources by $\mathscr{A}$, may strengthen the correlations between the sampling values in the image; they do so in the image of an incoherent object, and this strengthening of correlations reduces still further the amount of information; but in the image of a coherent object the aberrations do not increase the average correlation between different parts of the image $\hat{D}(x, y)$, and the amount of information is left unchanged (see § 3·231). Gabor's well-known method of image reconstruction† can be related to this last fact, since it depends on the harmlessness, so far as information is concerned, of the large amount of image-spread produced by strong defocusing of a coherent image.

3·13. *Distinguishable levels at a sampling point*

When a light-sensitive receiver is used to measure the brightness level at a sampling point in the image of an incoherent object it may, of course, indicate any brightness level between zero and its saturation value; but if two 'readings' are very nearly equal the difference between them may cease to have any appreciable significance in view of the inaccuracies of measurement induced by the noise which, as we have seen, must always be present.

An idea of the consequences of this may be gained by considering the particular case where the noise (defined as the actual intensity *minus* the statistical mean intensity) has, at the sampling point P, an effectively Gaussian probability distribution with a given mean-square value n^2. If what was known before the experiment indicated a Gaussian probability distribution at P for the 'signal' (in this case the actual intensity) with mean-square deviation s^2, then a result of Shannon (1948, p. 63, theorem 17) shows that the (mean) information gain‡ on making the measurement at P is

$$h = \tfrac{1}{2} \log \frac{s^2 + n^2}{n^2}.$$

Comparison with (3·7) shows that this is the same as the information gain when the effect of the observation is to select one out of m equally likely discrete values or 'levels', where

$$m = \left(\frac{s^2 + n^2}{n^2}\right)^{\frac{1}{2}}. \tag{3·10}$$

This result is intuitively acceptable, since it implies that the single measurement has reduced the uncertainty in the 'signal' value by a factor whose statistical average tends to s/n when the noise n is small compared with s and to 1 when n/s is large.

† Gabor (1949).
‡ When random processes are involved, so that to a given measured 'signal' there does not correspond one unique intensity in the object, the information gain on making a single measurement must itself be regarded as subject to statistical fluctuations. When one speaks of information gain in these circumstances, it is the statistical mean information gain which is ordinarily meant (cf. Woodward 1953, p. 54).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

A similar, though not identical, result holds for the distinguishable complex values of $\hat{D}(x, y)$ at a sampling point of a coherent image. In this case

$$m = \frac{s^2 + n^2}{n^2}. \tag{3·11}$$

For the immediate purpose, the important point is that in both cases the value of m at a sampling point is finite in any realizable experiment.

If m is the same at every sampling point P_{rs} in the image patch A, then the amount H of information in the image is $\log m$ multiplied by the number of independent parameters required to describe the image values at the sampling points.

3·2. *Noisy objects and their images*

3·21. *Analytical representation of optical noise*

Inasmuch as noise represents unpredictability from the point of view of the observer, it is represented mathematically by a statistical set† of functions having the maximum unpredictability (in the sense of maximum entropy) that is possible under the constraints which represent the observer's previous knowledge.

When all that the observer knows is that the noise originated in the object patch A and has reached the image surface through the optical system, the entropy-maximizing problem can be attacked, and a solution obtained in certain useful cases, by an application of the sampling theorem.

A function $f(x, y)$, confined to a rectangular area A in the (x, y)-plane with corner points $(x, y) = (\pm a, \pm b)$ and of integrable square over this area, possesses in A the Fourier expansion

$$\sum_{Q_{pq}} \alpha_{pq} \exp\left\{2\pi \mathrm{i}\left(\frac{px}{2a} + \frac{qy}{2b}\right)\right\}, \tag{3·12}$$

where Q_{pq} denotes the point $(u_p, v_q) = \left(\frac{p}{2a}, \frac{q}{2b}\right)$ in the frequency plane. The Fourier coefficient α_{pq} is given in terms of the Fourier transform

$$\epsilon(u, v) = F[f] = \iint_A \mathrm{e}^{-2\pi\mathrm{i}(ux+vy)} f(x, y)\, \mathrm{d}x\, \mathrm{d}y \tag{3·13}$$

by means of the equation

$$\alpha_{pq} = \frac{1}{|A|} \epsilon_{pq}, \tag{3·14}$$

where ϵ_{pq} stands for $\epsilon(u_p, v_q)$, the value of $\epsilon(u, v)$ at the 'sampling point' Q_{pq} in the frequency plane.

It follows that when $f(x, y)$ is confined to A we can write

$$f(x, y) = \sum_{Q_{pq}} \frac{1}{|A|} \epsilon_{pq} \exp\left\{2\pi \mathrm{i}\left(\frac{px}{2a} + \frac{qy}{2b}\right)\right\}. \tag{3·15}$$

Then

$$\begin{aligned} \epsilon(u, v) &= \iint_A \mathrm{e}^{-2\pi\mathrm{i}(ux+vy)} f(x, y)\, \mathrm{d}x\, \mathrm{d}y \\ &= \sum_{Q_{pq}} \epsilon_{pq}\, |A| \iint_A \exp\left\{-2\pi\mathrm{i}\left[\left(u - \frac{p}{2a}\right)x + \left(v - \frac{q}{2b}\right)y\right]\right\} \mathrm{d}x\, \mathrm{d}y \\ &= \sum_{Q_{pq}} \epsilon_{pq} \operatorname{sinc}(2au - p) \operatorname{sinc}(2bv - q). \end{aligned} \tag{3·16}$$

† This term has the same meaning as 'ensemble' in Shannon's paper (1948).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

If A has corner points $(x_0 \pm a, y_0 \pm b)$, $\epsilon(u, v)$ is $e^{-2\pi i(ux_0+vy_0)}$ times the right-hand side of (3·16) and each ϵ_{pq} is multiplied by $e^{-2\pi i(u_p x_0 + v_q y_0)}$.

Equation (3·16) establishes Shannon's sampling theorem for the function $\epsilon(u, v)$, which by Fourier's inversion theorem is frequency-limited to A. It asserts that when $f(x, y)$ is confined to A, $\epsilon(u, v) = F[f]$ has the form of an interpolation function and is completely determined by its values ϵ_{pq} at the 'sampling points' Q_{pq} corresponding to A.

By Parseval's theorem and (3·16)

$$\iint_A |f(x,y)|^2 \, dx \, dy = \iint_{-\infty}^{\infty} |\epsilon(u,v)|^2 \, du \, dv = \frac{1}{|A|} \sum_{pq} |\epsilon_{pq}|^2. \tag{3·17}$$

The condition that f shall be real is equivalent to the condition

$$\epsilon(-u, -v) = \epsilon^*(u, v), \tag{3·18}$$

or again, by (3·16) and the orthogality of the functions sinc $(2au-p)$ sinc $(2bv-q)$, to the condition

$$\epsilon_{-p,-q} = \epsilon^*_{p,q} \quad (-\infty < p, q < \infty). \tag{3·19}$$

In the case of an incoherent object we identify f with the intensity distribution $\sigma(x, y)$; then f is everywhere real and (3·18) holds. In the case of a coherent object we identify f with the complex displacement function $\hat{E}(x, y)$, and $\epsilon(u, v) = F[\hat{E}]$ is no longer restricted by (3·18). Partially coherent objects lie outside the scope of the present paper. There is no simple property of $\epsilon(u, v)$ which corresponds, in the incoherent case, to the condition $\sigma \geqslant 0$.

Coherent case. A statistical set of objects confined to A is represented by a set $\{f(x, y)\}$ of functions confined to A, together with a probability-density function defined in some way over the set.

As is well known from the theory of the microscope, spatial frequencies finer than $\frac{1}{2}\lambda$ in a (coherent or incoherent) object are not transmitted through an optical system. This means that the images of two object distributions $f(x, y)$, confined to A, whose ϵ_{pq} agree at all the points Q_{pq} inside a frequency rectangle $\mathfrak{M}$ with corner points $(\pm\alpha, \pm\beta)$, where $\alpha > 2/\lambda$, $\beta > 2/\lambda$, can be regarded as indistinguishable from each other and may therefore be represented in the analysis by the same image function. Thus the members $f(x, y)$ of a statistical set $\{f(x, y)\}$ of objects confined to A may be collected into subsets, each subset consisting of those f for which the ϵ_{pq} in $\mathfrak{M}$ have identical values, and the images of all the members of a subset may be regarded as indistinguishable. It is convenient to make the $\mathfrak{M}$ corresponding to a given A unique by defining α, β as the least numbers greater than $2/\lambda$ for which $4a\alpha$, $4b\beta$ are odd integers.

Each subset contains just one member for which $\epsilon_{pq} = 0$ at all Q_{pq} outside $\mathfrak{M}$; let $f^{(\mathfrak{M})}(x, y)$ denote this member. Then

$$f^{(\mathfrak{M})}(x,y) = \sum_{Q_{pq} \subset \mathfrak{M}} \frac{1}{|A|} \epsilon_{pq} \exp\left\{2\pi i\left(\frac{px}{2a} + \frac{qy}{2b}\right)\right\} \quad (x,y) \subset A,$$
$$= 0 \quad (x,y) \not\subset A \tag{3·20}$$

and the function

$$\epsilon^{(\mathfrak{M})}(u,v) = F[f^{(\mathfrak{M})}] = \sum_{Q_{pq} \subset \mathfrak{M}} \epsilon_{pq} \operatorname{sinc}(2au-p) \operatorname{sinc}(2bv-q). \tag{3·21}$$

If we group together all the $f(x, y)$ which correspond to each $f^{(\mathfrak{M})}(x, y)$, the original statistical distribution gives rise to a new probability-density distribution over the functions $f^{(\mathfrak{M})}(x, y)$ and the set $\{f^{(\mathfrak{M})}\}$ is likewise a statistical set. By virtue of (3·20), this last probability density 'induces' a joint probability among the parameters ϵ_{pq} (which are $|A||\mathfrak{M}|$ in number) and conversely. Because all the members of a subset yield images which are indistinguishable from each other, the set of images of the functions $f^{(\mathfrak{M})}(x, y)$ with their 'induced' probability density is statistically indistinguishable from the set of images of the functions $f(x, y)$ with the original probability density.

For every joint probability density of the ϵ_{pq} at the points Q_{pq} in $\mathfrak{M}$, there exists a statistical set of complex functions $f(x, y)$, confined to A, such that the corresponding set $\{f^{(\mathfrak{M})}(x, y)\}$ has its ϵ_{pq} distributed in accordance with the given probability density. (In fact $\{f^{(\mathfrak{M})}(x, y)\}$ is itself such a set.) The infinite set of ϵ_{pq} can be interpreted as the co-ordinates of a point in complex space of (exactly) $|A||\mathfrak{M}|$ dimensions, or their real and imaginary parts can be intepreted as the co-ordinates of a point in real space of $2|A||\mathfrak{M}|$ dimensions, and the continuous entropy H of the set $\{f^{(\mathfrak{M})}(x, y)\}$ relative to this latter co-ordinate system is changed only by an additive constant when the co-ordinates are transformed linearly.† It follows that, whether the entropy of the set $\{f^{(\mathfrak{M})}(x, y)\}$ is taken relative to the above real co-ordinates or to a linearly transformed set, it will be maximized, under variation of the probability-density distribution within the constraints which express the observer's prior knowledge, when the parameters ϵ_{pq} $(Q_{pq} \subset \mathfrak{M})$ have maximum joint entropy under these constraints.

When its entropy is thus maximized, the set $\{f^{(\mathfrak{M})}(x, y)\}$ is said to be random under the given constraints, and to represent *noise*.

Two particular cases are of special interest here. In the first, the constraints consist in a prescribed mean spectral power distribution for $\{f^{(\mathfrak{M})}(x, y)\}$, in the sense of prescribed values $\phi_{pq} > 0$ for the $|A||\mathfrak{M}|$ statistical means $\overline{|\epsilon_{pq}|^2} = \overline{|\epsilon^{(\mathfrak{M})}(u_p, v_q)|^2}$. Then it can be shown that the entropy of $\{f^{(\mathfrak{M})}(x, y)\}$ is maximized when

(i) the ϵ_{pq} in $\mathfrak{M}$ are statistically independent;

(ii) each ϵ_{pq} in $\mathfrak{M}$ has a Gaussian probability distribution with parameter

$$\xi_{pq} = +(\phi_{pq})^{\frac{1}{2}}.$$

It then follows from (3·20) that the statistical mean

$$\frac{1}{|G|}\overline{\iint_G |f^{(\mathfrak{M})}(x, y)|^2 \,\mathrm{d}x\,\mathrm{d}y} \simeq \frac{1}{|A|^2}\sum_{Q_{pq} \subset \mathfrak{M}} \overline{|\epsilon_{pq}|^2} \tag{3·22}$$

is approximately constant for all domains G in A which contain a circular area of diameter sufficiently large compared with λ.

In these circumstances we say that the set $\{f^{(\mathfrak{M})}(x, y)\}$ represents Gaussian noise, uniform over A and of prescribed power spectrum. The corresponding probability distribution

$$p(\epsilon) = \prod_{Q_{pq} \subset \mathfrak{M}} \frac{1}{2\pi\xi_{pq}^2} \exp\left(\frac{-|\epsilon_{pq}|^2}{2\xi_{pq}^2}\right) \tag{3·23}$$

represents the greatest randomness possible to the set under the given constraints.

† Shannon (1948, p. 55). The property holds for any real co-ordinate transformation with a constant Jacobian.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

The joint entropy of the ϵ_{pq} is then

$$H = -\int p(\epsilon) \log p(\epsilon) \, \mathrm{d}V, \tag{3·24}$$

where the $2|A||\mathfrak{M}|$-dimensional volume element

$$\mathrm{d}V = \prod_{Q_{pq} \subset \mathfrak{M}} \mathrm{d}\eta_{pq} \mathrm{d}\zeta_{pq} \quad (\eta_{pq}, \zeta_{pq} \text{ real}; \ \epsilon_{pq} = \eta_{pq} + \mathrm{i}\zeta_{pq}).$$

By (3·23), H is equal to the sum of the expressions

$$h_{pq} = -\int p(\epsilon) \log \left[\frac{1}{2\pi\xi_{pq}^2} \exp\left(\frac{-|\epsilon_{pq}|^2}{2\xi_{pq}^2}\right)\right] \mathrm{d}V$$

$$= -\iint_{-\infty}^{\infty} \frac{1}{2\pi\xi_{pq}^2} \exp\left(\frac{-\eta^2-\zeta^2}{2\xi_{pq}^2}\right) \log\left[\frac{1}{2\pi\xi_{pq}^2} \exp\left(\frac{-\eta^2-\zeta^2}{2\xi_{pq}^2}\right)\right] \mathrm{d}\eta \, \mathrm{d}\zeta, \tag{3·25}$$

where $\eta + \mathrm{i}\zeta = \epsilon_{pq}$

$$= \log(2\pi e \xi_{pq}^2). \tag{3·26}$$

It will be seen from (3·25) that h_{pq} is also the entropy of the complex value distribution of the single parameter ϵ_{pq}. Thus H is the sum of the separate entropies at the sampling points Q_{pq}. By (3·26)

$$H = \sum_{Q_{pq} \subset \mathfrak{M}} \log \xi_{pq}^2 + |A||\mathfrak{M}| \log 2\pi e, \tag{3·27}$$

and the *entropy per sampling point* is

$$\frac{1}{|A||\mathfrak{M}|} \sum_{Q_{pq} \subset \mathfrak{M}} \log \phi_{pq} + \log 2\pi e. \tag{3·28}$$

The last expression can be written in the approximate form

$$\frac{1}{|\mathfrak{M}|} \iint_{\mathfrak{M}} \log \phi(u, v) \, \mathrm{d}u \, \mathrm{d}v + \log 2\pi e$$

when there exists a smooth function $\phi(u, v) > 0$ which agrees with ϕ_{pq} at each point Q_{pq}.

In the second particular case, the only constraint is that the statistical mean power

$$\overline{P} = \iint_A \overline{|f^{(\mathfrak{M})}(x, y)|^2} \, \mathrm{d}x \, \mathrm{d}y \tag{3·29}$$

has a prescribed value. Then it can be shown that the most random (that is, entropy-maximized) set $\{f^{(\mathfrak{M})}(x, y)\}$ is the one in which

(i) the ϵ_{pq} in $\mathfrak{M}$ are statistically independent;

(ii) each ϵ_{pq} in $\mathfrak{M}$ has a Gaussian probability distribution with parameter

$$\xi_{pq} = \left(\frac{\overline{P}}{|\mathfrak{M}|}\right)^{\frac{1}{2}}.$$

It follows from (ii) that the spectral power is equally distributed among the different frequencies (u_p, v_q) in $\mathfrak{M}$ (not all of which can be transmitted by an optical system); thus we have here a special type of Gaussian noise, called 'uniform Gaussian noise', in which the power is uniformly distributed† over the rectangle $\mathfrak{M}$ in the frequency plane, as well as over the rectangle A in the (x, y)-plane.

† In the sense that the 'sampling values' $\overline{|\epsilon(u_p, v_q)|^2}$ in $\mathfrak{M}$ are all equal.

MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES
PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A

From (3·20) and (3·23) it follows that the distribution of values of $f^{(\mathfrak{M})}(x,y)$ is Gaussian, with the parameter $\xi = (\bar{P}/|A|)^{\frac{1}{2}}$, at every point of A and in particular at those points $P_{rs} = \left(\frac{r}{2\alpha}, \frac{s}{2\beta}\right)$ which lie in A. In the more special second case (uniform Gaussian noise) it can be shown that the distributions at the different points P_{rs} are statistically independent. In the first case (Gaussian noise of prescribed spectral power distribution) there will in general be correlations between the distributions at different P_{rs}.

Because of the manner in which $\mathfrak{M}$ was chosen, there are exactly $|A||\mathfrak{M}|$ of the points $P_{rs} = \left(\frac{r}{2\alpha}, \frac{s}{2\beta}\right)$ inside A. By (3·20), the values of $f^{(\mathfrak{M})}(x, y)$ at these points, namely the numbers

$$f_{rs} = f^{(\mathfrak{M})}\left(\frac{r}{2\alpha}, \frac{s}{2\beta}\right) \quad (P_{rs} \subset A), \tag{3·30}$$

are connected with the $|A||\mathfrak{M}|$ numbers ϵ_{pq} by the equations

$$f_{rs} = \sum_{Q_{pq} \subset \mathfrak{M}} \frac{1}{|A|} \epsilon_{pq} \exp\left\{2\pi i\left(\frac{pr}{4a\alpha} + \frac{qs}{4b\beta}\right)\right\}. \tag{3·31}$$

Thus the parameters f_{rs} are a linear transformation of the ϵ_{pq}, and the determinant

$$\left\| \exp\left\{2\pi i\left(\frac{pr}{4a\alpha} + \frac{qs}{4b\beta}\right)\right\} \right\| \tag{3·32}$$

of this transformation can be shown to be non-vanishing.

It follows that the $|A||\mathfrak{M}|$ 'sampling values' $f_{rs} = k_{rs} + il_{rs}$ form another set of parameters which uniquely define the statistical set $\{f^{(\mathfrak{M})}\}$, and that the entropy of this set relative to the system of $2|A||\mathfrak{M}|$ real co-ordinates $k_{rs}, l_{rs} (P_{rs} \subset A)$ differs only by an additive constant from its entropy relative to the $2|A||\mathfrak{M}|$ real co-ordinates η_{pq}, ζ_{pq} $(Q_{pq} \subset \mathfrak{M})$. Under given constraints, entropy maximizations of $\{f^{(\mathfrak{M})}(x,y)\}$ with respect to the 'sampling co-ordinates' k_{rs}, l_{rs} in A and with respect to the 'sampling co-ordinates' η_{pq}, ζ_{pq} in $\mathfrak{M}$ therefore lead to the same random set.

Incoherent case. In the incoherent case, the full set of η_{pq}, ζ_{pq} in $\mathfrak{M}$ no longer forms a suitable co-ordinate system, because in that system the conditions (3·19) are equivalent to an infinite amount of information.

More explicitly: the entropy per degree of freedom of a set S relative to a chosen co-ordinate system measures, on a logarithmic scale, the randomness of S relative to that of a comparison set S_0 which is uniformly distributed over a unit volume in the corresponding co-ordinate space. In the full (η_{pq}, ζ_{pq})-space, the conditions $\epsilon_{-p,-q} = \epsilon^*_{p,q}$ make the randomness of $\{f^{(\mathfrak{M})}\}$ infinitely small compared with that of the implied comparison set, and hence they make the entropy negatively infinite. To discuss entropy maximization we need a real co-ordinate system such that the implied comparison system S_0 has a randomness comparable with that of S itself, and the entropy of S relative to this coordinate system is consequently finite. Such a system is provided by those η_{pq}, ζ_{pq} in $\mathfrak{M}$ for which $q \geqslant 0$, and the statistical properties of the set $\{f^{(\mathfrak{M})}\}$ are then expressed by a probability-density distribution in complex space of $4a\alpha(2b\beta - \frac{1}{2})$ dimensions, or in real (η_{pq}, ζ_{pq})-space of dimensionality $4a\alpha(4b\beta - 1) \simeq |A||\mathfrak{M}|$.

Setting $f(x,y) = \sigma(x,y)$, $f^{(\mathfrak{M})}(x,y) = \sigma^{(\mathfrak{M})}(x,y)$ and disregarding for the present the condition $\sigma \geqslant 0$, we find that a random object set $\{\sigma^{(\mathfrak{M})}(x,y)\}$, confined to A and with prescribed mean power distribution $\phi_{pq} > 0$ over the points $Q_{pq} \subset \mathfrak{M}$ with $q \geqslant 0$, is represented by the probability distribution

$$p(\epsilon) = \prod_{Q_{pq} \subset \mathfrak{M},\, q \geqslant 0} \frac{1}{2\pi\phi_{pq}} \exp\left(\frac{-|\epsilon_{pq}|^2}{2\phi_{pq}}\right), \tag{3·33}$$

together with the equations $\epsilon_{-p,-q} = \epsilon^*_{p,q}$.

At every point (x, y) in A, the distribution of the (real) values of

$$\sigma^{(\mathfrak{M})}(x, y) = \sum_{Q_{pq} \subset \mathfrak{M}} \frac{1}{|A|} \epsilon_{pq} \exp\left\{2\pi i\left(\frac{px}{2a} + \frac{qy}{2b}\right)\right\} \tag{3·34}$$

is again Gaussian with the parameter $\xi = (\bar{P}/|A|)^{\frac{1}{2}}$; in particular, this holds at the $|A|\,|\mathfrak{M}|$ sampling points P_{rs}.

3·22. *Noisy objects*

As already stated, noise is represented mathematically by a statistical set of functions having the maximum unpredictability (in the sense of maximum entropy) that is possible under the constraints which represent the observer's previous knowledge.

We assume that all the observer knows in advance is that the noise originated in the object patch A and has reached the image surface through the optical system, and attack the entropy-maximizing problem by an application of the sampling theorem essentially similar to that already made in § 3·21. For the sake of brevity, only incoherent objects are considered here; the discussion of coherent objects is on essentially the same lines (see § 3·4).

We collect the members σ of the object set $\{\sigma\}$ into subsets, placing in a single subset those which yield indistinguishable images through an optical system of aperture $\mathscr{A}$. All that observations of an image formed by this system can tell us is the subset to which the corresponding object belongs. A subset consists of all σ with the same 'τ_0 cut-off' function σ^{τ_0}; we can use σ^{τ_0} to represent the subset.

Physically, two objects σ belong to the same subset if the transmissible parts of their spectral functions ϵ are identical. The function σ^{τ_0} which represents the subset is not strictly confined to A, as were the original functions σ, but it is everywhere $\geqslant 0$, and it is effectively confined to A in the sense that its values outside A are very small except close to the boundary of A. The error in treating σ^{τ_0} as though it were strictly confined to A is identical with that in treating the image of σ by an aberration-free system of aperture $\mathscr{A}$ as though it were strictly confined to A, as may be seen by writing equations (2·50), (2·37) in the forms

$$\sigma^{\tau_0} = F^*[\tau_0 \epsilon] = \sigma_* F^*[\tau_0], \quad I = F^*[\tau\epsilon] = \sigma_* F^*[\tau], \tag{3·35}$$

where $f_* g$ stands for the convolution $\iint_{-\infty}^{\infty} f(x', y')\, g(x-x', y-y')\, dx'\, dy'$.

Because $F^*[\tau_0]$ is nearly a δ-function when the greatest circular area contained in the region $\mathscr{F}$ is large (it is in fact the intensity distribution w in the image of a point source by the ideal aberration-free system of aperture $\mathscr{A}$), it follows that the convolution σ^{τ_0} of σ with $F^*[\tau_0]$ is, under the same conditions, effectively confined to the rectangle A in the object plane. The 'overspill' is harmlessly small for present purposes when the product of the diameters $d_{\mathscr{F}}$, d_A of the largest circular areas contained in $\mathscr{F}$ and in A respectively is large compared with 1.

Now consider the functions $\{n^{\tau_0}(x,y)\}$, where

$$n^{\tau_0} = \sigma^{\tau_0} - \sigma_0^{\tau_0}, \tag{3·36}$$

while σ^{τ_0}, ${\sigma_0}^{\tau_0}$ are the τ_0 cut-offs respectively of the object and of the object as it would be without the noise. They present in a convenient form all that can be learned about the object noise by observation through the optical system. With the same proviso as above, n^{τ_0} can also be regarded as effectively confined to the rectangle A. In determining the entropy-maximizing probability density in the set $\{n^{\tau_0}\}$ which is to represent greatest randomness of the observable noise, we may with only harmless inaccuracy replace each n^{τ_0} by its 'A cut-off' $n_A^{\tau_0} = \kappa_A n^{\tau_0}$ and proceed instead to maximize the entropy of the noise set $\{n_A^{\tau_0}\}$.

Because the functions $n_A^{\tau_0}$ of this last set are strictly confined to A, their transforms $\nu_A^{\tau_0} = F[n_A^{\tau_0}]$ satisfy the identical relationship

$$\nu_A^{\tau_0}(u,v) = \sum_{pq} \nu_{pq} \operatorname{sinc}(2au-p) \operatorname{sinc}(2bv-q) \tag{3·37}$$

obtained from (3·20) on giving f in § 3·21 the particular value $n_A^{\tau_0}$, and the probability distribution in the set can be expressed as a probability-density function

$$p(\nu) = p(\nu_{00}, \nu_{10}, \nu_{01}, \nu_{-1,0}, \nu_{0,-1}, \ldots)$$

of the variables ν_{pq}, which are in fact the values of $\nu_A^{\tau_0}$ at the 'sampling points' $Q_{pq} = \left(\dfrac{p}{2a}, \dfrac{q}{2b}\right)$.
From the identity

$$\nu_A^{\tau_0} = F[\kappa_A n^{\tau_0}] = F[\kappa_A] * (\tau_0 F[\sigma - \sigma_0]) \tag{3·38}$$

we see that ν^{τ_0} is the convolution of a function $F[\kappa_A]$ which is nearly a δ-function (if A is large compared with the resolution limit corresponding to the aperture of the system) and a function $\tau_0 F[\sigma-\sigma_0]$ which is strictly confined to $\mathscr{F}$. Thus the number of variables ν_{pq} which can differ appreciably from zero is equal, with only small percentage error, to the number of sampling points Q_{pq} which lie in $\mathscr{F}$; that is to say, it is approximately equal to $|A||\mathscr{F}|$.

By an extension of the arguments of § 3·21 we now infer that, if the product of the diameters d_A, $d_{\mathscr{F}}$ of the greatest circular areas contained respectively in A and in $\mathscr{F}$ is large compared with unity, the entropies of the noise sets $\{n^{\tau_0}\}$ and $\{n_A^{\tau_0}\}$ are both approximately maximized:

(1) under the sole constraint†

$$\overline{\iint_{-\infty}^{\infty} (n^{\tau_0})^2 \, dx \, dy} = N_0 \tag{3·39}$$

by setting $\nu_{pq} = 0$ for $Q_{pq} \not\subset \mathscr{F}$, while giving, for $q \geqslant 0$, $Q_{pq} \subset \mathscr{F}$, each complex parameter $\nu_{pq} = \nu^*_{-p,-q}$ an independent Gaussian probability distribution of parameter $\xi = (N_0/|\mathscr{F}|)^{\frac{1}{2}}$. That is to say, by choosing

$$\begin{aligned} p(\nu) &= \prod_{q\geqslant 0,\, Q_{pq}\subset\mathscr{F}} \left(\frac{1}{\sqrt{(2\pi)}\,\xi}\right)^2 \exp\left(\frac{-|\nu_{pq}|^2}{2\xi^2}\right) \quad \text{if} \quad \nu_{pq} = 0 \text{ for all } Q_{pq} \not\subset \mathscr{F} \\ &= 0 \quad \text{if} \quad \nu_{pq} \neq 0 \text{ for some } Q_{pq} \not\subset \mathscr{F}; \end{aligned} \tag{3·40}$$

† I.e. without the constraint $\sigma^{\tau_0} \geqslant 0$, which σ^{τ_0} obeys because it is the convolution of the two non-negative real functions σ and $F^*[\tau_0] = |F^*[\kappa_{\mathscr{A}}]|^2$.

(2) under the sole constraints

$$\overline{|\nu_{pq}|^2} = \phi_{pq} \quad (Q_{pq} \subset \mathscr{F})$$
$$= 0 \quad (Q_{pq} \not\subset \mathscr{F}) \tag{3·41}$$

by setting $\nu_{pq} = 0$ for $Q_{pq} \not\subset \mathscr{F}$, while giving for $q \geqslant 0$, $Q_{pq} \subset \mathscr{F}$, each parameter $\nu_{pq} = \nu^*_{-p,-q}$ an independent Gaussian probability distribution of parameter $\xi_{pq} = +\sqrt{\phi_{pq}}$. That is to say, by choosing

$$p(\nu) = \prod_{q\geqslant 0,\, Q_{pq}\subset \mathscr{F}} \frac{1}{2\pi\phi_{pq}} \exp\left(\frac{-|\nu_{pq}|^2}{2\phi_{pq}}\right) \quad \text{if} \quad \nu_{pq} = 0 \text{ for all } Q_{pq} \not\subset \mathscr{F}$$
$$= 0 \quad \text{if} \quad \nu_{pq} \neq 0 \text{ for some } Q_{pq} \not\subset \mathscr{F}. \tag{3·42}$$

The probability density (3·42) or (3·40) also determines the statistical behaviour of $n^{\tau_0}(x,y) = F^*[\nu^{\tau_0}(u,v)]$ in the rectangle A of the (x,y)-plane. Suppose in particular that $\mathscr{B}$ is a rectangle containing $\mathscr{F}$ in the (u,v)-plane and with sides, parallel to the (u,v)-axes, of lengths $2\alpha, 2\beta$ respectively. Because n^{τ_0} is frequency-limited to $\mathscr{B}$ an application of the sampling theorem in the (x,y)-plane gives the identity

$$n^{\tau_0}(x,y) = \sum_{rs} n^{\tau_0}_{rs} \operatorname{sinc}(2\alpha x - r)\operatorname{sinc}(2\beta y - s), \tag{3·43}$$

where $-\infty < r, s < \infty$ and $n^{\tau_0}_{rs}$ are the values of $n^{\tau_0}(x,y)$ at the sampling points

$$P_{rs} = (x_r, y_s) = \left(\frac{r}{2\alpha}, \frac{s}{2\beta}\right) \tag{3·44}$$

corresponding to $\mathscr{B}$.

It can be shown as in § 3·21 that if $\{n^{\tau_0}\}$ is governed by the probability law (3·42), each parameter $n^{\tau_0}_{rs}$ in (3·43) for which P_{rs} lies inside A has a Gaussian probability distribution of parameter $(N_0/|A|)^{\frac{1}{2}}$; when P_{rs} lies outside A the r.m.s. values of $n^{\tau_0}_{rs}$ can be taken as zero to a sufficient approximation. However, the $|A||\mathscr{B}|$ parameters $n^{\tau_0}_{rs}$ are not in general statistically independent; they are so only in the special case where ϕ_{pq} is constant inside $\mathscr{F}$ and $\mathscr{F}$ coincides with the rectangle $\mathscr{B}$. When ϕ_{pq} is constant inside $\mathscr{F}$ the noise may be called uniform Gaussian noise, but it is important to remember that the property of statistical independence of sampling values only holds in two dimensions when the Gaussian noise is uniform over a rectangle in frequency space. In this case, the entropy H of the set $\{n^{\tau_0}\}$ can be calculated from equation (3·42), which gives

$$H = \sum_{q\geqslant 0,\, Q_{pq}\subset\mathscr{F}} \log(2\pi e \xi^2_{pq}) \tag{3·45}$$

$$\simeq \tfrac{1}{2}|A| \iint_{\mathscr{F}} \log\phi(u,v)\,\mathrm{d}u\,\mathrm{d}v + |A||\mathscr{F}| \log 2\pi e \tag{3·46}$$

if a smooth function $\phi(u,v)$ exists with $\phi(u_p, v_q) = \phi_{pq}$ for all $Q_{pq} \subset \mathscr{F}$, and the entropy per degree of freedom is then

$$\frac{1}{|\mathscr{F}|}\iint_{\mathscr{F}} \log\phi(u,v)\,\mathrm{d}u\,\mathrm{d}v + \log\sqrt{(2\pi e)}. \tag{3·47}$$

The constraint $\sigma^{\tau_0} \geqslant 0$. Since Gaussian noise $n^{\tau_0}(x,y)$ of the type (3·43) can assume (though with very small probability) arbitrarily large negative values, it follows that not all of the real functions $\sigma^{\tau_0}_0 + n^{\tau_0}$ will satisfy the condition $\sigma^{\tau_0} \geqslant 0$. However, the reduction in entropy which results from imposing the condition is negligibly small in the case of an object $\{\sigma\}$ with Gaussian noise in which the r.m.s. noise $(\overline{(n_{rs})^2})^{\frac{1}{2}}$ at each sampling point P_{rs} in A is a few times smaller than the value of $\sigma_0^{\tau_0}$ at this point. If, for example, the parameter X_{rs} of the Gauss distribution at a particular P_{rs} in A does not exceed one-third of the (positive) value of $(\sigma_0^{\tau_0})_{rs}$ there, the imposition of the constraint $\sigma_{rs}^{\tau_0} \geqslant 0$ disturbs the probability distribution at this point alone by less than

$$\frac{1}{\sqrt{(2\pi)}\,X_{rs}}\int_{-\infty}^{-3X_{rs}} \exp\left(\frac{-t^2}{2X_{rs}^2}\right)\mathrm{d}t = \frac{1}{\sqrt{\pi}}\int_{3/\sqrt{2}}^{\infty} \mathrm{e}^{-u^2}\,\mathrm{d}u = 0{\cdot}0054. \tag{3·48}$$

The fractional change in the entropy per degree of freedom of the whole set is of the same order of smallness.

In these circumstances we can still say, with only harmless inaccuracy, that random object noise effectively confined to A and strictly confined to $\mathscr{F}$, of given mean total power N_0, is described by equation (3·40); while random object noise, similarly confined, of prescribed spectral power distribution ϕ_{pq} is described by (3·42). The mean total noise power in the last case is

$$N_0 = \frac{1}{|A|}\sum_{Q_{pq}\subset\mathscr{F}} \overline{|\nu_{pq}|^2} \simeq \iint_{\mathscr{F}} \phi(u,v)\,\mathrm{d}u\,\mathrm{d}v. \tag{3·49}$$

Adequacy of the approximations. The physical meaning of the proviso that the product $d_A d_{\mathscr{F}}$ of the diameters of the largest circular areas contained in A and in $\mathscr{F}$ respectively should be large compared with 1 is that the linear dimensions of the isoplanatism-patch A should be large compared with the resolution limit corresponding to the aperture of the optical system; that is, that the aberrations of the system shall change 'slowly' over the working field. How large the product $d_A d_{\mathscr{F}}$ needs to be in order that the approximations shall be accurate enough for present purposes depends on the signal-to-noise ratio in the image; when almost all the noise originates in the object surface, it suffices if $d_A d_{\mathscr{F}}$ is large enough to ensure that the statistical mean total power

$$\left(\iint_{-\infty}^{\infty} - \iint_A\right)\overline{(\sigma^{\tau_0})^2}\,\mathrm{d}x\,\mathrm{d}y$$

which is spilled out of the isoplanatism-patch A by diffraction on replacing σ by σ^{τ_0} is small compared with the mean transmissible noise power

$$\iint_A \overline{(\sigma^{\tau_0} - \sigma_0^{\tau_0})^2}\,\mathrm{d}x\,\mathrm{d}y$$

in A. When an appreciable part of the noise originates in the image surface, a somewhat less stringent condition is sufficient (see § 3·3).

3·23. *Information content of a noisy image*

Of the noise which affects an image, that part $\{n_1\}$ which originates in the object is frequency-limited to $\mathscr{F}$ since its spectrum has been multipied by τ in passing through the optical system. To this is added the noise $\{n_2\}$ associated with the radiation detector used to observe the image. Photon noise cannot of course be fully covered by a wave theory, but its effects on information content can be discussed by including it in the detector noise. This procedure takes into account the fact that its effect on the image is not frequency-limited to $\mathscr{F}$, as might have been expected on a wave theory if it were regarded as originating in the object. The noise $\{n_2\}$ may conveniently be subdivided into noise affected in magnitude by the presence of the arriving radiation and noise not so affected. Examples of noise not so affected are dark-current in photo-electric image tubes, and the granularity of fog in a photographic emulsion. Noise dependent on the incident radiation includes the 'shot'-noise of photo-current and the granularity of an emulsion not arising from fog; the noise in sensitive detectors under normal working conditions is often predominantly of this type. Even for very good present-day detectors this noise power is usually still some 10 to 100 times greater than the photon noise in the radiation, but it is of interest that in many cases the increase is most appropriately regarded as a multiplication relative to the signal of the photon noise, rather than as an addition to it. For example, a photocell or a television camera responds only to a certain random fraction ϵ $(0<\epsilon<1)$ of the incident photons; this multiplies the signal-to-noise power ratio by ϵ (Johnson 1948), and it is the noise so generated which predominates in many practical cases. Similarly, the (logarithmically) major part of photographic noise is accounted for by a similar 'quantum efficiency' ϵ together with the fact that several effective quantum hits are needed in order to make a grain developable (Silberstein & Trivelli 1938; Webb 1950), a developable grain being analogous to a single 'count' in a photon-counting photometer. The essential point is that in these cases noise, originating in the first place from the granular quantum structure of the light, is magnified relatively to the signal by the receiver responding much as if the quantum structure of the radiation had been coarsened by an increase in the value of Planck's constant. That similar considerations apply to the eye has been pointed out by Rose (1948). Noise of this second kind will clearly be correlated with the intensity in the image surface, and noise of the first kind may become so in effect if the response to radiation is non-linear, as in photography.

No formal distinction is usually necessary between noise generated in this way and noise of different physical origin, for example, that originating in unevenness of sensitivity over a photographic plate. In a well-made emulsion, the effect of such unevenness on the interpretation of fine detail is negligible; it may be perceptible, however, in the increase of Selwyn granularity for large scanning areas. We may accordingly represent the image by a statistical set $\{I_2(x,y)\}$ of intensity distributions $I_2(x,y)\geqslant 0$, where

$$I_2(x,y) = I_1(x,y) + n_2(x,y) \tag{3·50}$$

and the function $n_2(x,y)$ represents the effect of noise originating in the image surface. The set $\{n_1\} = \{I_1 - \overline{I}_1\}$ represents the effect of noise imaged through from the object surface; both n_1 and n_2 will in general show statistical correlation with I_1. Any frequencies of n_2 which lie in the part of the (u,v)-plane outside $\mathscr{F}$ can be distinguished from the 'signal', and their effect ignored by using $n_2^{\mathscr{F}} = F^*[\kappa_{\mathscr{F}} F[n_2]]$ in place of n_2; this has the same effect as supposing from the start that $F[n_2]$ is confined to $\mathscr{F}$.

In the case where object and image, confined as before to A, are of low contrast and the total noise power is small compared with the power corresponding to the mean brightness level, the correlations between n_1, n_2 and I_1 become negligible; and in this approximation maximum prior ignorance is expressed† by supposing $\{n_1\}$ and $\{n_2\}$ both to represent

† Compare Woodward (1953), §2·8.

Gaussian noise, effectively limited to A, strictly frequency-limited to $\mathscr{F}$, and statistically independent of each other and of I_1. This case is especially closely related to the problem of the assessment of optical images, since it is in rendering fine detail of low contrast that an optical system is most severely tested. It will accordingly be taken as the basis of the subsequent discussion. A discussion which includes the effects of correlation between signal and noise has been given in the one-dimensional case by Wiener (1949).

To the optical image spread may be added a further spread attributable to properties of the light-sensitive device which receives the image. In a photographic image, the effect of photographic spread is to replace the function $I = F^*[\tau F[\sigma]]$ by the convolution $I_1 = I_* w_1$, where $w_1(x, y)$ is a spread function (near to a δ-function) characteristic of the emulsion and the development process used. We suppose w_1 normalized by the condition

$$\iint_{-\infty}^{\infty} w_1(x, y)\, \mathrm{d}x\, \mathrm{d}y = 1.$$

Then the equations

$$\begin{aligned} I_1 = I_* w_1 &= F^*[F[I] \,.\, \tau_1] \\ &= F^*[\tau\tau_1 \epsilon], \end{aligned} \tag{3·51}$$

where $\tau_1 = F[w_1]$ and $|\tau_1| \leqq 1$, show that this effect can be expressed in terms of an 'acceptance factor' $\tau_1(u, v) = F[w_1]$ in frequency space.† Thus the photographic image of a low-contrast, incoherent object occupying an isoplanatism-patch A is statistically determined by the Duffieux transmission factor $\tau(u, v)$ of the optical system and by two functions $\tau_1(u, v)$, $\phi_1(u, v)$ which describe properties of the emulsion-development system; viz. its spread function and its noise-power spectrum under the given conditions of illumination and exposure. The relevant properties of the image of a single low-contrast object $\{\sigma\}$ can now be represented by a statistical set $\{I_1(x, y) + n_2^{\mathscr{F}}(x, y)\}$, where

$$\begin{aligned} I_1 &= F^*[\tau\tau_1 F[\sigma]] \\ &= F^*[\tau\tau_1 F[\sigma_0]] + n_1; \end{aligned} \tag{3·52}$$

the sets $\{n_1\}$, $\{n_2^{\mathscr{F}}\}$ being statistically independent of each other and of the spatial variations in σ_0 which represent the object structure. Here $n_1 = F^*[\tau\tau_1 F[n_0]]$ and $\{n_1\}$ is the part of the image noise which originates in the object surface.

3·231. *Statistical sets of images.* Suppose, for example, that we have (i) a set of laboratory photographs sufficient to define the statistics of the image noise, (ii) one external photograph. What is the information content of this photograph?

While it is possible to speak in a certain sense of the information gain when a particular object is presented and a particular image observed (see Woodward 1953, §§ 3·3, 3·5), this is unrealistic from the point of view of the designer, since by hypothesis he does not know exactly which object will be presented, or which image observed.

From the observer's point of view, the effect of the observation is to change the probability distributions which express his knowledge of the presumed object, and the information gain should be expressed in terms of these probability distributions. Shannon's general definition of information is explicitly in terms of these probabilities and the definitions of §§ 3·11 and 3·12 above are special cases of it. The notion of probability corresponds analytically to that of a statistical set and it is the average over such a set which was implied in the footnote on

† $\tau_1(0, 0) = 1$, in consequence of the normalization adopted for w_1.

p. 383. From Shannon's definition it then follows that, for given image noise and given prior knowledge of the object, the average amount of information in a single image is the mathematical expectation of the logarithm of the 'number of distinguishable images' which arise as the object varies statistically according to the constraints or probability densities which express the prior knowledge of the object class (cf. Shannon 1948, §§ 6, 20).

Given the statistical properties of the image noise and a statistically characterized object set, we can then ask, for example, what small traces of Seidel aberration will offset given fifth-order aberrations in the optical system in such a way as to maximize the (average) information content in the image of an object from the set. The answer represents that type of aberration balancing which optimizes optical performance when reinterpretation of images is not excluded. When reinterpretation is excluded, performance is more appropriately assessed by means of the average similarity, in some prescribed sense, between image and object. In both cases, the optimum design depends explicitly on the noise- and spread-characteristics of the light-sensitive receiving surface and the prior probabilities which specify the object set.

This example illustrates the point that it is hardly meaningful to speak of the information design of a system for one particular object but only for a class of objects. This class may in principle be arbitrarily narrow, but the amount of information transmitted by the optical system will always be very small if the range of the object class is reduced until that of the corresponding image class becomes comparable with the range in the noisy image of a single object.

3·232. *Distinguishability in terms of entropy*. Let $\{\sigma\}$ be a statistical set of intensity distributions $\sigma = \sigma^{(\mathfrak{M})}$ which represent objects to which the design is to be relevant, and let $p(\sigma)$ denote its probability density, defined relative to a suitable r-dimensional co-ordinate system S.

Inasmuch as 'noise' has been defined as the unpredictable disturbances to measurement, the distinction between a set of different objects and a single object subject to noise is not absolute, but may depend on the purpose of the observation. For example, the fluctuations in brightness associated with turbulence in the solar photosphere are properly regarded as 'noise' from the point of view of measurements of limb-darkening coefficients but not from that of investigations of solar granulation.

The drawing of a distinction between 'objects' and 'noise' in the set $\{\sigma\}$ corresponds analytically to the setting up of a law which assigns to every member σ a statistical subset $\{\sigma'\}_\sigma$ of $\{\sigma\}$ with a probability density $p_\sigma(\sigma')$. By the 'noisy object σ_0' in $\{\sigma\}$ is meant the subset $\{\sigma'\}_{\sigma_0}$. In the application, σ_0 is the mean value of σ' over the subset $\{\sigma'\}_{\sigma_0}$, but it is sufficient in the general discussion to suppose merely that no two different σ generate the same subset. Then each noisy object $\{\sigma'\}$ determines a unique σ_0, for which $\{\sigma'\} = \{\sigma'\}_{\sigma_0}$. We call this σ_0 the 'object without the noise' and the statistical set of functions $n_0 = \sigma' - \sigma_0$, with probability density $p_{\sigma_0}(n_0 + \sigma_0)$, is called the '$\sigma_0$-noise'.

The randomness $\rho(\sigma)$ of $\{\sigma\}$ relative to S is defined by the equation

$$\log \rho(\sigma) = H = -\int_S p(\sigma) \log p(\sigma)\, \mathrm{d}V, \tag{3·53}$$

where $\mathrm{d}V$ is the element of volume in the co-ordinate system S; that is, $\log \rho(\sigma)$ is the entropy of $\{\sigma\}$ relative to S. If $\{\sigma\}$ is uniformly distributed throughout a finite volume V in the r-dimensional S-space, then $\rho(\sigma) = V$. We call $[\rho(\sigma)]^{1/r}$ the 'spread' of $\{\sigma\}$ relative to the co-ordinate system S.

The effective number N of distinguishable noisy objects in $\{\sigma\}$ may be defined, relative to S in the first instance, as the continuous geometric mean of the ratio

$$\frac{\rho(\sigma')}{\rho(n_0)} = \left(\frac{\text{spread of } \{\sigma'\} \text{ in } S}{\text{spread of } \sigma_0\text{-noise in } S}\right)^r \tag{3·54}$$

as σ_0 runs over $\{\sigma\}$. Or, what is equivalent, $\log N$ is the continuous arithmetic mean of the difference

$$(\text{entropy of } \sigma' \text{ relative to } S) - (\text{entropy of } \sigma_0\text{-noise relative to } S)$$

as σ_0 runs over $\{\sigma\}$.

If the probability distribution $p_{\sigma_0}(n_0)$ is the same in all the noisy objects, that is, if object σ_0 and noise n_0 are uncorrelated, this gives at once

$$\log N = (\text{entropy of } \sigma') - (\text{entropy of noise}), \tag{3·55}$$

both entropies being evaluated relative to S.

In the general case we represent σ' as a point in a second co-ordinate space S', identical with S, and obtain

$$\begin{aligned}\log N &= -\int p(\sigma') \log p(\sigma') \,\mathrm{d}V' + \int p(\sigma) \,\mathrm{d}V \int p_\sigma(\sigma') \log p_\sigma(\sigma') \,\mathrm{d}V' \qquad (3·56)\\ &= -\int p(\sigma') \log p(\sigma') \,\mathrm{d}V' \int p(\sigma) \,\mathrm{d}V + \int p(\sigma) \,\mathrm{d}V \int p_\sigma(\sigma') \log p_\sigma(\sigma') \,\mathrm{d}V'\\ &= -\iint \mathrm{d}V \mathrm{d}V' p(\sigma, \sigma') \log p(\sigma') + \iint \mathrm{d}V \mathrm{d}V' p(\sigma, \sigma') \log p_\sigma(\sigma'),\end{aligned}$$

where $p(\sigma, \sigma') = p(\sigma)\, p_\sigma(\sigma') = p(\sigma')p_{\sigma'}(\sigma)$ is the joint probability density of (σ, σ') in the product space $S \times S'$,

$$= \iint \mathrm{d}V \mathrm{d}V' p(\sigma, \sigma') \log \frac{p(\sigma, \sigma')}{p(\sigma)p(\sigma')}. \tag{3·57}$$

The first term on the right of (3·56) is equal to the entropy of σ' relative to S', and the second term becomes equal to the entropy of the noise relative to S' in the special case where the noise is uncorrelated with the object.

A transformation of co-ordinates in S and in S' evidently leaves the expression (3·57) unaltered. Thus the effective number of distinguishable noisy objects is invariant under a co-ordinate transformation of non-vanishing Jacobian.

To calculate the effective number N of distinguishable levels at a single sampling point (§ 3·12) in the (x, y)-plane we have only to set $r = 1$ and to suppose that $\{\sigma_0\}$ and $\{n_0\}$ have independent Gaussian distributions with known statistical mean-square deviations s^2 and n^2 from the mean values σ_0 and 0 respectively. Then $\sigma' = \sigma_0 + n_0$ has a Gaussian distribution of mean-square deviation $s^2 + n^2$, and $\{n_0\}$ is statistically uncorrelated with $\{\sigma_0\}$ in the sense that its probability density distribution is the same for all σ_0. Now the entropy of a Gaussian distribution of standard deviation ξ is $\log \surd(2\pi e)\, \xi$, by an easy calculation (Shannon 1948, p. 54). Hence, by (3·55),

$$\log N = \sqrt{\frac{s^2 + n^2}{n^2}}, \tag{3·58}$$

which is equivalent to Shannon's result quoted in § 3·12.

3·233. *Application to low-contrast images with Gaussian noise.* Each individual observed image I_2 may be regarded as belonging to the set

$$\{I_2\} = I_1 + \{n_2\}_{I_1} = F^*[\tau\tau_1 F[\sigma]] + \{n_2\}_{I_1}$$

which corresponds to the object intensity $\sigma = \sigma_0 + n_0$ prevailing during the observation. (The observer of course knows only a probability distribution for σ). The set $\{I_2\}$ in turn is a subset of the set of images $\{I_0\} = \{I_1 + \{n_2\}_{I_1}\}$ corresponding to the individual object σ_0 which was presented, together with its noise, and $\{I_0\}$ is itself a subset of the statistical set $\{I\}$ of all possible images corresponding to $\{\sigma\} = \{\sigma_0 + \{n_0\}_{\sigma_0}\}$.

As already noted in § 3·23, it is permissible to suppose from the start that $\{n_2\}$ is frequency-limited to $\mathscr{F}$, since only the '$\mathscr{F}$ cut-off' of n_2 affects the inferred probability distribution of the object.

We now suppose that the prior knowledge of the object class and of the object noise is expressed by saying that the statistical sets $\{\sigma_0\}$ and $\{n_0\}_{\sigma_0}$ are confined to A and have prescribed spectral power densities but are otherwise random.

The spectral power density of $\{\sigma\}$, which depends on the individual spectral power-densities of $\{\sigma_0\}$, $\{n_0\}$ and on their correlation properties, is determined if these are specified; in particular it is the sum of the two individual power densities when the correlation is zero. In the low-contrast case, $\{\sigma_0\}$ and $\{n_0\}$ are uncorrelated and the statistical mean

$$\overline{\sigma(x, y)} = B\kappa_A(x, y),$$

where B is a positive constant; it then follows from what has been said in § 3·22 that $\{\sigma - B\kappa_A\}$ has the statistical structure of Gaussian noise, uniform over A and of prescribed spectral power density.

The set $\{I_2\}$ to which the observed image I_2 is regarded as belonging is then the set $\{\sigma\}$ transformed by the operator $T = F^*[\tau\tau_1 F[\dots]]$ and combined additively with the set $\{n_2\}$, which is statistically independent of it in the low-contrast case, and which also has the statistical structure of Gaussian noise, uniform over A and of prescribed spectral power density. If we use the notation

$$\epsilon_0(u, v) = F[\sigma_0 - \bar{\sigma}_0], \quad \nu_0(u, v) = F[n_0];$$

$$\epsilon_1(u, v) = F[I_1 - \bar{I}_1] = \tau\tau_1 \epsilon_0(u, v), \quad \nu_1(u, v) = F[n_1] = \tau\tau_1 \nu_0(u, v);$$

$$\epsilon_2(u, v) = F[I_2 - \bar{I}_2], \quad \nu_2(u, v) = F[n_2];$$

where the bars denote statistical means over the appropriate sets, the spectral powers of $\{\sigma_0\}$, $\{n_0\}$, $\{n_2\}$ at the sampling points Q_{pq} can be written as $\overline{|\epsilon_0|^2_{pq}}$, $\overline{|\nu_0|^2_{pq}}$ and $\overline{|\nu_2|^2_{pq}}$ respectively, while that of $\{I_2 - \bar{I}_2\}$ is

$$\phi_{pq} = |\tau\tau_1|^2_{pq} (\overline{|\epsilon_0|^2_{pq}} + \overline{|\nu_0|^2_{pq}}) + \overline{|\nu_2|^2_{pq}}. \tag{3·59}$$

The function $\overline{|\nu_2(u, v)|^2}$ is the same as ϕ_1 in § 3·22; it is zero outside $\mathscr{F}$ when, as here, $\{n_2\}$ is supposed to be frequency-limited to $\mathscr{F}$.

To evaluate the effective number of distinguishable images in $\{I_2\}$, we use the procedure already applied to $\{\sigma\}$. In the complex co-ordinate system S_1 provided by the parameters ϵ_{pq} at those sampling points Q_{pq} for which $q \geq 0$ and $(\tau\tau_1)_{pq} \neq 0$, the set $\{I_2 - \bar{I}_2\}$ has a Gaussian distribution of spectral power density ϕ_{pq} and it contains Gaussian noise, uncorrelated with I_2, of spectral power density

$$\omega_{pq} = |\tau\tau_1|^2_{pq}\,\overline{|\nu_0|^2_{pq}} + \overline{|\nu_2|^2_{pq}}. \tag{3·60}$$

Therefore the effective number N of distinguishable images I_2 is given by the equation

$$\log N = (\text{entropy of } \{I_2\}) - (\text{entropy of image noise})$$
$$= (\text{entropy of } \{I_2 - \bar{I}_2\}) - (\text{entropy of } \{n_1 + n_2\}),$$

both entropies being relative to the real co-ordinate system derived from S_1,†

$$= \sum_{q\geq 0,\,(\tau\tau_1)_{pq}\neq 0} \log(2\pi e\phi_{pq}) - \sum_{q\geq 0,\,(\tau\tau_1)_{pq}\neq 0} \log(2\pi e\omega_{pq}), \tag{3·61}$$

by (3·45),

$$= \sum_{q\geq 0,\,(\tau\tau_1)_{pq}\neq 0} \log \frac{\phi_{pq}}{\omega_{pq}}$$

$$\simeq \tfrac{1}{2}|A| \iint_{\mathscr{F}} \log \frac{\phi(u,v)}{\omega(u,v)}\, du\, dv \tag{3·62}$$

if a smooth function $\dfrac{\phi(u,v)}{\omega(u,v)} > 0$ exists which agrees with $\dfrac{\phi_{pq}}{\omega_{pq}}$ at the sampling points Q_{pq} in $\mathscr{F}$.‡ When $\overline{|\epsilon_0(u,v)|^2}$, $\overline{|\nu_0(u,v)|^2}$ and $\overline{|\nu_2(u,v)|^2}$ are smooth positive functions inside $\mathscr{F}$, (3·61) and (3·62) can be written in the form

$$\log N = \sum_{q\geq 0,\,(\tau\tau_1)_{pq}\neq 0} \log\left(1 + \frac{|\tau\tau_1|^2_{pq}\,\overline{|\epsilon_0|^2_{pq}}}{|\tau\tau_1|^2_{pq}\,\overline{|\nu_0|^2_{pq}} + \overline{|\nu_2|^2_{pq}}}\right) \tag{3·63}$$

$$\simeq |A| \iint_{\mathscr{F}} \log \surd\left(1 + \frac{|\tau\tau_1|^2\,\overline{|\epsilon_0|^2}}{|\tau\tau_1|^2\,\overline{|\nu_0|^2} + \overline{|\nu_2|^2}}\right) du\, dv \tag{3·64}$$

and the *information per unit area* in the image is seen to be

$$\frac{1}{f^2} \iint_{\mathscr{F}} \log \surd\left(1 + \frac{|\tau\tau_1|^2\,\overline{|\epsilon_0|^2}}{|\tau\tau_1|^2\,\overline{|\nu_0|^2} + \overline{|\nu_2|^2}}\right) du\, dv, \tag{3·65}$$

where f is the focal length of the system.

When the noise $\{n_2\}$ is negligibly small, so that $\overline{|\nu_2|^2_{pq}}$ can be taken as zero for all p, q, the quotient ϕ_{pq}/ω_{pq} reduces to $(\overline{|\nu_0|^2} + \overline{|\epsilon_0|^2})/\overline{|\nu_0|^2}$ and (3·51) shows that $\log N$ is independent of the actual values of the non-zero factors $(\tau\tau_1)_{pq}$. In this case the number of distinguishable images I_2 is equal to the number of distinguishable functions σ^{τ_0} (see (2·50)), whatever the aberrations. Thus aberrations alone do not reduce the information in the image; it is the interaction between the aberrations and the noise n_2 which reduces information. From (3·63), remembering that $|\tau| \leq \tau_0$, we see that in the presence of image noise n_2 the value of $\log N$ (that is, the mathematical expectation of the amount of information in a single image I_2) is greatest when the system is aberration-free.

3·3. *Optical consequences of equation* (3·64)

Some interesting consequences follow from the fact that the (statistical mean) information in the image depends on τ, τ_1, ϵ_0, ν_0, ν_2 only through their squared moduli.

† The substitution of $\{I_2 - \bar{I}_2\}$ for $\{I_2\}$ is equivalent to a parallel bodily shift of the corresponding probability distribution in this co-ordinate system.

‡ Inside $\mathscr{F}$, $\tau(u,v) \neq 0$ except at a negligibly small proportion of the points Q_{pq}, and we can suppose $\tau_1(u,v) \neq 0$.

A set in which each member represents the same point-object situated in A (but with fluctuating brightness) and a uniform Gaussian set $\{\sigma^{(\mathfrak{M})}-B\kappa_A\}$ corresponding to a random low-contrast set $\{\sigma^{(\mathfrak{M})}\}$ of objects frequency-limited to $\mathscr{F}$ possess the same spectral power function $\overline{|\epsilon(u,v)|^2}$ in $\mathscr{F}$ (namely one which has the same values $\overline{|\epsilon|^2_{pq}}$ at all the sampling points Q_{pq} in $\mathscr{F}$) if they have the same total mean square fluctuation. It then follows from (3·64) that the condition for maximizing the (statistical mean) information in the image can be expressed as a condition on the intensity distribution in the image of a point-object situated at an arbitrary point (x,y) inside A.

The above conclusion was based on the assumption of a uniform Gaussian object set $\{\sigma^{(\mathfrak{M})}\}$, confined to A; this corresponds to the absence of prior information about the mean spectral power distribution in the object set. When the object set has a known mean power spectrum and a known (uncorrelated) noise power spectrum, an extension of the same result appears as an analytical consequence of the relation

$$\tau\tau_1 = \tau_1 C[\mathscr{E},\mathscr{E}^*] = F[w_* w_1] \tag{3·66}$$

(see § 3·23, equation (3·51)), which allows us to write (3·64) in the form

$$\log N \simeq |A| \iint_{\mathscr{F}} \log \sqrt{\left(1+\frac{\overline{|\epsilon_0|^2}\,|F[w_* w_1]|^2}{|F[w_* w_1]|^2\,\overline{|\nu_0|^2}+\overline{|\nu_2|^2}}\right)}\,\mathrm{d}u\,\mathrm{d}v. \tag{3·67}$$

Here $w_* w_1$ is the intensity distribution in the noise-free image I_1 of a point-object of unit power and w_1 is the 'spread function' in the receiving surface.

(3·67) gives the values of $\log N$ for a single isoplanatism-patch A. The corresponding result for the whole image is

$$\log N \simeq \iint_F \mathrm{d}x\,\mathrm{d}y \iint_{\mathscr{F}} \log \sqrt{\left(1+\frac{\overline{|\epsilon_0|^2}\,|F[w_* w_1]|^2}{|F[w_* w_1]|^2\,\overline{|\nu_0|^2}+\overline{|\nu_2|^2}}\right)}\,\mathrm{d}u\,\mathrm{d}v, \tag{3·68}$$

where now $\epsilon = \epsilon(u,v;x,y)$, $\nu_0 = \nu_0(u,v;x,y)$, $\nu_2 = \nu_2(u,v;x,y)$ are calculated, for each (x,y), from the values of σ in the largest available isoplanatism-patch surrounding (x,y), and where the (x,y)-integration is over the working field F.

A number of interesting special results can be deduced from (3·68). First we note in passing that, whatever the aberrations may be, information is increased by increasing field or aperture,† though it may have to be extracted by elaborate interpretation or image-reconstruction processes if, for example, an increase in aperture has introduced heavy aberrations. Next, in any optical system:

(1) For given field and aperture, the optimum camera design from the information point of view is determined by the four functions $\overline{|\epsilon_0|^2}$, $\overline{|\nu_0|^2}$, w_1 and $\overline{|\nu_2|^2}$, that is to say, by the mean-power spectra of the expected objects and the expected object noise on the one hand, and by the two characterizing functions (spread and noise power) of the receiver on the other.

(2) When $\overline{|\nu_2|^2}$ is negligibly small; that is, when nearly all the noise is in the object, the (statistical mean) amount of information in an image is independent of the aberrations, as noted in § 1, and is

$$\log N = \iint_F \mathrm{d}x\,\mathrm{d}y \iint_{\mathscr{F}} \log \sqrt{\left(1+\frac{\overline{|\epsilon_0|^2}}{\overline{|\nu_0|^2}}\right)}\,\mathrm{d}u\,\mathrm{d}v. \tag{3·69}$$

† Unless, indeed, the r.m.s. noise level in the image surface increases more rapidly than in simple proportion to the mean brightness of the low-contrast image.

If, in particular, n_2 is negligible and there is no prior information about object-structure or object-noise power spectrum, the 'most random' (that is, the entropy-maximized) sets $\{\sigma_0\}$ and $\{n_0\}$ are obtained by giving $\overline{|\epsilon_0|^2}$ and $\overline{|\nu_0|^2}$ constant values s^2, n^2 respectively in $\mathscr{F}$, and (3·69) becomes

$$\log N = |F|\,|\mathscr{F}|\log\sqrt{\left(1+\frac{s^2}{n^2}\right)}.$$

In this case the information is simply the product of field area, aperture area, and a simple function of the signal-to-noise ratio.†

(3) The inner integral in (3·68) tends formally, as the signal-to-noise ratio becomes small, to the analogue of the expression used in electronics to evaluate the 'detectability' of a pulse in the presence of noise; there the maximum squared ratio of peak pulse height to r.m.s. noise level obtainable by suitable filtering is proportional to the frequency integral of the power ratio in the spectra of received signal and noise. Detection is possible at a lower signal-to-noise ratio than is shape evaluation, so that it is satisfactory to find the information equation for weak signals taking a form which suggests that it is concerned with detection.

(4) When $\overline{|\nu_2|^2} \gg \overline{|\nu_0|^2}$; that is, when nearly all the noise is either quantum noise or noise originating in the image surface, we have

$$\log N = \iint_F \mathrm{d}x\,\mathrm{d}y \iint_{\mathscr{F}} \log\sqrt{\left(1+\frac{\overline{|\epsilon_0|^2}\,|\tau\tau_1|^2}{\overline{|\nu_2|^2}}\right)}\,\mathrm{d}u\,\mathrm{d}v, \tag{3·70}$$

and it is apparent that both aberrations and 'image spread' reduce information; the former by reducing $|\tau|$, the latter by reducing $|\tau_1|$. By (3·66), $|\tau|$ is greatest when $\mathscr{E} = 1$ in $\mathscr{A}$ and 0 elsewhere; that is, when the aberrations are zero. Thus modifications of the images by phase shift in the exit pupil can never improve the information-passing capacity of the camera. They can, of course, present the information which does get through in a more convenient form.

(5) When object noise is negligible but

$$\iint_{\mathscr{F}} |\tau\tau_1|^2\,\overline{|\epsilon_0|^2}\,\mathrm{d}u\,\mathrm{d}v \ll \iint_{\mathscr{F}} \overline{|\nu_2|^2}\,\mathrm{d}u\,\mathrm{d}v, \tag{3·71}$$

so that the image detail is almost everywhere smothered in noise, we obtain from (3·66) and (3·68)

$$\log N = \iint_F \mathrm{d}x\,\mathrm{d}y \iint_{\mathscr{F}} \frac{\overline{|\epsilon_0|^2}\,|\tau\tau_1|^2}{2\,\overline{|\nu_2|^2}}\,\mathrm{d}u\,\mathrm{d}v. \tag{3·72}$$

In the special case where object structure and image noise are both uniform Gaussian over $\mathscr{F}$, (3·72) becomes

$$\log N = \frac{s^2}{2n^2}\iint_F \mathrm{d}x\,\mathrm{d}y \iint_{\mathscr{F}} |\tau\tau_1|^2\,\mathrm{d}u\,\mathrm{d}v. \tag{3·73}$$

The inner integral

$$\iint_{\mathscr{F}} |\tau\tau_1|^2\,\mathrm{d}u\,\mathrm{d}v = \iint_{-\infty}^{\infty} |\tau\tau_1|^2\,\mathrm{d}u\,\mathrm{d}v = \iint_{-\infty}^{\infty} (I_1(x',y'))^2\,\mathrm{d}x'\,\mathrm{d}y'; \tag{3·74}$$

where $I_1(x',y')$ is the normalized intensity distribution in the image (with 'spread') of a point source of unit power situated at (x,y) in the object plane, the normalization being given by the equation

$$\iint_{-\infty}^{\infty} I_1(x',y')\,\mathrm{d}x'\,\mathrm{d}y' = 1. \tag{3·75}$$

† This result may be compared with Shannon's quoted above (§3·13).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Dropping the normalization, we obtain for the (statistical mean) amount of information in a field patch A', not necessarily isoplanatic, of the noise-smothered image of a random low-contrast object the evaluation

$$\frac{s^2}{2n^2}\iint_{A'} \mathrm{d}x\,\mathrm{d}y\,\frac{\iint_{-\infty}^{\infty} I_1^2\,\mathrm{d}x'\,\mathrm{d}y'}{\left(\iint_{-\infty}^{\infty} I_1\,\mathrm{d}x'\,\mathrm{d}y'\right)^2}.$$

The expression
$$\iint_{-\infty}^{\infty} I_1^2\,\mathrm{d}x'\,\mathrm{d}y'\Big/\left(\iint_{-\infty}^{\infty} I_1\,\mathrm{d}x'\,\mathrm{d}y'\right)^2$$

under the integral sign is of the same form as one which has been discussed (Fellgett 1953, equation (18)) in connexion with the evaluation of 'sharpness' in photographic images and grain.

3·4. *The coherent case*

Essentially similar results follow if we suppose that the object is coherently lit, that resolved components of the complex displacement $\hat{D}$ over the image can be measured in some way, and that the relevant properties of the receiver are described by a complex spread function $\hat{w}_1$ and a complex noise set $\{\hat{n}_2\}$, so that the quantities 'observed' are components of $\hat{D}_2 = \hat{D}_* \hat{w}_1 + \hat{n}_2$. If $\hat{n}_2$ is uncorrelated with $\hat{D}$, the information gain for any given component of $\hat{D}_2$ is then of a similar form to (3·63) and (3·64); the mean-square signal and noise intensities are replaced by the mean-square values of the corresponding components of $\hat{D}_2$ and of $\hat{n}_1+\hat{n}_2$, where $\hat{n}_1 = F^*[\hat{\tau}\hat{\tau}_1 F[\hat{n}_0]]$, and the integration is over $\mathscr{A}$ instead of over $\mathscr{F}$. Since observation of a single resolved component of $\hat{D}_2$ cannot distinguish the separate contributions from the members of a pair of 'opposite' frequencies (u, v) and $(-u, -v)$, the information content of each resolved component of the image depends on the members of such a pair only through their sum. For the average information content $\log N_{(\Re)}$ of the real part of $\hat{D}_2$ we obtain, as the analogue of (3·64), the equation

$$\log N_{(\Re)} = |A| \iint_{\mathscr{A}} \log \sqrt{\left(1+\frac{(\Re\hat{\tau}\hat{\tau}_1\hat{\epsilon}_0 \between u, v)}{(\Re\hat{\tau}\hat{\tau}_1\hat{\nu}_0 \between u, v)+(\Re\hat{\nu}_2 \between u, v)}\right)}\,\mathrm{d}u\,\mathrm{d}v, \tag{3·76}$$

where $\hat{\nu}_0 = F[\hat{n}_0]$, $\hat{\nu}_2 = F[\hat{n}_2]$, $\hat{\epsilon}_0 = F[\hat{E}-\hat{B}\kappa_A]$ and the notation $(\Re\hat{f}\between u, v)$ denotes the statistical mean square

$$\overline{(\Re[\hat{f}(u,v)+\hat{f}(-u,-v)])^2}.$$

An evaluation of the average information content $\log N_{(\Im)}$ of the imaginary part of $\hat{D}_2$ is obtained on replacing the symbol $\Re$ (real part) by $\Im$ (imaginary part) in (3·76).

Measurements of the above kind are not normally possible with the more conventional types of optical instruments; however, they are relevant to a discussion of the information content of images by interference microscopes and phase-contrast microscopes.

If the presumed object set has the statistical structure of Gaussian noise with known power spectrum, then at each sampling point Q_{pq} in $\mathfrak{M}$

$$(\Re\,\mathrm{e}^{\mathrm{i}\phi}\,\hat{\tau}\hat{\tau}_1\hat{\epsilon}_0 \between u_p, v_q) = |\hat{\tau}\hat{\tau}_1|_{pq}^2\,\overline{|\hat{\epsilon}_0|_{pq}^2},$$

$$(\Re\,\mathrm{e}^{\mathrm{i}\phi}\,\hat{\tau}\hat{\tau}_1\hat{\nu}_0 \between u_p, v_q) = |\hat{\tau}\hat{\tau}_1|_{pq}^2\,\overline{|\hat{\nu}_0|_{pq}^2}, \quad (\Re\hat{\nu}_2 \between u_p, v_q) = \overline{|\hat{\nu}_2|_{pq}^2} \tag{3·77}$$

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

for any fixed value of the phase angle ϕ. It follows that the mean information content is the same in each resolved component of $\hat{D}_2$, namely,

$$\log N = |A| \iint_{\mathscr{A}} \log \sqrt{\left(1 + \frac{|\hat{\tau}\hat{\tau}_1|^2 \overline{|\hat{\epsilon}_0|^2}}{|\hat{\tau}\hat{\tau}_1|^2 \overline{|\hat{\nu}_0|^2} + \overline{|\hat{\nu}_2|^2}}\right)} \, du \, dv. \tag{3·78}$$

The right-hand side of (3·78) depends on the transmission factor $\hat{\tau}$ only through $|\hat{\tau}|$, which by (2·38) is independent of the aberrations; thus aberrations do not alter the mean information content in a given component of the images of coherent objects belonging to a set of this type. Non-uniform aperture shading can alter the information content, since it alters $|\hat{\tau}|$.

A more restricted type of object set is of some specialized interest in microscopy; namely, one in which the objects transmit the light with negligible absorption and with small phase variation. In this case it is known in advance that the resolved part of the first-order small quantity $\hat{\epsilon}_0(u, v) + \hat{\epsilon}_0(-u, -v)$ in the direction parallel to the vector $\hat{\epsilon}(0, 0)$ is of the second order of smallness. If the observer is restricted (as is approximately the case when an object of this type is viewed in an ordinary microscope with axial illumination) to observation of the resolved part of $(\hat{\tau}\hat{\tau}_1\hat{\epsilon}_0)_{u,v} + (\hat{\tau}\hat{\tau}_1\hat{\epsilon}_0)_{-u,-v}$ in the direction parallel to the vector $(\hat{\tau}\hat{\tau}_1\hat{\epsilon})_{0,0}$, then the mean information content of the observed images is greater when aberrations are present (i.e. when $\hat{\tau}$ is not everywhere real) than when they are absent. A similar property is turned to practical account when a well-corrected microscope is deliberately defocused in order to infer the structure of a fully transparent object under critical or Köhler illumination.

4. Image assessment

4·1. *Assessment by similarity*

The calculus established in § 2 for the similarity of image to object, and that established in § 3 for the information content of the image, now make it possible to formulate specific definitions which relate the image quality, from these two points of view, to the ikonal function $e(\xi, \eta; x, y)$ of the optical system.

In ray theory, one of the most elementary measures of image defect is that based on the 'spread' or maximum angular extent of the geometrical image. It is well known that this measure is unsatisfactory in cases where, for example, the geometrical image consists of a bright central nucleus containing most of the light and a much more widely spread outer region of very low ray density. With images of this kind, a marked improvement in assessment can be obtained in a very simple way by using the r.m.s. deviations of the rays instead of their maximum deviations; that it to say, by calculating the radius of gyration of the geometrical image instead of its maximum radius. This method, adopted by Gauss in 1801 to assess the effects of primary spherical aberration, and used again more recently by Linfoot & Wayman (1949) in a discussion of the aberrations of field-flattened Schmidt cameras,† is inapplicable to diffraction images. The essential difficulty is that, whether or not aberrations are present, an optical system with a sharply defined aperture of the usual kind gives diffraction images in which the contribution to the radius of gyration from an annulus of given width centred on the principal point of the image is of the same order of

† This paper contains two errors; a corrected version of the analysis appears in Linfoot (1955).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

magnitude whether the radius of the annulus is large or small. It follows that a satisfactory assessment of diffraction images must give less weight to the 'distant diffracted light' and more to the central part of the image than does that by means of the radius of gyration of the intensity distribution.

Assessment by fidelity

An assessment which meets this requirement, and at the same time retains some of the good points of the assessment by radius of gyration, is suggested by the analysis of § 2. In § 2·4, the fidelity defect of a noise-free image was defined as the normalized mean-square distance between this image and the corresponding object. It can be pictured in geometrical terms as the distance between two points in multi-dimensional space which represent the object and the image respectively (compare § 3·21), divided by the normalizing factor $\iint_A \sigma^2 \,dx\,dy$ or $\iint_A |\hat{E}|^2 \,dx\,dy$ in the incoherent and coherent cases respectively.

When the situation is complicated by the presence of noise, the analysis takes a very similar form. For shortness we consider only the imaging of an incoherent object contained in an isoplanatism-patch A. The extension to a full working field F can be made by dividing it into isoplanatism-patches. The r.m.s. distance $\bar{d}(\sigma, I_2) \geqslant 0$ between the noisy object $\sigma = \sigma_0 + n_0$ and the noisy image $I_2 = I_1 + n_2$ is defined by the equations

$$\bar{d}^2(\sigma, I_2) = \overline{\iint_A (\sigma - I_2)^2 \,dx\,dy} = \iint_A \overline{(\sigma - I_2)^2} \,dx\,dy. \qquad (4·1)$$

If we assume that the object structure is statistically independent of the total object brightness, and that the noise is uncorrelated with both of them, (4·1) gives, by Parseval's theorem,

$$\bar{d}^2(\sigma, I_2) = \iint_A \overline{(\sigma - I_1)^2} \,dx\,dy + \iint_A \overline{n_2^2} \,dx\,dy$$

$$= \iint_{\mathscr{F}} |1 - \tau\tau_1|^2 (\overline{|\epsilon|^2} + \overline{|\nu_0|^2}) \,du\,dv + \iint_{C(\mathscr{F})} (\overline{|\epsilon|^2} + \overline{|\nu_0|^2}) \,du\,dv + \iint_{-\infty}^{\infty} \overline{|\nu_2|^2} \,du\,dv,$$

where $\epsilon = F[\sigma_0]$, while ν_0 and ν_2 have the same meaning as in § 3,

$$= \iint_{-\infty}^{\infty} (\overline{|\epsilon|^2} + \overline{|\nu_0|^2}) \,du\,dv - \iint_{\mathscr{F}} (1 - |1 - \tau\tau_1|^2)(\overline{|\epsilon|^2} + \overline{|\nu_0|^2}) \,du\,dv$$

$$+ \iint_{-\infty}^{\infty} \overline{|\nu_2|^2} \,du\,dv. \qquad (4·2)$$

(4·2) is applicable whenever the object set is of the type discussed in § 3·233; the relation between ϵ here and ϵ_0 in that section is given by the equations

$$\epsilon = F[\sigma_0] = F[\sigma_0 - B\kappa_A] + BF[\kappa_A]$$
$$= \epsilon_0 + B|A| \operatorname{sinc} 2au \operatorname{sinc} 2bv.$$

In (4·2) the first term, which is equal to $\iint_A \overline{\sigma^2} \,dx\,dy$, depends only on the object set; in the notation just introduced, it can be written as $\bar{d}^2(\sigma, 0)$. Then the fidelity defect in A may be defined as

$$\frac{\bar{d}^2(\sigma, I_2)}{\bar{d}^2(\sigma, 0)} = 1 - \frac{1}{\bar{d}^2(\sigma, 0)} \iint_{\mathscr{F}} (1 - |1 - \tau\tau_1|^2)(\overline{|\epsilon|^2} + \overline{|\nu_0|^2}) \,du\,dv + \frac{\bar{d}^2(n_2, 0)}{\bar{d}^2(\sigma, 0)}. \qquad (4·3)$$

The assessment of image quality through fidelity defect in the above sense penalizes image distortion just as heavily as lack of sharpness. In bringing out this consequence of the proviso that image reconstruction is excluded, our analysis has shown that its strict application penalizes distortion much more heavily than is usually considered appropriate, and it may be concluded that a rudimentary form of reconstruction is involved when an image is looked at in the ordinary way. It is the habitual use of reconstruction or interpretation processes during the act of seeing which explains the fact that even fairly large amounts of smooth distortion in an image do not prevent the easy recognition of familiar object patterns. A modified assessment of image quality which takes account of this fact can be obtained by allowing for the calculated distortion of the optical system before matching the image against the object. Strictly speaking, this elimination of distortion is a rudimentary reconstruction or interpretation process which should be used whenever the recognizability of patterns or 'symbols' is given precedence as a requirement over the strict geometrical similarity of image to object. It occurs automatically if, in defining the coordinate mesh (x', y') in the image surface S' (see figure 1) we assign to the principal point of the geometrical image patch of the object-point (x, y) the coordinate numbers $x'=x$, $y'=y$. With the definition adopted in § 2·1, it amounts to replacing, in the function $w(x', y'; x, y)$, the arguments x', y' by $x'-\frac{x'}{r'}Z(r)$, $y'-\frac{y'}{r'}Z(r)$ respectively, where $r'=+\sqrt{(x'^2+y'^2)}$, $r=+\sqrt{(x^2+y^2)}$ and the distortion function $Z(r)$ is calculated from the design data of the system.

When image quality is assessed by means of fidelity defect in the sense of (4·3), the problem of optimizing a camera design relative to a given object set can be interpreted as that of choosing the available design parameters so as to minimize the statistical mean of $d^2(\sigma, I_2)$ or, what is equivalent, the expression

$$-\iint_{\mathscr{F}}(1-|1-\tau\tau_1|^2)\,(\overline{|\epsilon|^2}+\overline{|\nu_0|^2})\,\mathrm{d}u\,\mathrm{d}v \tag{4·4}$$

for the corresponding image set $\{I_2\}$.

Expression (4·4) provides, for example, a means of determining in what circumstances an increase in aperture, with consequent worsening of the aberrations, may improve the image fidelity over a given field through reduction of the diffraction spread. When, on the other hand, the aperture is given in advance, the expression

$$\iint_{\mathscr{F}}(\overline{|\epsilon|^2}+\overline{|\nu_0|^2})\,\mathrm{d}u\,\mathrm{d}v$$

is unaffected by changes in optical design and the problem reduces to the minimization of the expression

$$\iint_{\mathscr{F}}|1-\tau\tau_1|^2\,(\overline{|\epsilon|^2}+\overline{|\nu_0|^2})\,\mathrm{d}u\,\mathrm{d}v \tag{4·5}$$

under variation of the available design parameters. The optimal solution will depend on the statistics of the expected object class through $\overline{|\epsilon_0|^2}$ and $\overline{|\nu_0|^2}$; prior ignorance of this class is expressed by taking $\overline{|\epsilon_0|^2}+\overline{|\nu_0|^2}$ to be constant† throughout $\mathscr{F}$ (compare §§ 3·231

† More accurately, by taking $\overline{|\epsilon_0|^2_{pq}}+\overline{|\nu_0|^2_{pq}}$ to be constant over the sampling points Q_{pq} in $\mathscr{F}$.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

to 3·233). Then the part of the image defect corresponding to (4·5) takes the simple form

$$\frac{1}{|\mathscr{F}|}\iint_{\mathscr{F}} |1-\tau\tau_1|^2 \,\mathrm{d}u\,\mathrm{d}v. \tag{4·6}$$

Since $|\tau| \leqslant \tau_0$ everywhere, (4·6) can never fall below the value

$$\frac{1}{|\mathscr{F}|}\iint_{\mathscr{F}} (1-\tau_0|\tau_1|)^2 \,\mathrm{d}u\,\mathrm{d}v, \tag{4·7}$$

where, in the case of a circular aperture of radius a in (u,v)-units,

$$\pi a^2 \tau_0 = C[\kappa_{\mathscr{A}}] = 2a^2 \arc\cos\frac{r}{2a} - r\sqrt{(a^2-\tfrac{1}{4}r^2)} \quad (u^2+v^2<4a^2),$$
$$= 0 \quad (u^2+v^2 \geqslant 4a^2).$$

If the spread function w_1 in the receiving surface is near to a δ-function, so that its Fourier transform τ_1 is substantially equal to 1 throughout $\mathscr{F}$, (4·7) has the value 0·615. It follows that in a system of circular aperture the part of the image defect corresponding to (4·5) never falls below this value.

The spread function w_1 is usually symmetrical in practice, so that τ_1 is real for all (u,v); it need not, however, be always of positive sign. In a photographic emulsion, both w_1 and τ_1 are ordinarily positive everywhere if the intensity I_2 is taken to mean photographic density. The equations $\tau(0,0)=1$, $\tau_1(0,0)=1$ show that the appropriate normalizations are already incorporated in the notation. In the case of a photographic negative, the effect of the w_1-normalization is to replace it by a normalized 'positive' which can be directly compared with the object.

4·2. *Assessment by information*

We have seen in § 4·1 that the adoption of object-image similarity as the guiding principle in image assessment does not lead naturally to a unique figure of merit, but leaves room for the formulation of a variety of analytical criteria. Assessment by information content, on the other hand, can conform to intuitive notions only if the information is measured in the manner which Shannon showed to be uniquely prescribed by a set of assumptions embodying these notions.

We can define the 'information defect' of the image in the isoplanatism-patch A as the information loss per unit area in this part of the image as compared with the loss in an aberration-free system of the same aperture. The measure of this information loss is $\frac{1}{|A|}\log(N_0/N)$, where

$$\log N = \tfrac{1}{2}|A|\iint_{\mathscr{F}} \log\left(1+\frac{\overline{|\epsilon_0|^2}\,|\tau\tau_1|^2}{|\tau\tau_1|^2\,\overline{|\nu_0|^2}+\overline{|\nu_2|^2}}\right)\mathrm{d}u\,\mathrm{d}v \tag{4·8}$$

is the 'mean information content' (in the sense of § 3) for the actual system and the value of $\log N_0$, the corresponding content for an ideal aberration-free system of the same aperture, is obtained on writing τ_0 for τ in (4·8).†

The optimum design from the information point of view is that which maximizes $\log N$; when the aperture is fixed in advance, this is equivalent to minimizing the information loss $\log(N_0/N)$.

† (4·8) may be compared with Shannon's result quoted above (§3·13).

If object noise is small compared with image noise and $\overline{|\epsilon_0|^2}/\overline{|\nu_2|^2} \gg 1$, we have, approximately,

$$\log N = \tfrac{1}{2}|A| \iint_{\mathscr{F}} \log \frac{\overline{|\epsilon_0|^2}\,|\tau\tau_1|^2}{\overline{|\nu_2|^2}}\,\mathrm{d}u\,\mathrm{d}v, \tag{4·9}$$

$$\log N_0 = \tfrac{1}{2}|A| \iint_{\mathscr{F}} \log \frac{\overline{|\epsilon_0|^2}\,|\tau_0\tau_1|^2}{\overline{|\nu_2|^2}}\,\mathrm{d}u\,\mathrm{d}v, \tag{4·10}$$

and the information defect is

$$\frac{1}{|A|}\log\frac{N_0}{N} = \iint_{\mathscr{F}} \log\left|\frac{\tau_0}{\tau}\right|\mathrm{d}u\,\mathrm{d}v. \tag{4·11}$$

This equation, applicable when object noise is negligible and image structure is strong compared with image noise, provides a measure of the 'information quality' of the optical system in terms of its aperture and its ikonal function alone. It is only in this case that the information quality is independent of ϵ_0, ν_0, ν_2 and τ_1.

For the image over the whole field F (again on the assumptions $\overline{|\nu_0|^2} \ll \overline{|\nu_2|^2} \ll \overline{|\epsilon_0|^2}$), we have

$$\log N = \iint_F \mathrm{d}x\,\mathrm{d}y\,\frac{1}{2}\iint_{\mathscr{F}} \log\frac{\overline{|\epsilon_0|^2}\,|\tau|^2}{\overline{|\nu_2|^2}}\,\mathrm{d}u\,\mathrm{d}v, \tag{4·12}$$

$$\log N_0 = \iint_F \mathrm{d}x\,\mathrm{d}y\,\frac{1}{2}\iint_{\mathscr{F}} \log\frac{\overline{|\epsilon_0|^2}\,|\tau_0|^2}{\overline{|\nu_2|^2}}\,\mathrm{d}u\,\mathrm{d}v, \tag{4·13}$$

where now $\tau = \tau(u, v; x, y)$ and, if vignetting is also allowed, $\tau_0 = \tau_0(u, v; x, y)$. Thus we obtain the equation

$$\log\frac{N_0}{N} = \iint_F \mathrm{d}x\,\mathrm{d}y \iint_{\mathscr{F}} \log\left|\frac{\tau_0(u, v; x, y)}{\tau(u, v; x, y)}\right|\mathrm{d}u\,\mathrm{d}v \tag{4·14}$$

for the (statistical mean) information loss over the whole field due to the combined effects of aberration and emulsion noise.

Although this approximate result is formally independent of the actual size of $\overline{|\nu_2|^2}$, the presence of the noise n_2 plays an essential part. If, for example, $\overline{|\nu_2|^2} \ll \overline{|\nu_0|^2}$, equation (4·8) takes the form

$$\log N = \tfrac{1}{2}|A| \iint_{\mathscr{F}} \log\left(1 + \frac{\overline{|\epsilon_0|^2}}{\overline{|\nu_0|^2}}\right)\mathrm{d}u\,\mathrm{d}v, \tag{4·15}$$

which as in § 3·3 shows that when noise associated with the image surface is negligible the effect of aberrations is to reduce fidelity without destroying information.

It is of special interest that (4·14) involves only properties of the optical system and is formally independent of the statistics of the presumed object set and of the noise and also of the spread function of the image surface; and, consequently, camera designs can be optimized (in the sense of passing the greatest statistical average amount of information) over a wide range of object and film types by choosing the aberration balancing to minimize (4·14). No analogous simplification occurs in connexion with fidelity optimization.

Chromatism

A detailed discussion of the effects of chromatism on image assessment would extend this paper unduly, and we therefore only remark that in the polychromatic case the ikonal function $e(\lambda u, \lambda v; x, y)$ becomes $e(\lambda u, \lambda v; x, y; \lambda)$, where λ is the wave-length of the light.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

The image assessments can then be made in terms of λ-means weighted in accordance with the expected spectral brightness distribution in the object set and the spectral sensitivity of the receiver.

References

Blanc-Lapierre, A. 1953 *Symposium on microwave optics*, **2**, 46. McGill University.
Duffieux, P. M. 1946 *L'Intégral de Fourier et ses applications à l'Optique*. Rennes.
Elias, P., Grey, D. S. & Robinson, D. Z. 1952 *J. Opt. Soc. Amer.* **42**, 127.
Fellgett, P. B. 1953 *J. Opt. Soc. Amer.* **43**, 271.
Gabor, D. 1949 *Proc. Roy. Soc.* A, **197**, 454.
Gabor, D. 1952 *Tech. Rep. Electron. Mass. Inst. Tech.* no. 238.
Johnson, H. L. 1948 *Astrophys. J.* **107**, 34.
Linfoot, E. H. 1946 *Proc. Roy. Soc.* A, **186**, 72.
Linfoot, E. H. 1955 *Recent advances in optics*. Oxford: Clarendon Press.
Linfoot, E. H. & Wayman, P. A. 1949 *Mon. Not. R. Astr. Soc.* **109**, 535.
Rose, A. 1948 *J. Opt. Soc. Amer.* **38**, 196.
Rose, A. 1953 *J. Opt. Soc. Amer.* **43**, 715.
Shannon, C. 1948 *Bell Syst. Tech. J.* **27**, 379, 623.
Shannon, C. 1949 *Proc. Inst. Radio Engrs, N.Y.*, **37**, 10.
Silberstein, L. & Trivelli, A. P. H. 1938 *J. Opt. Soc. Amer.* **28**, 441.
Webb, J. H. 1950 *J. Opt. Soc. Amer.* **40**, 3.
Wiener, N. 1949 *Stationary time series*. New York: John Wiley & Sons.
Woodward, P. M. 1953 *Probability and information theory, with applications to radar*. London: Pergamon Press.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES