Coherence Problems in High-Resolution Microdensitometry

G. B. PARRENT

Joyce, Loebl & Co. Ltd., Gateshead

My background is neither astronomy nor instrumentation but rather image analysis. Hence, I am not going to describe new equipment nor new astronomical findings, but rather some difficulties which have arisen in other fields where images must be interpreted and manipulated, and which are sure to arise in astronomy as well.

I shall limit my attention to problems arising from coherence of the radiation in the analysing instrument and attempt to isolate their source and give guidance on how to recognize them and avoid them.

Advanced uses of the microdensitometer, including astronomical uses, tend more and more to exploit the digital computer as a tool for image processing. Inherent in all such applications is the assumption that the basic instrument is both linear and stationary in transmission. That is, the starting point in such an analysis is the relationship

$$T_M(x) = \int T_A(x')S(x' - x)dx' \tag{1}$$

Here $T_M(x)$ is the measured transmission; $T_A(x)$ the actual transmission, and $S(x)$ is the impulse response of the instrument. More often the equivalent frequency domain analysis is used, *i.e.*

$$\tilde{T}_M(\mu) = \tilde{T}_A(\mu)\tilde{S}(\mu) \tag{2}$$

In (2) the tilde denotes the Fourier transform. In point of fact, these equations are never strictly valid for transilluminated objects. In this paper we will discuss the factors leading to this conclusion as well as the instrumental and other considerations which can be used to circumvent the difficulties arising from this fact. There are two principal sources of difficulty. In systems which use the entire field of the lens, the imaging is never stationary. That is, the impulse response varies with the location of the object point in a manner determined by the aberration balance of the lens. This effect arises principally in flying spot scanner instruments and may be further complicated by non-uniformities in the phosphor of the tube. Under these conditions equation (1) becomes

$$T_M(x) = \int T_A(x')S(x', x)dx' \tag{3}$$

and there is no equation analogous to (2). This effect is only serious for extended objects. In the case of objects occupying a small field of view, this non-stationarity is not too serious since a computer program can be written to compensate. One normally divides the field of view into small areas, called isoplanatic patches, over which the impulse response varies slowly compared with the detail being studied. The system is then treated as stationary over these regions.

While such a program is cumbersome, requiring large amounts of computer storage, it is possible except in cases of very high resolution. The second effect is far more serious in that even in principle it cannot be corrected for. This effect is one of non-linearity. Since it is fundamentally impossible for light to be incoherent over regions on the order of a wavelength,[1] it is not possible to express the relationship between T_M and T_A even in the cumbersome form of equation (3). In order to obtain the correct expression for the relationship between object and image in systems with resolutions of the order of a wavelength it is necessary to appeal to the theory of partial coherence. Because a number of authors have treated this subject each with a slightly different notation, it will be necessary to review briefly the formalism used here. We shall follow the notation introduced by Wolf[2] and use without proof a number of theorems developed by Parrent and presented in Beran and Parrent.

BASIC DEFINITIONS

Mutual Coherence Function

The basic entity in the theory of partial coherence is the mutual coherence function, $\Gamma_{12}(\tau)$, which may be defined by

$$\Gamma_{12}(\tau) \equiv \Gamma(x_1, x_2, \tau) = \langle V(x_1, t)V^*(x_2, t+\tau)\rangle . \tag{4}$$

Here the underscore denotes position vector, the asterisk a complex conjugate, and the sharp brackets a long time average,* *i.e.*

$$\langle f\rangle \equiv \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} f dt. \tag{5}$$

In (4), V is the analytic signal associated with the optical disturbance, which we assume to be a single Cartesian component of the electric field vector. In terms of the mutual coherence function, the complex degree of coherence, $\gamma_{12}(\tau)$ is defined as

$$\gamma_{12}(\tau) = \frac{\Gamma_{12}(\tau)}{\sqrt{\Gamma_{11}(0)\Gamma_{22}(0)}}. \tag{6}$$

It should be noted that the complex degree of coherence, like the mutual coherence function, is a function of seven variables, six position coordinates, and the time-delay coordinate τ. The physical significance of these parameters is illustrated by the example discussed below (p. 164). The treatment of problems involving partially coherent light involves the solution of the two wave equations:

$$\Delta_s^2\Gamma_{12}(\tau) = \frac{1}{c^2}\frac{\delta^2\Gamma_{12}(\tau)}{\delta\tau^2}, \qquad (s=1, 2), \tag{7}$$

where Δ_s^2 denotes the Laplacian operator in the coordinates of the point x_s. A typical problem involves determining the mutual coherence in the source or object plane, solving (7) to obtain the mutual coherence on a later surface, such as the image plane, and then recovering the intensity, I, in the plane of interest from the relation

$$I(x_1) = \Gamma(x_1, x_1, 0). \tag{8}$$

Equation (8) follows directly from the definition of the mutual coherence function and the properties of the analytic signal.

For a large class of problems the theory outlined in the preceding paragraph may be greatly simplified. These problems are characterized by the quasi-monochromatic approximations, which are stated as

$$\left\{\begin{matrix} \Delta\nu < \bar{\nu} \\ 1/\Delta\nu < < |\tau| \end{matrix}\right\}.$$

where $\Delta\nu$ is the spectral width. Of these two constraints, the second is obviously the more significant White light may often be treated as quasi-monochromatic if the path differences, $c|\tau|$, involved in the experiment are suitably small. In those circumstances for which the approximations above are applicable, the mutual coherence function may be replaced by the mutual intensity function, $\Gamma(x_1, x_2)$,

$$\Gamma(x_1, x_2) \equiv \Gamma_{12} = \Gamma(x_1, x_2, 0). \tag{9}$$

The complex degree of coherence reduces to $\gamma_{12}(0) \equiv \gamma_{12}$ and the wave equations (7) reduce to the two Helmholtz equations

$$\Delta_s^2\Gamma_{12} + k^2\Gamma_{12} = 0 \qquad (s=1, 2). \tag{10}$$

where k is the wave number.

In high resolution microdensitometry the quasi-monochromatic approximations are nearly satisfied and we will accordingly limit our attention to this case.

Coherent and Incoherent Fields

Equations (4) through (10) provide the basis of the theory of partial coherence as introduced by

* Equation (5) is equivalent to the definition introduced by Wolf, though in a slightly different form.

https://doi.org/10.1017/S0252921100032425 Published online by Cambridge University Press

Wolf. To apply this theory to the imaging problem, and recover the familiar limiting forms, several theorems due to Parrent are required. Principal among these are:

1. A field is coherent if and only if the mutual intensity function describing it can be factored in the form

$$\Gamma_{12} = U(x_1)U^*(x_2),$$

where

$$\Delta^2 U(x_1) + k^2 U(x_1) = 0 \tag{11}$$

2. An incoherent field cannot exist in free space; however, an incoherent source consistent with this result may be defined.[1]

(For the proof of these theorems and their extensions to polychromatic fields the reader is referred to Beran and Parrent.[1]) Of particular significance for the problem of image evaluation is the second of these theorems. We shall reserve a discussion of the significance of the incoherent limit for a later point; (a comprehensive treatment may be found in Beran and Parrent, chap. 2 and 3).

The Van Cittert-Zernike Theorem

An additional theorem is required before attacking the treatment of the image formation problem. The van Cittert-Zernike theorem may be stated as follows:

The mutual intensity of the illumination derived from a distant incoherent source may be expressed in the form

$$\Gamma(x_1, x_2) = \int I(\xi) e^{\frac{2\pi i}{\lambda R}\xi\cdot(x_1 - x_2)} d\xi. \tag{12}$$

Here I is the intensity distribution across the source, and R is the distance from the source plane to the observation plane. If the source is placed in the focal plane of a lens and the coherence of the emergent beam examined, it is found to follow the same law with the R replaced by the focal length f.

THE IMAGING PROBLEM

We may now direct our attention to the formulation of the general imaging problem. As will become clear in the following discussion, a basic description of image formation (at least as far as the lenses are concerned) already exists in coherence theory and, in fact, may be found in Refs. 1 and 2. This theory has not however been applied to the significant problems of image evaluation. Indeed, the theory has been applied to very few problems. In the next section the basic theory is outlined and those pertinent problems that have been solved are reviewed and discussed.

Review of Image Theory

In coherence theory an object is described by its mutual intensity* (or mutual coherence) distribution rather than its intensity distribution. Thus the object is described by $\Gamma_0(\xi_1, \xi_2)$ and the relationship between object and image, $\Gamma_1(x_1, x_2)$, is developed by solving the two Helmholtz equations (10) subject to the appropriate boundary conditions. The general solution is (see Ref. 1, chap. 7 and 8):

$$\Gamma_i(x_1, x_2) = \int\int \Gamma_o(\xi_1, \xi_2) K(x_1 - \xi_1) K^*(x_2 - \xi_2) d\xi_1 d\xi_2. \tag{13}$$

Here K denotes the amplitude impulse response of the lens; *i.e.*, denoting the complex transmission of the aperture by $A(\alpha)$ we may write

$$K(\xi) = K\left(\frac{\beta}{\lambda f}\right) = \int A(\alpha) e^{\frac{2\pi i}{\lambda f}\alpha\cdot\beta} d\alpha. \tag{14}$$

The two familiar limits may be recovered from (13) by using the theorems of the previous section. Thus, in the coherent limit $\Gamma_{12} = U_1 U_2^*$, and (13) reduces to

$$\Gamma_i(x_1, x_2) = \int U_0(\xi_1) K(x_1 - \xi_1) d\xi_1 \int U_0^*(\xi_2) K^*(x_2 - \xi_2) d\xi_2. \tag{15}$$

From (15) and theorem 1 of the preceding section, it is clear that the image of a coherently illuminated object is coherent. A somewhat more surprising result (and certainly more interesting in the image

* Our discussion in this section will be limited to quasi-monochromatic radiation. This serves to introduce the concepts, and at the same time keeps the development tractable.

evaluation problem) is obtained in the incoherent limit. Thus, we may take* $\Gamma_{12} = \mathrm{I}(\xi_1)\delta(\xi_1 - \xi_2)$ to describe the object. The general image, Eq. (13), then reduces to

$$\Gamma(x_1, x_2) = \int \mathrm{I}(\xi)\mathrm{K}(x_1 = \xi)\mathrm{K}^*(x_2 - \xi)\mathrm{d}\xi. \tag{16}$$

From (16) it is clear that the image mutual intensity is no longer of the same form as the object mutual intensity; *i.e.*, the image of an incoherent object is not incoherent but is partially coherent. This result will be seen to have rather far-reaching implications in the problems of image evaluation.

For most applications, the primary exposing radiation may be safely taken as incoherent. For example, sunlight is coherent only over a distance of approximately 1/20 mm. Thus, even a reconnaissance system which resolved an inch on the ground could probably be safely described by the incoherent limit of Eq. (13). In this case, the intensity in the image can be obtained by setting $x_1 = x_2$ in (16), thus

$$\mathrm{I}_i(x) = \int \mathrm{I}_0(\xi)|\ \mathrm{K}(x - \xi)\ |^2\mathrm{d}\xi. \tag{17}$$

Equation (17) will be recognized as the familiar incoherent imaging equation. The difficulty arises, of course, when the scale of the mutual coherence function becomes comparable with the resolution of the optical instrument. (This point will be discussed at length in a later section.) While this condition is not likely to arise in the original taking system in the near future, it becomes serious in viewing and analysing equipment such as microscopes, enlargers, and microdensitometers at the present state of the art. If one envisions improvements in taking equipment of a factor of two or more, it will become even more serious. This point will become clear as we analyse transilluminated objects.

While (13) represents the general solution to the partially coherent imaging problem, a more useful form for application to image analysis is obtained by considering the object to be a transparency that is transilluminated. This is, of course, the case in almost all viewing of reconnaissance imagery, and certainly in all uses of microscopes and microdensitometers in image evaluation. To describe this class of problems, it is necessary to describe the object in terms of its complex transmission $\mathrm{t}(\xi)$. For transilluminated objects Eq. (13) may be expressed as

$$\Gamma_i(x_1, x_2) = \iint \Gamma_0(\xi_1, \xi_2)\mathrm{t}(\xi_1)\mathrm{t}^*(\xi_2)\mathrm{K}(x_1 - \xi_1)\mathrm{K}^*(x_2 - \xi_2)\mathrm{d}\xi_1\mathrm{d}\xi_2. \tag{18}$$

In most cases, one is interested in the intensity of the image, which may be obtained from (18) by setting $x_1 = x_2$. Thus,

$$\mathrm{I}_i(x) = \iint \Gamma_0(\xi_1, \xi_2)\mathrm{t}(\xi_1)\mathrm{t}^*(\xi_2)\mathrm{K}(x - \xi_1)\mathrm{K}^*(x—\xi_2)\mathrm{d}\xi_1\mathrm{d}\xi_2. \tag{19}$$

In (18) and (19) $\Gamma_1(\xi_1, \xi_2)$ must be interpreted as the coherence of the illumination incident on the transparency. The illumination in most cases of practical interest will be derived from a primary incoherent source. In this case $\Gamma_0(\xi_1, \xi_2)$ takes a special form (because of the van Cittert-Zernike theorem):

$$\Gamma_0(\xi_1, \xi_2) \equiv \Gamma_0(\xi_1 - \xi_2); \tag{20}$$

that is, it becomes a function of coordinate differences only. Under these circumstances (19) becomes

$$\mathrm{I}_i(x) = \iint \Gamma_0(\xi_1 - \xi_2)\mathrm{t}(\xi_1)\mathrm{t}^*(\xi_2)\mathrm{K}(x - \xi_1)\mathrm{K}^*(x - \xi_2)\mathrm{d}\xi_1\mathrm{d}\xi_2. \tag{21}$$

From (21) it is clear that for transilluminated objects the transition from object intensity $|\ t(\xi)\ |^2$ *to* image intensity is nonlinear. The significance of this conclusion is that the customary image evaluation techniques and criteria are not, in general, applicable to such systems. For example, knowing how such a system images sine waves or edges does not permit us to describe how it images other objects. Furthermore, the same optical system could be expected to yield different results if the coherence of the illumination varied. At high resolutions a small variation in the scale of the coherence function can produce dramatic changes in the image. This may account, in part, for the difficulty encountered in intercalibrating instruments in different laboratories, or in the cross checking of microdensitometers that have essentially equivalent optical components but produce different results in edge trace analysis.

Since systems of this type are inherently nonlinear, it is impossible to characterize them by a

* Actually this form for the incoherent limit is only an approximation and must be used with care. However, it is sufficiently precise to illustrate the present problem.

https://doi.org/10.1017/S0252921100032425 Published online by Cambridge University Press

transfer function. This point is easily established by taking the Fourier transform of both sides of (21). Thus,

$$\tilde{I}(\mu) = \int \tilde{t}(\beta)\tilde{t}^*(\mu - \beta \left\{ \int \tilde{\Gamma}\left[\mu - (\sigma + \beta) \right] \tilde{K}(\mu - \sigma \tilde{K}(\sigma)^* d\sigma \right\} d\beta. \tag{22}$$

In (22) the inner integral is characteristic of the instrument only, while the factors $\tilde{t}(\beta)$ and $\tilde{t}^*(\mu - \beta)$ are determined solely from the object spectrum. However, (22) is not in the form of "object spectrum times transfer function equals image spectrum." The inner integral has been referred to as a generalized transfer function, but that nomenclature is rather misleading since the function is not used as a transfer

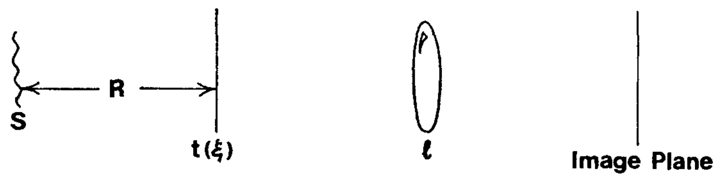


Fig. 1

function at all. A better terminology is the more cumbersome one introduced by Wolf, the "*transmission cross coefficient*," which emphasizes that it is a function of two frequencies.

While the preceding analysis provides a basis or structure for the complete description of partially coherent imaging systems, it is formidable enough to make intuitive interpretation rather difficult. Therefore, to gain some insight into the significance of these developments in image evaluation, some examples will now be given.

Examples of Partially Coherent Imaging

We shall apply the theory developed in the preceding section to the description of two typical experiments which might be used to "measure the transfer function of an imaging system."

Cosine Targets. First let us consider the approach of direct imaging of cosine targets. The experimental arrangement is illustrated schematically in Figure 1. Here a source S is located R units to the left of a transparency $t(\xi)$, which is then imaged by the lens l onto the image plane. The nature of our calculation will be such that the results are equally applicable for all R and for the case where a lens is used to collimate or image the source onto $t(\xi)$. Since the radiation will be partially coherent, it is necessary to describe the experiment very carefully. We shall be interested in calculating the ratio of the image contrast in terms of intensity to the object contrast, also measured in intensity. This contrast ratio will, of course, be identical with the transfer function in the incoherent limit, but otherwise will be an apparent transfer function describing the result of the measurement of cosine waves but of little or no use in describing the measurement of other objects.

To formally solve the problem, it is, of course, necessary to specify the complex transmission of the object. To avoid the complexity of phase objects we shall consider the transmission to be of the form

$$t(\xi) = 1 + \cos 2\pi\mu_0\xi. \tag{23}$$

Here, and throughout this section, we shall limit our analysis to one dimension. The corresponding intensity in the object is

$$|\, t(\xi)\,|^2 = \frac{3}{2} + 2\cos 2\pi\mu_0\xi + \frac{1}{2}\cos 2\pi 2\mu_0\xi. \tag{24}$$

The resulting image may be shown to be an intensity function of the form

$$I(x) = A + B\cos 2\pi\mu_0\xi + C\cos 2\pi 2\mu_0\xi, \tag{25}$$

where, of course, A, B, and C are yet to be determined. The contrast ratios in the image divided by the contrast ratios in the object will be plotted versus μ_0.

Substituting from (25) into (22) the integrals may be evaluated to give a result which is plotted in Figure 2.

To further illustrate the complexity of the problem, we could consider determining the apparent transfer function from an edge trace analysis. If the system were linear, this would, of course, produce the same result as the sine wave example. Proceeding as before, that is assuming a perfect edge and

https://doi.org/10.1017/S0252921100032425 Published online by Cambridge University Press

no errors in measurement, one can calculate the result that would be obtained from "measuring the transfer function" of a diffraction limited instrument by the edge trace method. The calculations are straightforward but cumbersome; therefore, we will skip directly to the result which is plotted in Figure 3.

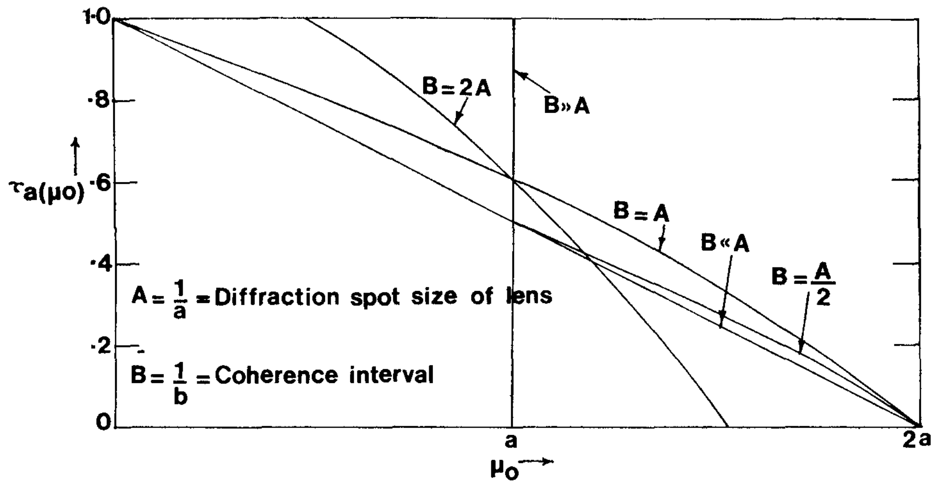


Fig. 2

With this review in mind we require one additional result which will also be stated without proof. The theorem is simply that an incoherent source cannot exist. (Beran and Parrent).[1] It may be shown that the limiting coherence interval is equal to the mean wavelength of the radiation employed. Thus for visible light 0.5 micron is a reasonable estimate of the coherence interval directly on the surface of

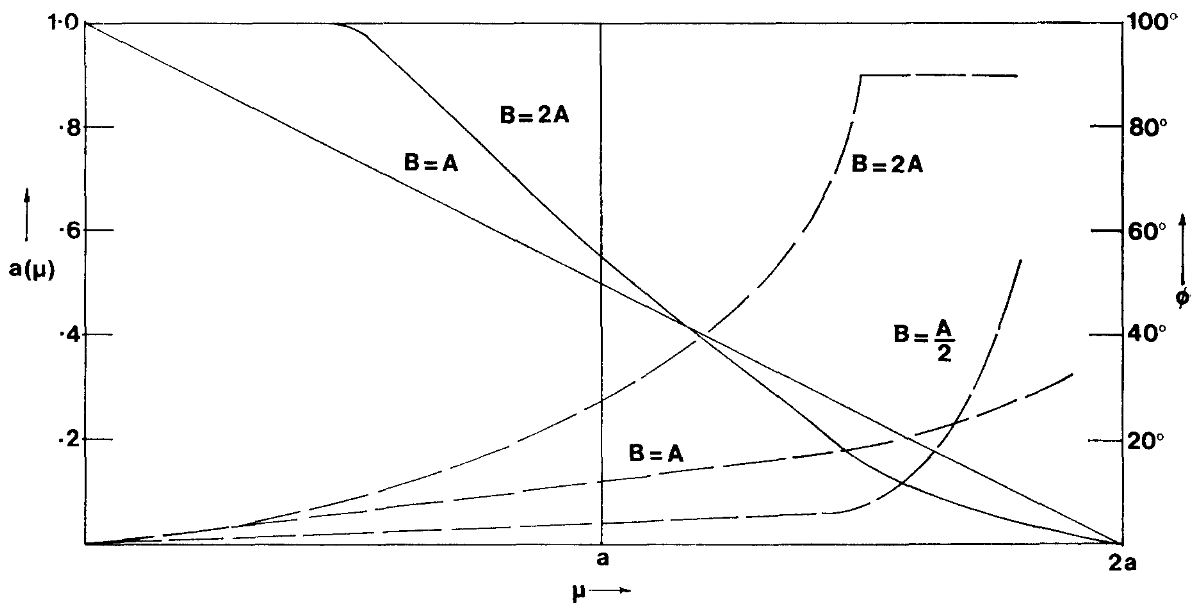
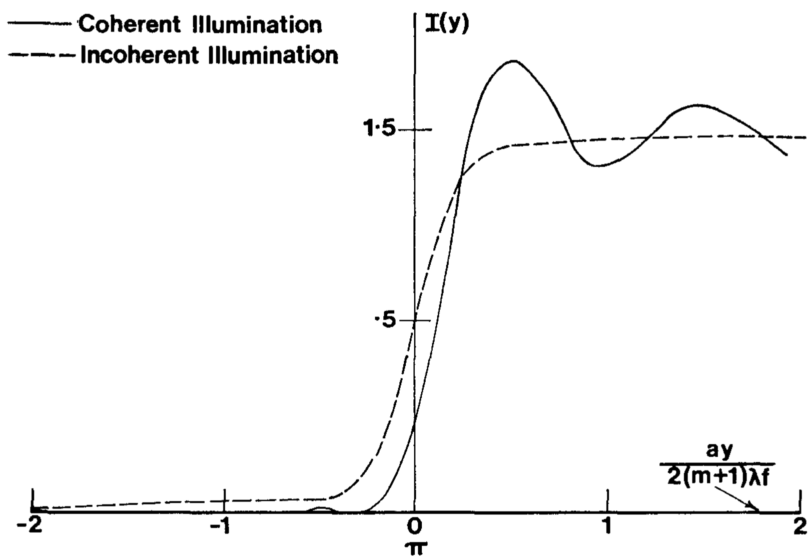


Fig. 3

a black body. From the preceding mathematical analysis it is clear that propagation or indeed imaging can only increase the coherence of the radiation, not decrease it.

Let us now consider the practical implications of these factors in the areas of applied micro-densitometry. Firstly, since the film being examined is never in practice placed in contact with a black body, it is clear that we are always employing partially coherent light with a coherence interval of 0.5 microns or more. Thus, unless the effective slit widths of the instrument are considerably greater than this, the system will behave in a fundamentally non-linear fashion. The nature of this non-linearity, that is, the collapsing of a function of two variables to a function of one variable, is such that it is fundamentally impossible to invert the equations, even in the computer. Furthermore, a partially coherent system is extremely sensitive to phase variations, such as arise from the swelling of emulsions during processing.

Secondly, since the primary light source is always imaged on to the film and since the exact



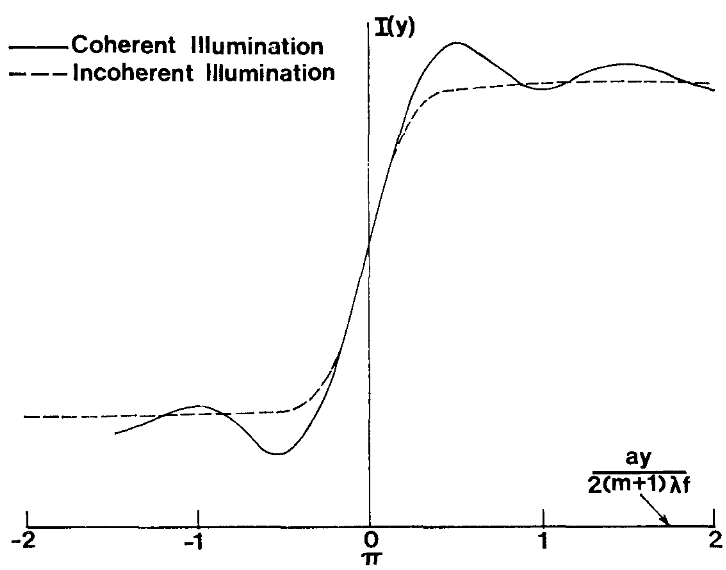


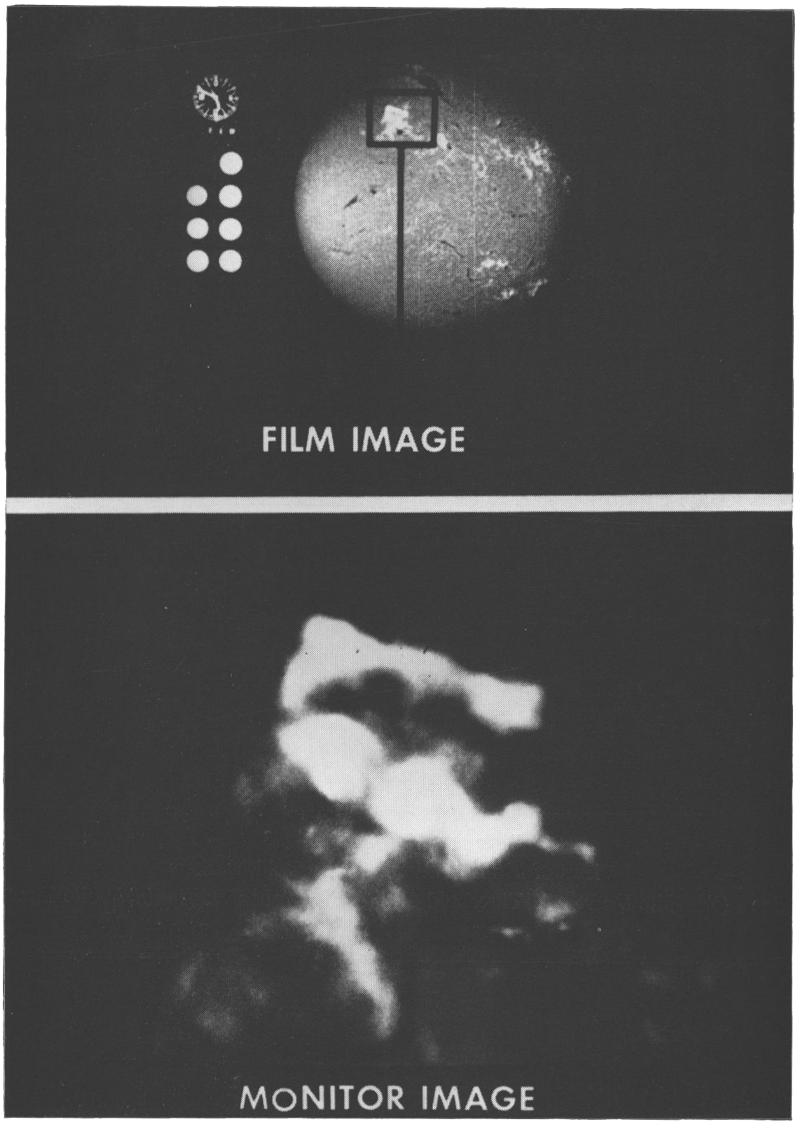
Fig. 4, Fig. 5

coherence interval depends in detail on the nature of this substage illumination system, it is not possible to know *a priori* what slit size can be used to avoid these difficulties.

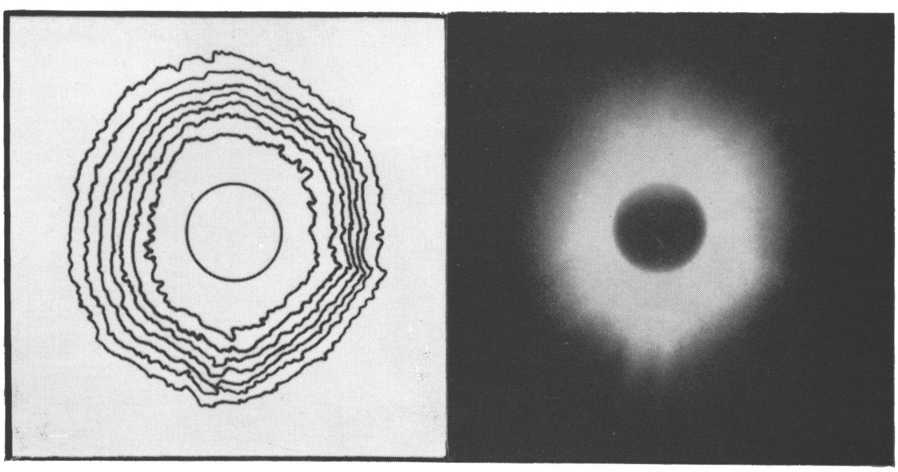
Thirdly, the computer is incapable of detecting the presence of such non-linear effects from the raw data. Thus a computer programmed to produce the transfer function of the instrument from an edge trace (or other target) would produce the meaningless results just illustrated; and, of course, the use of such results in any sort of image processing would only increase further the deviation from the truth.

These considerations are not meant to imply that quantitative microdensitometry is impossible

https://doi.org/10.1017/S0252921100032425 Published online by Cambridge University Press

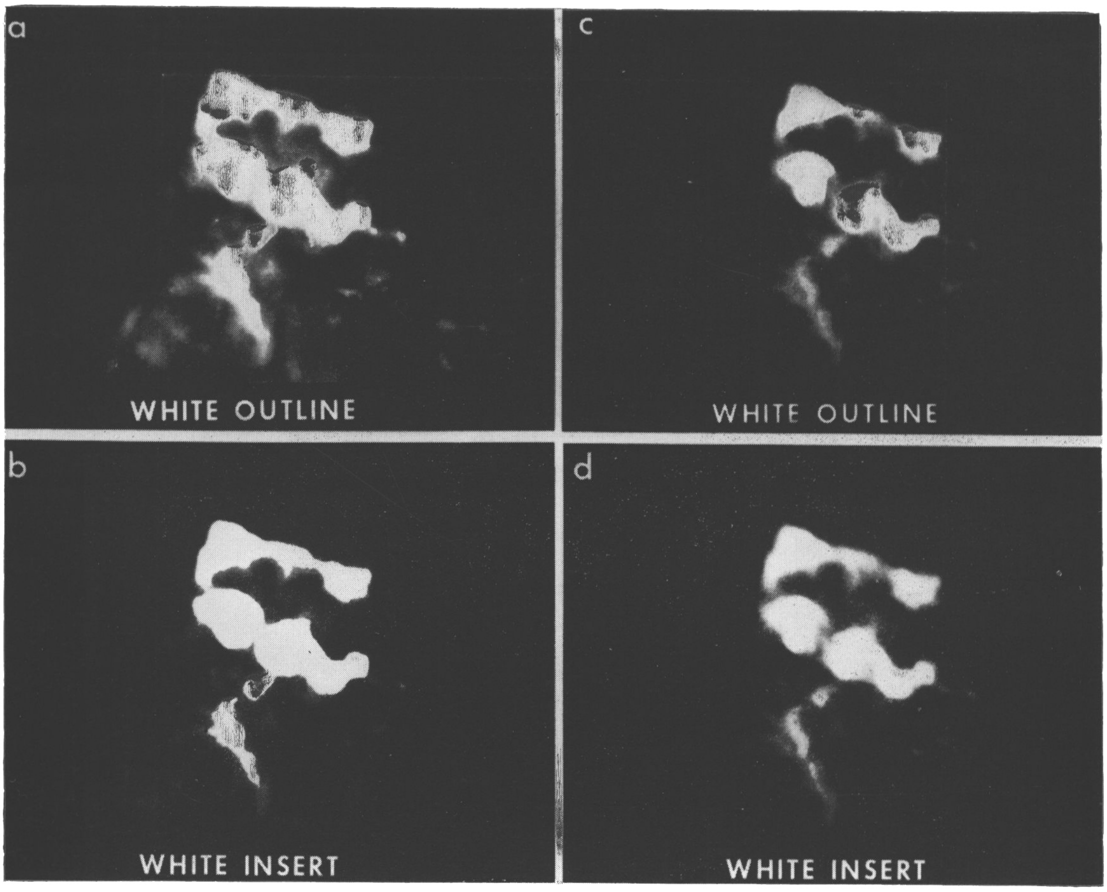


REID AND VORHABEN Fig. 2
Rear Projection Screen View and Monitor View

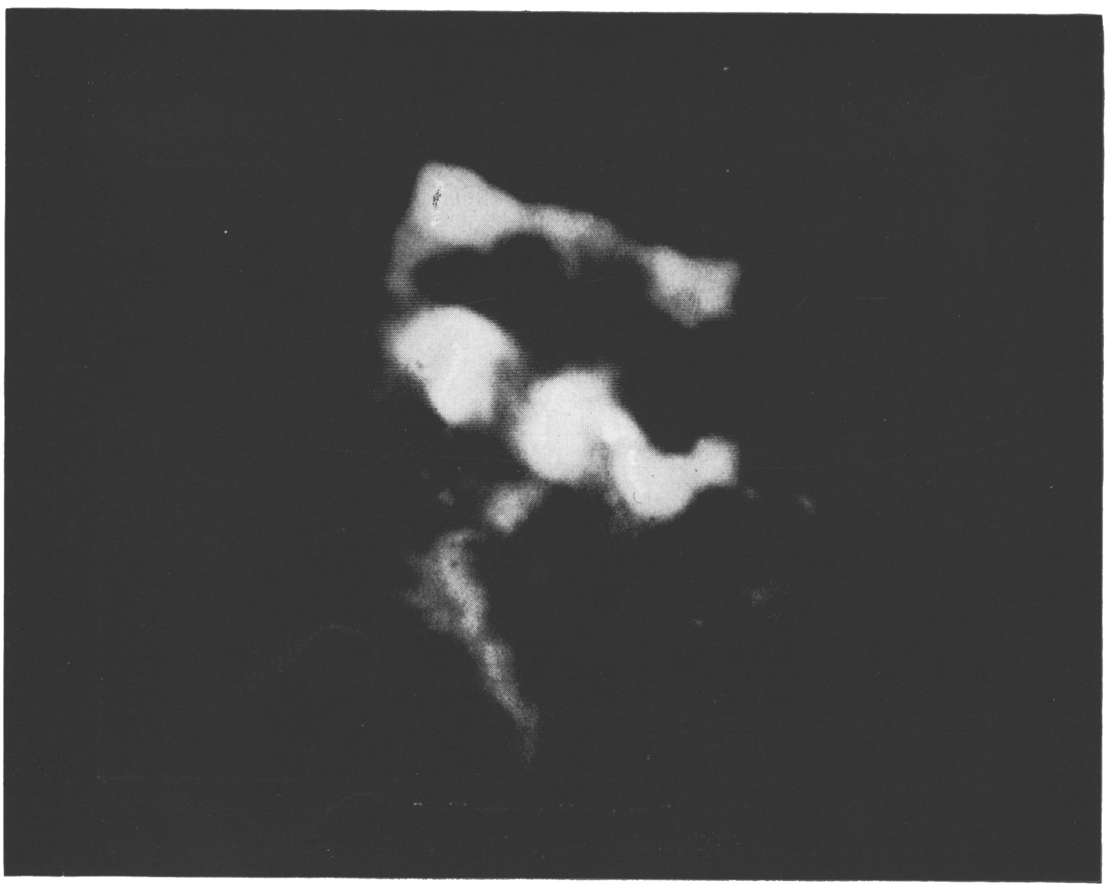


REID AND VORHABEN Fig. 7
Coronal Photograph and Isophotes

https://doi.org/10.1017/S0252921100032425 Published online by Cambridge University Press



REID AND VORHABEN Fig. 3
Views of Monitor Showing Threshold Markers
(*a*) Threshold Outline (*b*) White Insert (*c*) Threshold Outline for Higher Value (*d*) White Insert for Higher Value



REID AND VORHABEN Fig. 5
Peak Intensity Markers

https://doi.org/10.1017/S0252921100032425 Published online by Cambridge University Press

but rather to point out that it must be done with care. Let us now examine some of the qualitative and quantitative checks that the investigator can make to determine if he has coherence problems with his instrument as set up. Firstly, it is obvious that he has such problems if his effective slit size is of the order of 0.5 micron. But suppose he is operating with larger slits (2 - 10 microns). How may he be sure he is safe? A simple test is to scan a high contrast, high quality edge. Chrome on quartz with a density of one through the evaporated metal makes an excellent test target. Also the commercially available metallic long line resolution targets of Diffraction Limited are quite satisfactory. If, when such an edge is scanned, one observes any ringing in the trace, such as that shown in Figures 4 and 5, he has coherence problems. An even more sensitive and certainly more dramatic test is to prepare a phase image and scan this with the microdensitometer. Of course, a microdensitometer linear in transmission should not "see" a phase image; if it does, the instrument as set up has coherence problems. Useful phase images may be prepared, for example, by photographing a Ronchi ruling prism and then bleaching the resultant image. One hundred to two hundred lines per millimeter is a useful frequency range. Using a periodic phase object, of course, helps the operator to detect its presence in the resultant trace. To obtain the smallest effective slit usable in his particular instrument, the operator should repeatedly scan the phase image continuously, varying the slit size until the modulation in the output is zero. (Then open it a little larger). The only way to be really sure is to actually measure the coherence of the instrument. This is relatively easily done with a Wollaston prism. It should also be remembered that the microdensitometer is a completely different instrument each time one changes the optics, condensers and objectives, and, therefore, should be re-examined when such changes are made. One may also measure the coherence by measuring the apparent transfer function by edge trace techniques and comparing the results with the Bechere-Parrent equations. This is, of course, the most simple approach if one already has a program for edge trace analysis.

We have reviewed a few of the ways in which the operator may detect the presence of coherence problems. What now, if he finds that the set up parameters required to remove the problem make it impossible to obtain the desired resolution? Firstly, the image being examined may be liquid gated, *i.e.* the film covered with an index matching liquid, thus removing phase errors. Since phase is the source of the most deleterious effects of partial coherence, this precaution will frequently allow the operator to go a few per cent small in his selection of slit size. Secondly, if it is possible, he may prepare his image in a lower contrast form, since this also tends to minimize the effect. Thirdly, if the nature of the object being scanned is well known, *e.g.* one wants to determine the separation of two lines which have a rectangular shape in transmission, one can employ the coherence theoretic treatment presented here and match the traces with the calculated results to determine the unknown parameters. This is a sort of curve fitting approach. In extreme cases, where in fact resolution of fractions of a micron are literally required, one can measure not the transmission or density but the mutual intensity function. In this case, the system is completely linear and the equation can be inverted. This seems the obvious approach but it must be remembered that there is as yet no routinely usable method of measuring the mutual intensity and each problem becomes a research project.

In the second place, the equation is cumbersome to invert.

REFERENCES

1. Beran, M. and Parrent, G. B., 1963. *Theory of Partial Coherence* (Eaglewood Cliffs, New Jersey: Prentice Hall, Jrnl.)
2. Born, M. and Wolf, E., 1959. *Principles of Optics*, London: Pergamon Press.

DISCUSSION

J. HAMBLETON: You have indicated that coherence effects in a microdensitometer can cause displacement of a recorded edge trace. Does this same displacement occur on the lagging edge of a bar?

G. B. PARRENT: Yes, it produces a non-linear shift; it will produce the same sort of shift on the other edge, with the net result that the measured width of a line or gap will be wrong. It is easy to calculate the answer for sharp edges; the same apparent shrinking occurs for non-sharp edges but is more difficult to calculate.