

UNIVERSITY OF LONDON

Imperial College of Science and Technology,
Physics Department,
Applied Optics Section.

OPTIMISATION METHODS IN OPTICAL DESIGN.

by

ROBIN WILLIAM GOSTICK B.Sc. A.R.C.S.

Thesis submitted for the degree of Doctor of Philosophy
of the University of London.

1974

Abstract.

Computer-aided optical design usually requires that the performance of an optical system is described in a single figure of merit, or merit function. The first part of the thesis discusses and compares several merit functions in current use in the light of optical transfer^{function}/theory. A new merit function, based on the geometric optical transfer function, is investigated by model aberration balancing techniques and it is shown that this merit function is valid over the whole range of practical aberration balancing. For high aberrations and spatial frequencies it is shown that this new merit function is markedly superior to any of the other proposed functions.

An optimisation programme has been written which incorporates the new merit function, and the second part of the thesis gives details of the implementation of the merit function.

Examples are given of practical lens optimisations carried out with this new programme, and it is shown that large improvements in performance can be made to systems previously optimised with conventional programmes.

Contents

	Pages
Chapter 1 Introduction	1.1-1.3
Chapter 2 Optical Imagery and the optical transfer function.	2.1-2.14
Chapter 3 Aberration balancing and merit functions.	3.1-3.21
Chapter 4 Numerical techniques.	4.1-4.8
Chapter 5 Low frequency axial aberration balancing.	5.1-5.41
Chapter 6 Coma aberration balancing.	6.1-6.21
Chapter 7 High frequency aberration balancing.	7.1-7.16
Chapter 8 Practical optimisation techniques.	8.1-8.17
Chapter 9 Practical lens optimisation.	9.1-9.29
Appendix Tables of design data.	A.1-A.7
	Acknowledgements 1 page.
	References 1 page.

Chapter 1. Introduction

The advent of high-speed computers provides the optical designer with a powerful tool for use during the final stages of design. By the process of automatic, or semi-automatic lens optimisation programmes the computer is able to vary the design parameters of the system to bring it as close as possible to the desired conditions. The designer has to give a starting design and has to specify the detailed requirements of the performance.

In general terms the quality of the system may be described by a function Ψ of the aberrations, and the task of the programme is to find the minimum of this function, subject to the physical constraints of the system and any constraints set by the designer. The function Ψ is usually termed the merit function, and this thesis is concerned with the detailed structure of merit functions in relation to image formation theory.

Since the start of the use of optimisation programmes much discussion has taken place on the subject of the merit function (see the Conference on Optimisation with Large Computers 1964). In the early part of the optimisation process the detailed structure of the merit function is not of critical importance since even a merit function based on simple aberration theory will normally produce a large improvement in image quality. The purpose of the merit function at this stage is to steer the system into a region where the aberrations are sufficiently small to allow the use of more advanced merit functions.

Perfect aberration correction is generally impossible, so the problem in the final stage of design becomes one of aberration balancing rather than one of aberration reduction. The

residual aberrations must be balanced against each other to provide the best imaging properties for the proposed use of the system.

The first merit functions (Wynne 1959) were constructed from the finite ray aberrations found by tracing several rays through the system. The merit function is defined as the weighted sum of the squares of these aberrations,

$$\psi = \sum_i (x_i \omega_i)^2 \quad 1.1$$

These ray aberrations are selected by the designer, and may include transverse, longitudinal and wave aberrations. They may also include values for effects such as astigmatism and distortion. Additional variables may be added into the function to allow for constraints on the system, such as thickness, and for variables with required, or target values. Then if x_0 represents a target value for the given variable,

$$\psi = \sum_i [(x_i - x_{0,i}) \omega_i]^2 \quad 1.2$$

This merit function is now in a form suitable for optimisation by a damped least squares technique, which is the most common method in use (Levenberg 1944, Wynne 1959).

The merit function based on a simple sum of squares of the aberrations suffers from two limitations in the final stage of design. Firstly, to produce a good aberration balance requires considerable experience on the part of the designer, since the balance depends critically on the choice of aberrations and weights used. Secondly, to find the balance giving optimum image-forming properties, the merit function should be considered from the point of view of image formation and information transfer. The diffraction theory of image formation and the optical transfer function are discussed in later chapters and their relation to optim-

isation criteria is given. In most cases of practical design the final aberrations are large compared with the wavelength of light, and in these cases the full diffraction treatment is not necessary. It will be shown, however, that even when these aberrations are of the order of the wavelength an approach based on the geometrical transfer function is still valid for optimisation.

Several authors have given details of the use of optimisation criteria based on diffraction theory, and on approximations to the diffraction transfer function. These criteria are compared with the proposed geometrical criterion, and the useful limits of the criteria have been examined. It is found that in nearly all regions of aberration size the geometric criterion gives aberration balances most nearly matching those given by the full diffraction theory.

A merit function based on the geometrical transfer function has been incorporated into the existing Imperial College SLAMS design program, and details of its successful use are given in the final sections of this thesis.

Optical Imagery and the Optical Transfer
Function

The image forming properties of an optical system may be investigated by considering the effect the system has on a wavefront passing through the system. A 'point' object emits spherical wavefronts, and for perfect imagery the wavefront emerging from the system should also be spherical. The departure of the wavefront from its spherical shape gives rise to aberrations in the image.

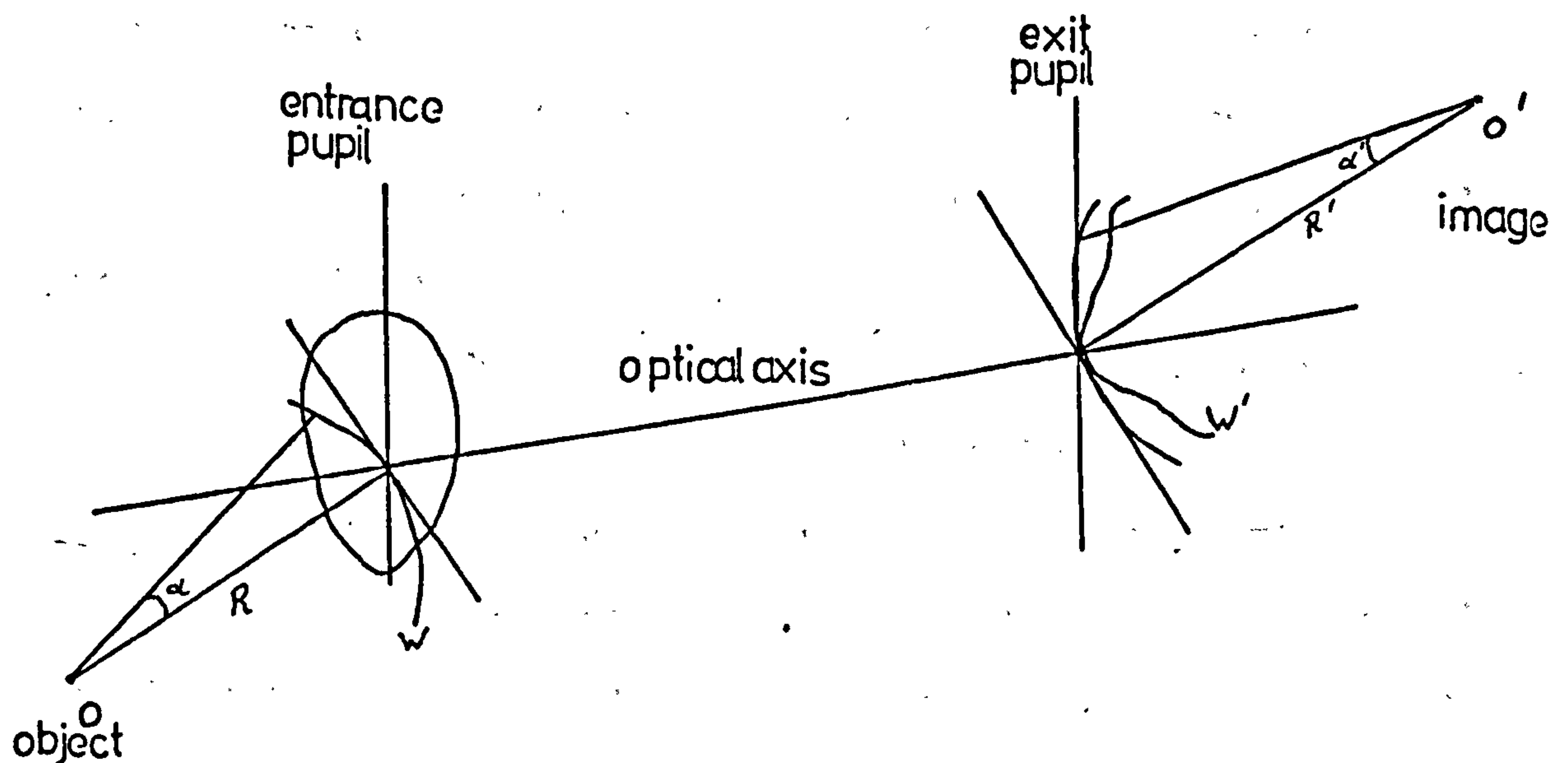


Fig 2.1

Figure 2.1 shows the general optical system, where the wavefront from the object O meets the optical system at the entrance pupil and emerges as an aberrated wavefront from the exit pupil. The aberrations are calculated with respect to a sphere centred on the geometrical, unaberrated image point O' .

The coordinate system used in the discussion is set out in fig. 2.2.

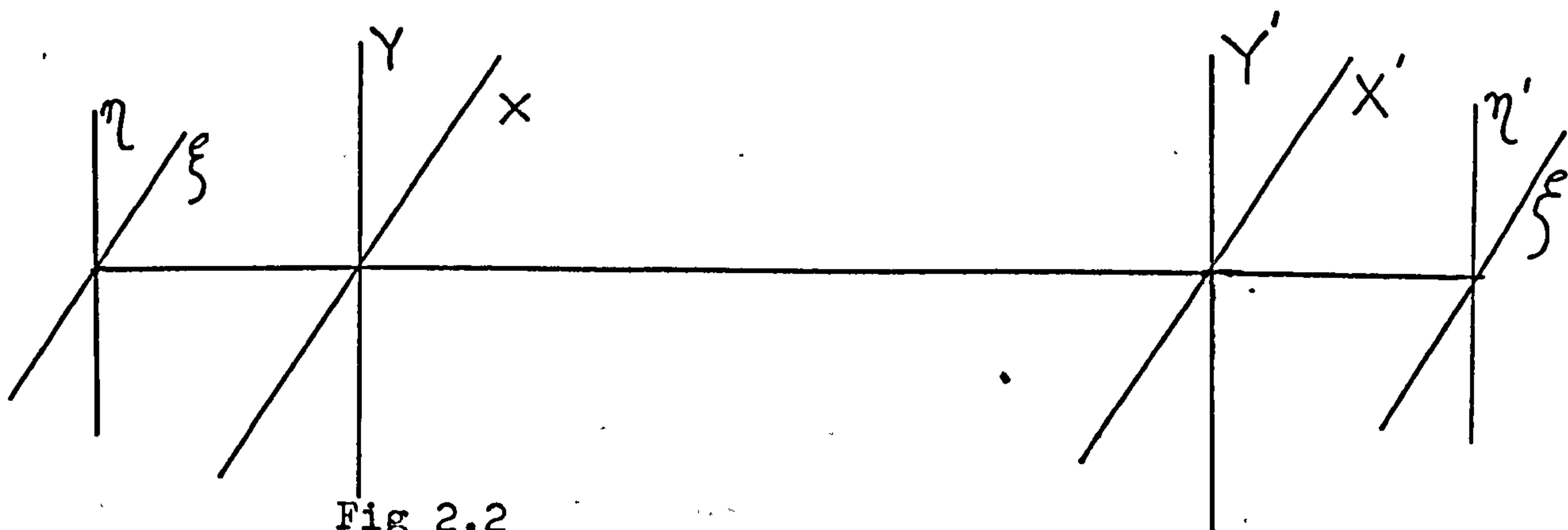


Fig 2.2

For the centred optical system, in which all the surfaces have rotational symmetry about a common axis, symmetry allows us to consider objects in one direction only in the object plane. Conventionally this is taken as the η direction. The plane containing the object and the axis of symmetry is called the meridian plane.

The treatment is further simplified by the use of normalised canonical coordinates for the optical system. In the entrance and exit pupils these are denoted by lower case letters (x, y) , (x', y') , and in the object and image planes by (u, v) , (u', v') .

For the axial object, if h, h' are the heights of the marginal ray at the entrance and exit pupils, then

$$\begin{aligned} x &= \frac{X}{h} & y &= \frac{Y}{h} \\ x' &= \frac{X'}{h} & y' &= \frac{Y'}{h} \end{aligned} \quad 2.1$$

The canonical object and image coordinates are given in terms of the numerical aperture of the system $n \sin \alpha$, $n' \sin \alpha'$ where n, n' are the refractive indices in the object and image spaces and $\sin \alpha = h/R$, $\sin \alpha' = h'/R'$, where R, R' are the distances from the object and image to the entrance and exit pupils. R and R' are the radii of the reference spheres as shown in fig. 2.1. The canonical coordinates are then

2.3

$$u = \frac{n \sin \alpha}{\lambda} \xi \quad u' = \frac{n' \sin \alpha'}{\lambda} \xi'$$

$$v = \frac{n \sin \alpha}{\lambda} \eta \quad v' = \frac{n' \sin \alpha'}{\lambda} \eta' \quad 2.2$$

Hopkins (1966) gives a full treatment of the canonical coordinates for the axial and extra-axial objects, and shows that if the isoplanatic, or space invariant condition is valid

$$\begin{aligned} x &= x' & y &= y' \\ u &= u' & v &= v' \end{aligned} \quad 2.3$$

The isoplanatic condition, as applied to optical imagery, states that the aberrations of the system must be slowly varying, such that a small movement of the object gives rise to a corresponding movement of the image. This is true for most lens systems, and is assumed to hold in the rest of this treatment.

With this canonical convention the general form of the aberration of the emergent wavefront may be expressed as a polynomial in the pupil coordinates (x, y) and the field variable v . The symmetry permits the use of the alternative pupil coordinates (ρ, ϕ) where

$$\rho^2 = x^2 + y^2, \quad \phi = \tan^{-1}(x/y) \quad 2.4$$

The aberration polynomial may then be written

$$W = \sum_{i,j,k} W_{ijk} \rho^i \cos^j \phi v^k \quad 2.5$$

In considering aberration balancing it is usual to consider the aberrations from a given object point, and the field

variable contribution is then incorporated into the coefficient to give the simpler polynomial

$$W = \sum w_{ij} \rho^i \cos^j \phi \quad 2.6$$

Symmetry considerations permit only certain combinations of i and j , and the first few terms of the expanded polynomial are

$$\begin{aligned} W = & W_{00} + \\ & W_{20} \rho^2 + W_{11} \rho \cos \phi + \\ & W_{40} \rho^4 + W_{31} \rho^3 \cos \phi + W_{22} \rho^2 \cos^2 \phi \\ & W_{60} \rho^6 + W_{51} \rho^5 \cos \phi + W_{42} \rho^4 \cos^2 \phi + W_{33} \rho^3 \cos^3 \phi \end{aligned} \quad 2.7$$

The significance of the various terms will be discussed in later sections.

For an unaberrated wavefront the image will not be an exact point, but will be the Airy disc, as found by diffraction theory. The light intensity $I(u)$ is given by

$$I(u) = \left| 2 J_1(2\pi u) / (2\pi u) \right|^2 \quad 2.8$$

where the intensity is normalised to unity at the focus ($u=0$). This distribution is shown in fig.2.3.

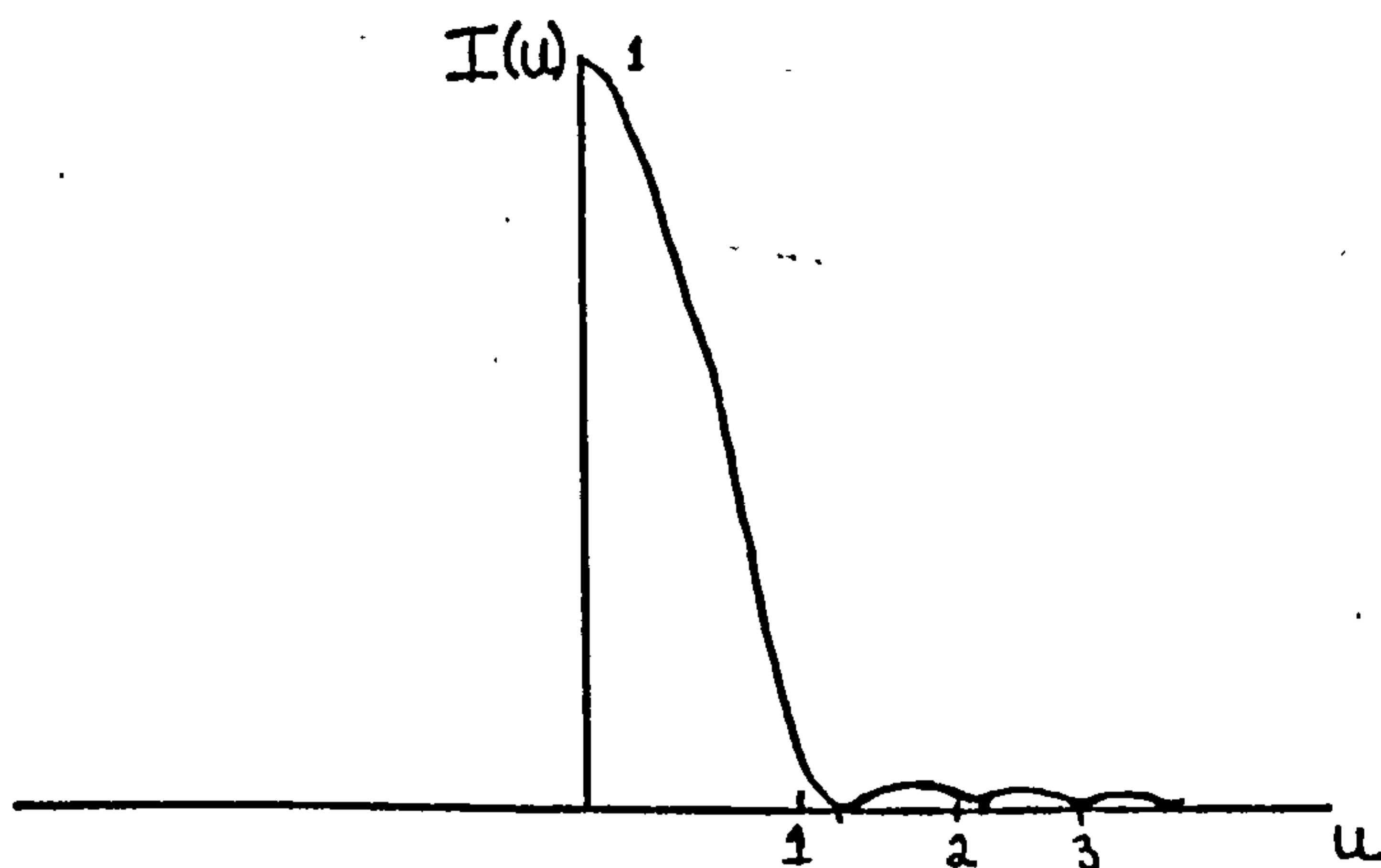


Fig 2.3 The Airy Distribution

The effect of small aberrations will be to reduce the central intensity of the distribution, with an increase of energy in the outer region or halo. Strehl (1902) uses this reduction in the central intensity as a single figure of merit to describe the effect of aberrations. This figure of merit, called the Strehl ratio or Definitionshelligkeit, is defined as the ratio of the image irradiance at the diffraction focus with the aberration present to that without aberration. The intensity in the presence of the aberration is given by the diffraction formula

$$I = \left| \iint_A \exp(ikW(x,y)) dx dy \right|^2 \quad 2.9$$

where $k = 2\pi/\lambda$.

Maréchal (1940) considers the expansion of the Strehl ratio. If $\bar{W}$ denotes the mean value of the wavefront aberration over the pupil A then eqn.2.9 may be expressed as

$$I = \left| \exp(ik\bar{W}) \iint_A \exp(ik(W-\bar{W})) dx dy \right|^2 \quad 2.10$$

Maréchal showed that when $I > 0.8$ the integrand may be approximated by the first three terms of the exponent, giving

$$I = \left| 1 - (2\pi^2/\lambda^2)E \right|^2 \quad 2.11$$

where

$$E = \iint_A W^2 dx dy - \left| \iint_A W dx dy \right|^2 \quad 2.12$$

This approximation to the Strehl ratio provides a simple diffraction-based criterion for aberration balancing, as discussed by Maréchal. This criterion is further investigated in this thesis, and it will be shown that it provides a valuable method for obtaining an overall frequency balance.

The Optical Transfer Function (o.t.f.)

Fourier theory leads to the concept of synthesising the object distribution into spatial frequency components. By this method the object may be considered as the superposition of a number of sinusoidal distributions of varying spatial frequency, amplitude and azimuth. Providing the isoplanatic condition holds, it will be shown that each sinusoidal component in the object is imaged as a sinusoidal distribution, so that the system acts as a passive linear filter of spatial frequencies. If the amplitude of a given spatial frequency component S in the object is $b(S)$ the effect of a linear filter will be to reduce the amplitude by a factor $M(S)$ and to introduce a phase shift $\theta(S)$. The image spectrum $b'(S')$ will then be given by

$$b'(S') = M(S)b(S)\exp(i\theta(S)) \quad 2.13$$

For a system of magnification m

$$S' = S / m \quad 2.14$$

The optical transfer function is then defined as

$$D(S) = M(S)\exp(i\theta(S)) \quad 2.15$$

where $M(S)$ is the modulation transfer function, m.t.f.

For incoherent light we can consider each spatial frequency as giving an intensity fluctuation about a mean intensity where the distribution has the form

$$B(\xi) = \alpha + \beta \sin(2\pi S \xi) \quad 2.16$$

The ratio of the fluctuation to the mean, β/α , gives the contrast of the object. The effect of the aberrated system will be to reduce this contrast by $M(S)$ and to introduce the phase shift $\theta(S)$, giving the image distribution

$$B'(\xi) = \alpha + \beta M(S) \sin(2\pi S' \xi' + \theta(s)) \quad 2.17$$

This is illustrated in figure 2.4.

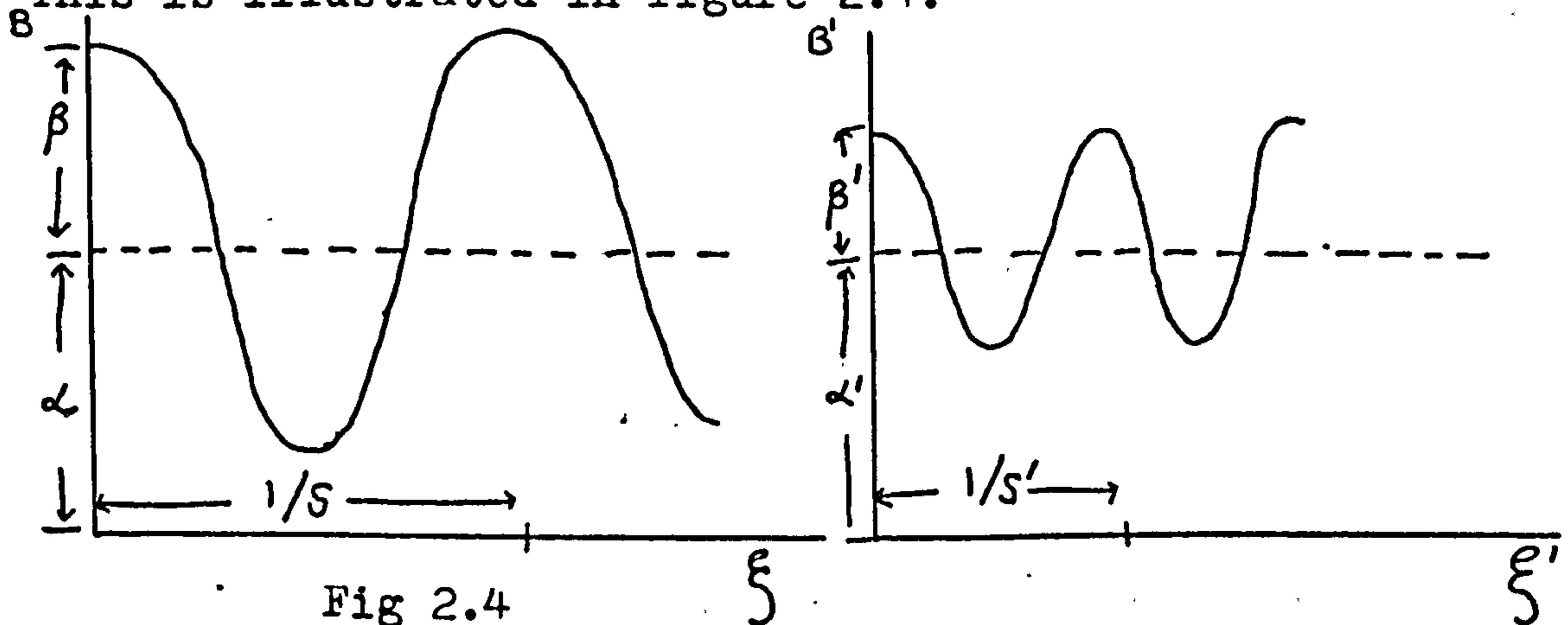


Fig 2.4

Conventionally we take S to be the frequency of a line structure parallel to the y -axis and T to be the frequency of a line structure parallel to the x -axis, the sagittal and tangential lines. In terms of the canonical coordinates we can introduce the canonical spatial frequency variables, (s, t) (s', t') , where

$$s = \frac{\lambda}{n \sin \alpha} S$$

$$s' = \frac{\lambda}{n' \sin \alpha'} S'$$

$$t = \frac{\lambda}{n \sin \alpha} T$$

$$t' = \frac{\lambda}{n' \sin \alpha'} T' \quad 2.18$$

Once again there is the correspondence between the object and image spaces, and $s = s'$, $t = t'$. We may consider without loss of generality a single distribution of frequency s at an angle ϕ to the y -axis and may rotate the axis to lie in the direction of the line structure.

We now introduce the pupil function $f(x, y)$ defined by

$$f(x, y) = \exp(ikW(x, y)) \text{ within the pupil}$$

$$= 0$$

elsewhere

2.19

From the diffraction theory of image formation the complex amplitude $F(u, v)$ due to a point object at $(0, 0)$ in the object plane is given by

$$F(u,v) = \iint_{-\infty}^{\infty} f(x,y) \exp(2\pi i(ux+vy)) dx dy \quad 2.20$$

which is just the Fourier transform of the pupil function. The intensity $G(u,v)$ is given by the square of the complex amplitude, so

$$G(u,v) = |F(u,v)|^2 \quad 2.21$$

$G(u,v)$ is known as the point spread function, and for the unaberrated pupil gives the Airy pattern, fig.2.3.

The intensity at $(u,0)$ due to a line object along $u = 0$ is given by the integral of the point spread function

$$G(u) = \int_{-\infty}^{\infty} G(u,v) dv \quad 2.22$$

$G(u)$ is known as the line spread function. For an isoplanatic system in incoherent light the intensity due to a series of line structures $B(u)$ will be given by

$$B'(u) = \int_{-\infty}^{\infty} B(u) G(u'-u) du' \quad 2.23$$

This has the form of a convolution integral, so if the Fourier transforms of B , B' and G are given by

$$b(s) = \int_{-\infty}^{\infty} B(u) \exp(-2\pi ius) du$$

$$b'(s) = \int_{-\infty}^{\infty} B'(u) \exp(-2\pi ius) du \quad 2.24$$

$$g(s) = \int_{-\infty}^{\infty} G(u) \exp(-2\pi ius) du$$

then the convolution theorem gives

$$b'(s) = b(s)g(s) \quad 2.25$$

Equation 2.25 has the form of the o.t.f. equation 2.13, and since we have shown that the image spectrum due to a one dimensional object is simply the object spectrum multiplied by the factor $g(s)$, the optical system is acting as a linear spatial filter. For normalisation the factor $g(s)$ is divided by $g(0)$ to give the one-dimensional o.t.f.

$$D(s) = g(s)/g(0) \quad 2.26$$

Thus the one dimensional o.t.f. is the Fourier transform of the line spread function.

Returning to the two-dimensional case, similar equations hold, with

$$g(s,t) = \iint_{-\infty}^{\infty} G(u,v) \exp(-2\pi i(us+vt)) du dv \quad 2.27$$

and
$$D(s,t) = g(s,t)/g(0,0) \quad 2.28$$

The two-dimensional o.t.f. is then the transform of the point spread function. Using equations 2.20 and 2.21, equation 2.28 can be transformed to give

$$g(s,t) = \iint_{-\infty}^{\infty} f(x,y) f^*(x-s,y-t) dx dy \quad 2.29$$

The normalised o.t.f. becomes, after a shift of origin

$$D(s,t) = \frac{1}{A} \iint_{-\infty}^{\infty} f(x+s/2,y+t/2) f^*(x-s/2,y-t/2) dx dy \quad 2.30$$

Where, as before, A is the pupil area.

Here the o.t.f. is expressed as the autocorrelation of the pupil function $f(x,y)$.

In terms of the wavefront aberrations the o.t.f. integral eqn.2.30 becomes

$$D(s,t) = \frac{1}{A} \iint_A \exp(ik(W(x+s/2, y+t/2) - W(x-s/2, y-t/2))) dx dy \quad 2.31$$

The area of integration of the autocorrelation integral is the area common to the two pupils displaced by $(\pm s/2, \pm t/2)$ with respect to each other, as shown in fig.2.5.

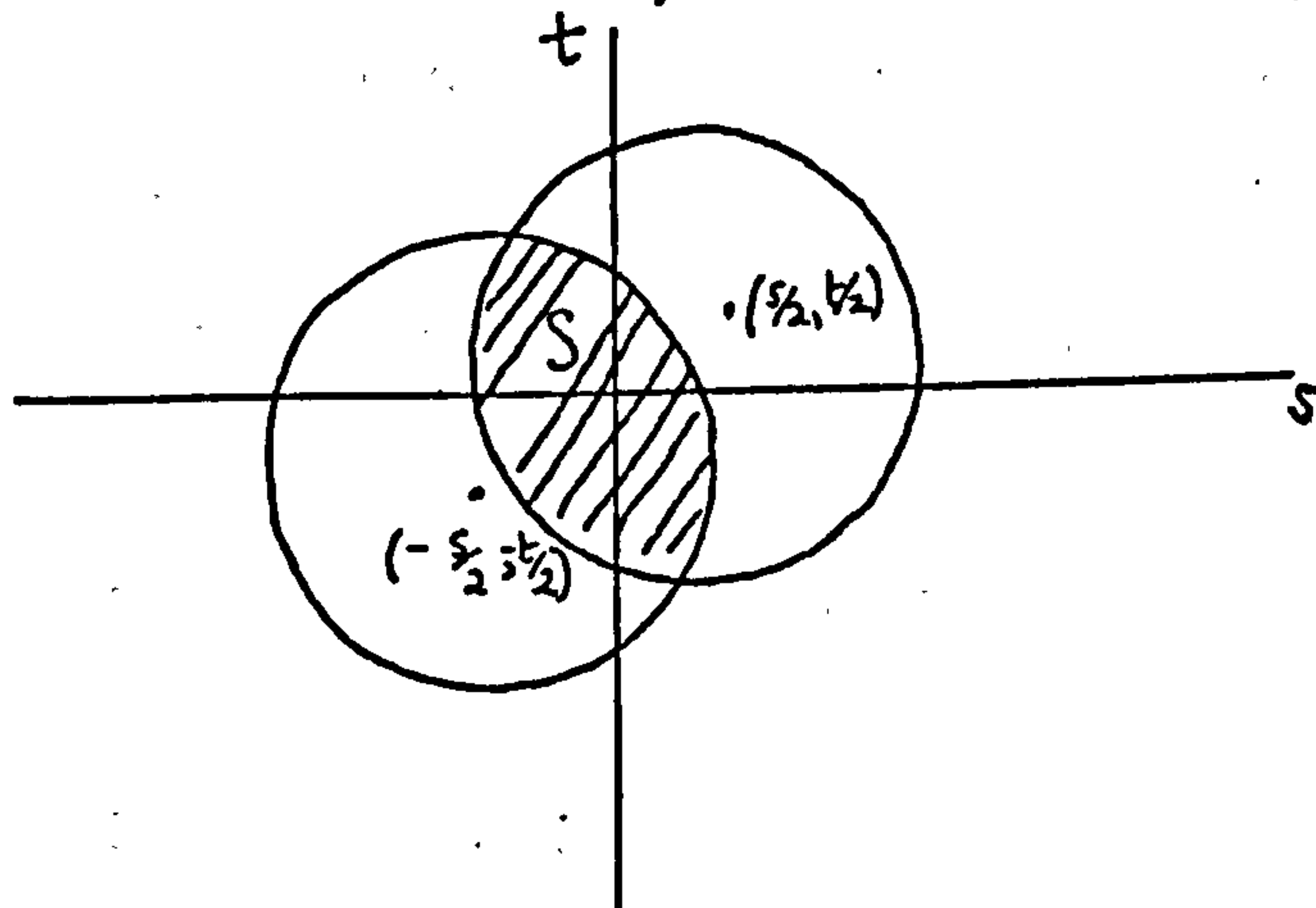


Fig 2.5

This common area, which will be denoted by S , limits the reduced spatial frequencies to lie in the range ± 2 . For a typical photographic lens this corresponds to 1000 cycles per millimetre for an $f/2$ aperture.

An important result is found by considering the integrated response function. Integrating equation 2.27 gives

$$\begin{aligned} \iint_0^2 g(s,t) ds dt &= \iint_0^2 \iint_A G(u',v') \exp(-2\pi i(u's + v't)) du' dv' ds dt \\ &= \iint_A \left\{ \iint_0^2 \exp(-2\pi i(u's + v't)) ds dt \right\} G(u',v') du' dv' \\ &= \iint_A \delta(u',v') G(u',v') du' dv' \\ &= G(0,0) \end{aligned} \quad 2.32$$

But $G(0,0)$ is just the intensity at the centre of the point spread function, and thus represents the Strehl intensity.

The Strehl intensity ratio may be considered as the integrated response, and may be used as an optimisation criterion for

overall frequency response. An increase in the Strehl ratio should give an increase in the overall frequency response, but it should be determined which parts of the frequency spectrum contribute most to the Strehl ratio, and hence which parts will be improved by optimising this. This has been investigated in later sections of the thesis, and it will be shown that the Strehl ratio provides a useful method for optimisation of low and medium spatial frequencies.

The full diffraction integral, 2.31, is complicated to compute, and it has not yet been found possible to use it for direct practical optimisation. In this thesis a new merit function has been developed, based on the geometrical approximation to the diffraction o.t.f., and its use is justified both theoretically and practically in later sections. The geometric o.t.f., GOTF, has been used for many years as an easily computed approximation to the full diffraction o.t.f., DOTF, although its use has been restricted to the analysis of systems designed by other criteria, and it has not previously been used for direct optimisation.

The geometrical o.t.f. may be derived by considering the Fourier transform of the point spread function computed from ray intersection densities (Hopkins 1957), as in equation 2.27 for the diffraction o.t.f. It is more convenient and informative here to use the equivalent derivation which considers the effect on the diffraction o.t.f. of taking the geometrical limit of zero wavelength.

Considering the exponent of the integrand in equation 2.30

we may write

$$sV = W(x+s/2, y+t/2) - W(x-s/2, y-t/2) \quad 2.33$$

If we consider again the case of the single spatial frequency s , we may expand W in a Taylor series about (x, y) to give

$$sV = sW'(x, y) + \frac{1}{3!} \frac{s^3}{4} W'''(x, y) + \frac{1}{5!} \frac{s^5}{16} W^{(5)}(x, y) + O(s^7)$$

$$\text{Then } V = W'(x, y) + \frac{1}{3!} \left(\frac{s}{2}\right)^2 W'''(x, y) + O(s^4) \quad 2.34$$

From equation 2.18 as the wavelength λ tends to zero, s tends to zero, so in the limit

$$V = W'(x, y) \quad 2.35$$

Then the geometrical o.t.f. $D_g(s)$ becomes

$$\begin{aligned} D_g(s) &= \frac{1}{A} \iint \exp(iksV) dx dy \\ &= \frac{1}{A} \iint \exp(iksW'(x, y)) dx dy \end{aligned} \quad 2.36$$

In the geometrical limit the area of integration becomes the area of the pupil, which in the absence of vignetting is the unit circle for the normalised coordinates. Although both s and λ become zero the product $ks = 2\pi s/\lambda$ remains finite. The differences between the geometric and diffraction o.t.f. come from the truncation of the Taylor series in equation 2.34 and from the difference in the areas of integration. The importance of the difference in areas is easily shown in the absence of aberrations. The diffraction o.t.f. in the absence of aberrations $D_o(s)$ has the form

$$D_o(s) = \frac{2}{\pi} \left(\cos^{-1} \left(\frac{s}{2} \right) - \frac{s}{2} \sin \left(\cos^{-1} \left(\frac{s}{2} \right) \right) \right) \quad 2.37$$

This is shown in fig.2.6



Fig 2.6

The Diffraction O.T.F. of an unaberrated system

The geometric o.t.f. is unity for all spatial frequencies in the absence of aberrations, and the slope of the GOTF curve is zero at zero frequency for all aberrations. Both of these results are due to the difference in the areas of integration, but both of these effects may be overcome by using the mixed treatment due to Hopkins (1957). If relative modulation values are considered we can write

$$M(s,t)\exp(i\theta(s,t))=D(s,t)/D_0(s,t) \quad 2.38$$

Since, as shown, the geometric o.t.f. is unity at zero aberrations the values for $D_0(s,t)$ are the diffraction o.t.f. values. Then for the geometric case we have

$$M_g(s,t)\exp(i\theta(s,t))=D_g(s,t)/D_0(s,t) \quad 2.39$$

Values for the mixed treatment agree more closely with the diffraction o.t.f. values than the geometric o.t.f. values alone. (Hopkins 1957, Miyamoto 1957)

Miyamoto, Hopkins and other authors have considered the accuracy of the geometric and mixed responses. Miyamoto showed that the mixed treatment gave results almost identical to the diffraction values for aberrations less than 0.1λ

and for aberrations greater than 2λ . This latter result agrees with those of Hopkins for small spatial frequencies.

While these and similar investigations provide useful details of the accuracy of the geometric calculation, they are to a large extent irrelevant for the purposes of aberration balancing. For a criterion to be valid for aberration balancing, all that is required is that a parameter change giving an improvement in the chosen criterion also gives an improvement in the diffraction o.t.f. Strict numerical comparison between the values is not, therefore, significant.

It will be shown that for most practical cases the geometrical criterion produces aberration balances sufficiently close to those which would be produced by direct optimisation of the diffraction o.t.f. for it to be a valid merit function. It will further be shown that in nearly all cases where an o.t.f. based criterion is required the geometric criterion is superior to criteria already used in optimisation programmes or proposed as optimisation criteria.

Aberration Balancing and Merit Functions

Since the first introduction of optimisation programmes many aberration balancing merit functions have been suggested and several of these have been incorporated into working programmes. The complexity of these programmes, and the fact that many of the proposed criteria have not yet been programmed makes direct comparison of these criteria from practical results impossible.

It is possible, however, to consider the effects of the criteria on simplified models, and this approach produces sets of aberration balances which are easy to interpret and compare. The basic method is to take certain terms from the general aberration polynomial in equation 2.7, give fixed values to one or more coefficients, then allow the other terms to vary to optimise the chosen criterion. Several criteria give analytic formulae for the optimum balances for the simple models, as given by Maréchal (1940), Hopkins (1957) and others. Other criteria, such as the diffraction and geometric o.t.f. integrals, do not give analytic balances without making approximations so these have to be studied by numerical methods. Goodbody (1958) and Barakat (1963) provide numerical studies of the diffraction o.t.f., but neither paper uses direct optimisation techniques and consist of sets of o.t.f. values for ranges of aberration coefficients. Their results, although limited in range, agree with the results found in this present study.

In this study numerical methods are used to directly optimise both the diffraction and geometric o.t.f. with respect to the aberration coefficients, and in this way gives

balances which can be compared with those given by analytic methods. A wider range of values of aberrations has been used than in previous studies. As well as providing an illustration of the relative merits of the various criteria the results found from the diffraction o.t.f. optimisation give valuable details of the effect of aberration balancing on the o.t.f. One of the most significant effects is the variation of the optimum wavefront shape with both spatial frequency and the size of the aberrations. This behaviour is expected from practical experience, but the exact form of the variation had not been studied. Any proposed merit function which attempts to optimise, directly or indirectly, the o.t.f. must be able to take account of this variation. None of the analytic or other previously proposed criteria is capable of following the variation due to aberration size because, as will be shown, these criteria are based on homogeneous quadratic forms of the aberration coefficients, and as such their optimum wavefront shapes are independent of aberration size. This effect limits the range of usefulness of these criteria to regions of small aberrations. The new merit function proposed by this study, based on the geometric o.t.f. does not suffer from this drawback, and it will be shown that it is most efficient for large aberrations, where the other criteria have ceased to be useful for o.t.f. optimisation.

In selecting the criteria to be investigated the original type of merit function, equation 1.2, based on individually weighted aberrations has been omitted. With this merit function the experience of the designer enables the function to be modified to give different weights to the various areas of the pupil to produce balances according to the required

use of the system. Effectively the merit function, by a suitable choice of weights, can be made similar to many of the other criteria. This study is concerned with merit functions based on more automatic methods, where the weighting factors are partially set by the programme, and the designer need only set relative weights, to the various field angles and colours.

Listed below are the merit functions considered, although it will be shown that they are not all independent.

- a. Root-mean-square image spot size
- b. Minimum image spot size
- c. Mean-square value of the wave aberration
- d. Maréchal approximation to the Strehl ratio
- e. Hopkins approximation to the diffraction o.t.f.
- f. Hopkins approximation to the geometric o.t.f.
- g. The geometric o.t.f.
- h. The diffraction o.t.f.

To compare these criteria two cases of aberration balancing have been considered, spherical aberration and coma.

For the axial object the only terms remaining in the aberration polynomial 2.7 are the spherical aberration terms and the defocus term. These are

$$\begin{aligned}
 W &= W_{20} \rho^2 && \text{defocus} \\
 &+ W_{40} \rho^4 && \text{primary spherical aberration} \\
 &+ W_{60} \rho^6 && \text{secondary spherical aberration} \\
 &+ W_{80} \rho^8 && \text{higher order terms}
 \end{aligned}
 \tag{3.1}$$

For practical aberration balancing the low order terms, the primary aberrations, vary rapidly with system parameter changes while the higher order terms are slowly varying. For this

reason in the simplified model of aberration balancing the high order coefficients are usually given fixed values, and the lower order coefficients are allowed to vary. For the axial case W_{60} is given a fixed value and each criterion is optimised with respect to W_{20} and W_{40} .

As the object moves away from the axis, the first terms to have significant values are the coma and transverse focal shift terms, which give

$$\begin{aligned} W = & W_{11} \rho \cos \phi \text{ transverse focal shift} \\ & + W_{31} \rho^3 \cos \phi \text{ primary coma} \\ & + W_{51} \rho^5 \cos \phi \text{ secondary coma} \\ & + W_{71} \rho^7 \cos \phi \text{ higher order terms} \end{aligned} \quad 3.2$$

In this case W_{51} is given a fixed value and W_{11} and W_{31} are allowed to vary. All terms of higher orders are omitted.

In this part of the study we are primarily concerned with optimisation of systems which require a high o.t.f. value, so an o.t.f. based criterion is appropriate, and the true optimisation criterion is, of course, the diffraction o.t.f. In this respect the criteria under investigation are all compared with the diffraction o.t.f. It must be realised that many optical systems do not require an o.t.f. criterion. Astronomical systems, for instance, normally require a system with the optimum resolving power, or minimum image spot size. The choice of optimisation criterion must be closely related to the proposed use of the system, and this is discussed in greater detail in chapter 8.

The two polynomials 3.1 and 3.2 are conveniently written in the alternative forms

$$W = W_{60}(B_{26}\rho^2 + B_{46}\rho^4 + \rho^6) \quad 3.3$$

and

$$W = W_{51}(B_{15}\gamma + B_{35}\rho^2\gamma + \rho^4\gamma) \quad 3.4$$

where

$$\begin{aligned} B_{26} &= \frac{W_{20}}{W_{60}} & B_{46} &= \frac{W_{40}}{W_{60}} \\ B_{15} &= \frac{W_{11}}{W_{51}} & B_{35} &= \frac{W_{31}}{W_{51}} \end{aligned} \quad 3.5$$

Details of the criteria are given below, and their application to the two cases are given in the next two chapters.

a. Root-mean-square image spot size

For a given wavefront aberration $W(x,y)$ the transverse aberrations $(\delta\xi, \delta\eta)$ are given by

$$\begin{aligned} n \sin \alpha \cdot \delta\xi &= \delta\zeta = -\partial W / \partial x \\ n \sin \alpha \cdot \delta\eta &= \delta H = -\partial W / \partial y \end{aligned} \quad 3.6$$

$$\text{We then define } r^2 = \delta\zeta^2 + \delta H^2 \quad 3.7$$

and the root-mean-square image spot size is then given by

$$\begin{aligned} \langle r^2 \rangle &= \frac{1}{A} \iint r^2 dA \\ &= \frac{1}{A} \iint (\delta\zeta^2 + \delta H^2) dA \\ &= \frac{1}{A} \iint \left[\left(\frac{\partial W}{\partial y} \right)^2 + \left(\frac{\partial W}{\partial x} \right)^2 \right] dA \end{aligned} \quad 3.8$$

Differentiation of the wavefront polynomials 3.3 and 3.4 and substitution into equation 3.8 enables the optimum balance to be evaluated by differentiating with respect to the B_{ij} coefficients of equation 3.5. This gives rise to analytic solutions for the best wavefront shape, but as stated previously this shape will be independent of aberration size. This treatment, based on transverse aberrations, does not take account of diffraction effects, and its use should

be restricted to the case of large aberrations relative to the wavelength of light. This restriction applies to all criteria based on transverse aberrations, but it is not always clear at which point an aberration may be considered to be 'large'. Accordingly this investigation has considered each criterion over a wide range of aberration sizes, and it will be shown that in the case of the geometric o.t.f., which is based on transverse aberrations, that the restriction to large aberrations is not a limiting factor.

b. Minimum image spot size

The r-m-s spot size criterion provides images consisting of bright central regions surrounded by a diffuse halo. For some applications it is required that the image spot size is as small as possible, and a minimum spot size criterion should be used. This is obtained by minimising the distance between the largest positive and negative transverse aberrations.

From equation 3.6 there will be a maximum or minimum transverse aberration when $\frac{\partial^2 W}{\partial \rho^2} = 0$. In general, since the aperture limits ρ to be less than unity, the value of the aberration at the edge of the pupil, $\rho = 1$, may also be numerically large, even though it is not a differential extremum. This criterion would be difficult to incorporate into an automatic optimisation program working on a limited selection of rays, but has been included here for completeness.

The next four criteria fall into the category of variance methods.

c. Mean-square value of the wave aberration

If the wavefront aberration is measured with respect to a spherical surface, radius R , centred on the geometrical image point O' , as shown in fig.2.1, we may define the mean square wave aberration $\bar{w}^2$ as

$$\bar{w}^2 = \frac{1}{A} \iint_A w^2 dA \quad 3.9$$

This criterion is analagous to the r-m-s spot size criterion. Both of these criteria have been used in standard optimisation programmes for many years, since they are the basis of the general merit function (1.2) when the selected rays are evenly distributed over the pupil, and the weighting factors are constant over each field angle.

d. Variance of the wavefront aberration

The r-m-s wavefront aberration criterion may be usefully modified by incorporating a facility to enable the reference radius, R , to vary. If R is chosen to make the average wavefront aberration zero, the previous criterion leads to the variance criterion E , where

$$\begin{aligned} E &= \bar{w}^2 - (\bar{w})^2 \\ &= \frac{1}{A} \iint_A w^2 dA - \left\{ \frac{1}{A} \iint_A w dA \right\}^2 \end{aligned} \quad 3.10$$

This facility has been available in the Imperial College programmes for some time, and it can be seen from equation 3.10 that it is simply the variance used in the Maréchal approximation to the Strehl ratio in equation 2.12. If the Maréchal approximation is accurate, therefore, this criterion should optimise the overall frequency balance, as shown in equation 2.32. Maréchal took the approximation to be accurate for systems having a Strehl ratio greater than 0.8. Since it

is the simple Maréchal variance that will be used in this study rather than the more complex Strehl integral, it is appropriate at this stage to give a comparison of the two criteria outside the strict limits set by Maréchal. The numerical comparison between the exact integral, 2.9, and the approximation 2.11 is shown in fig.3.1, where the integrals have been evaluated with varying amounts of second order spherical aberration W_{60} . King (1968) investigated the accuracy of the Maréchal approximation from aberration balancing considerations, when as has been stated the strict numerical comparison is unimportant. Although direct optimisation techniques were not used in his study, the results provide some useful general features. King concluded that the approximation was a valid criterion for optimising the Strehl intensity when this was greater than 0.5. This conclusion appears valid from similar trials carried out as part of this study, but since this study is concerned with optimisation of the o.t.f. it was decided to use the variance as a criterion even outside this region because it is the variance which has been used in several optimisation programmes rather than the Strehl ratio itself. It will be shown that the variance has significance as an o.t.f. optimising criterion even for systems with 20 wavelengths of high order aberration, when the Strehl ratio has decreased to 0.03.

Once again the criterion contains only quadratic powers of the aberration polynomial coefficients, and will be unable to compensate for the change of optimum o.t.f.-balanced wavefront shape with aberration size. However the Strehl criterion attempts to provide an overall frequency response, and this investigation shows that the variance criterion is still

Marechal Approximation: Strehl Intensity

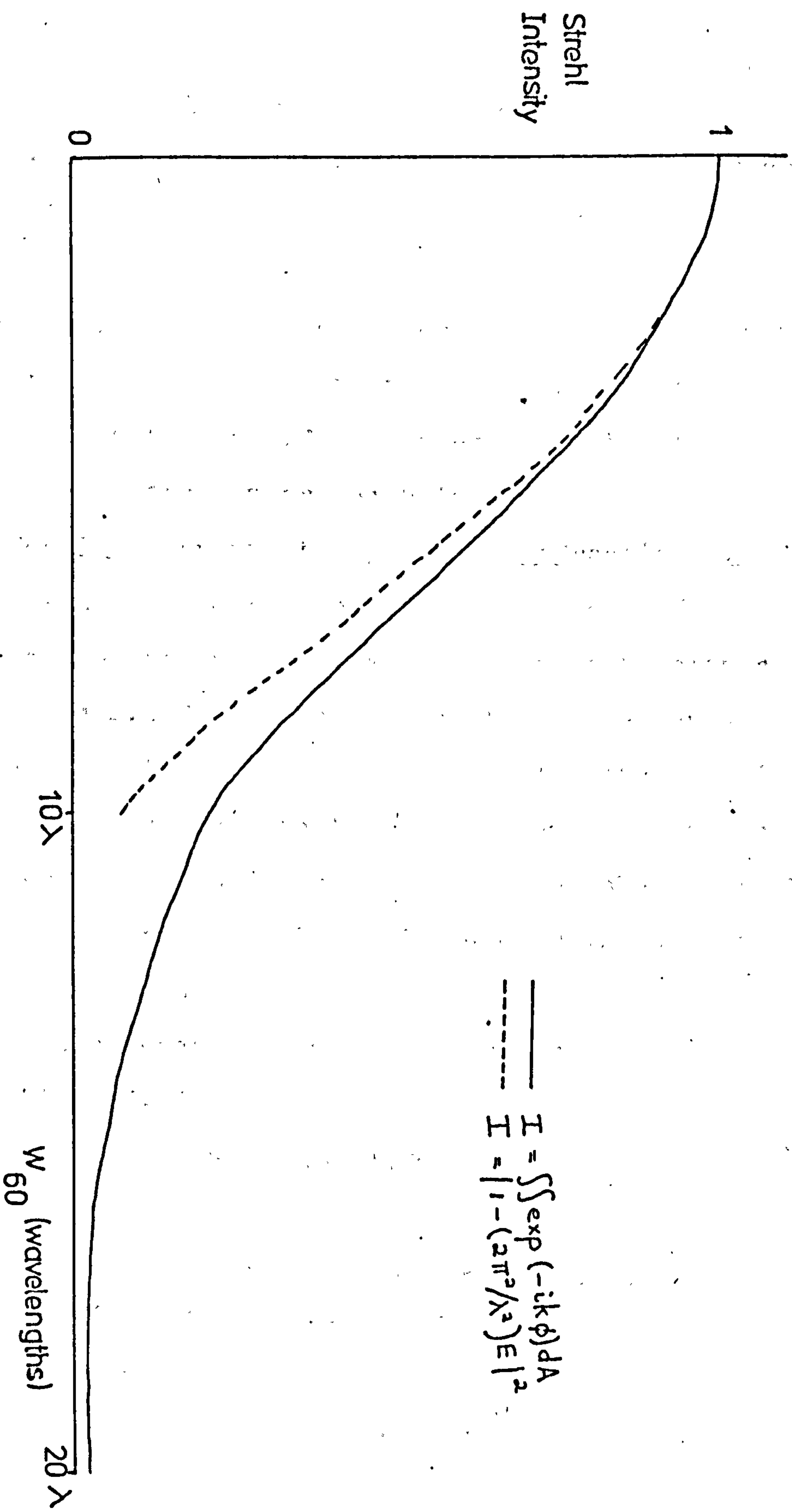


FIG 31

useful for this purpose for high aberrations. When high o.t.f. is required, either at a high spatial frequency or overall, it will be shown that for practical optimisation it is best to use a variance method for initial optimisation before the direct o.t.f. criterion is used for final balancing.

Hopkins (1957) investigated approximations to the geometric and diffraction o.t.f. for aberration balancing, and his treatments are summarised below.

e. Hopkins approximation to the diffraction o.t.f.

Hopkins considered aberration balancing for the purpose of aberration tolerancing, and takes as a tolerance that the relative m.t.f. of the system at the frequency of interest should not fall below 0.8. Since his investigations did not consider systems with larger aberrations than the tolerance values, it is to be expected that the aberration balancing formulae derived will be of use only in the limited range of high performance.

For the general case of a single spatial frequency, s , from equation 2.31 we can write

$$D(s) = 1/A \iint_S \exp(iksV) dA \quad 3.11$$

where

$$sV = W(x+s/2, y) - W(x-s/2, y) \quad 3.12$$

Then if $\bar{V}$ is the average of V over the region S , we may write equation 3.11 as

$$D(s) = \exp(iks\bar{V})/A \iint_S \exp(iks(V-\bar{V})) dA \quad 3.13$$

Putting $D(s)$ into modulus $M(s)$ and phase $\phi(s)$ form, and expanding the exponential inside the integral of 3.13 gives

$$\begin{aligned}
 M(s)\exp(i\theta(s)) &= \exp(iks\bar{V}) \iint_S \{1 + iks(V-\bar{V}) - \frac{1}{2}k^2s^2(V-\bar{V})^2\} dA \\
 &= \exp(iks\bar{V}) \left\{1 - \frac{2\pi^2s^2}{\lambda^2} K(s)\right\}
 \end{aligned}
 \tag{3.14}$$

Since we are now taking relative m.t.f. values the factor $1/A$ has disappeared, and we may now separate equation 3.14 into modulus and phase terms, giving

$$M(s) = 1 - \frac{2\pi^2s^2}{\lambda^2} K(s) \tag{3.15}$$

$$\theta(s) = ks\bar{V}$$

where

$$K(s) = \iint_S V^2 dA - \left\{ \iint_S V dA \right\}^2 \tag{3.16}$$

The relationship between $M(s)$ and $K(s)$ is analogous to the relationship between I and E for the Maréchal approximation to the Strehl ratio. In this case K is the variance of the wave aberration difference function V , and is dependant on the spatial frequency s . It should be noted that the area of integration is now the reduced autocorrelation area S .

Examination of the variance shows that once again the criterion contains a quadratic form of the aberration coefficients, so the optimised balances will be independant of aberration size. It will be shown that this seriously limits this criterion when it attempts to optimise the o.t.f. at high aberrations for all but very low spatial frequencies.

King (1969) investigated the use of this variance in an identical manner to his investigation of the Maréchal approximation. For axial aberration balancing he showed that the variance is useful in the range where the m.t.f. is greater than 0.69. This appears consistent with the results reported later in this thesis, where the analytic balances derived by Hopkins are compared with those found by direct

optimisation techniques, for the first time.

f. Hopkins approximation to the geometric o.t.f.

Hopkins (1957) uses a similar approximation to the one above for the case of aberration balancing with the geometric o.t.f. From equation 2.36 we may write, for the relative o.t.f.,

$$M_g(s)\exp(i\theta(s)) = \iint_A \exp(iksW'(x,y)) dA \quad 3.17$$

Then, taking $\overline{W'}$ to be the mean value of W' over the pupil we have

$$M_g(s) = 1 - \frac{2\pi^2}{\lambda^2} s^2 K_g$$

$$\theta_g(s) = ks\overline{W'} \quad 3.18$$

where

$$K_g = \iint_A W'(x,y)^2 dA - \left\{ \iint_A W'(x,y) dA \right\}^2 \quad 3.19$$

Since $W'(x,y) = \partial W / \partial x$, equations 3.19 and 3.6 show that K_g is effectively the variance of the x-component of the transverse aberration. Although, as with the previous merit function K_g is independent of aberration size, it is also independent of spatial frequency, since the area of integration is just the pupil area for all frequencies.

Since these balances will be frequency-independent we may calculate them for the limit of zero frequency, when from equations 2.31 and 2.36 the geometric and diffraction integrals become identical. Thus the balances from this approximation to the geometric o.t.f. represent the low-frequency limit of those calculated from the diffraction o.t.f.

Hopkins gives details of the evaluation of the variances K and K_g for the general aberration polynomial in terms of the integrals

$$H_{kl}(s) = \iint_S x^k y^l dA \quad 3.20$$

In the geometric limit this reduces to

$$H_{kl}(0) = \frac{\Gamma(\frac{1}{2}(k+1)) \Gamma(\frac{1}{2}(l+1))}{\Gamma(\frac{1}{2}(k+l+4))} \quad 3.21$$

The balances given by the use of these approximations will be shown in the later sections.

It is useful to compare the latter criterion with the r-m-s spot size criterion. On axis the wavefront is symmetrical about the axis, so the mean transverse aberration is zero. This means that equations 3.8 and 3.19 are identical, so that for the axial case optimisation of the r-m-s spot size is equivalent to optimisation of the low frequency o.t.f. Similar relationships for the off-axis objects are demonstrated later.

g and h. The Diffraction and Geometric o.t.f.

The previous treatments of the diffraction and geometric o.t.f. were based on approximations to the respective integrals. These approximations lead to limiting restrictions on the regions of use of the criteria, and lead to analytic formulae for the optimum balances for simple models. The full diffraction and geometric integrals do not give analytic expressions for the best balances, but as will be shown they are valid optimisation criteria over vastly increased ranges. The full integrals must be studied by numerical techniques, and by using computational techniques described in chapter 4 it has been possible to directly optimise equations 2.31 and 2.36 with respect to the polynomial coefficients. The results from these optimisations are given in chapters 5-7 and are compared with the balances given by the other criteria listed above.

Practical Merit Functions

Although many authors have written about merit functions, little has been published giving details of optimisation programmes based on merit functions other than the original type mentioned in chapter one. Many computational schemes have been described for the various merit functions, but few of these have been used in practical optimisations. This section summarises the proposed methods, and evaluates the few published results from the use of these methods. Finally, details are given of the construction of a merit function based on the geometrical o.t.f., which has been incorporated into the Imperial College optimisation programme.

The r-m-s spot size, r-m-s wave aberration and the variances of the transverse and wave aberrations are all essentially modifications of the original merit function, and all of these have been in use for some time at Imperial College. In practice these merit functions are not used in the simple forms defined above, and the designer usually modifies the weighting factors after each optimisation.

Meiron (1968) gives an example of a system optimised using r-m-s wave aberrations, wave aberration variance, and the variance of the wave aberration difference function $K(s)$.

For the first two cases a series of rays are traced at each field angle, to give a set of wavefront aberrations W_i . The first merit function takes the equally weighted aberrations aW_i to form the merit function E_1 for each field, where

$$E_1 = \sum_i a^2 W_i^2 \quad 3.22$$

Subtracting the mean aberration from each wavefront leads to the second merit function E'_1 , where

$$E'_1 = \sum_i a^2 W_i^2 - a^2 \left(1/n \sum_i W_i \right)^2 \quad 3.23$$

In a similar manner the variance of the wave aberration

difference function is found by taking contributions V_i from each of the n rays, where

$$V_i = 1/s \left(W_i(x+s/2, y) - W(x-s/2, y) \right) \quad 3.24$$

The merit function is then given by replacing W_i by V_i in equation 3.23. The merit function E_1 is taken over all the field angles to give the complete merit function, Ψ , as

$$\Psi = \sum_1 E_1 \quad 3.25$$

The merit function is then suitable for optimisation by the normal least-squares method.

Examination of Meiron's results show that there is a great improvement from all three methods over the original design, which had been optimised by the standard spot-size criterion. The results, however, are inconclusive, since the final aberration balances could have been obtained by appropriate alterations to the original weighting factors. In addition the system is well corrected initially, and is optimised at a low spatial frequency ($s=0.12, M=0.7$), and in this region all three merit functions are expected to give a good final o.t.f. performance. As expected the third merit function, $K(s)$, gave the highest o.t.f. performance at the frequency of optimisation, while the other two gave a better overall o.t.f. balance. Meiron concludes that the extra complication and expense of the third merit function does not justify its use in preference over the other two functions. For the limited example given, this conclusion is reasonable, but it will be shown that for regions of poorer correction a criterion which optimises at a specified frequency, assuming it is a valid criterion, should be preferred for final balancing, even though the computing costs are higher.

Hopkins (1966) describes numerical methods for evaluation

of the variances E (wavefront aberration) and $K(s)$ (wavefront difference function) which lead to merit functions based on the Maréchal theory and on the approximation to the diffraction o.t.f. The two methods are similar in that they both use a technique of polynomial fitting to a small number of rays in standard patterns. Although the theory has been well developed since 1966 there is no evidence that the technique has been used for a practical optimisation programme.

The method may be illustrated by taking the axial case, where the wavefront is represented by the polynomial

$$W(y) = W_{20}y^2 + W_{40}y^4 + W_{60}y^6 \quad 3.26$$

If only three rays are traced at heights y_i , they will give three wavefront aberrations w_i ,

$$w_i = W_{20}y_i^2 + W_{40}y_i^4 + W_{60}y_i^6 \quad 3.27$$

These three equations may be inverted to give the three terms in the polynomial 3.26, W_{n0} , as

$$W_{n0} = k_{n0,1}w_1 + k_{n0,2}w_2 + k_{n0,3}w_3 \quad 3.28$$

for $n = 2, 4, 6$.

The variance E can be expressed in a polynomial in the aberration coefficients W_{n0} , and substitution of the terms from equation 3.28 gives the variance in terms of the individual aberrations w_i as the polynomial

$$E = \sum_{ij} P_{ij} w_i w_j \quad 3.29$$

The coefficients P_{ij} are constant for a standard set of rays in the entrance pupil, so the calculation of E requires merely the evaluation of wavefront aberrations on only three rays.

In a similar manner the variance K of the difference function may be expressed in terms of the wavefront aberrations, but

the coefficients will now be frequency dependant. We may then write

$$K(s, \phi) = \sum_{ij} P_{ij}(s, \phi) w_i w_j \quad 3.30$$

Similar equations hold for the off-axis objects, where the aberration coefficients used must be selected from the general polynomial 2.7.

King (1968) gives details of the evaluation of the P_{ij} coefficients for both variances, and of the construction of suitable optimisation merit functions, although no practical results have been published.

The use of the two variances as optimisation criteria is discussed in later chapters, but it is appropriate at this stage to consider the use of the polynomial method. To effectively optimise a system using a wavefront aberration polynomial, the polynomial must accurately represent the full wavefront. King suggests the use of between 3 and 5 rays for the axial case, to give up to 5 terms in the aberration polynomial. The number of rays is chosen before the start of the optimisation and the appropriate set of P_{ij} coefficients is used. There is no way, therefore, to check the accuracy of the polynomial approximation during the optimisation. Although tests have shown that most wavefronts may be represented sufficiently by three or four terms on axis, these terms should be determined by least-squares fitting to a larger number of aberrations if reliable values are required.

As stated by King, and as discussed here, these merit functions are only valid for systems with high o.t.f. values. So far only two merit functions have been proposed which attempt to extend this limited range. King (1972) describes

a merit function based on the diffraction o.t.f. integral, where the integral is expanded as a truncated Taylor series with respect to small parameter changes. This appears highly complex to programme, and no results of its use have yet been published.

Offner (1970) also describes an extended range merit function, based on a method of numerical integration of the diffraction o.t.f. integral due to Hopkins (1957). This method, which appears to need a large number of rays, requires a modified form of the normal least squares equations, since the merit function contains a linear term as well as a quadratic term. Once again, no practical results have been published.

The new merit function proposed in this thesis has the advantage, as will be shown in later sections, of a wider range than any of the previous functions. It also has the advantage of being relatively simple to incorporate into an existing optimisation programme.

If we consider the geometrical o.t.f. equation 2.36 we can write the separate sagittal and tangential equations as

$$D_g(s,0) = \iint_A \exp(iksW'_x(x,y)) \, dx dy \quad 3.31$$

$$D_g(0,t) = \iint_A \exp(iktW'_y(x,y)) \, dx dy \quad 3.32$$

where $W'_x(x,y) = \partial W / \partial x$ $W'_y(x,y) = \partial W / \partial y$ 3.33

From equations 2.2, 2.18 and 3.6 we may return to un-normalised coordinates,

$$\begin{aligned} ksW'_x(x,y) &= -2\pi S' \\ ktW'_y(x,y) &= -2\pi T' \end{aligned} \quad 3.34$$

The geometric integrals may now be written in terms of the transverse aberrations as

$$D_g(S,0) = \iint_A \exp(-2\pi i S \delta \xi) \, dx dy \quad 3.35$$

$$D_g(0,T) = \iint_A \exp(-2\pi i T \delta \eta) \, dx dy \quad 3.36$$

For the sagittal case we may expand 3.35 to give

$$D_g(S,0) = \iint_A \cos(2\pi S \delta \xi) \, dx dy - i \iint_A \sin(2\pi S \delta \xi) \, dx dy \quad 3.37$$

Since there is symmetry about the y-axis, the imaginary part of the integral is zero, and hence the sagittal o.t.f. has no phase term. Then

$$D_g(S,0) = \iint_A \cos(2\pi S \delta \xi) \, dx dy$$

$$D_g(S,0) = \iint_A (1 - \sin^2(\pi S \delta \xi)) \, dx dy \quad 3.38$$

Expressing the integral in the form 3.38 leads to a merit function of the usual quadratic form, if the merit function is taken as

$$G_s(S) = \iint_A \sin^2(\pi S \delta \xi) \, dx dy \quad 3.39$$

For a practical optimisation programme, where n finite rays are traced at this field angle, the merit function becomes

$$G_s(S) = \sum_i \sin^2(\pi S \delta \xi)_i \quad 3.40$$

For the tangential case, equation 3.32 may be expanded as

$$D_g(0,T) = \iint_A \cos(2\pi T \delta \eta) \, dx dy - i \iint_A \sin(2\pi T \delta \eta) \, dx dy \quad 3.41$$

The first term in the equation may be expressed in a similar form to the sagittal case, but in this case there is no longer symmetry about the axis, so the imaginary part does not disappear. This imaginary part, which gives rise to the

phase term of the o.t.f. cannot be expressed in the required quadratic form for optimisation. On physical and practical considerations it may be shown that the imaginary term is of little consequence in most cases. These arguments are given in more detail in the section on phase in a later chapter. For optimisation, therefore, we take as the tangential merit function

$$G_t(T) = \sum_i \sin^2(\pi T \delta\gamma)_i \quad 3.42$$

The two merit functions G_s and G_t may be combined into one merit function, and summations must be taken for each field angle and wavelength and for each spatial frequency of interest. The merit function may also contain terms to control the other aberrations and targets, such as the distortion and the required focal length.

The final merit function, ψ , may be written

$$\psi = \sum_l (G_s(S) + G_t(T))_l + \sum_i w_i^2 (f_i - t_i)^2 \quad 3.43$$

This new merit function has several important advantages.

Firstly, since it is composed of a sum of squares of the various terms, it is suitable for minimisation by the damped least squares technique, and hence may be easily incorporated into an existing programme. The program requires relative weights between the various field angles and wavelengths, and these weights are easily selected by considering the required spectral response and proposed use of the system. The designer no longer needs to individually weight each ray for the optimisation, and so the program is considerably easier to use than the conventional type.

The most important advantage of this merit function is that it provides direct optimisation of the geometric o.t.f.

since no approximations have been made to the o.t.f. integral. The improvement to the diffraction o.t.f. will depend on the range of validity of the geometric o.t.f. as an optimisation criterion, and as will be shown in the later sections, this range is extremely wide.

Numerical Techniques

Numerical evaluation of the diffraction and geometric o.t.f. integrals 2.31 and 2.36 has been possible for some time with even small computers. To achieve a reasonable degree of accuracy, however, requires a considerable amount of computer time, and it is found that the amount of time needed increases as the size of aberration increases or as the spatial frequency increases, i.e. as the o.t.f. decreases.

Most previous studies of model aberration balancing have, therefore, confined themselves to regions where the o.t.f. is still high. Most of these studies have consisted merely of the evaluation of o.t.f. curves for small sets of parameter changes about a pre-determined starting point. This starting point is usually given by one of the analytic formulae given earlier, such as the Maréchal formulae. These studies provide useful results within their limited ranges, but most of them fail to show the dependance of optimum wavefront shape on aberration size, which becomes of increasing importance at high aberrations.

The extensive computer facilities at Imperial College provided the computer power necessary to conduct a large-scale numerical investigation into o.t.f. aberration balancing. Previous methods would have been too lengthy for the extended range proposed, so programmes were developed to provide direct optimisation of the diffraction and geometric o.t.f. integrals with respect to the terms in the aberration polynomial. As mentioned in the previous chapter, the use of direct optimisation provides two sets of results. Firstly the effects of aberration balancing on the

diffraction o.t.f. can be investigated over a much wider range than has been previously possible, and secondly the relative merits of the various other optimisation criteria, and in particular the geometric o.t.f., may be easily compared over the complete range.

All the computations were performed on the Control Data CDC 6400 computer, and even on this fast machine one optimisation could take as much as 200 seconds for large aberrations at high spatial frequencies.

For efficiency and accuracy Gaussian quadrature was used for the numerical integration. The first trials made use of a routine from the extensive library available on the computer. This routine, written in machine code, COMPASS, was capable of performing multi-dimensional integration, and was thus easily adapted to the two-dimensional o.t.f. integrals. The routine divides each region of integration into a number of equal intervals, and performs six-point quadrature in each interval. The wavefront aberration difference function V and the differentiated wavefront aberration W' were supplied as functions to the integration routine, and the appropriate limits of integration were set by the controlling programme. For each case the numerical accuracy was ensured by increasing the number of intervals until successive values agreed to within a pre-determined limit. For all the computations for o.t.f. in this investigation, all programmes were set to work to an accuracy of 0.2%. The results produced by these programmes were compared with the standard routines from the Applied Optics group, and the results were consistent to within the limit set.

After the first few trials this straightforward method was

discarded, for two reasons. Firstly, to produce the increased numerical accuracy needed for low o.t.f. calculations the programme increases the number of intervals for integration. For double integration the time needed varies as the square of the number of intervals used. It is important, therefore, to have as few intervals as possible, and the method used needs many more intervals than would be required by performing Gaussian quadrature over the complete range. For a low number of intervals the first method is both easy to use and takes little computer storage. To perform quadrature with variable numbers of Gaussian intervals the coefficients must be stored for each interval, and to provide a wide range of use this can be a significant amount of storage. However this direct method is much faster in execution and more accurate than the previous method, particularly when many intervals are needed. This method also has the advantage that, for o.t.f. calculations, the pupil function is most rapidly varying near the edge of the pupil, and Gaussian quadrature weights the integrand most heavily near the edges of the function.

The second disadvantage of the library function became apparent when off-axis cases were considered. The routine took the form of a function subroutine, which meant that it could only return one value on each call. For the cases where both amplitude and phase are required for the tangential o.t.f. this method becomes highly impractical.

It was decided, therefore, to adapt one of the efficient o.t.f. routines written by Michael Kidger of the optical design group. In its basic form the routine evaluates the diffraction o.t.f. integral by setting a number of Gaussian intervals, and uses this

set of intervals to perform the double integral twice, once integrating over the x-direction first, then over the y-direction and then in the reverse order. If these values do not agree to within the required accuracy limit, the programme increases the number of intervals and repeats the process. The method is explained in greater detail in a paper by Kidger (1974), and shows the various use made of the particular symmetries of the o.t.f. integral.

Although the routine was written to evaluate the diffraction o.t.f., only minor modifications are needed to enable it to evaluate the geometric o.t.f. instead. This modified routine gave results which were in agreement with those from the other geometric routines.

The self-checking method described above to ensure accuracy caused certain problems when the routines were used in an optimisation programme. The routines calculate the o.t.f. at least twice on each call, which is obviously expensive on time, but a more serious problem arose when the routine increased the number of intervals within an optimisation. In the course of the optimisation, the controlling programme uses finite differences to calculate the direction for parameter changes, and the change caused by an alteration in the intervals caused errors in the derivatives. These problems were overcome by including in the routine a parameter such that on the first call to the routine the number of intervals is increased to provide the required accuracy, and this number is kept constant throughout the rest of the optimisation. Omitting the error-checking process does not lead to any loss of accuracy, as during the optimisation the o.t.f. is increasing, and so the routine is then working to a

higher degree of accuracy, as described previously.

When the initial set of intervals is being calculated, the routine sometimes finds that the two integrals agree with a small number of intervals, when their value is not the true value of the o.t.f. To eliminate this cause of error, once the routine has found a suitable set of intervals it increases the number and compares the values again. If these results agree the routine assumes that the first set of intervals is correct, but if they do not agree the routine continues to increase the number of intervals. This simple method ensures that the routine is always working to at least the limit set by the controlling programme.

Another programme from the library was used to perform the optimisation. This programme, which had been extensively tested, uses a method of conjugate directions to provide efficient optimisation. The function to be evaluated f is supplied to the programme with a set of parameters x_1 . The function is then minimised with respect to any or all of the parameters, subject to any constraints set by a control routine.

The parameters were taken as the coefficients of the wavefront aberration polynomial, which allows any high order aberration, or combination of aberrations, to be balanced against lower order aberrations. For the first method of Gaussian quadrature and for the preliminary axial study the function was taken as

$$f = 1 - \text{OTF}(s) \quad 4.1$$

for the spatial frequency s . The program could take either the geometric or diffraction o.t.f. as required.

For the off-axis cases, where sagittal, tangential and phase values were required, the merit function was taken as

$$f = 1 - a.OTF_s - b.(OTF_r^2 + c.OTF_i^2)^{\frac{1}{2}} \quad 4.2$$

where;

OTF_s is the sagittal o.t.f.

OTF_r is the tangential real o.t.f..

OTF_i is the tangential imaginary o.t.f.

a, b, and c are weighting factors which could be set at the beginning of each optimisation. With c set at 1 the programme optimises the modulus of the tangential o.t.f., and with c set at zero the real part only is optimised. In this latter form the merit function becomes identical to the geometric merit function incorporated into the lens optimisation programme, as will be shown in later chapters. The effect of the phase contribution may also be investigated by suitable choice of c.

In considering several optimisation criteria over a range of spatial frequencies and a range of aberration sizes there is a considerable amount of data to be presented and analysed. As well as tabular data, graphical representation is very valuable, especially to illustrate wavefront shapes. For this purpose, a general graph plotting routine was developed, which was capable of presenting several graphs on one set of axes. As each set of balances is produced the programme computes the diffraction o.t.f. curves for each criterion and plots them on the line printer, followed by a plot of the wavefront shapes produced by the criteria. In later stages the balances produced at a given size of high-order aberration but optimised at several spatial frequencies were compared, and a new plotting procedure was developed. This new programme computed the o.t.f. curves for the various criteria, and all the curves were plotted on the microfilm output system. In this way all the curves could easily

be compared at the same time. These graphs are shown in the later sections.

As well as comparing the different criteria, the other purpose of the study is to analyse the variation of optimum wavefront shape for varying spatial frequencies and wavefront aberration size. It was seen from the first results that the contours of o.t.f. in parameter space for a given spatial frequency appear to form long, narrow valleys. This is to be expected from practical results with optimisation programmes, where this type of multi-dimensional valley is nearly always found (Kidger 1971). To investigate this further, more use was made of the microfilm system to produce contour plots. As will be seen in later sections each optimisation takes one fixed, high-order aberration term and varies two lower-order terms to produce the optimum balance. The optimisation is thus a two-parameter system, which leads very simply to the use of contour graphs. A contour plotting routine is provided as part of the software associated with the microfilm system, and this proved suitable for o.t.f. contours. The routine required values of the function to be plotted over a rectangular grid, and contours are plotted at intervals defined by the controlling programme. The largest part of the execution time for the programme was taken in producing the values across the grid, and after a series of trials a grid of 41×41 points was chosen as a compromise between accuracy and computation time. From the considerations of choice of Gaussian intervals given earlier, the number of Gaussian intervals was chosen at the point where the lowest o.t.f. was expected, usually at one of the corners of the plot.

The positions corresponding to the balances obtained from the various criteria were plotted on the graphs, and these illustrate very clearly the way in which the criteria behave in the regions considered.

Even with a grid as coarse as 41×41 points, the time taken to produce one contour plot at high aberration and spatial frequency was over 1000seconds on the CDC 6400, which represents a considerable amount of computer resources.

Low Frequency Axial Aberration Balancing

In this section the effects of aberration balancing will be investigated for the case of axial aberrations at low spatial frequencies. Balances will be evaluated for a wide range of aberration size, and for several spatial frequencies.

Using the numerical techniques described in the last section the diffraction and geometric o.t.f. will be optimised at each frequency and aberration size. Using the balance produced by the geometric optimisation and by the analytic formulae of chapter 2 the diffraction o.t.f. may be calculated and hence a direct comparison of the various merit functions may be made over the chosen range.

The wavefront polynomial used is the axial secondary spherical aberration polynomial 3.1 and 3.3

$$\begin{aligned} W &= W_{20}\rho^2 + W_{40}\rho^4 + W_{60}\rho^6 \\ &= W_{60}(B_{26}\rho^2 + B_{46}\rho^4 + \rho^6) \end{aligned} \quad 5.1$$

All aberration polynomials are given in units of wavelength, λ .

The balances will be evaluated for values of W_{60} of 5, 10, 20 and 40 wavelengths, and for ^{reduced} spatial frequencies 0.01, 0.05, and 0.10.

In considering the merit functions in section 3 it was seen that not all the functions are independent. The r.m.s. spot size and the Hopkins approximation to the geometric o.t.f. both represent the low-frequency limit of the diffraction o.t.f., and are thus both covered by the Hopkins approximation to the diffraction o.t.f., and are the zero-frequency limits of the formulae which Hopkins derived. Hence only this later criterion will be considered.

The minimum spot-size criterion and the r.m.s. wavefront aberration criterion are rarely used in their exact forms. The minimum spot size criterion has few practical applications, and the

wavefront aberrations are either used with unequal weights across the pupil, as chosen manually by the designer, or the reference pupil radius is varied to produce the Maréchal variance. The balances for these two criteria are evaluated in this section, for completeness, but the o.t.f. values and wavefronts are not presented.

The analytic formulae, where applicable, are derived below.

a. Minimum Spot Size

The wavefront aberration polynomial, 3.1, is of cubic form in ρ^2 and the general shape of the wavefront and associated transverse aberration will be as shown in figure 5.1a, and 5.1b.

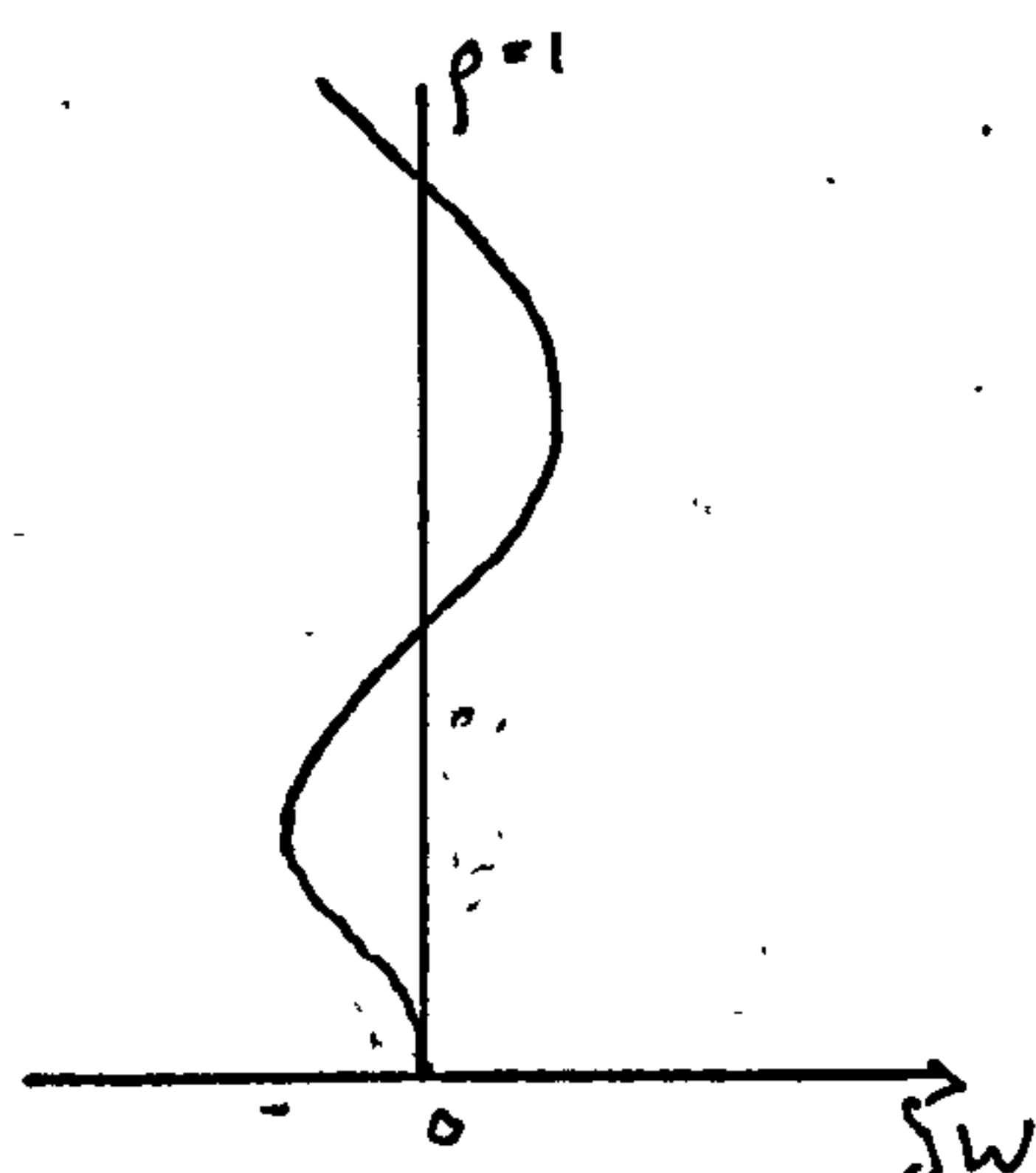


Fig 5.1a

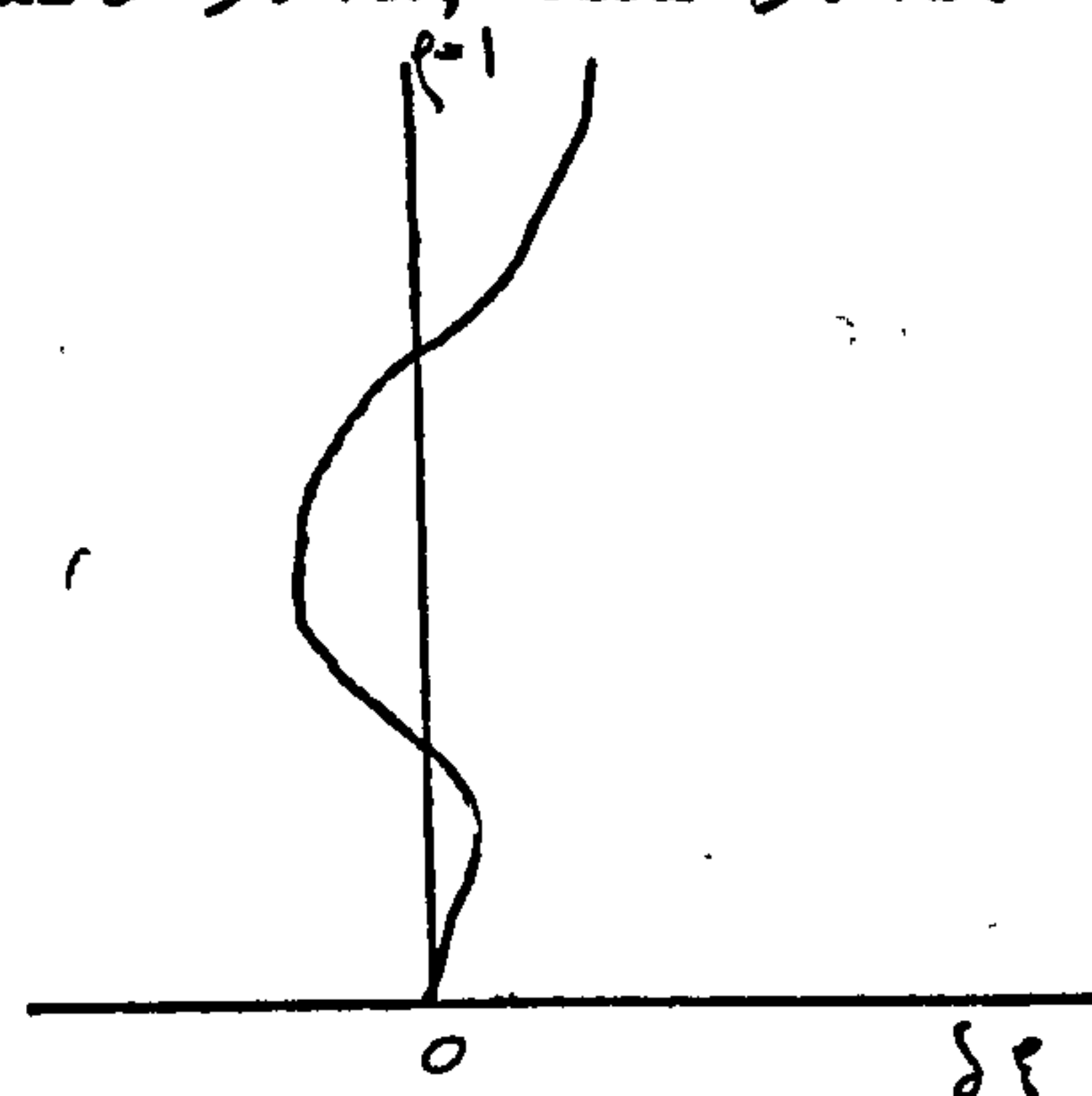


Fig 5.1b

Wavefront and transverse aberrations for spherical aberration

As described in section 3b, the three maximum transverse aberrations usually occur at the edge of the pupil and at the two intermediate extrema, where $\omega'' = 0$. It is obvious from figure 5.1b that the minimum spot size will be produced when all three aberrations have the same numerical value, with the centre one of opposite sign to the outer two, as shown.

The equation of the transverse aberration across the pupil is

$$\delta \xi(\rho) = 2W_{20}\rho + 4W_{40}\rho^3 + 6W_{60}\rho^5 \quad 5.2$$

and the maxima occur when $\partial(\delta \xi)/\partial \rho = 0$.

The two values of ρ satisfying this condition are substituted back into equation 5.2 to give the aberrations $\delta\xi_1, \delta\xi_2$.

The aberration at the edge of the pupil, $\delta\xi_0$, is simply given by

$$\delta\xi_0 = 2W_{20} + 4W_{40} + 6W_{60} \quad 5.3$$

The minimum spot size criterion now reduces to the solution of

$$\delta\xi_1 = -\delta\xi_2 = \delta\xi_0 \quad 5.4$$

The solutions of these equations give

$$W_{20} = 15W_{60} / 16 \quad W_{40} = -15W_{60} / 8 \quad 5.5$$

and thus

$$B_{26} = 0.9375 \quad B_{46} = -1.825 \quad 5.6$$

b. Mean-square wavefront aberration

Substitution of the polynomial 5.1 into equation 3.9 gives the r.m.s. wave aberration as

$$\begin{aligned} W^2 &= \iint_A (W_{20}\rho^2 + W_{40}\rho^4 + W_{60}\rho^6)^2 dA \\ &= 2\pi \int_0^1 (W_{20}\rho^2 + W_{40}\rho^4 + W_{60}\rho^6)^2 \rho d\rho \end{aligned} \quad 5.7$$

Evaluation of the integral and differentiation with respect to the variable coefficients W_{20} and W_{40} leads to the simultaneous equations

$$\begin{aligned} W_{20} &= 0.35 W_{60} & W_{40} &= -1.33 W_{60} \\ B_{26} &= 0.35 & B_{46} &= -1.33 \end{aligned} \quad 5.8$$

These values differ significantly from those obtained when the reference sphere radius is allowed to vary, to give the Maréchal variance formulae, shown below.

c. Maréchal wavefront aberration variance

From equations 3.10 and 5.1 the variance of the wavefront aberration is given by

$$E = \langle W^2 \rangle - \langle W \rangle^2$$

$$= 1/A \iint_A (W_{20} \rho^2 + W_{40} \rho^4 + W_{60} \rho^6)^2 dA - \left\{ 1/A \iint_A (W_{20} \rho^2 + W_{40} \rho^4 + W_{60} \rho^6) dA \right\}^2$$

5.9

Setting the derivatives of E to zero, as in the last section, leads to the solutions

$$\begin{aligned} W_{20} &= 0.6 W_{60} & W_{40} &= -1.5 W_{60} \\ B_{26} &= 0.6 & B_{46} &= -1.5 \end{aligned}$$

5.10

d. Hopkins approximation to the diffraction o.t.f.

This criterion differs from the previous balances in that it is the only criterion which gives frequency-dependant analytic balances. As shown in section 3e, the diffraction o.t.f. may be expressed in terms of the variance $K(s)$ of the wavefront aberration difference function $V(x,y,s)$, defined in equation 3.12 as

$$K(s) = 1/A \iint_S V^2 dA - \left(1/A \iint_S V dA \right)^2$$

5.11

For a given aberration polynomial $K(s)$ may be expanded with the $H_{kl}(s)$ integrals of equation 3.20, which then gives $K(s)$ as a homogeneous quadratic form of the aberration coefficients. Hopkins (1957) gives this expansion for the spherical aberration polynomial, and the expansion may be differentiated with respect to the aberration coefficients to give the optimum balance for any given spatial frequency. In general this gives a complex expression for B_{26} and B_{46} , but for spatial frequencies less than 0.2 Hopkins showed that it may be accurately represented by the formulae

$$\begin{aligned} B_{26} &= 0.9 - 1.4 s + 2.1 s^2 \\ B_{46} &= -(1.8 - 1.5 s + 1.5 s^2) \end{aligned}$$

5.12

These formulae give values for B_{26} and B_{46} which are accurate to within 0.1 wavelengths for the minimum values of $K(s)$.

Minimising $K(s)$, however, will not always lead to the optimum diffraction o.t.f., since this variance technique is based on an approximation to the o.t.f.. Hopkins showed that the formula for the relative m.t.f., equation 3.15, is accurate for regions where the m.t.f. is greater than 0.8. This sets a condition on the maximum spatial frequency for any given size of aberration, and this tolerance is given by the formula

$$W_{60} = \pm (0.58/s_0 + 1.5)\lambda \quad 5.13$$

which may be inverted to give, for $W_{60} < 0$,

$$s_0 = 0.58/(W_{60}/\lambda + 1.5) \quad 5.14$$

where s_0 is the frequency for which the relative m.t.f. is 0.8.

As stated previously, the absolute numerical accuracy of the approximation is unimportant for aberration balancing, and the use of this variance criterion has been examined over a much wider range than given by the tolerances above. Table 5.1 gives the tolerance frequencies for the values of W_{60} which will be investigated.

W_{60}	s_0	
-5	0.166	Table 5.1
-10	0.068	Values of W_{60} , s_0 giving
-20	0.031	m.t.f. of 0.8.
-40	0.015	

Since the balancing formulae, 5.12, are independant of the size of aberration, it is expected that these balances will be poor outside these strict tolerances, and this is shown in the later results.

Examination of the formulae for the minimum r.m.s. spot size

criterion and the Hopkins approximation to the geometric o.t.f. gives balances in both cases of $B_{26} = 0.9$ and $B_{46} = -1.8$, which are the zero-frequency limits of equations 5.12 as expected.

e. The diffraction and geometric o.t.f.

The full diffraction and geometric o.t.f. formulae do not give analytic formulae for their optimum balances, and so the balances for these two criteria were obtained by the numerical optimisation techniques described in the previous chapter.

These results are summarised in table 5.2, where the balances found by the numerical techniques are presented, together with the balances from the Strehl-Maréchal and Hopkins formulae. Table 5.3 repeats these balances in terms of the B_{26} and B_{46} coefficients. The Maréchal formulae give just one set of coefficients, since they are both frequency and size-independant, while the Hopkins formulae give three sets of balances for the three frequencies used.

The B coefficients effectively describe the shape of the wavefronts, and table 5.3 clearly shows the expected variation of the optimum diffraction o.t.f. wavefront with aberration size. This is illustrated in figure 5.2^{*}, where the diffraction-balanced wavefronts are shown. For each of the four aberration sizes two wavefronts have been plotted, one at the low-frequency limit, and one at the balance for $s = 0.10$. The low-frequency balances have the same shape in all cases, as given by the Hopkins theory, but the balances for $s = 0.1$ vary greatly between the four cases. It is seen that as the aberration size increases the tendency is for the aberration at the edge of the pupil to increase greatly, so

* The signs of figures 5.2 to 5.6 have been reversed, for clarity.

Low Frequency Axial Balances

ζ_5	S	Strehl		Hopkins		Diffraction		Geometric		
		W_{20}	W_{40}	W_{20}	W_{40}	W_{20}	W_{40}	W_{20}	W_{40}	W_{60}
	.01	-3.00	7.50	-4.43	8.93	-4.44	8.94	-4.50	9.00	
	.05	-3.00	7.50	-4.18	8.64	-4.14	8.63	-4.49	8.99	-5.
	.10	-3.00	7.50	-3.91	8.32	-3.82	8.29	-4.44	8.93	
	.01	-6.00	15.00	-8.86	17.85	-8.89	17.88	-9.01	18.01	
	.05	-6.00	15.00	-8.35	17.29	-8.23	17.19	-8.89	17.86	-10.
	.10	-6.00	15.00	-7.81	16.65	-4.71	13.10	-5.80	14.42	
	.01	-12.00	30.00	-17.72	35.70	-17.76	35.75	-18.02	36.01	
	.05	-12.00	30.00	-16.71	34.58	-10.02	26.92	-11.46	28.68	-20.
	.10	-12.00	30.00	-15.62	33.30	-6.87	22.36	-5.62	20.25	
	.01	-24.00	60.00	-35.45	71.41	-35.19	71.11	-35.76	71.68	
	.05	-24.00	60.00	-33.41	69.15	-14.32	45.23	-13.76	44.38	-40.
	.10	-24.00	60.00	-31.24	66.66	-6.68	31.28	-6.39	30.08	

Low Frequency Axial Balances

5.8

S	Strehl		Hopkins		Diffraction		Geometric		W_{60}
	26	46	26	46	26	46	26	46	
.01	-.600	1.500	-.886	1.785	-.888	1.788	-.900	1.800	
.05	-.600	1.500	-.835	1.729	-.818	1.723	-.898	1.798	-5.
.10	-.600	1.500	-.781	1.665	-.762	1.658	-.888	1.786	
.01	-.600	1.500	-.886	1.785	-.889	1.788	-.901	1.801	
.05	-.600	1.500	-.835	1.729	-.823	1.719	-.889	1.786	-10.
.10	-.600	1.500	-.781	1.665	-.471	1.310	-.580	1.442	
.01	-.600	1.500	-.886	1.785	-.888	1.787	-.901	1.801	
.05	-.600	1.500	-.835	1.729	-.501	1.346	-.573	1.433	-20.
.10	-.600	1.500	-.781	1.665	-.343	1.118	-.281	1.012	
.01	-.600	1.500	-.886	1.785	-.879	1.777	-.894	1.792	
.05	-.600	1.500	-.835	1.729	-.358	1.131	-.349	1.109	-40.
.10	-.600	1.500	-.781	1.665	-.147	.782	-.159	.752	

TABLE 53

DIFFRACTION BALANCED WAVEFRONT

— AXIAL

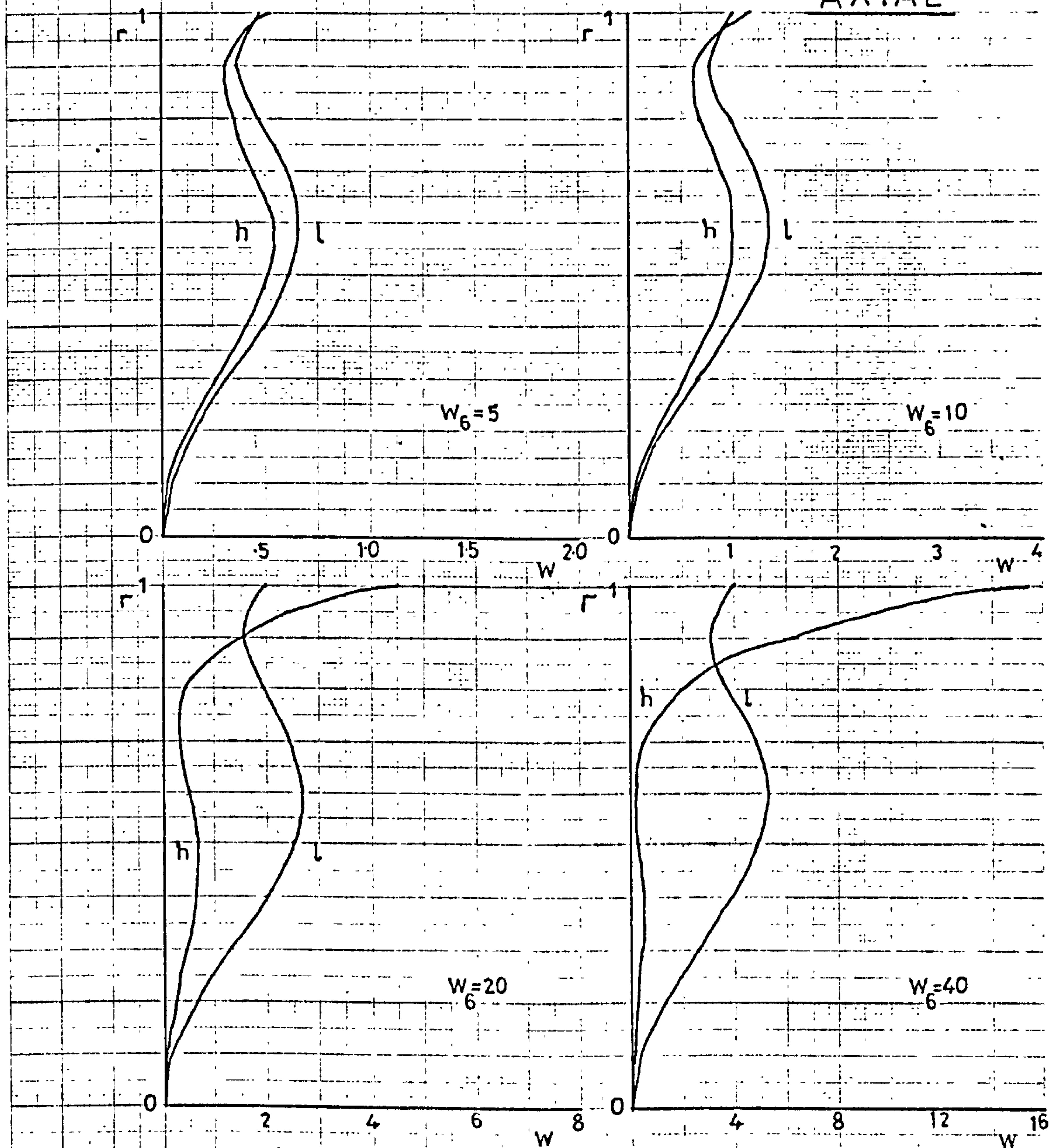
 r = normalised pupil radius W = wavefront aberration

FIG 52

WAVEFRONT ABERRATION BALANCES

AT $W_6 = 5\lambda$

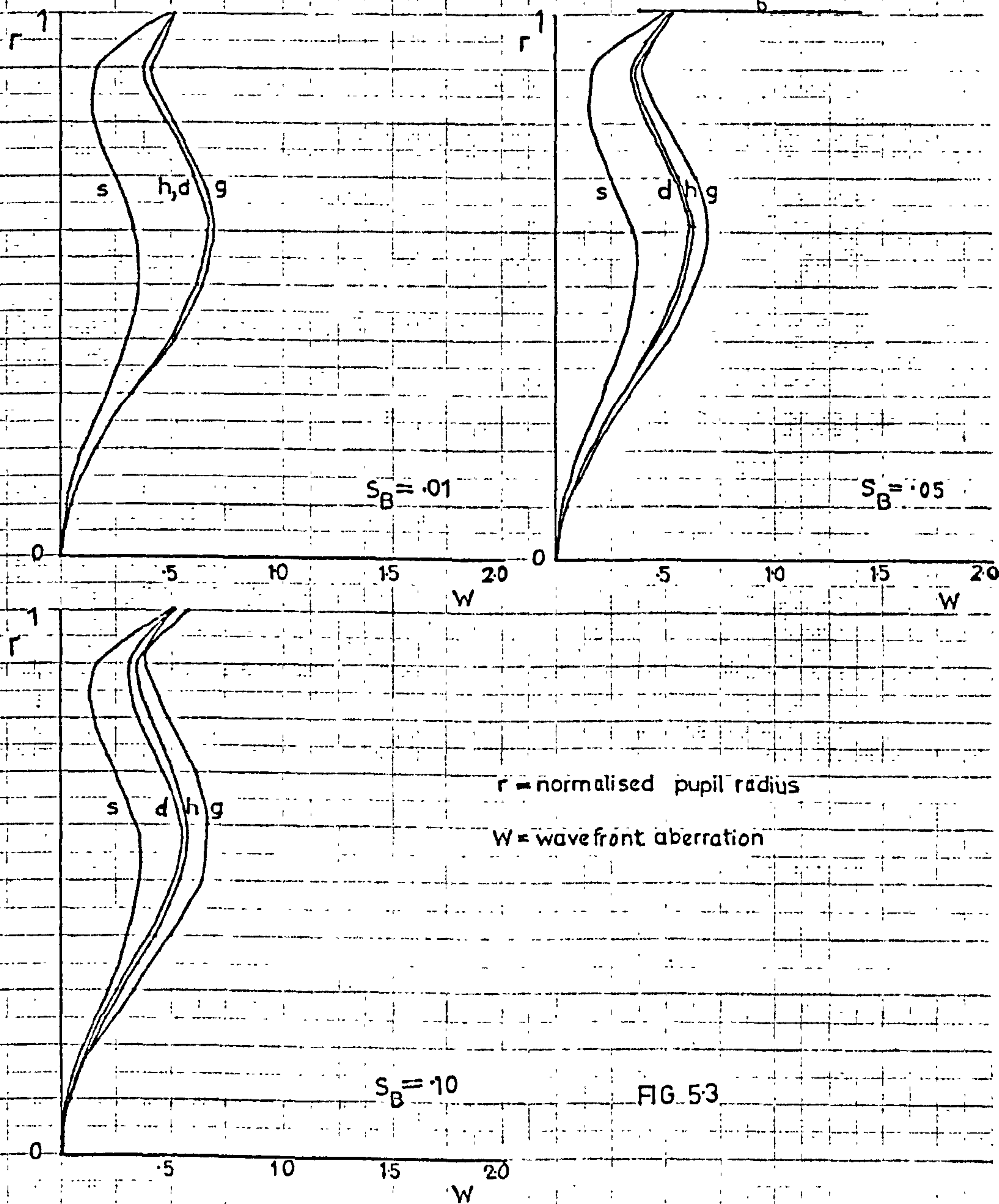
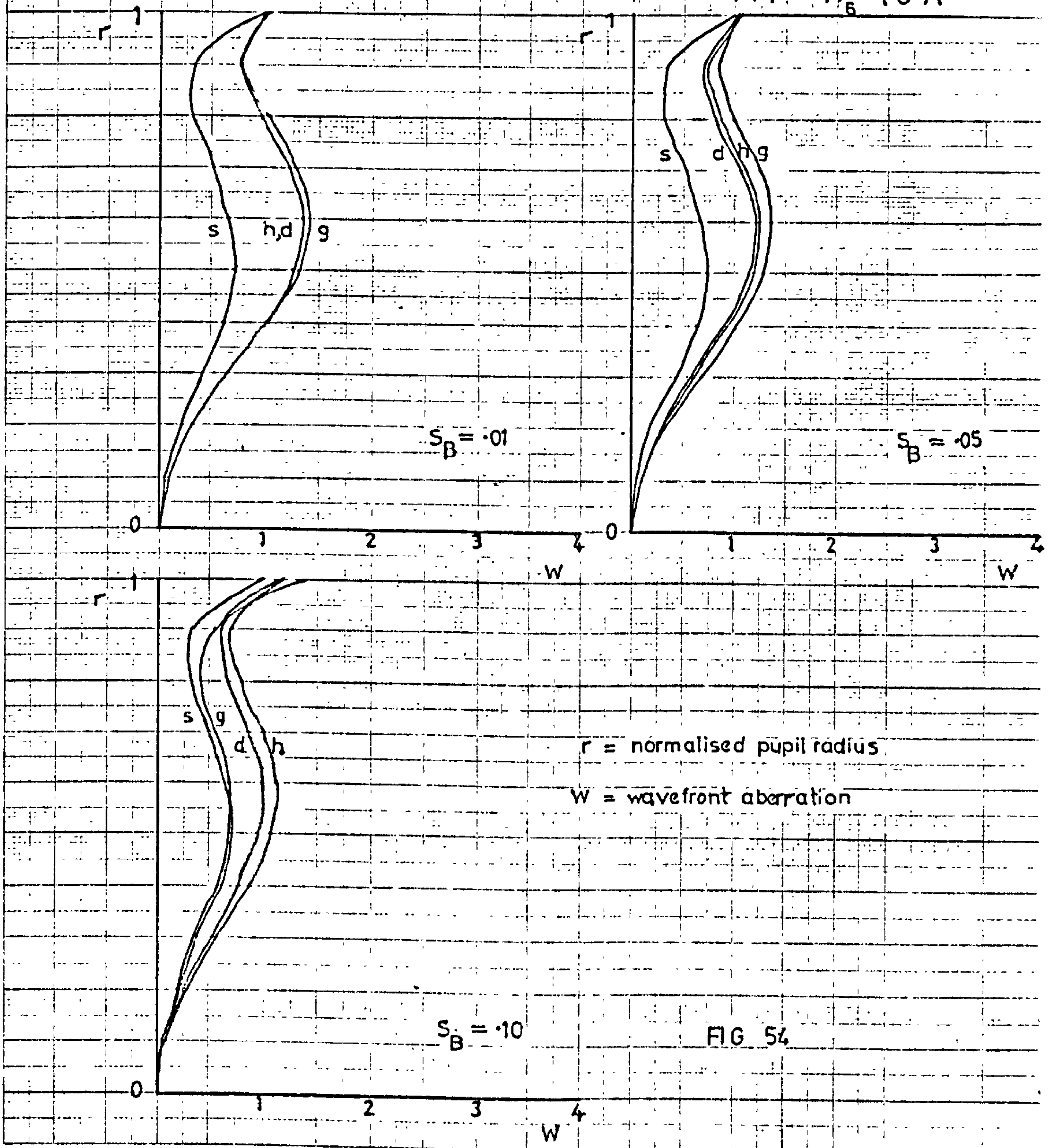


FIG 53

WAVEFRONT-ABERRATION BALANCES

AT $W_6 = 10\lambda$



WAVEFRONT ABERRATION BALANCES

AT $W_6 = 20\lambda$

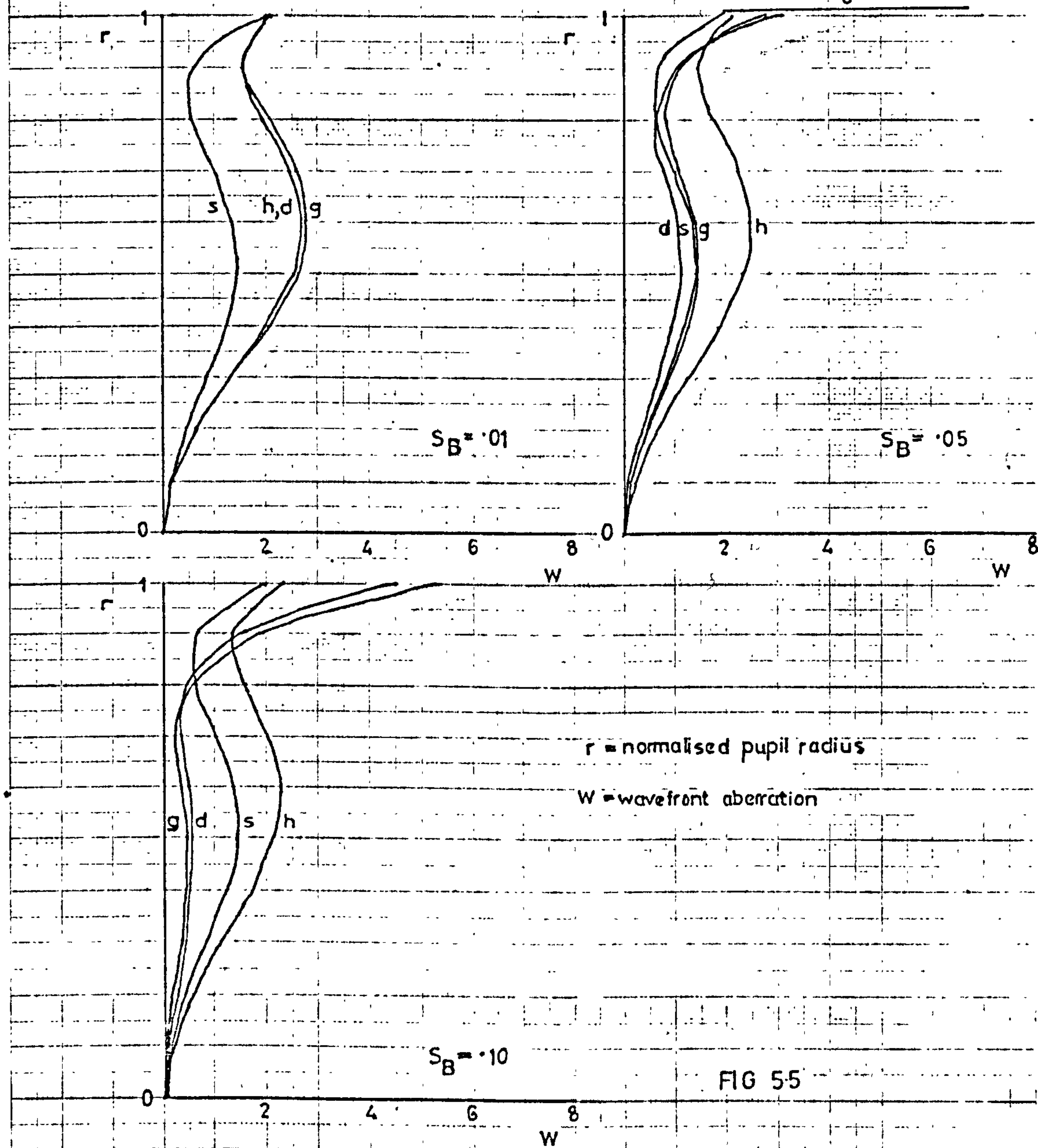


FIG 55

WAVEFRONT ABERRATION BALANCES

AT $W_0 = 40\lambda$

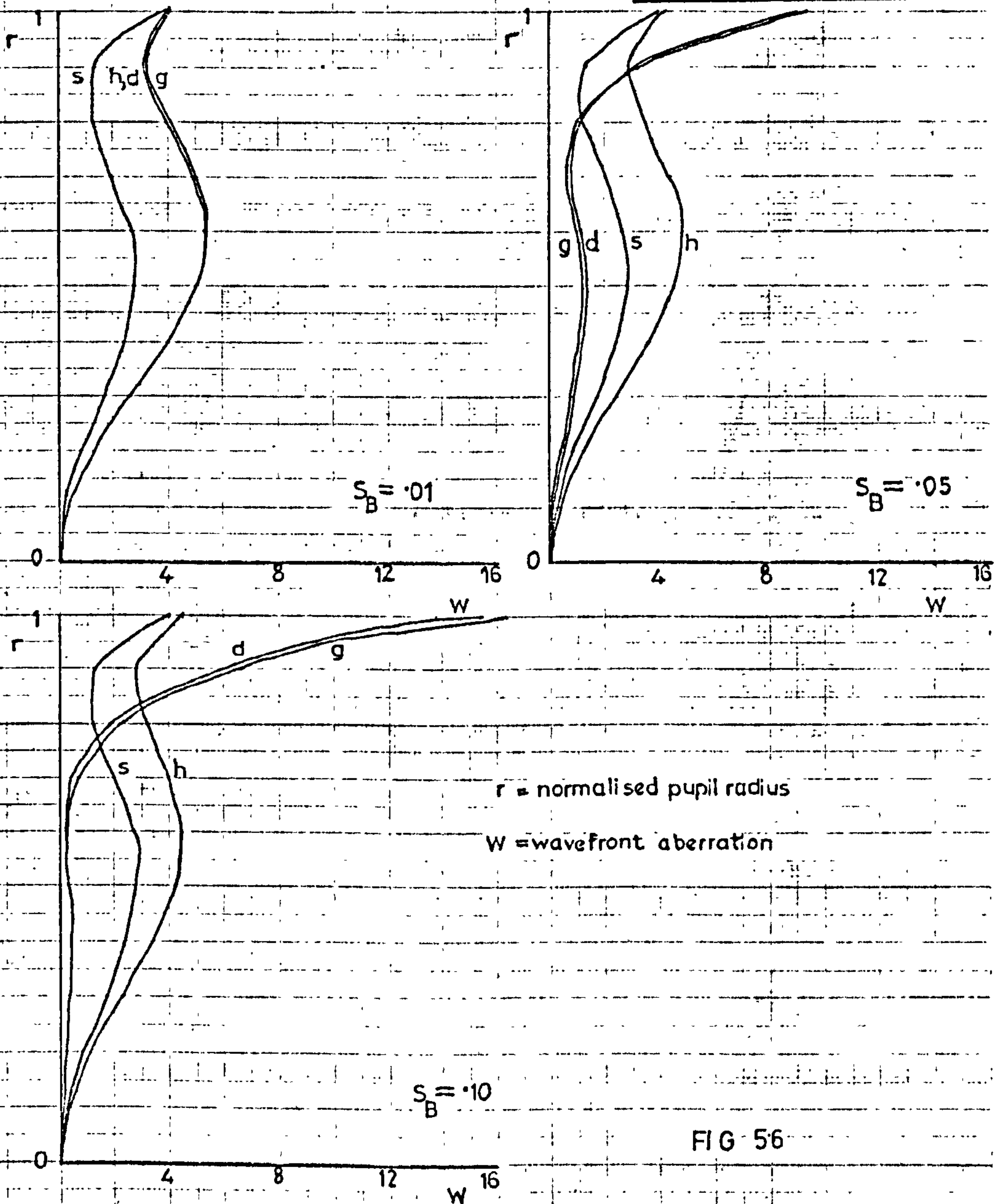


FIG 56

DIFFRACTION OTF AT M6=-5

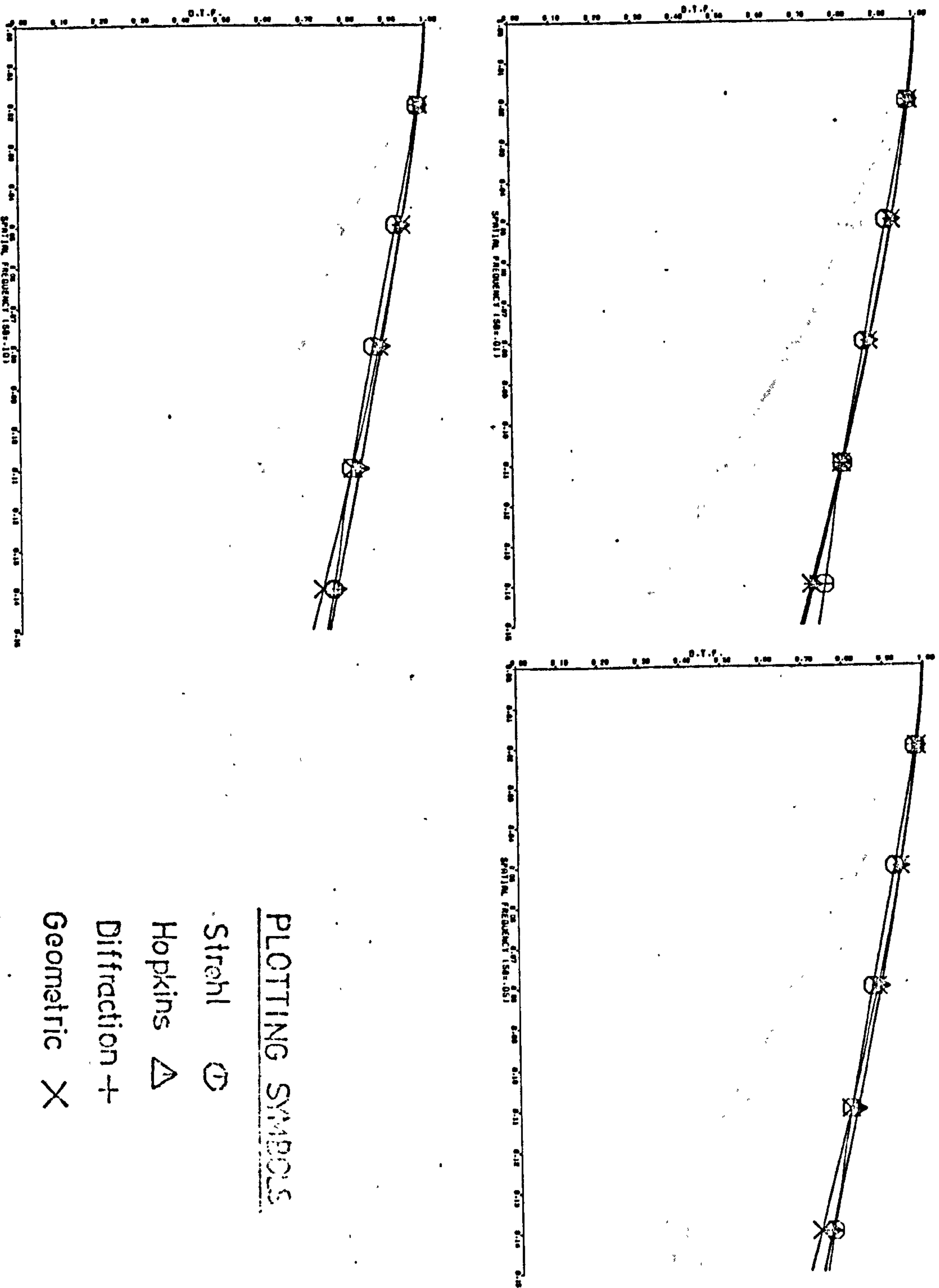
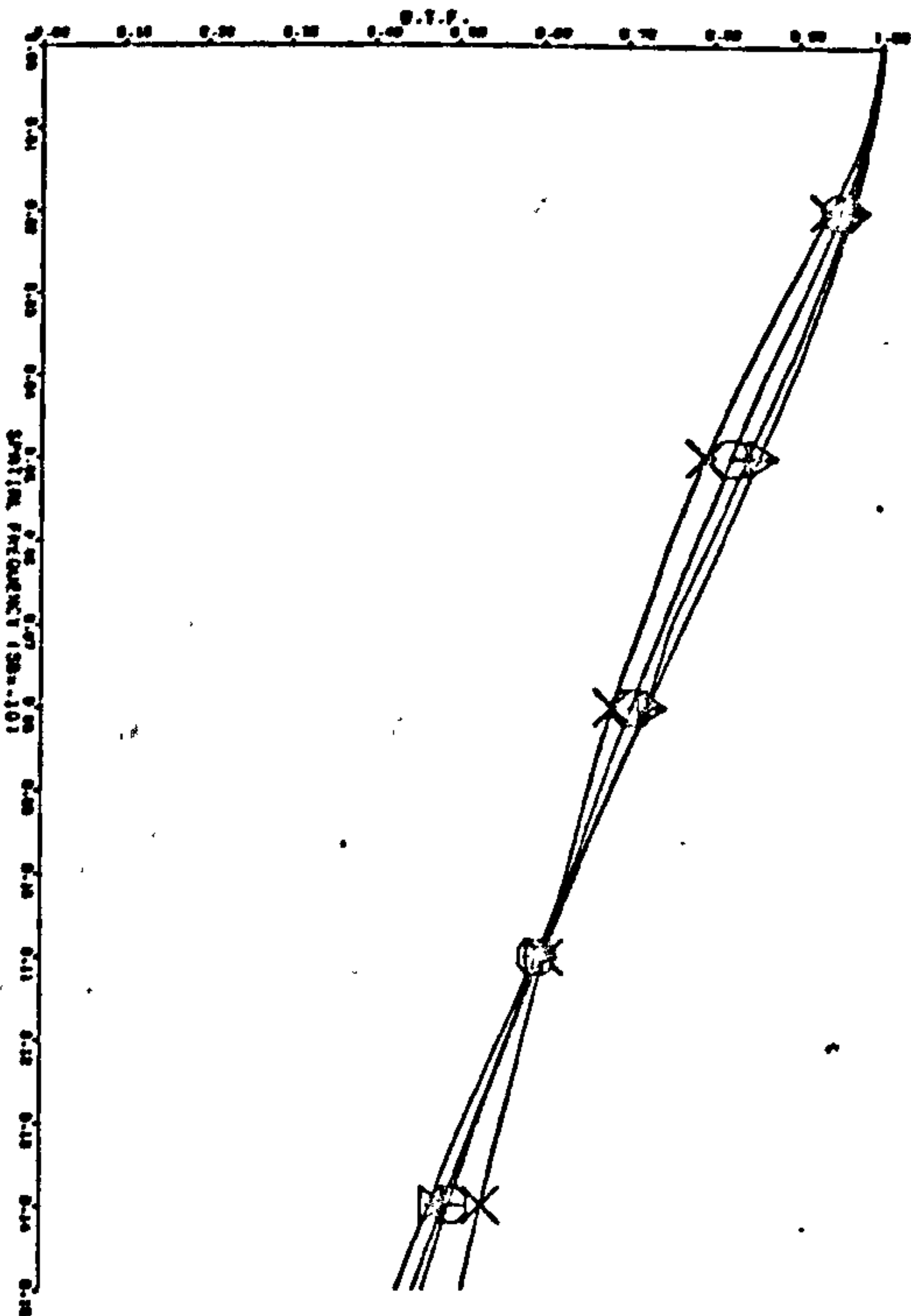
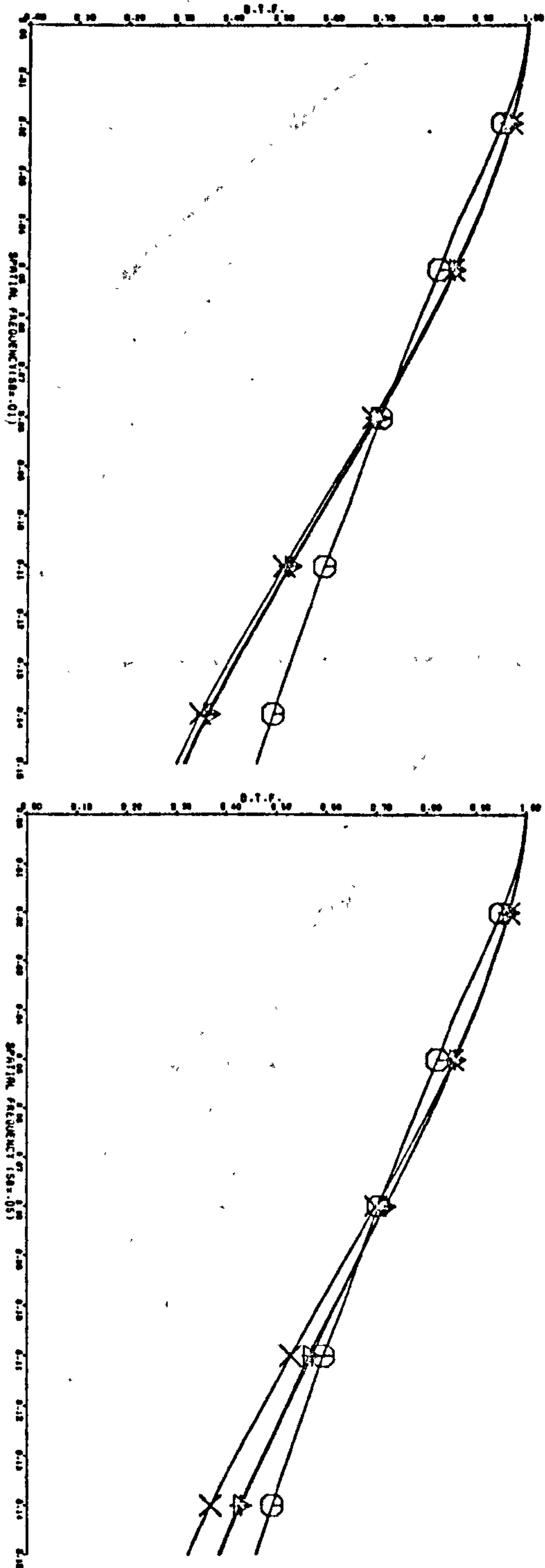


FIG 57

PLOTTING SYMBOLS

- Strehl ○
- Hopkins △
- Diffraction +
- Geometric X

DIFFRACTION OTF AT $M_6 = -10$



PLOTTING SYMBOLS

- Strehl ○
- Hopkins △
- Diffraction +
- Geometric X

FIG 58

DIFFRACTION OTF AT $M_6 = -20$

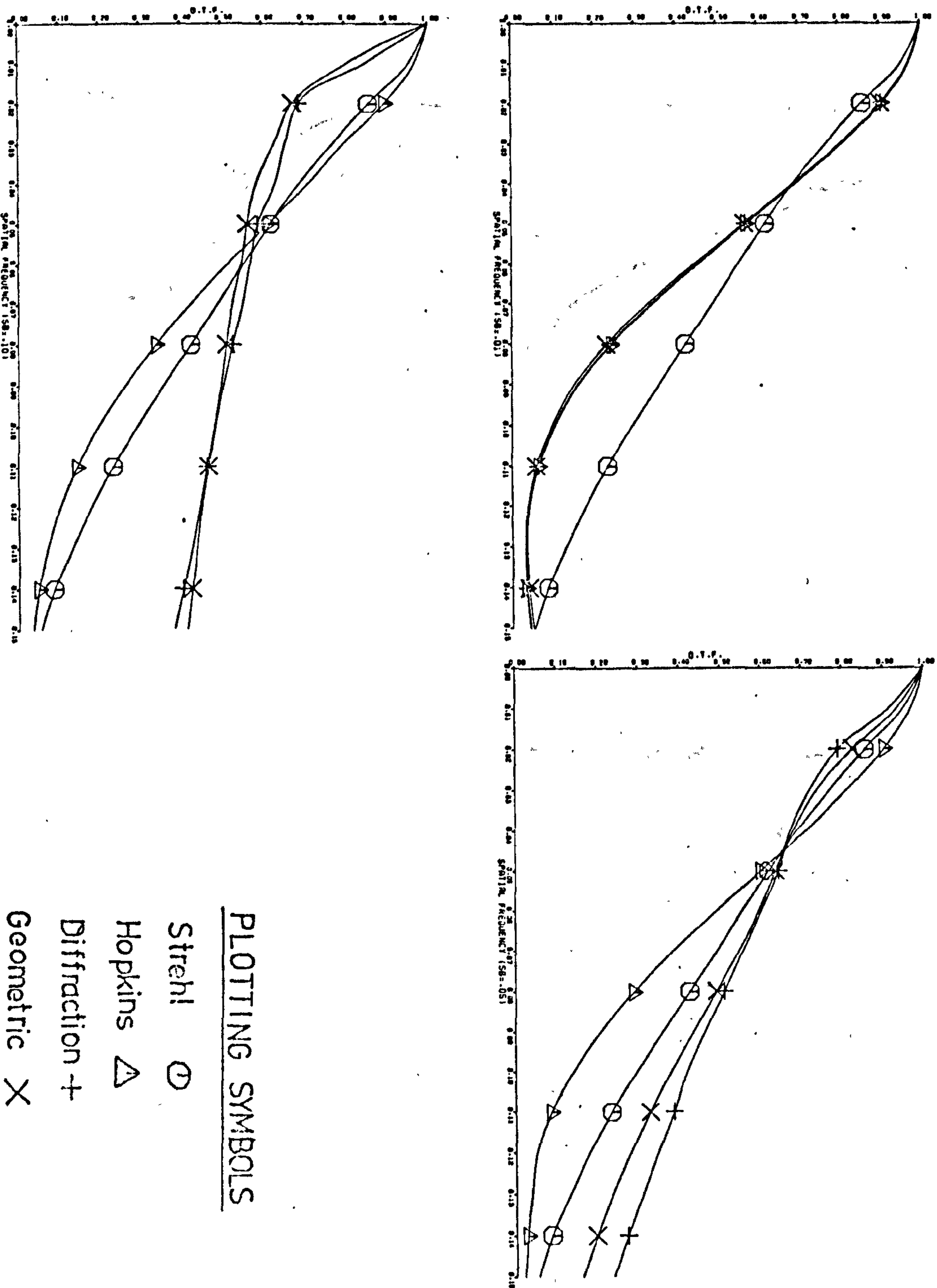


FIG 5.9

DIFFRACTION OTF AT $M_6 = -40$

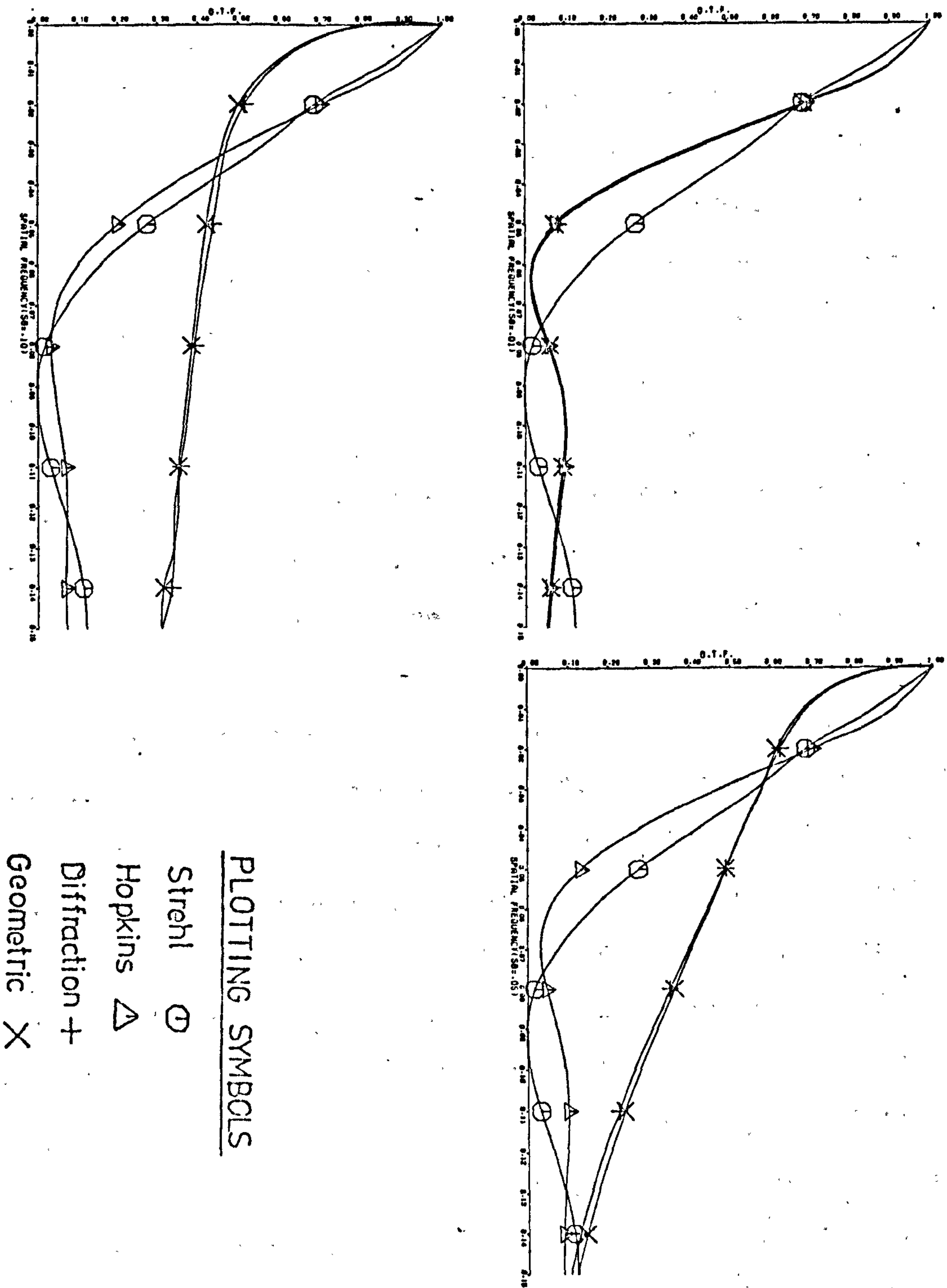


FIG 510

that the aberration near the centre of the pupil remains as small as possible. For the $W_{60} = -40$ case, the aberration is less than 0.25 as far out as 0.7 of the pupil radius, and thus is inside the Rayleigh limit. Outside this well-controlled region the aberration rapidly increases to -16 wavelengths. For the low aberration case, the variation of shape with frequency is much smaller than in the high aberration case, which is an indication of the higher tolerance at low aberrations.

This variation of optimum balance with both size and frequency cannot, of course, be predicted by any of the analytic formulae, but examination of table 5.3 shows the accuracy with which the geometric balances follow the diffraction values over the complete range.

To illustrate further the accuracy of the various criteria, the wavefronts and diffraction o.t.f. values for each of the four criteria are plotted in figures 5.3 to 5.10. The Maréchal criterion produces the minimum variance of the wavefront, and hence may be considered to be "corrected" for spherical aberration. The wavefront graphs show that for low frequencies all the criteria except the Maréchal one overcorrect the high-order aberration. As the frequency and aberration size increase, the diffraction and geometric balances move from this state of overcorrection to a state of undercorrection. The Hopkins balances follow this trend at low aberrations, but at high aberrations the change of shape is much smaller than that of the diffraction and geometric balances, and the Hopkins balance never passes through the corrected state.

As can be seen from the wavefronts and from table 5.3, the Maréchal balance represents an average of the high and low frequency diffraction balances, which is to be expected from the property of the Strehl ratio as the integrated o.t.f., shown in equation 2.32. This is also shown by the o.t.f. plots. In each plot the frequency of optimisation is marked, and for the lowest frequency balance, $s = 0.01$, the Maréchal-Strehl balance gives the lowest o.t.f. value at $s=0.01$, but is higher than the other three criteria at higher frequencies.

Within the tolerance limits, equation 5.14, the Hopkins balances are virtually identical to the diffraction balances, as expected, and the o.t.f. curves are indistinguishable. It is important to notice that even within this region, where the geometric o.t.f. might be expected to give poor results, in no case is the o.t.f. produced by the geometric balance more than 2% lower than the diffraction-balanced value. Thus the geometric o.t.f. remains a valid criterion in this highly corrected region.

Outside this region, the Hopkins-balanced o.t.f. values are in all cases lower than the other three values at the particular balance frequency. In this extended region the geometric balance values are all within 1% of the diffraction values at the balance frequency, and hence the geometric o.t.f. may be used as a merit function to optimise the diffraction o.t.f. over the whole region of investigation.

As has been shown, the Strehl-Maréchal criterion gives reasonable o.t.f. values over the complete region, and is preferable to the more complex Hopkins criterion for use as a preliminary merit function in the initial stages of design.

The contour graphs, figures 5.11 to 5.22, were produced by the method described in chapter 4, and show the variation of o.t.f. with aberration balance. The balances produced by the four criteria are plotted on the graphs to show how closely the Maréchal, Hopkins and geometric criteria approach the theoretical maximum o.t.f., which is, of course, produced by the diffraction o.t.f. optimisation. The series of graphs also demonstrates the way in which the criteria move as the aberration size and spatial frequency vary.

Each graph represents the o.t.f. variation for a fixed value of W_{60} and spatial frequency, s . The contour separation has been varied between the graphs, and its value is given at the top of each graph. Also given is a figure representing the lowest contour plotted. Due to the way the programme calculates the contour values, the minimum contour plotted is actually one contour interval above this stated minimum. In the three cases with $W_{60} = -5, -10$ and -20 wavelengths the scales on the graphs have been fixed for each case for ease of comparison over the three spatial frequencies, and so that all four criteria may be plotted. For the case of $W_{60} = -40$ wavelengths the amount of computation to do this would have been excessive, and so the scales have been changed for each spatial frequency. This means that the Maréchal and Hopkins balances can only be plotted on the low frequency, $s=0.01$, case.

It can be seen that, as expected, the form of the contours is approximately an elongated ellipse. As the frequency and size of aberration increase the ellipse becomes more elongated, and the centre of the ellipse, the diffraction o.t.f. maximum, moves towards lower absolute values of W_{20} and W_{40} .

PLOTTING SYMBOLS
X STREHL + HOPKINS
O DIFFRACTION * GEOMETRIC

M6=-5 S=.01 CONTOUR EVERY .002 TO .90

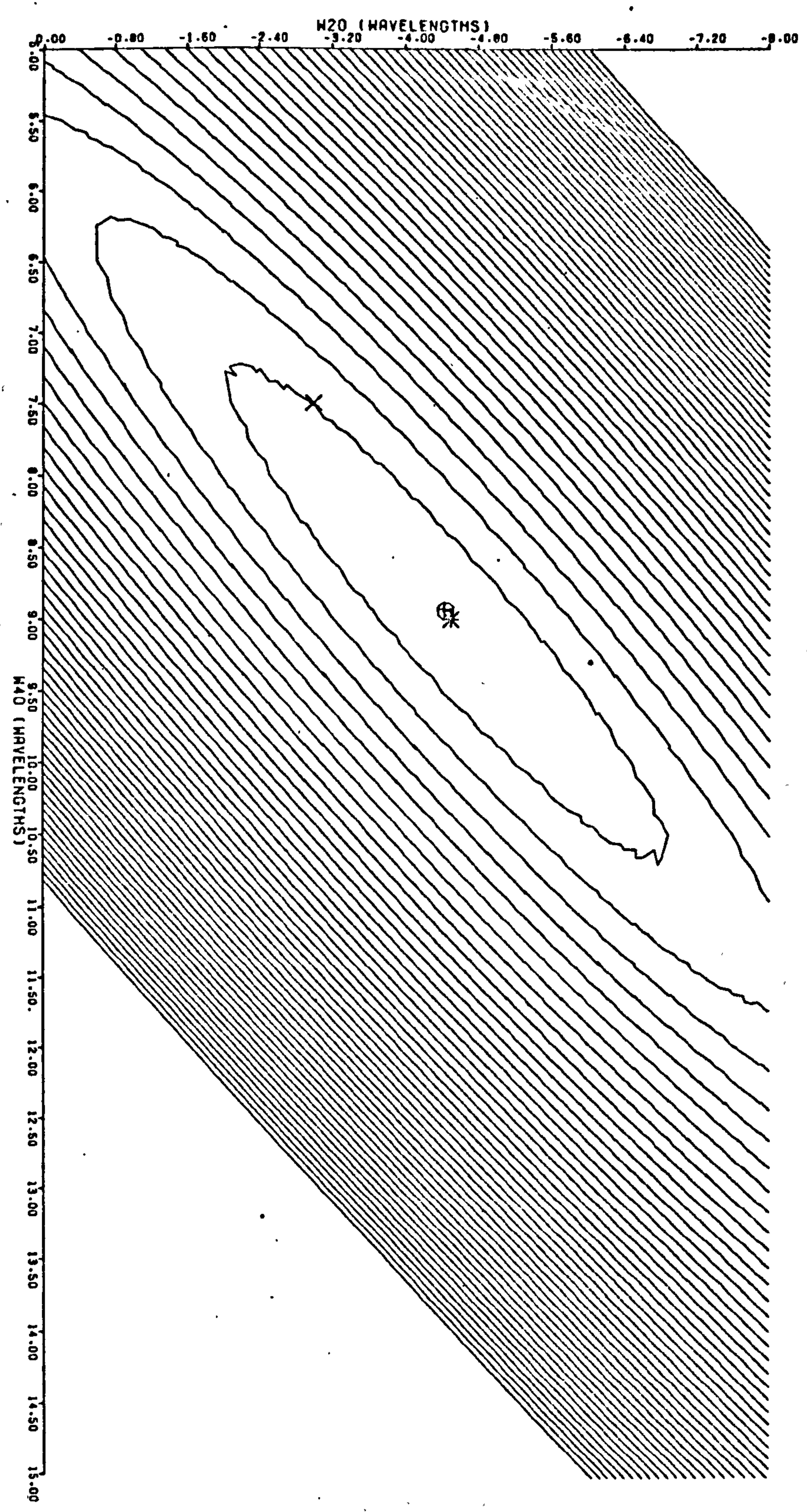


FIG. 511

PLOTTING SYMBOLS

X	STREHL	+	HOPKINS
⊙	DIFFRACTION	*	GEOMETRIC

$M_6 = -5$ $S = .05$ CONTOUR EVERY .005 TO .8

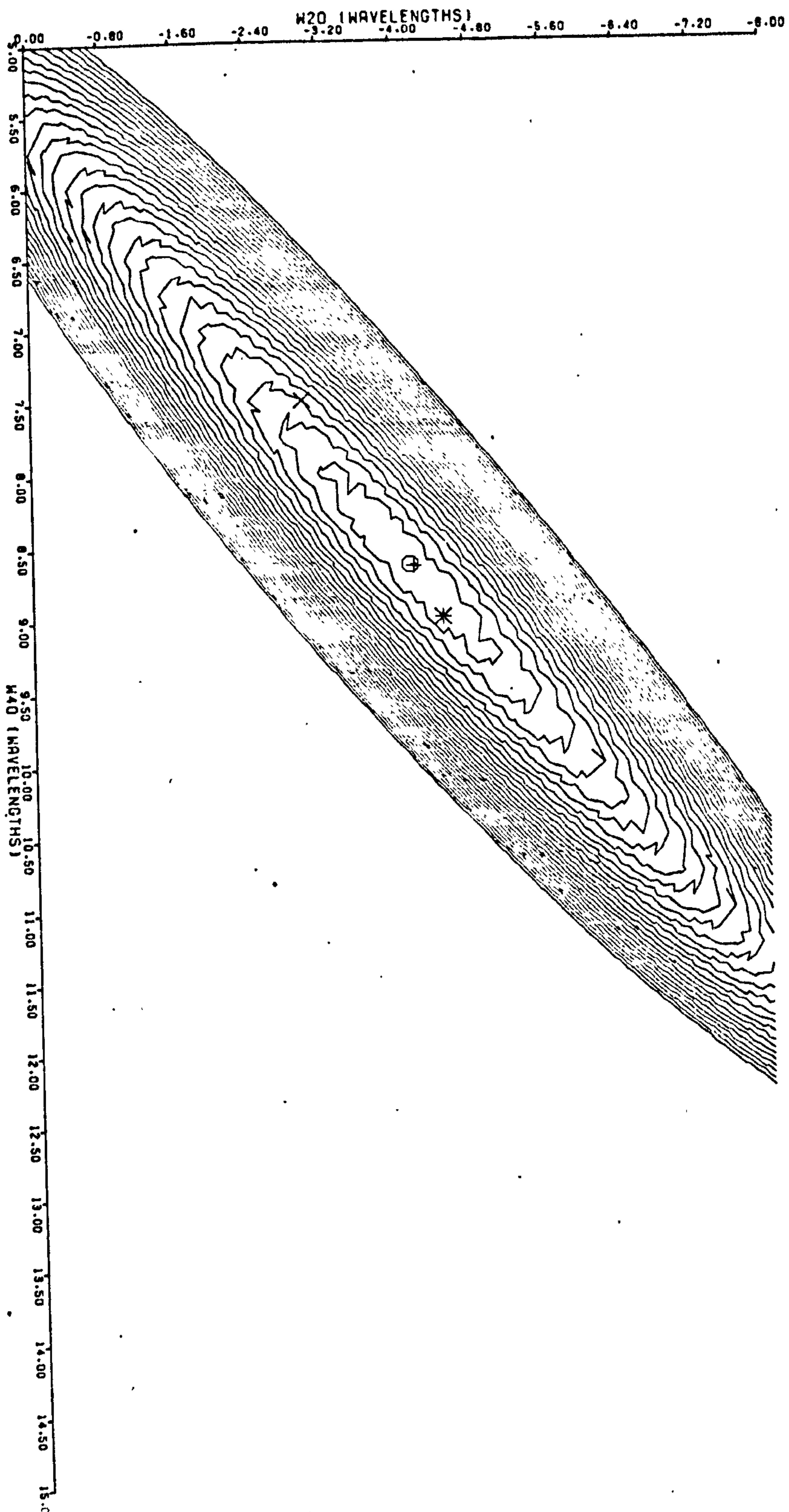
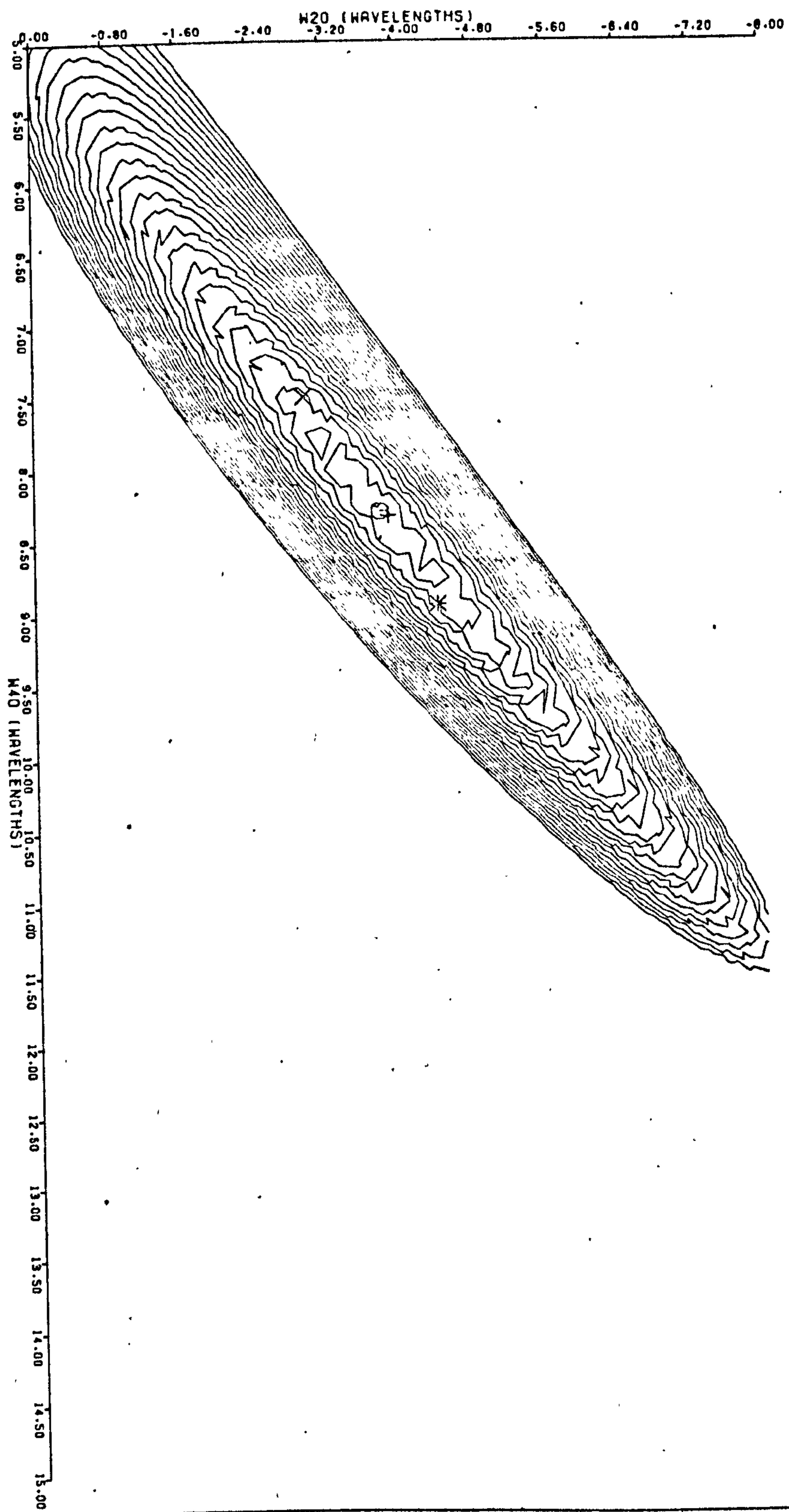


FIG 512



W20 = -5 S = -1 CONTOUR EVERY .01 TO .6

FIG 513

PLOTTING SYMBOLS
 X STREHL + HOPKINS
 O DIFFRACTION * GEOMETRIC

M6-10 S=.01 CONTOUR EVERY .001 TO .90

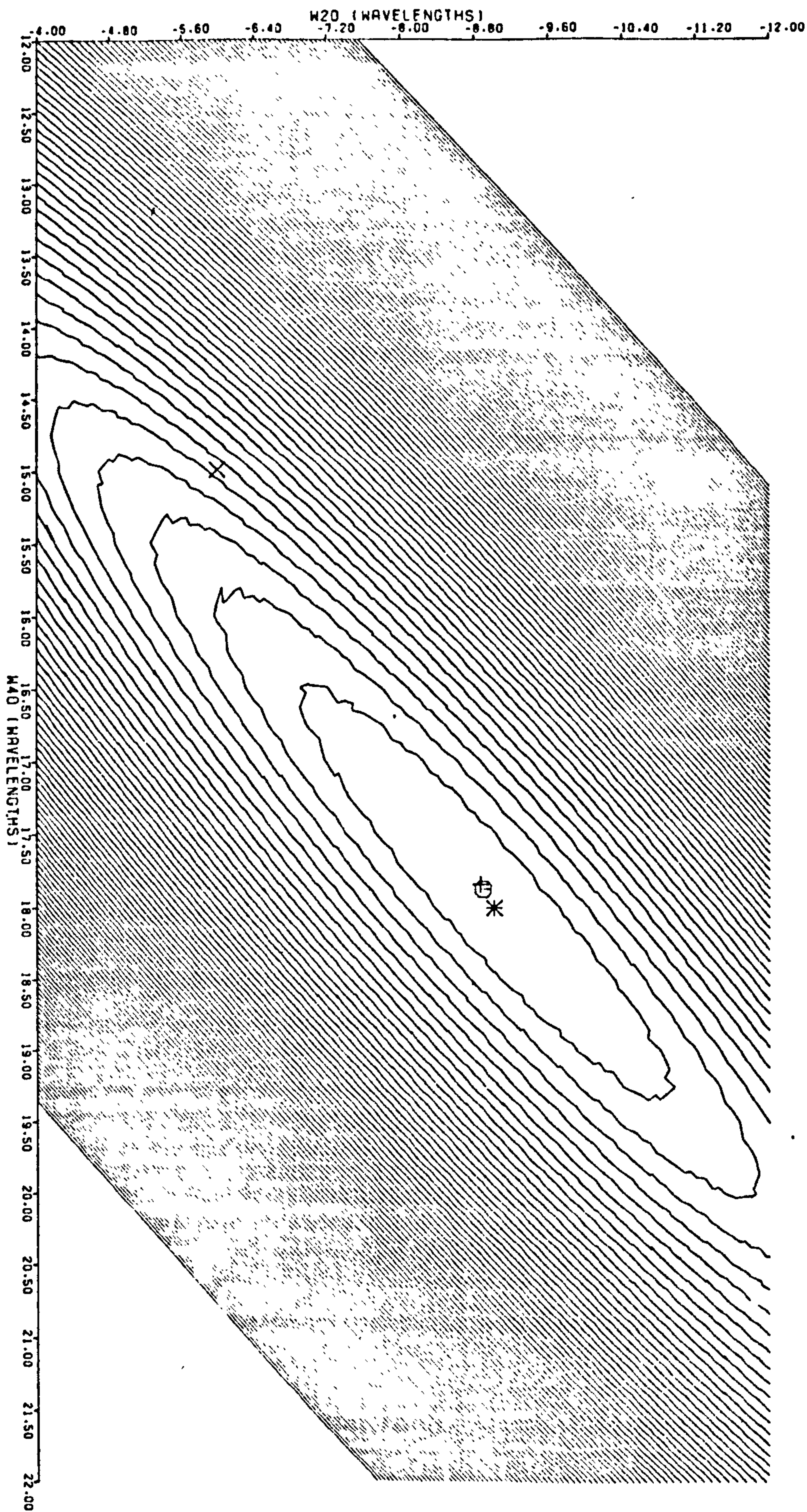
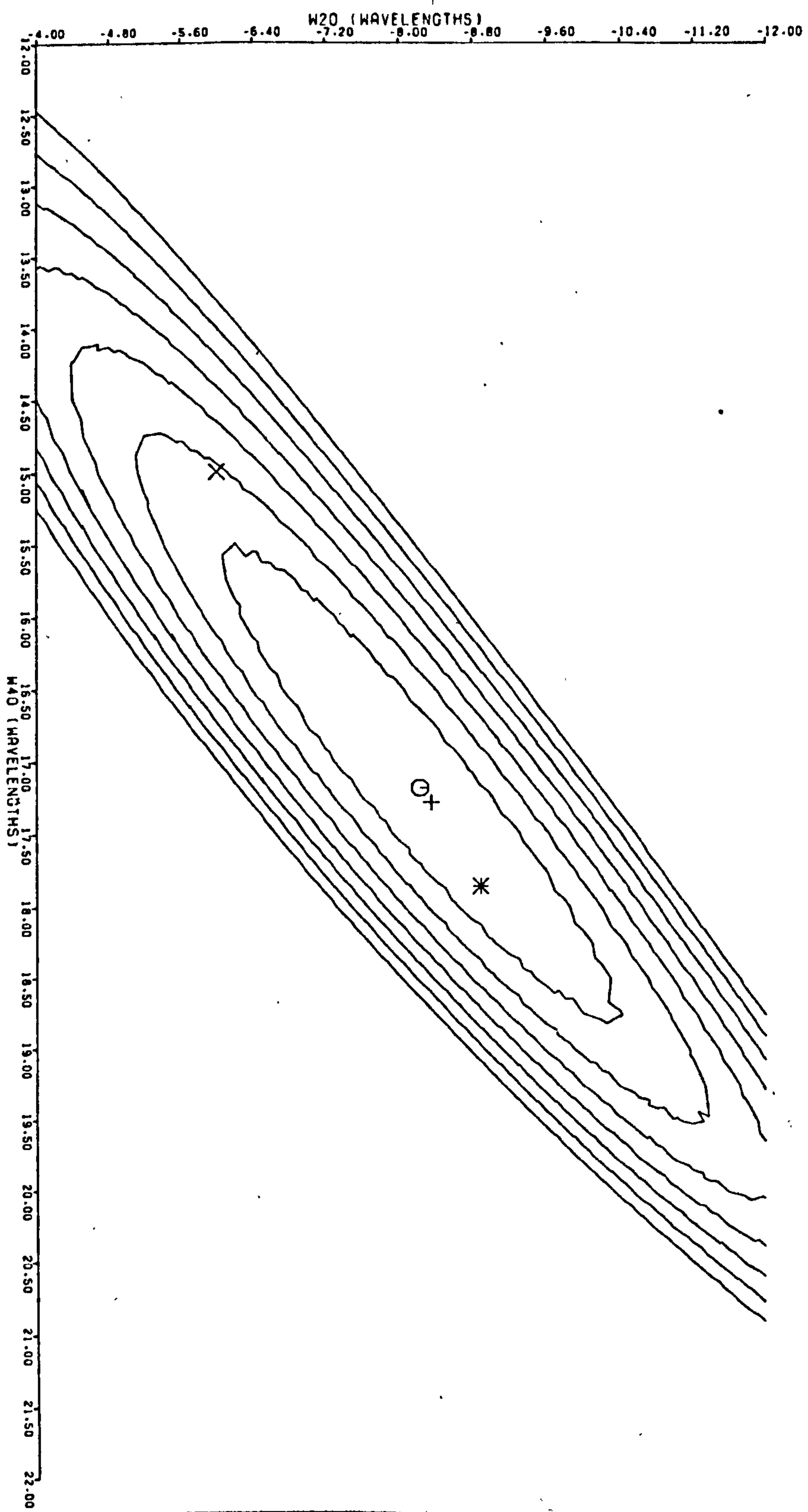


FIG 514



PLOTTING SYMBOLS
X STREHL + HOPKINS
⊖ DIFFRACTION * GEOMETRIC

W6-10 S:-05 CONTOUR EVERY .02 TO .70

FIG 515

PLOTTING SYMBOLS

X STREHL + HOPKINS
 O DIFFRACTION * GEOMETRIC

M6=-10 S=.1 CONTOUR EVERY .02 TO .40

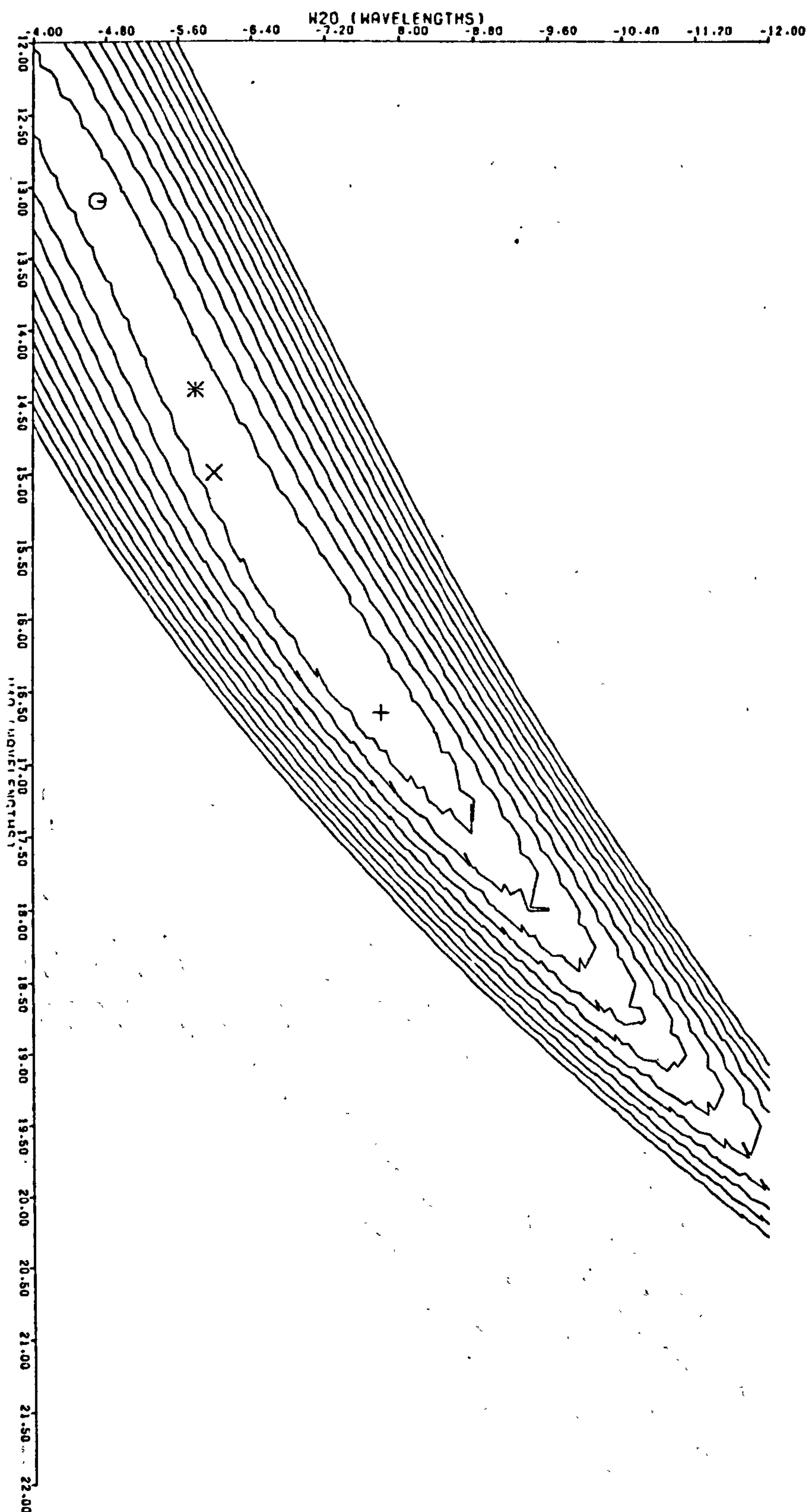


FIG 516

PLOTTING SYMBOLS
X STREHL + HOPKINS
⊖ DIFFRACTION * GEOMETRIC

M6--20 S=.01 CONTOUR EVERY .01 TO .70

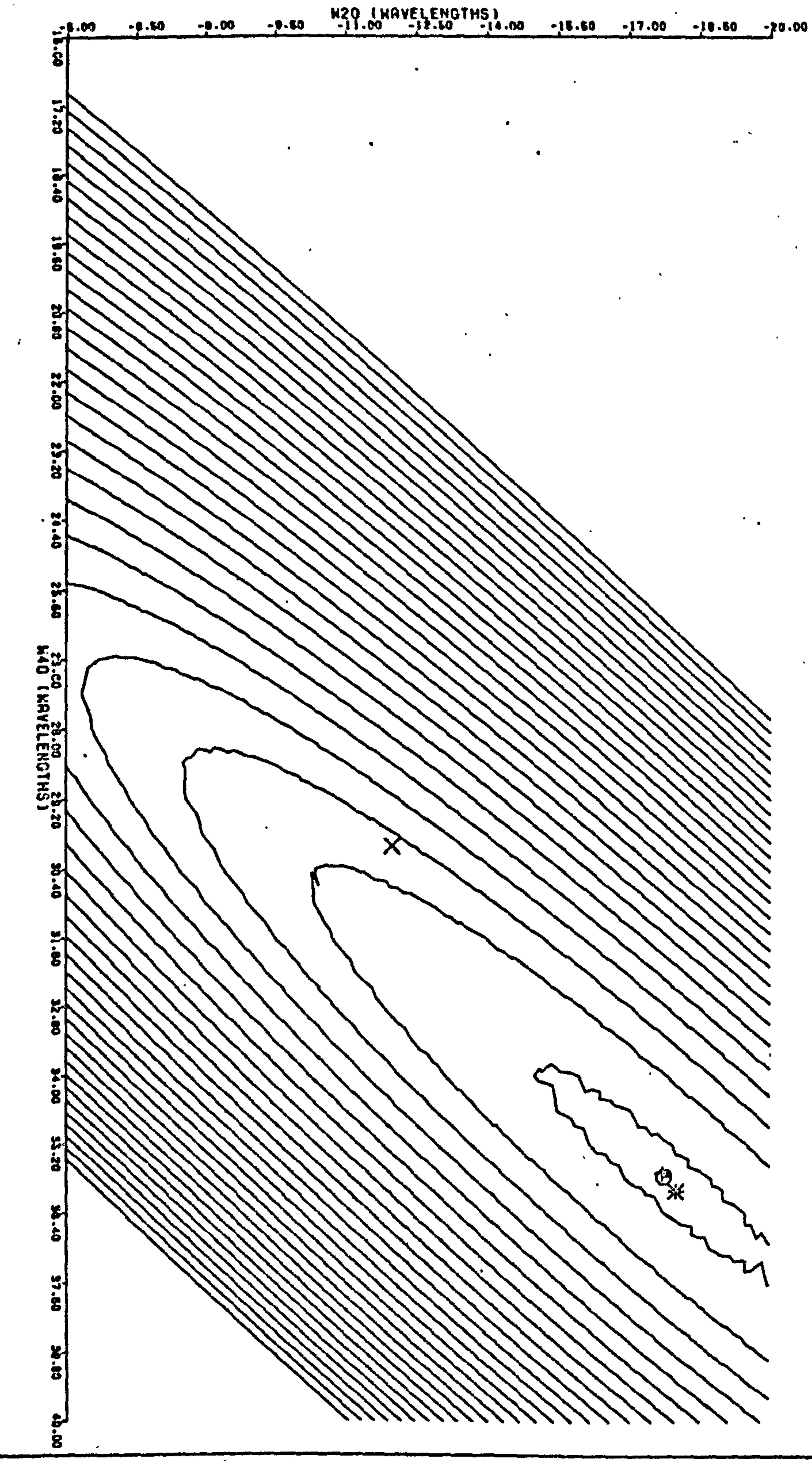


FIG 5-17

PLOTTING SYMBOLS
X STREHL + HOPKINS
O DIFFRACTION * GEOMETRIC

M6--20 S=.05 CONTOUR EVERY .02 TO .30

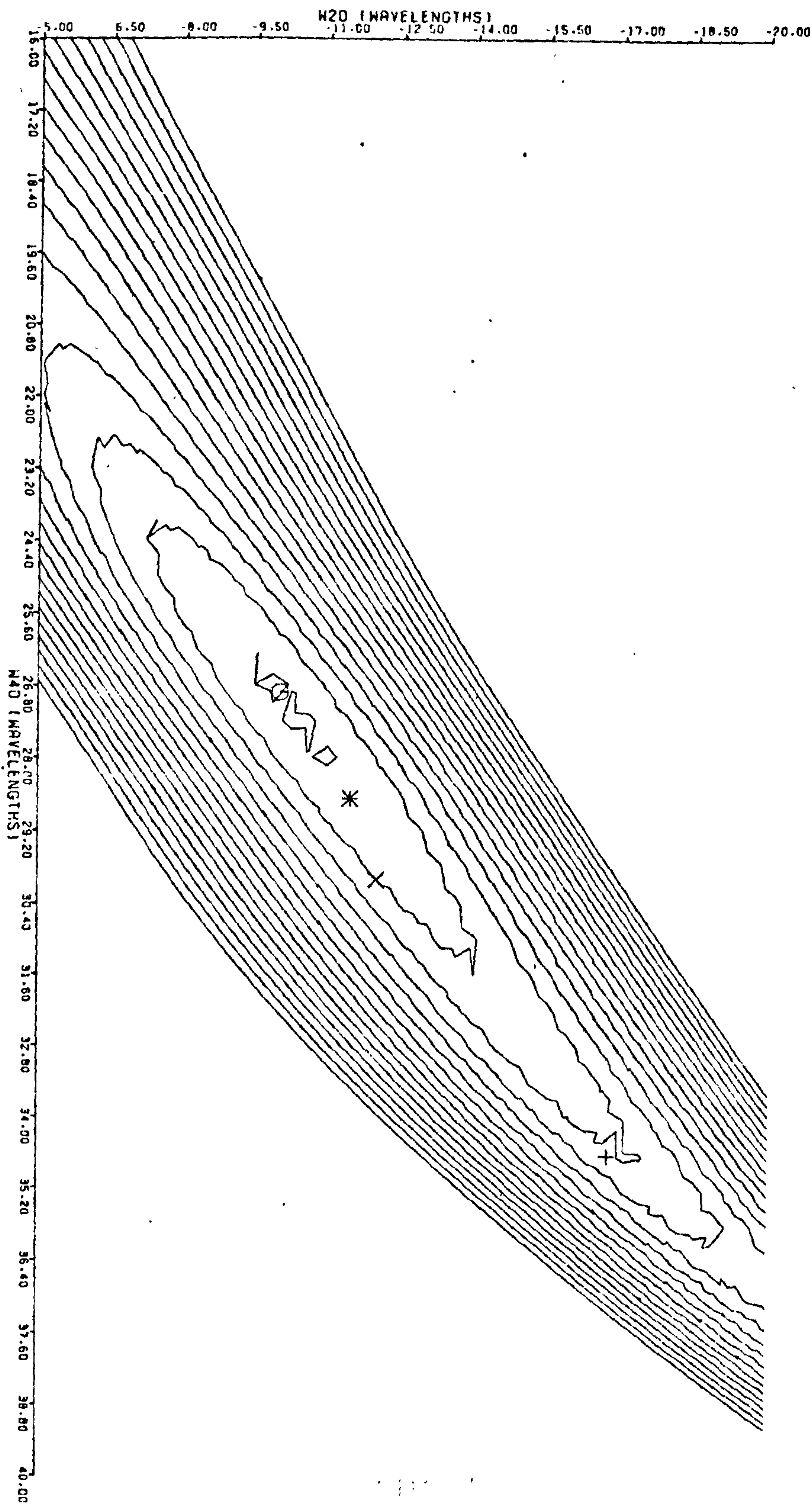


FIG 518

PLOTTING SYMBOLS
 X STREHL + HOPKINS
 O DIFFRACTION * GEOMETRIC

M6-20 S=.1 CONTOUR EVERY .02 TO .2

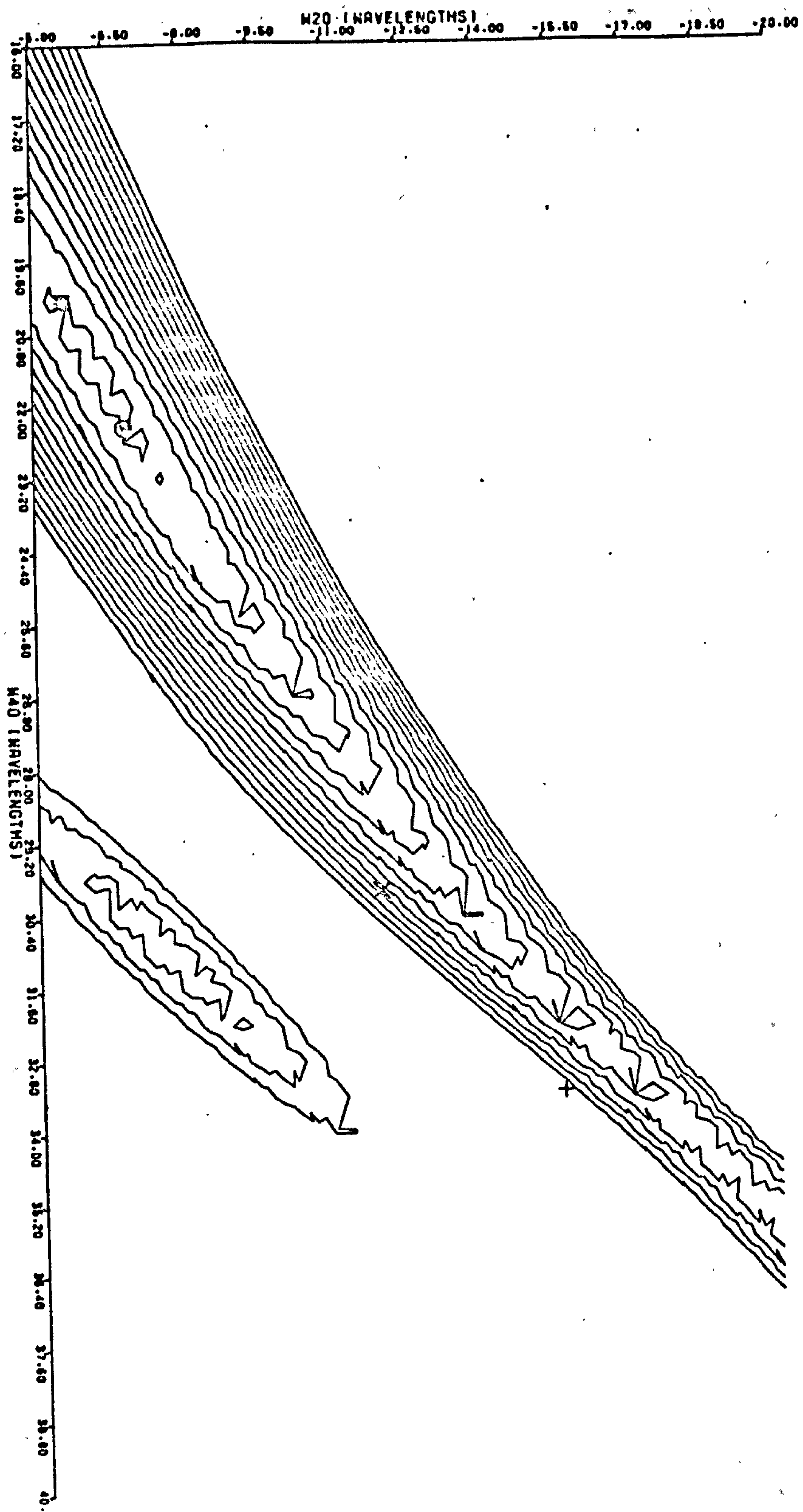


FIG 519

PLOTTING SYMBOLS

X	STREHL	+	HOPKINS
⊖	DIFFRACTION	*	GEOMETRIC

M6=-40 S=.01 CONTOUR EVERY .002 TO .80

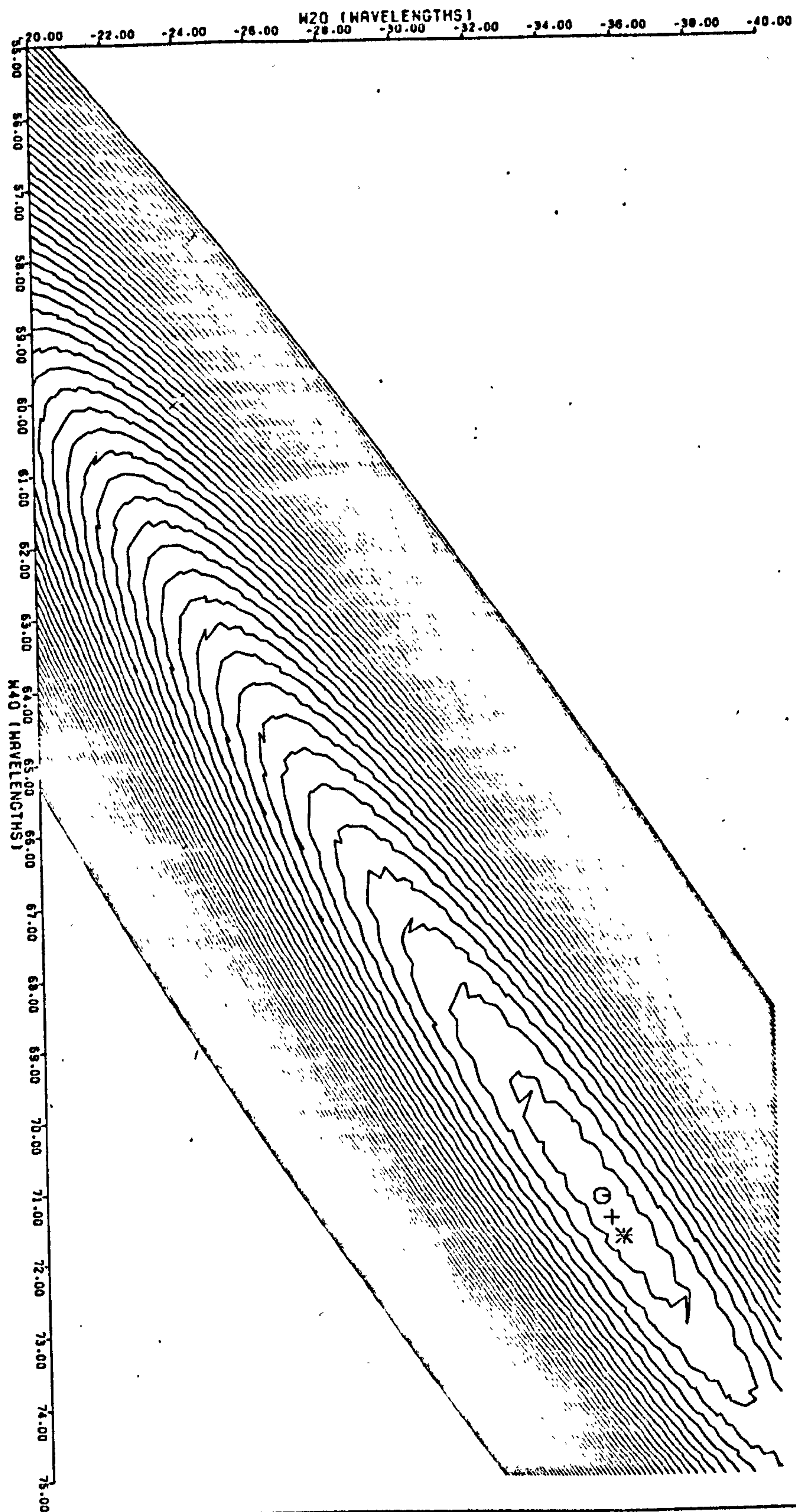


FIG 5-20

PLOTTING SYMBOLS
 X STREHL + HOPKINS
 O DIFFRACTION * GEOMETRIC

MS--40 S=.05 CONTOUR EVERY .02 TO .20

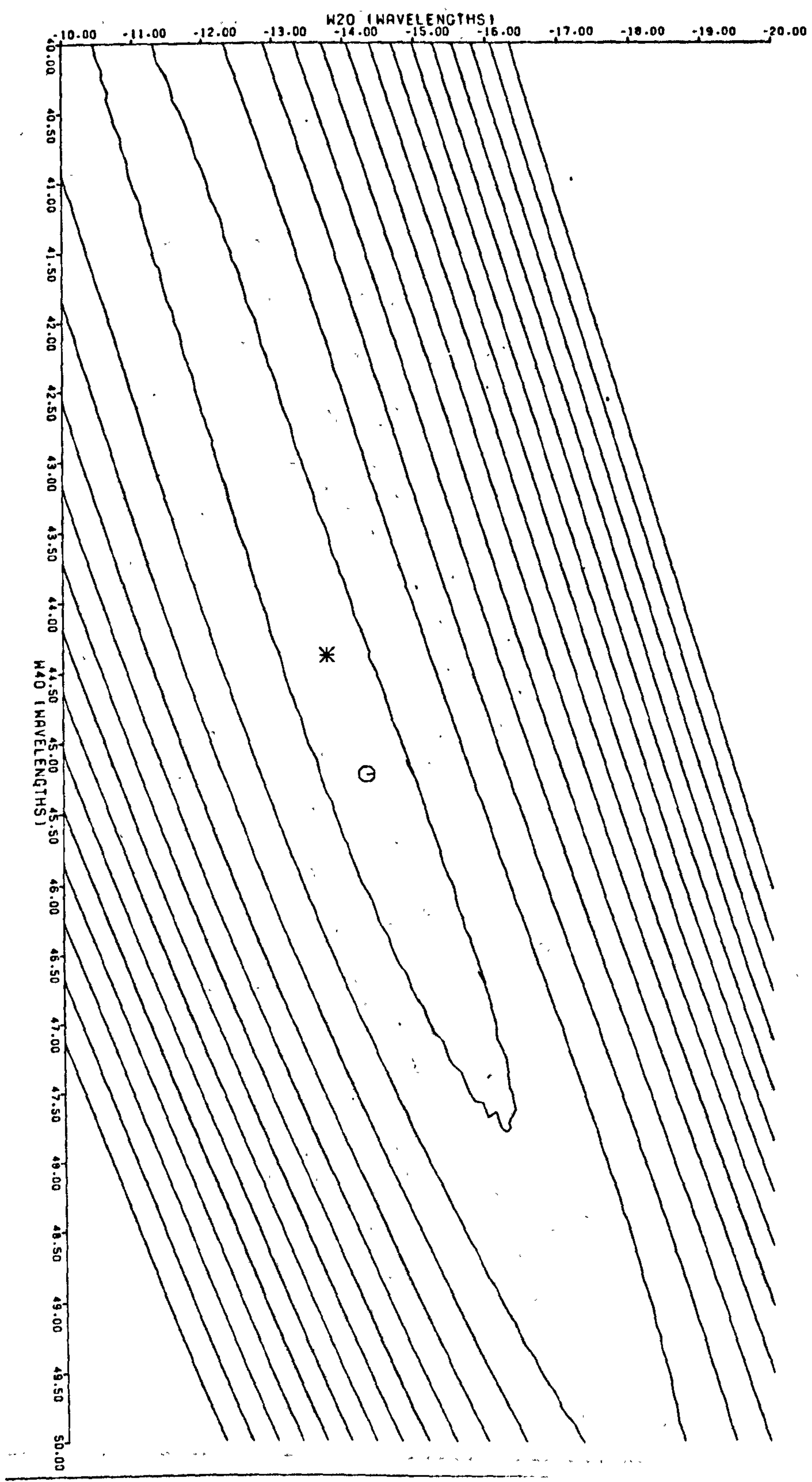
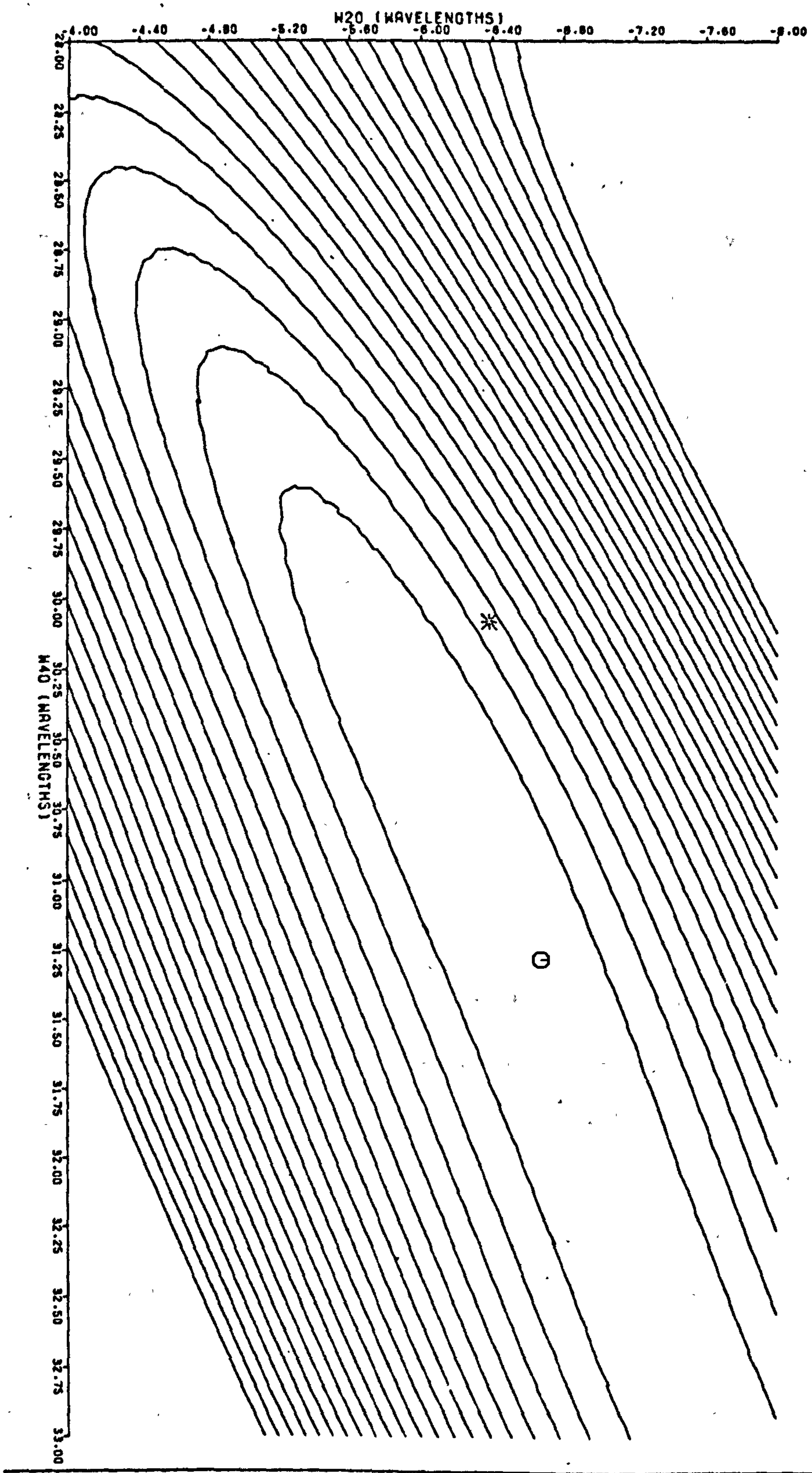


FIG 521



M5--40 S=.1 CONTOUR EVERY .01 TO .2

PLOTTING SYMBOLS
X STREHL + HOPKINS
O DIFFRACTION * GEOMETRIC

FIG 5.22

These plots, particularly those for -10 and -20 wavelengths, show very clearly the way in which the Maréchal balance attempts to form an overall balance, and the way in which the Hopkins balance fails to move with the diffraction balance at the higher aberrations. They also show the closeness with which the geometric balance follows the diffraction balance.

It is to be expected that for any given aberration balancing situation there is a possibility of more than one local o.t.f. maximum existing in parameter space. One such case is clearly shown in the contour plot, figure 5.19, for $W_{60} = -20$, $s = 0.10$. In this case the secondary maximum is much lower than the true maximum, and well removed from the expected best balance. A more interesting example was found when the balance at $W_{60} = -10$, $s = 0.1$ was investigated.

The diffraction optimisation programme was given an arbitrary starting point and produced the balance $W_{40} = 13.1$, $W_{20} = -4.7$. These values seemed reasonable when compared with other values, and on the contour plot, figure 5.16, the point lies within the highest contour plotted, at 0.62. When the diffraction o.t.f. values were calculated for the four criteria it was noticed that the value produced by the diffraction balance, 0.626, was lower than that produced by the Hopkins balance, 0.637, and the Strehl balance, 0.634. Since the optimisation programme had been extensively tested and was working correctly, these results suggested the existence of more than one maximum.

The optimisation programme was run again, starting from a balance point on the opposite side of the expected true maximum. This run produced a balance point at $W_{40} = 16.11$, $W_{20} = -7.27$, with

an o.t.f. value of 0.649, which indicated that the true maximum $\times$ had been found. The o.t.f. curve wavefront and contour plot are shown in figures 5.23 to 5.25. The contour spacing on figure 5.25 is too coarse to show the double maximum, so an enlarged scale plot was produced, and this is shown in figure 5.26. The o.t.f. routine worked to an error of less than 0.002, so to obtain the most sensitive plot possible the contour interval was set at 0.002. Although the contours are not smooth, due to the grid spacing, the double maximum is clearly shown.

The evidence of the contour plots and the values in tables 5.2 and 5.3 leads to further investigation of the relationship between B_{26} and B_{46} . One thorough treatment by Rosenhauer et al (1965) considers the problem by giving W_{40} a fixed value and varying W_{20} and W_{60} . The authors take, as an optimisation criterion

$$I = \int_0^{s_0} D(s) ds \quad 5.15$$

where $D(s_0) = 0.2$.

This function represents an average over all spatial frequencies in a similar manner to the Strehl criterion, with a cut-off at an o.t.f. value of 0.2.

The wavefront aberration is expressed in the form

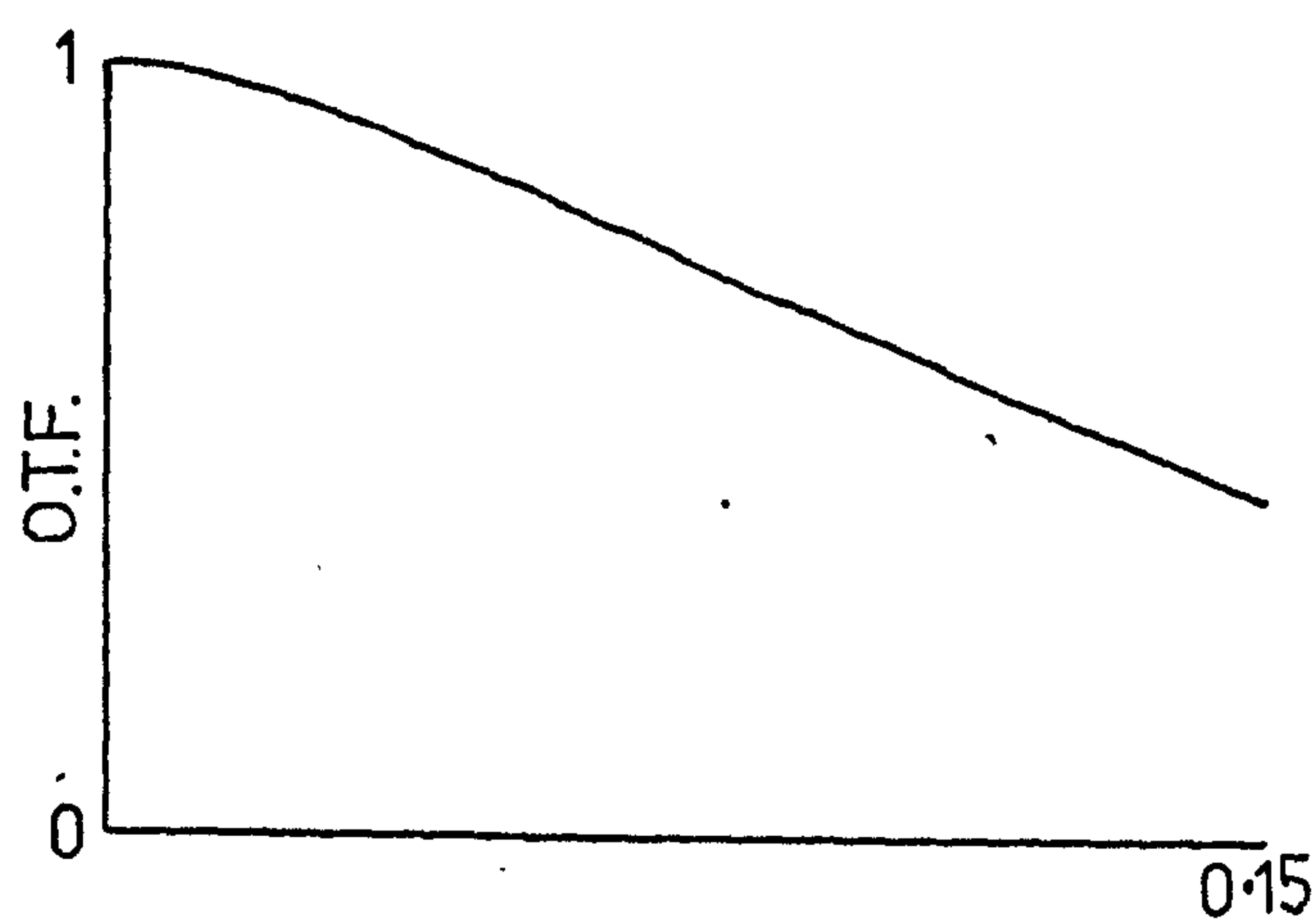
$$W = A(a\rho^2 - 0.5\rho^4 + c/3\rho^6) \quad 5.16$$

In terms of the coefficients used in this thesis

$$c = 3/(2B_{46}) \quad a = B_{26}/(2B_{46}) = B_{24}/2 \quad 5.17$$

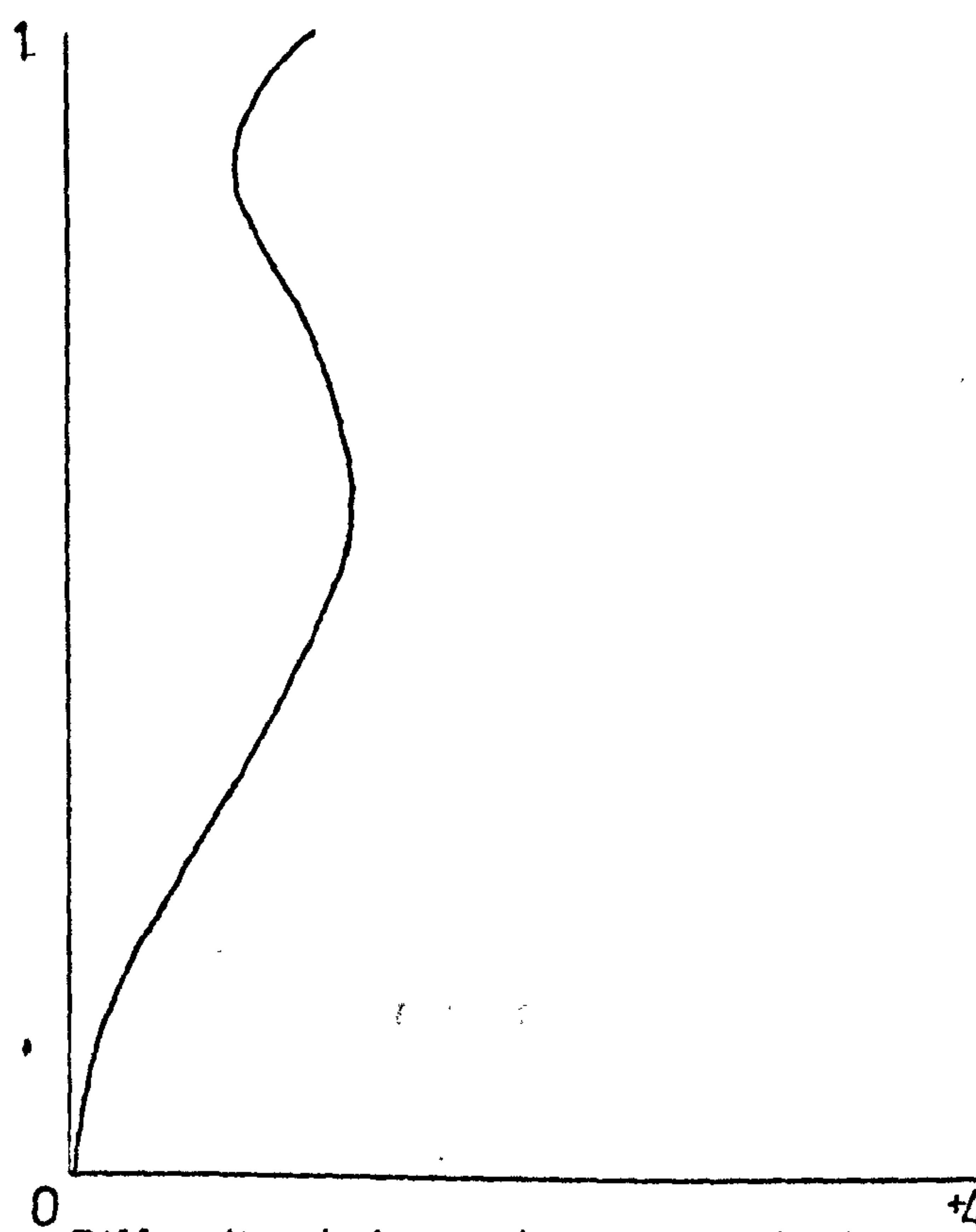
where B_{24} is defined as B_{26}/B_{46} , and absolute values of B_{24} , B_{26} and B_{46} are used.

For a value $c = 1.0$ the system is corrected, with $B_{46} = 1.5$, which is the Maréchal condition. The general results agree with the results in this investigation, although many of the results



Diffraction o.t.f. at $W_6 = -10$ after rebalancing

FIG 5.23



Diffraction-balanced wavefront at $W_6 = +10$ after rebalancing

FIG 5.24

PLOTTING SYMBOLS
 X STREHL + HOPKINS
 O DIFFRACTION * GEOMETRIC

W6=-10 S=.1 CONTOUR EVERY .02 TO .40

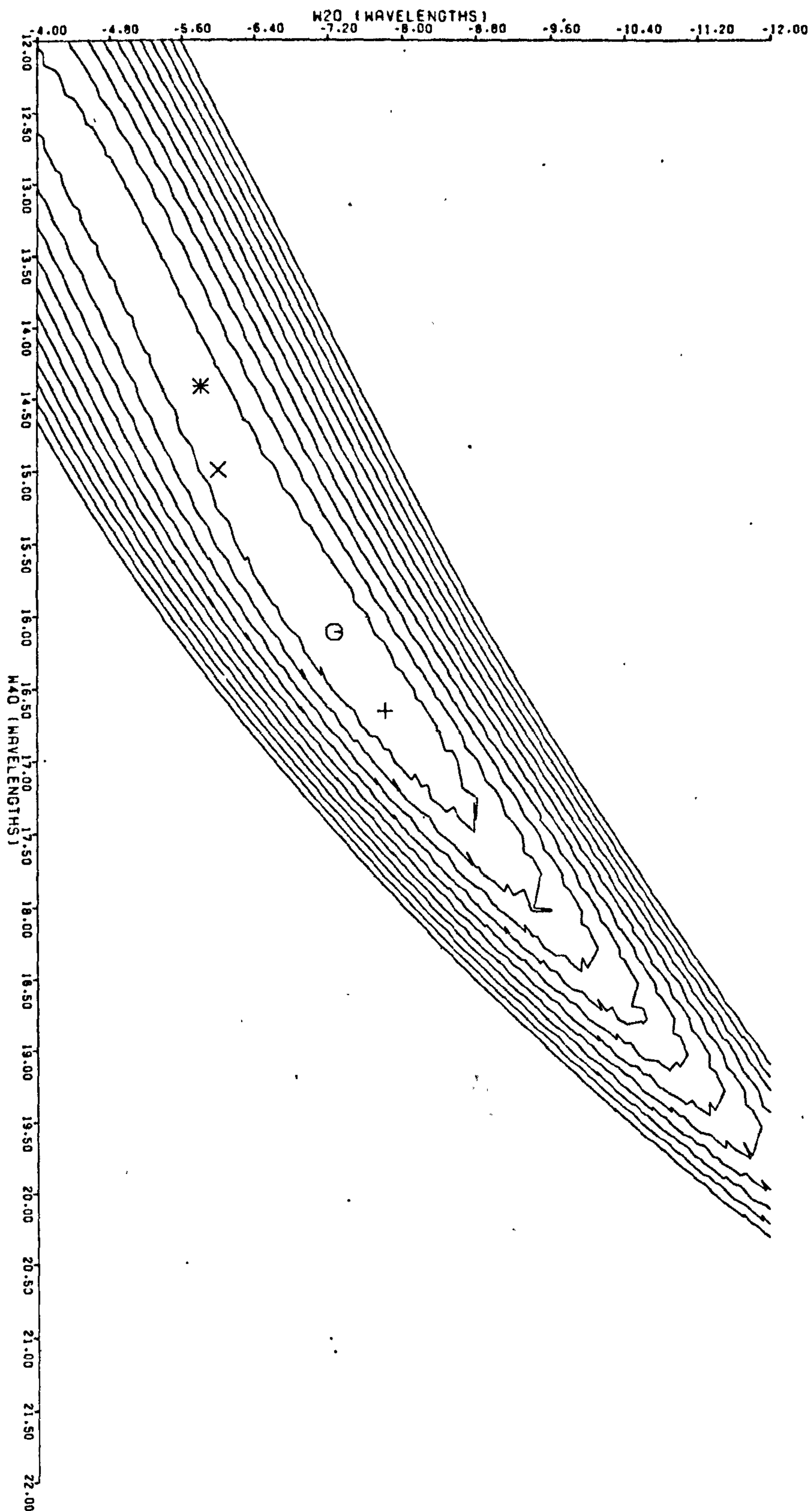
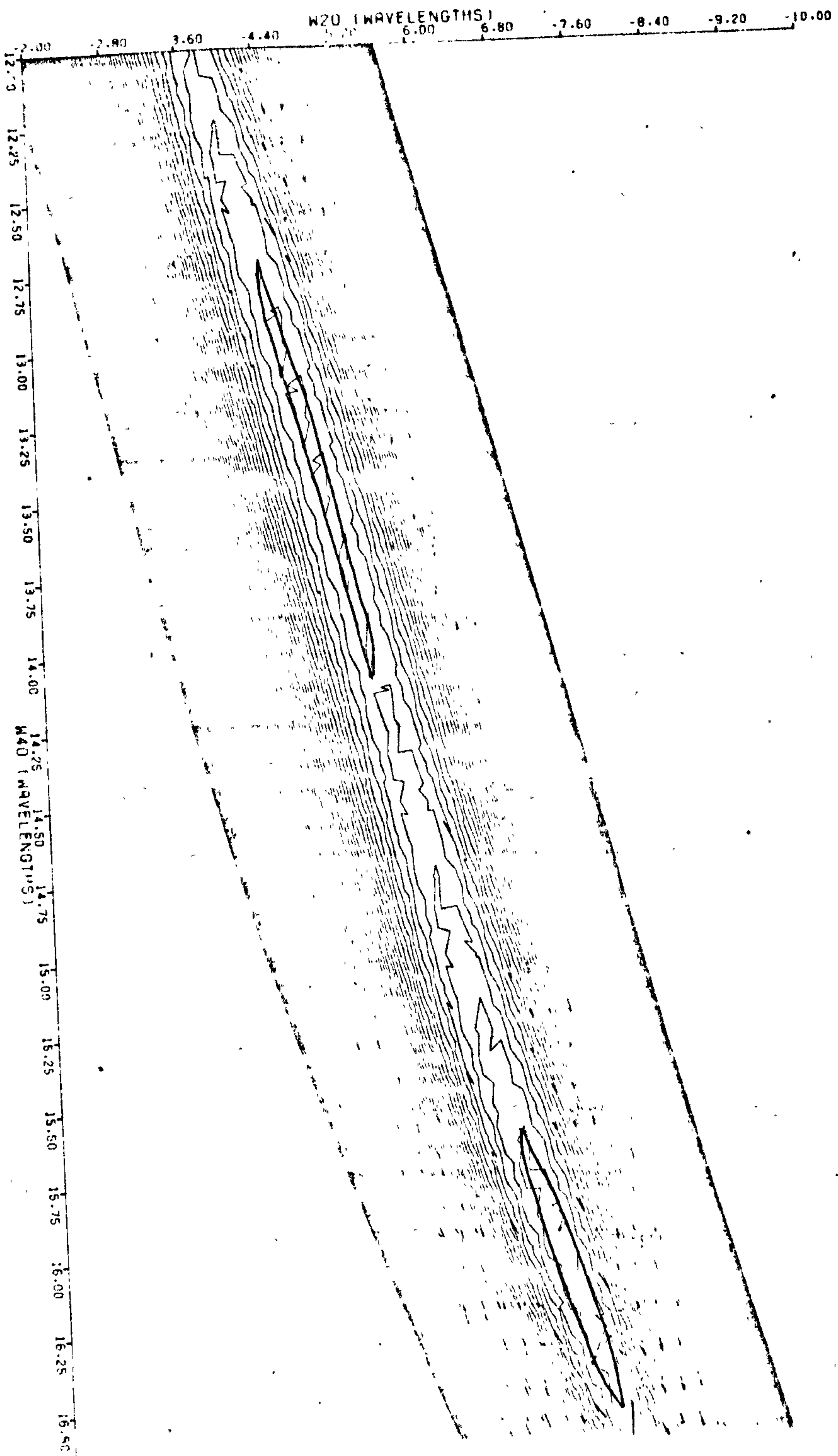


FIG 5.25



W20 - 10 S = 1 CONTOUR EVERY .003 TO .40

FIG 526

are not directly comparable, since the criteria are different.

For a given value of the shape coefficient, c , Rosenhauer gives the defocus term, a , as

$$a = 0.5 - 0.3c \quad 5.18$$

Then
$$\frac{B_{26}}{2B_{46}} = 0.5 - 0.3 \times \frac{3}{2B_{46}} \quad 5.19$$

hence
$$B_{26} = B_{46} - 0.9 \quad 5.20$$

If we take $B_{46}=1.5$, then $B_{26}=0.6$, which is again the value given by the Maréchal formulae, and is expected from the definition of the criterion 5.15.

Rosenhauer also quotes a balance evaluated by Lukosz (1958) where

$$a = 2/3 - c/2 \quad 5.21$$

and
$$B_{26} = 4B_{46}/3 - 1.5$$

This equation is valid only for small frequencies and aberrations, but Rosenhauer states that in general the best balance will lie between the two lines 5.21 and 5.20.

These two lines are plotted together with the balances found by the Hopkins, geometric and diffraction criteria in figure 5.27. It can be seen that most of the high-aberration diffraction and geometric balances lie outside the Rosenhauer-Lukosz region, but this is expected since the Rosenhauer equation is based on low aberration cases.

Rosenhauer also considered the effect of stopping down on the various balances he produced. From equation 2.18 the reduced spatial frequency depends on the numerical aperture, and so as the system is stopped down the frequency scale must be altered,

Plot of $\beta_{x6} : \beta_{z6}$ For Axial Aberration Balancing

- SYMBOLS
- + DIFFRACTION BALANCES AT $S = 0.1$
 - x GEOMETRIC BALANCES AT $S = 0.1$
 - o MINIMUM R-M-S WAVEFRONT

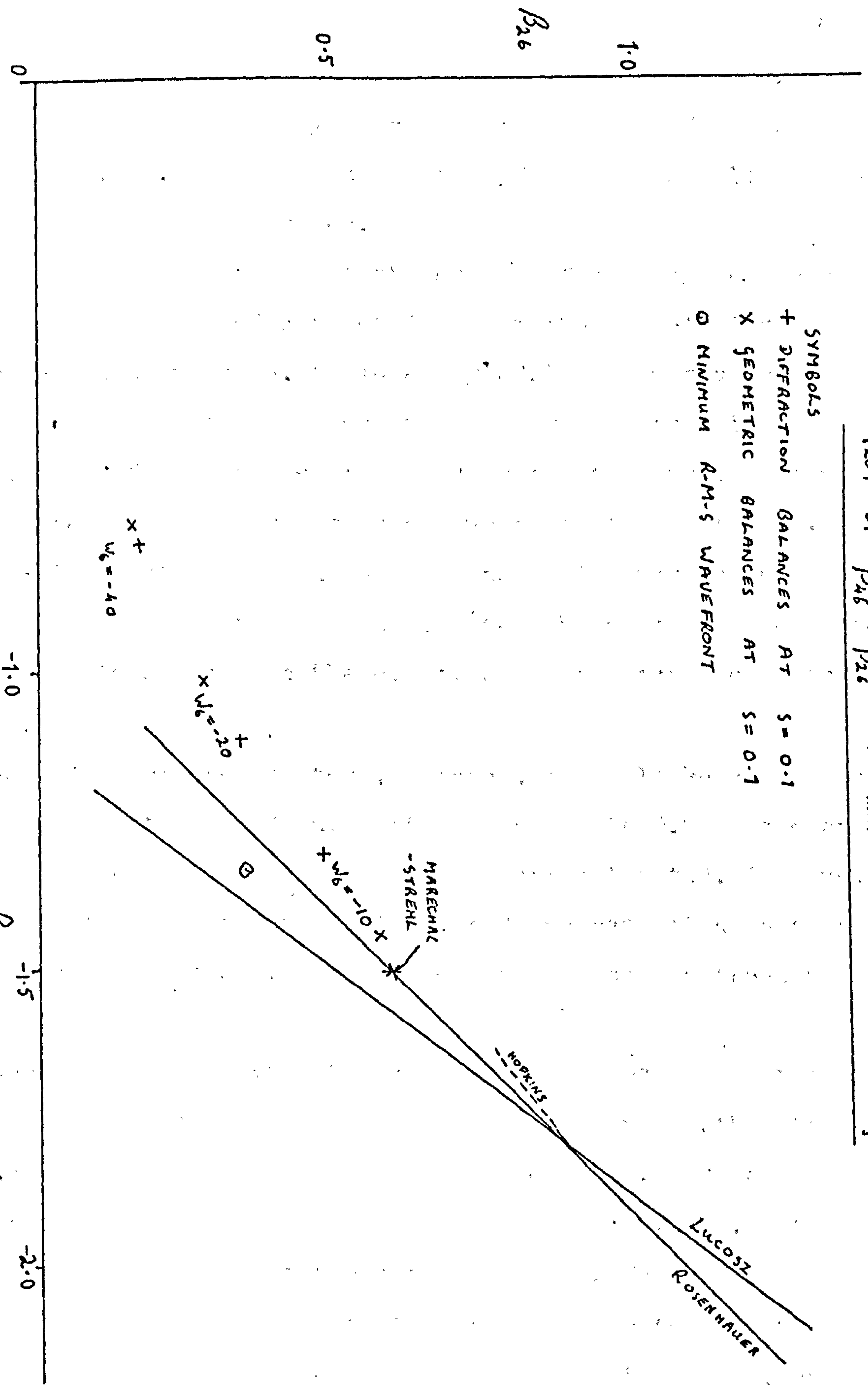


FIG 5.27

and the cut-off frequency is reduced. Thus even for a perfect lens the o.t.f. for all real spatial frequencies will be lower for the stopped down system.

Rosenhauer concluded that the overall frequency response for the overcorrected and corrected systems would deteriorate at first on stopping down, followed by a slight improvement until the effects of the reduction in scale again worsen the performance. This effect can be predicted by considering the shape of the corrected and overcorrected wavefronts in figure 5.2. As the pupil is stopped down the variance will increase until the aperture is 0.6 of the original, after which it will decrease, giving an improvement in the o.t.f. until the scale effects dominate.

The undercorrected systems, however, will all be considerably improved on stopping down, until the pupil is 0.7 of the original, when the system is effectively diffraction-limited. The system is then similar to the perfect lens, and any reduction in aperture will reduce the o.t.f. due to the scale change in frequency.

Conclusions

There are several valuable conclusions to be drawn from this section. The most important for the later work in this thesis is the success with which the geometric o.t.f. acts as a merit function for optimisation of the diffraction o.t.f., both in the low aberration and high aberration systems. The Hopkins criterion, while marginally superior to the geometric criterion within its limited range, rapidly deteriorates outside its range, and for a wide-ranging merit function the simpler Strehl-Maréchal criterion proves superior. As will be shown later, the geometric function

is only used in the final stage of design, when it is markedly the best function for o.t.f. optimisation.

The results also indicate that for high-frequency, high-aberration correction an undercorrected wavefront has the best frequency response. This was generally expected from practical experience, but the degree of undercorrection was surprising. This result is of use in the construction of the traditional type of merit function, operating on wavefront aberrations, since much greater weight can be given to the rays at the centre of the pupil, allowing the wavefront aberrations at the edge of the pupil to increase.

Coma Aberration Balancing

The first terms in the general aberration polynomial to become important as the object moves away from the optic axis are the coma terms, since these vary as the first power of the object height or field angle. The terms considered in this section are the primary and secondary coma, given by

$$W = W_{31} \rho^3 \cos(\phi) \quad \text{primary coma} \\ + W_{51} \rho^5 \cos(\phi) \quad \text{secondary coma} \quad 6.1$$

In the spherical aberration case a longitudinal defocus term was included, and for the coma case it is useful to consider a transverse focal shift in the polynomial. This shift gives a term

$$W = W_{11} x = W_{11} \rho \cos(\phi) \quad 6.2$$

Since there is no longer rotational symmetry about the optic axis, as shown in figure 6.1, both the sagittal and tangential frequency response must be considered. In general the tangential o.t.f. will be complex, with an associated phase shift, and this makes the aberration balancing more complicated to analyse.

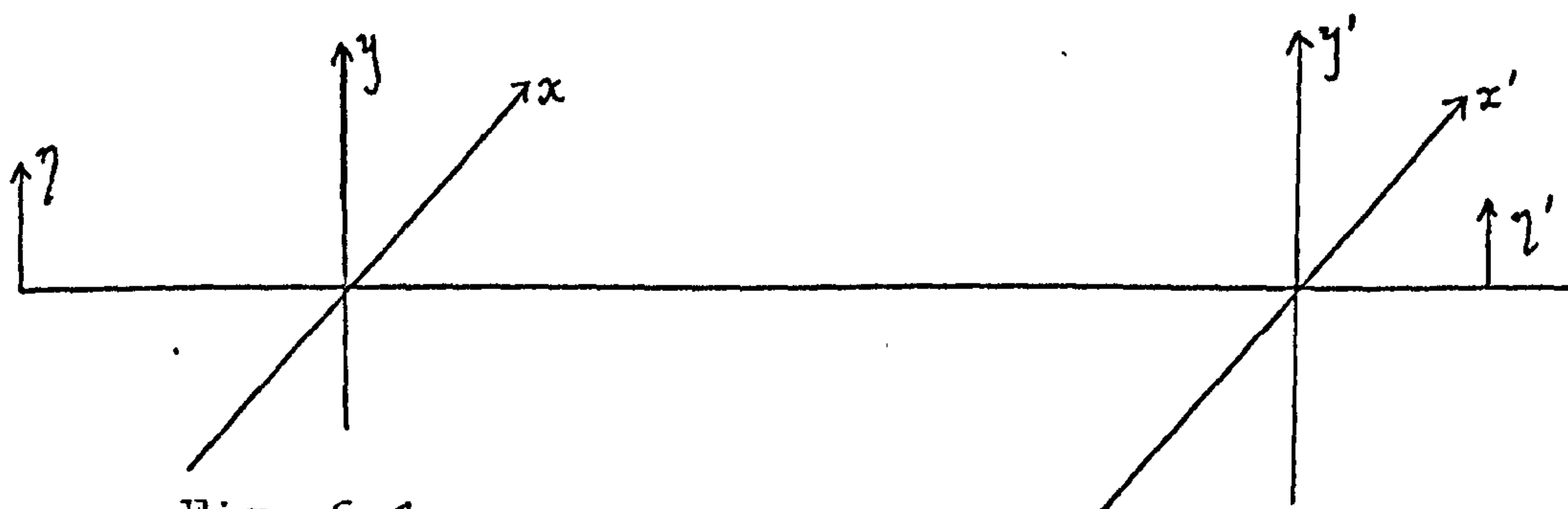


Fig. 6.1

Although the transverse focal shift, W_{11} , has been included, it will be shown that this does not affect the modulus of the o.t.f., but only the phase of the tangential case.

From Fourier theory, if a function $f(x)$ transforms to $F(s)$, i.e.

$$f(x) \rightleftharpoons F(s)$$

then the shift theorem shows that

$$f(x+a) \rightleftharpoons \exp(-2\pi ias).F(s) \quad 6.3$$

Thus the transform of $f(x+a)$ remains the same in modulus, but contains a phase shift, $2\pi as$. This phase shift is proportional to the frequency, and hence there is no relative phase shifting between the components.

The diffraction o.t.f. equation, 2.31, may be separated into the sagittal and tangential parts as

$$D_s(s,0) = \iint_S \exp(2\pi i(W(x+s/2,y)-W(x-s/2,y))) \, dA \quad 6.4$$

$$D_t(0,t) = \iint_S \exp(2\pi i(W(x,y+t/2)-W(x,y-t/2))) \, dA$$

For the transverse focal shift, $W_{11}y$, the contributions to the two cases are

$$\begin{aligned} W(x+s/2,y)-W(x-s/2,y) &= 0 \\ W(x,y+t/2)-W(x,y-t/2) &= W_{11}t \end{aligned} \quad 6.5$$

Thus the transverse shift leaves the sagittal o.t.f. unaltered, and adds a constant phase term $\exp(2\pi iW_{11}t)$ to the tangential o.t.f., as predicted from Fourier theory.

To illustrate this effect, and as a test for the o.t.f. programme, the programme was used to calculate the phase at a given spatial frequency as a function of W_{11} . The other terms in the polynomial were taken as $W_{31} = 7.155$ and $W_{51} = -5$, for a

spatial frequency, $t=0.15$. The values of phase calculated are given below

W_{11} (wavelengths)	-3	-2	-1	0	1
Phase ϕ	-.540	.4025	1.345	2.287	3.232

Phase effects of transverse shift

These values give an accurate linear graph with the equation

$$\phi = 0.943(W_{11} + 2.462) \quad 6.6$$

If the spatial frequency, t , is introduced into the equation, 6.6 may be rearranged to give

$$\phi = 0.943(W_{11} + 2.462)t/0.15$$

$$\phi = 6.283(W_{11} + 2.462)t$$

$$\phi = 2\pi(W_{11} + 2.462)t \quad 6.7$$

This has the expected form for the phase effect of the transverse focal shift. The factor 2.462 represents the linear part of the phase shift introduced by the other terms in the polynomial, which indicates that introducing a focal shift $W_{11}=-2.462$ would reduce the average phase shift to zero. As will be seen later, the Strehl-Maréchal method of aberration balancing requires the introduction of such a focal shift.

Optimisation Criteria

As with the previous chapter, the formulae derived from the analytic criteria are given below. The balances are evaluated for a series of aberration sizes and spatial frequencies, and

these are compared with those found by the numerical techniques.

For the coma polynomial, equation 6.1, all the terms contain the factor $\rho \cos(\phi)$, and integration over the pupil gives the average wavefront aberration as zero. Thus the root-mean-square wavefront aberration is identical to the variance of the wavefront aberration, the Strehl-Maréchal criterion, and so only this latter criterion will be considered.

a. Strehl-Maréchal variance

Including the transverse focal shift term, the aberration polynomial becomes

$$W = W_{11}\rho \cos\phi + W_{31}\rho^3 \cos\phi + W_{51}\rho^5 \cos\phi \quad 6.8$$

and the mean square aberration is given by

$$W^2 = \iint (W_{11}\rho + W_{31}\rho^3 + W_{51}\rho^5)^2 \cos^2\phi \rho d\rho d\phi \quad 6.9$$

$$= \pi \int_0^A (W_{11}\rho + W_{31}\rho^3 + W_{51}\rho^5)^2 \rho d\rho \quad 6.10$$

Differentiating with respect to W_{11} and W_{31} leads to the simultaneous equations

$$\frac{W_{11}}{2} + \frac{W_{31}}{3} + \frac{W_{51}}{4} = 0$$

6.11

$$\frac{W_{11}}{3} + \frac{W_{31}}{4} + \frac{W_{51}}{5} = 0$$

These give the solutions

$$W_{11} = 0.3 W_{51}$$

$$W_{31} = -1.2 W_{51}$$

6.12

Introducing, as in the previous case, the B-coefficients we have

$$B_{15} = W_{11}/W_{51} = 0.3$$

$$B_{35} = W_{31}/W_{51} = -1.2$$

6.13

b. Hopkins approximation to the Diffraction o.t.f.

As with the axial case, we can find similar relationships between other of the criteria. The minimum r.m.s. spot size is given by equation 3.8 as

$$\begin{aligned} \langle r^2 \rangle &= 1/A \iint_A r^2 dA \\ &= 1/A \iint_A (\int \xi^2 + \int \eta^2) dA \end{aligned} \quad 6.14$$

This may be separated to give

$$\langle r^2 \rangle = \frac{1}{A} \iint_A \left(\frac{\partial W}{\partial x} \right)^2 dA + \frac{1}{A} \iint_A \left(\frac{\partial W}{\partial y} \right)^2 dA \quad 6.15$$

Considering the Hopkins approximation to the geometric o.t.f. in section 3f, we may separate the sagittal and tangential values to give

$$\begin{aligned} K_g(s) &= \iint_A \left(\frac{\partial W}{\partial x} \right)^2 dA - \left\{ \iint_A \frac{\partial W}{\partial x} dA \right\}^2 \\ K_g(t) &= \iint_A \left(\frac{\partial W}{\partial y} \right)^2 dA - \left\{ \iint_A \frac{\partial W}{\partial y} dA \right\}^2 \end{aligned} \quad 6.16$$

For the sagittal case, since there is symmetry about the y-axis, the average x-component of transverse aberration is zero, and by a suitable transverse focal shift the average y-component can be reduced to zero. As has been shown, the transverse focal shift does not affect the modulus of the o.t.f., and taking as the merit function the average o.t.f. between the two azimuths, the criterion reduces to

$$K_g = K_g(s) + K_g(t) = \iint_A \left\{ \left(\frac{\partial W}{\partial x} \right)^2 + \left(\frac{\partial W}{\partial y} \right)^2 \right\} dA \quad 6.17$$

which is identical to the r.m.s. spot size criterion in equation 6.15.

It was shown in section 3f that this approximation to the geometric o.t.f. is identical to the true geometric and diffraction o.t.f. integrals at zero frequency, and so balances given by both the above criteria will once again be just the low frequency limits given by the analytic Hopkins diffraction o.t.f. formulae which are given below.

Hopkins (1957) applied the variance treatment of section 3c to the case of coma, by expanding the variance in terms of the $H_{kl}(s)$ coefficients described previously. In order to maximise the o.t.f. modulus only the second and higher order terms need be considered, and so the balance is taken between W_{51} and W_{31} . Differentiating the variance formulae with respect to W_{31} gives, for the sagittal case, the approximate formula

$$W_{31} = -(1.49 - 0.94s + 0.41s^2) \cdot W_{51} \quad 6.18$$

and for the tangential case

$$W_{31} = -(1.49 - 1.26t + 0.85t^2) \cdot W_{51} \quad 6.19$$

These formulae have similar restrictions to the spherical aberration formulae, and once again Hopkins sets the tolerance o.t.f. at 0.8. Taking this tolerance, the maximum permissible values of W_{51} for the sagittal and tangential are given by

$$W_{51}(\text{sag}) = \pm(0.37/s + 0.83) \quad 6.20$$

$$W_{51}(\text{tan}) = \pm(0.64/t + 0.96)$$

For this variance treatment it is found that the phase shift for the coma polynomial in the tangential section is simply

given by

$$\theta(t) = 1.16 + 0.35t \quad 6.21$$

This phase shift may be corrected for any given spatial frequency by an appropriate transverse focal shift.

In order to compare the various criteria, the two formulae 6.18 and 6.19 have been combined to give aberration balances for the average sagittal and tangential o.t.f. This gives

$$W_{31} = -(1.49 - 1.1s + 0.63s^2) \cdot W_{51} \quad 6.22$$

or

$$B_{35} = -(1.49 - 1.1s + 0.63s^2) \quad 6.23$$

As with the axial case these formulae have been examined outside the strict limits set by Hopkins, since, as stated, the strict numerical accuracy is unimportant. The tolerance values for the second order coma used are given in table 6.1

W_{51}	s_0	<u>Table 6.1</u> Values of s_0 , W_{51} giving m.t.f. of 0.8
-5	0.123	
-10	0.055	
-20	0.026	
-40	0.0129	

The balance formulae 6.18 and 6.19, as with all the Hopkins formulae, are independant of aberration size, and as with the axial formulae it will be shown that these balances are poor outside the tolerance limits.

c. The diffraction and geometric o.t.f.

The balances for the diffraction and geometric o.t.f. integrals were evaluated by the numerical techniques described previously. The merit function used in both cases was taken from the basic form given in equation 4.2 with $a=b=c=1.0$, to give

$$\psi(s) = 1 - \text{OTF}_s(s) - (\text{OTF}_r^2(s) + \text{OTF}_i^2(s))^{\frac{1}{2}} \quad 6.24$$

$$\psi(s) = 1 - \text{MTF}_s(s) - \text{MTF}_t(s) \quad 6.25$$

where MTF_s and MTF_t are the moduli of the sagittal and tangential o.t.f. integrals, and both integrals are evaluated for the same spatial frequency s .

The results for this section are presented in a similar form to those for the axial case. The high order aberration, W_{51} , takes the values $-5, -10, -20$, and -40 wavelengths, for spatial frequency s of $0.01, 0.05$, and 0.10 . Since the merit functions have been chosen to maximise the average m.t.f. between the two azimuths only the average values have been plotted. It was noted, however, that for each balance the sagittal m.t.f. values were consistently higher than the tangential values, as expected.

All the terms in the coma aberration polynomial 6.1 are of the form $\rho^n y$, and so the aberration in the x-direction is always zero. Only the wavefronts in the y-direction, therefore, need to be plotted.

The balances computed for the four independant criteria, the Strehl-Maréchal variance, the Hopkins variance and the two numerical o.t.f. integrals, have been tabulated as absolute values in table 6.2 and as B-coefficients in table 6.3. Since the Strehl

Low Frequency Coma Balances

Table 6.2

S		Strehl		Hopkins		Diffraction		Geometric
	W_{51}	W_{11}	W_{31}	W_{31}		W_{31}		W_{31}
.01		-1.5	6.0	7.39		7.435		7.499
.05	-5.	-1.5	6.0	7.19		7.155		7.422
.10		-1.5	6.0	6.95		6.788		7.167
.01		-3.0	12.0	14.8		14.863		15.006
.05	-10.	-3.0	12.0	14.35		13.984		14.352
.10		-3.0	12.0	13.85		11.285		10.323
.01		-6.0	24.0	29.58		29.613		29.854
.05	-20.	-6.0	24.0	28.75		22.786		20.654
.10		-6.0	24.0	28.22		14.749		15.505
.01		-12.0	48.0	59.1		58.224		58.632
.05	-40.	-12.0	48.0	57.4		29.447		31.010
.10		-12.0	48.0	55.4		22.490		26.590

TABLE 6.2

Low Frequency Cona Balances

Table 6.3

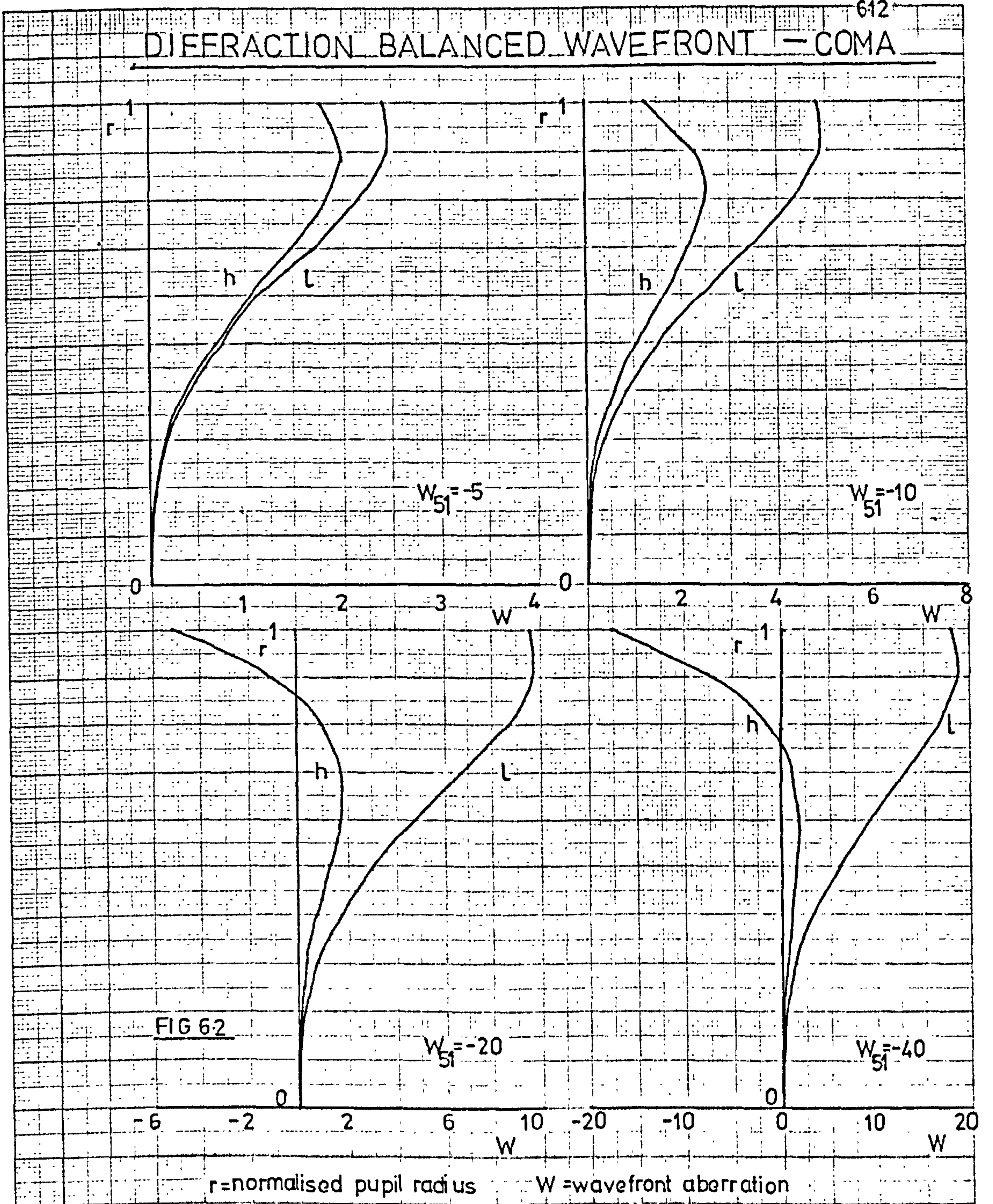
S		Strehl		Hopkins		Diffraction		Geometric
	W_{51}	B_{15}	B_{35}	B_{35}		B_{35}		B_{35}
.01		0.3	-1.2	-1.48		-1.487		-1.490
.05	-5.	0.3	-1.2	-1.435		-1.431		-1.484
.10		0.3	-1.2	-1.385		-1.357		-1.433
.01		0.3	-1.2	-1.48		-1.486		-1.500
.05	-10.	0.3	-1.2	-1.435		-1.398		-1.435
.10		0.3	-1.2	-1.385		-1.128		-1.032
.01		0.3	-1.2	-1.48		-1.485		-1.492
.05	-20.	0.3	-1.2	-1.435		-1.139		-1.032
.10		0.3	-1.2	-1.385		-0.737		-0.775
.01		0.3	-1.2	-1.48		-1.455		-1.463
.05	-40.	0.3	-1.2	-1.435		-0.736		-0.775
.10		0.3	-1.2	-1.385		-0.562		-0.665

TABLE 63

criterion is the only one to include the transverse focal shift term, the graphical representation of the wavefronts becomes more complex than for the axial case. To provide the simplest comparison between the various wavefront shapes, the Strehl wavefronts have been plotted without the focal shift. As before, the diffraction-balanced wavefronts have been plotted at the high and low frequency limits ($s=0.10$, $s=0$) in figure 6.2. As with the axial case, the shape of the low-frequency curves are the same in all four cases, but the high-frequency curves once again show the expected variation with aberration size. The absence of the lowest-order term in the wavefront polynomial means that the curves are more difficult to interpret, since the axis no longer provides an easy reference for the aberrations. However, since the focal shift term may be considered as a simple rotation of the axis, an appropriate shift may easily be added to each curve. If this process is applied to the large aberration cases, it can be seen that the curvature of the high frequency curves is reduced in the region close to the origin, and that these curves lie close to the rotated axis as far as 0.6 of the pupil radius. This behaviour is identical to that of the axial wavefronts, where as the high order aberration and frequency are increased, the optimum wavefront becomes increasingly under-corrected.

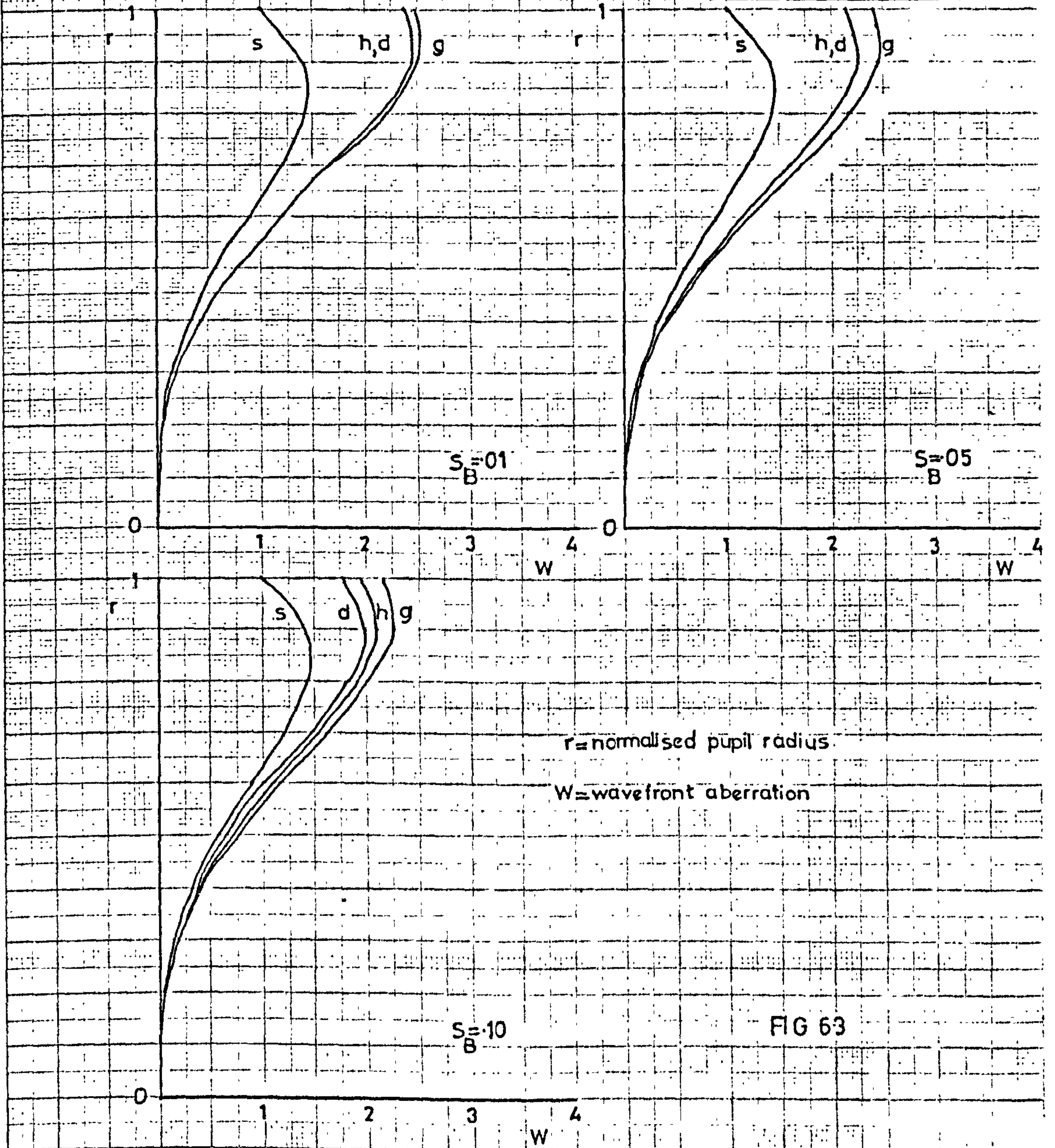
The wavefront aberration curves and diffraction o.t.f. curves for all four criteria have been plotted in figures 6.3 to 6.10. Allowing for the focal shift term, these curves have many features in common with the axial curves. Taking the shape of the Strehl-Maréchal curve ($B_{35} = -1.2$) as the corrected wavefront, it can

DIFFRACTION BALANCED WAVEFRONT — COMA



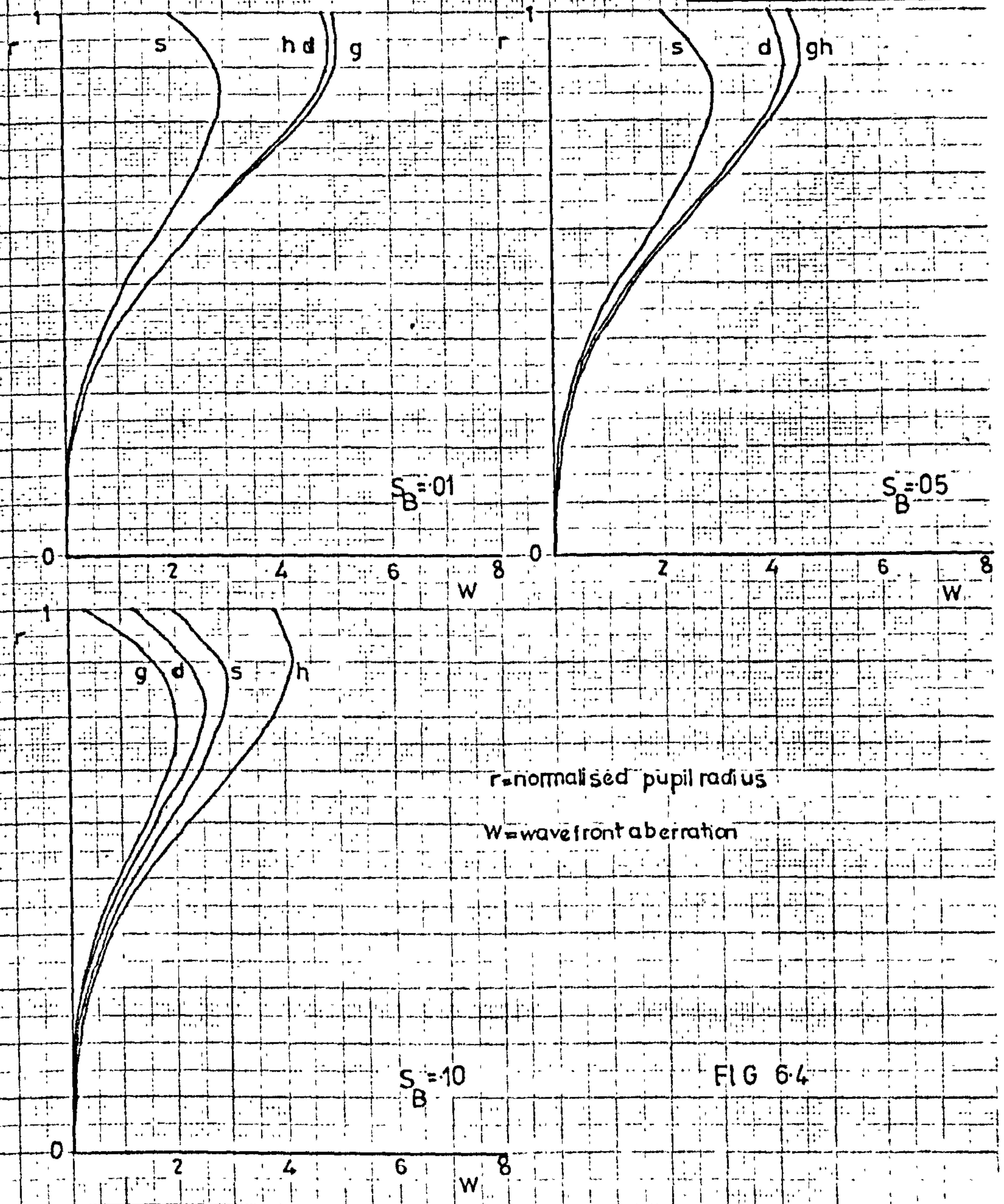
WAVEFRONT ABERRATION BALANCES

AT $W_{s1} = -5\lambda$

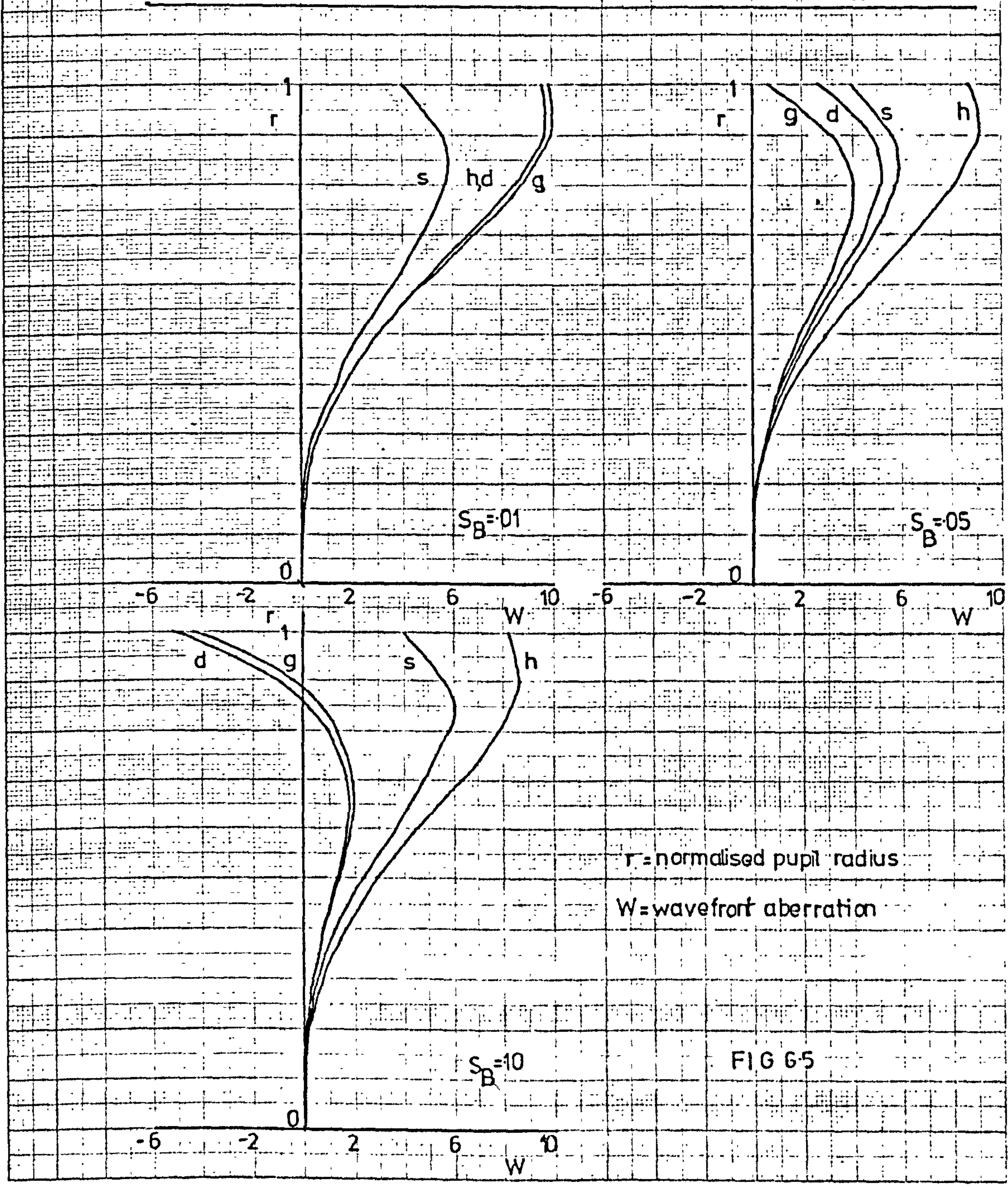


WAVEFRONT-ABERRATION BALANCES

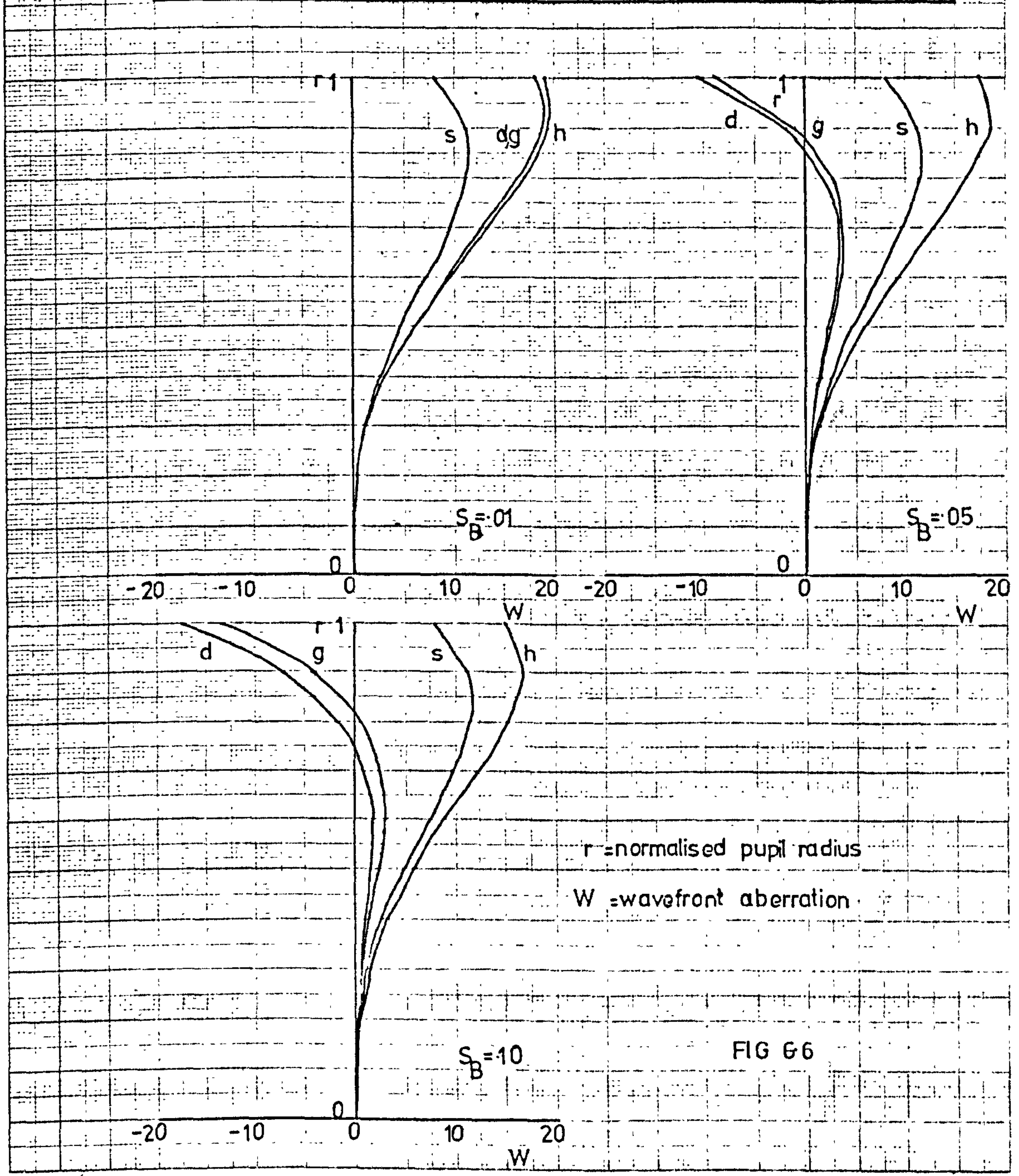
AT $W_{51} = -10\lambda$



WAVEFRONT ABERRATION BALANCES AT $W_{S1} = -20\lambda$



WAVEFRONT ABERRATION. BALANCES AT $W_{51} = -40\lambda$



DIFFRACTION 0.T.F. (COMA) H1 W51=-5

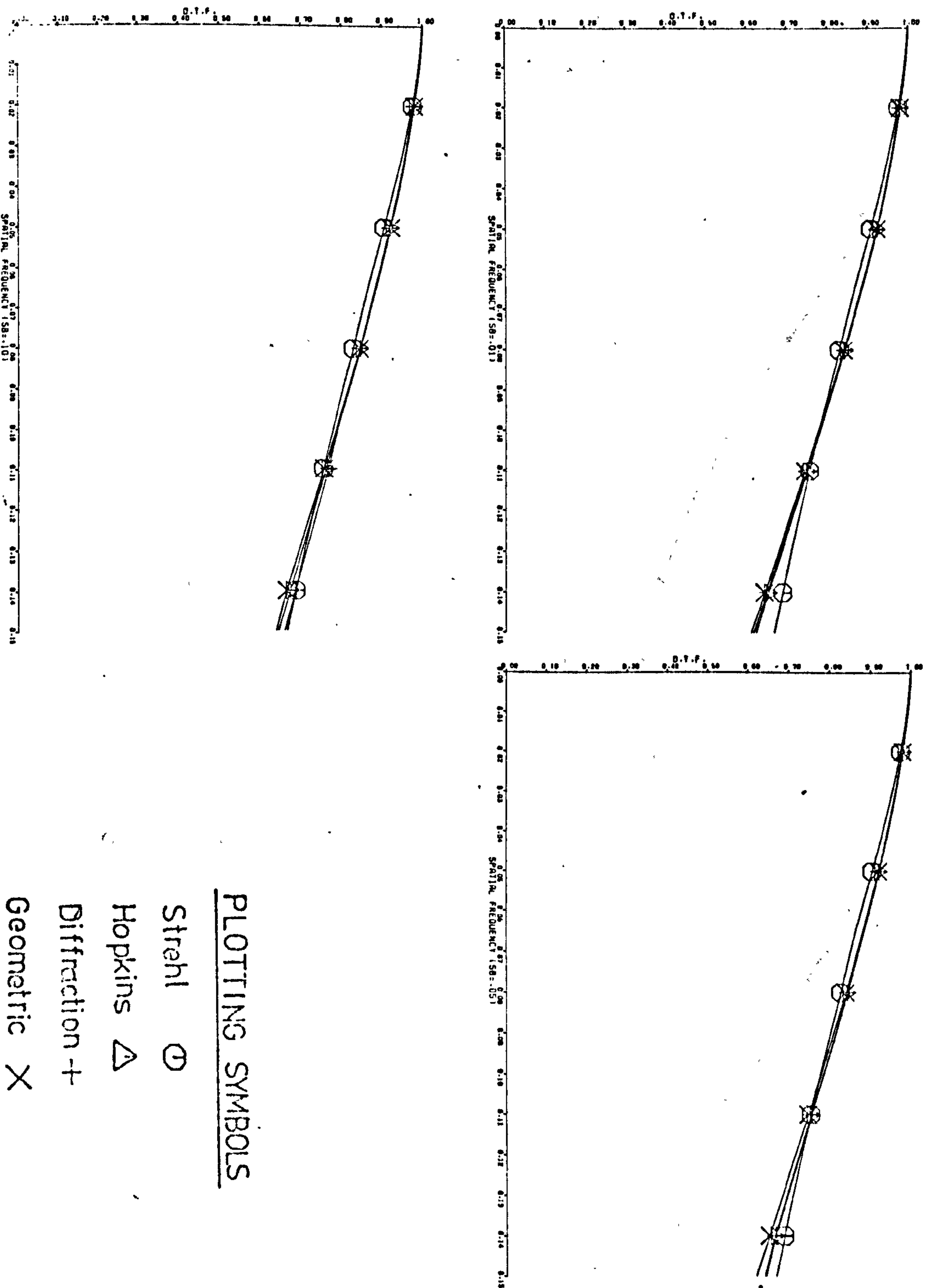


FIG 6.7

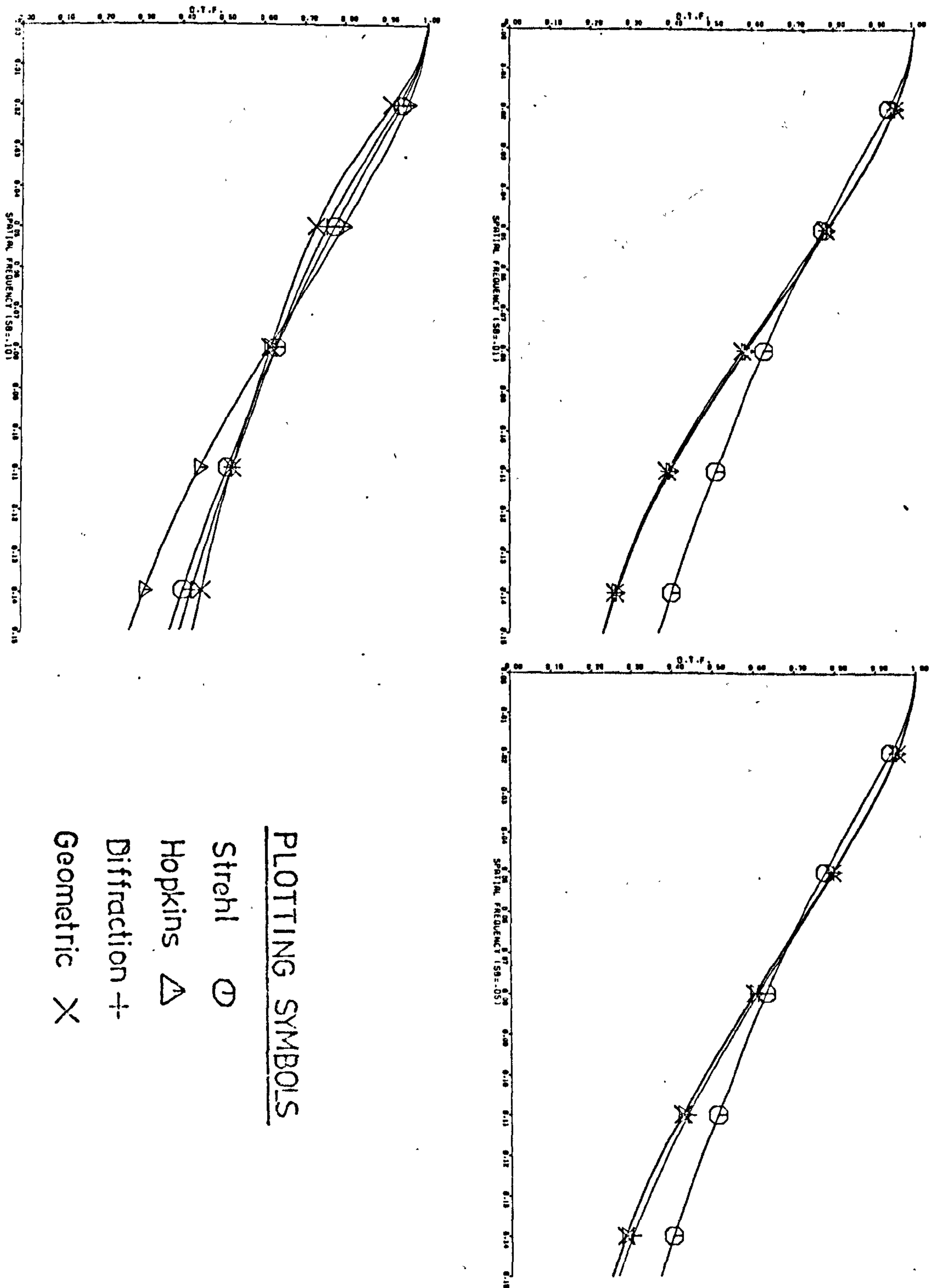


FIG 6-8

DIFFRACTION 0.T.F. (COMA) AT W51=-20

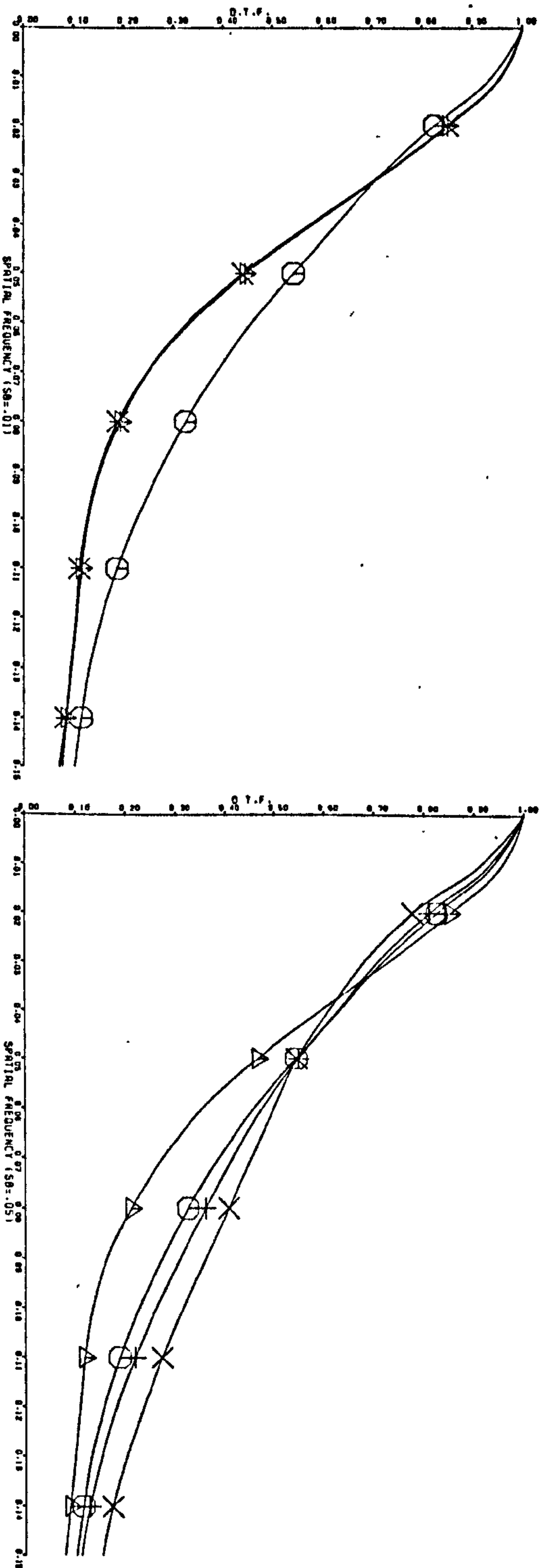
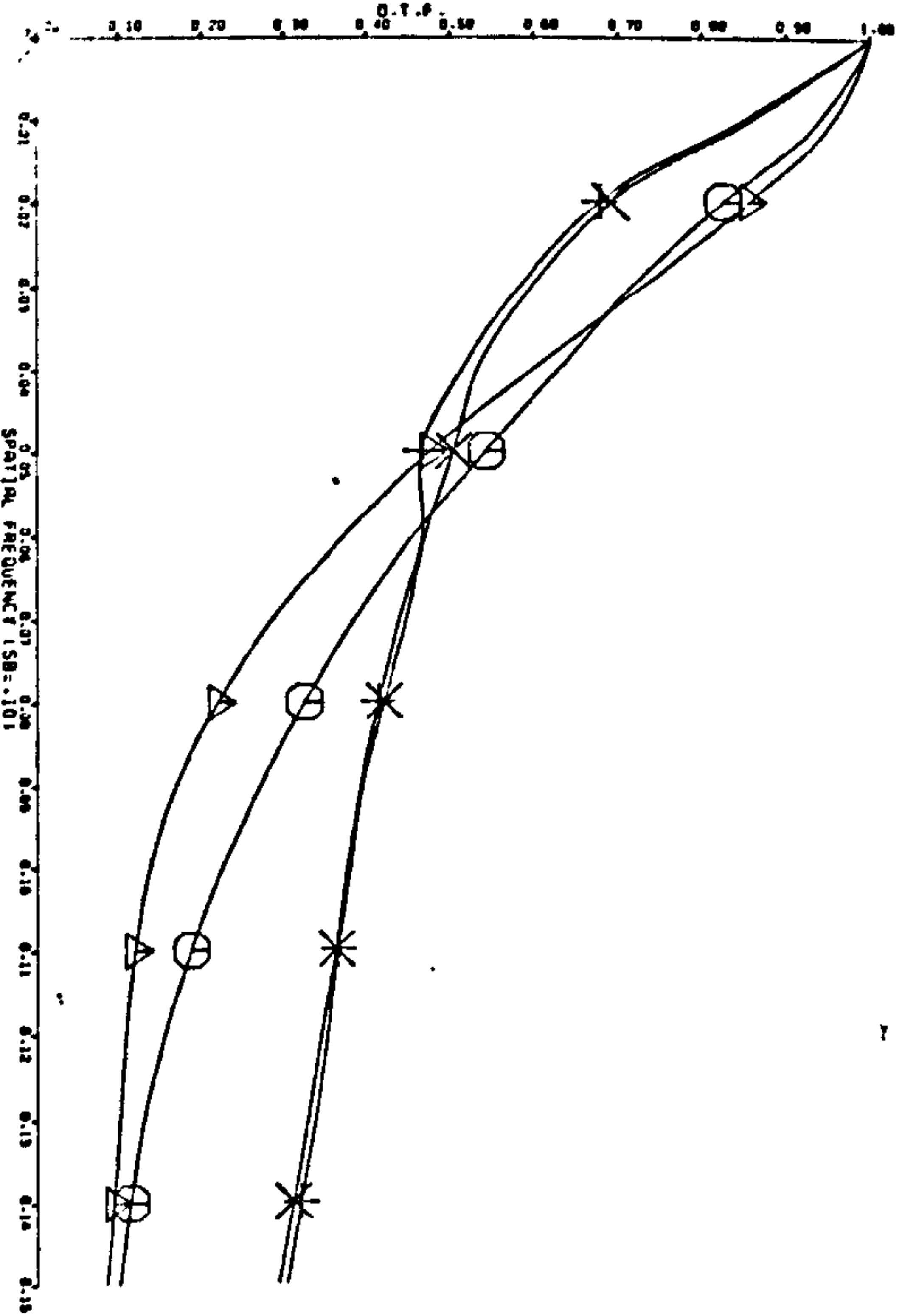


FIG69



PLOTTING SYMBOLS

- Strehl ○
- Hopkins △
- Diffraction +
- Geometric X

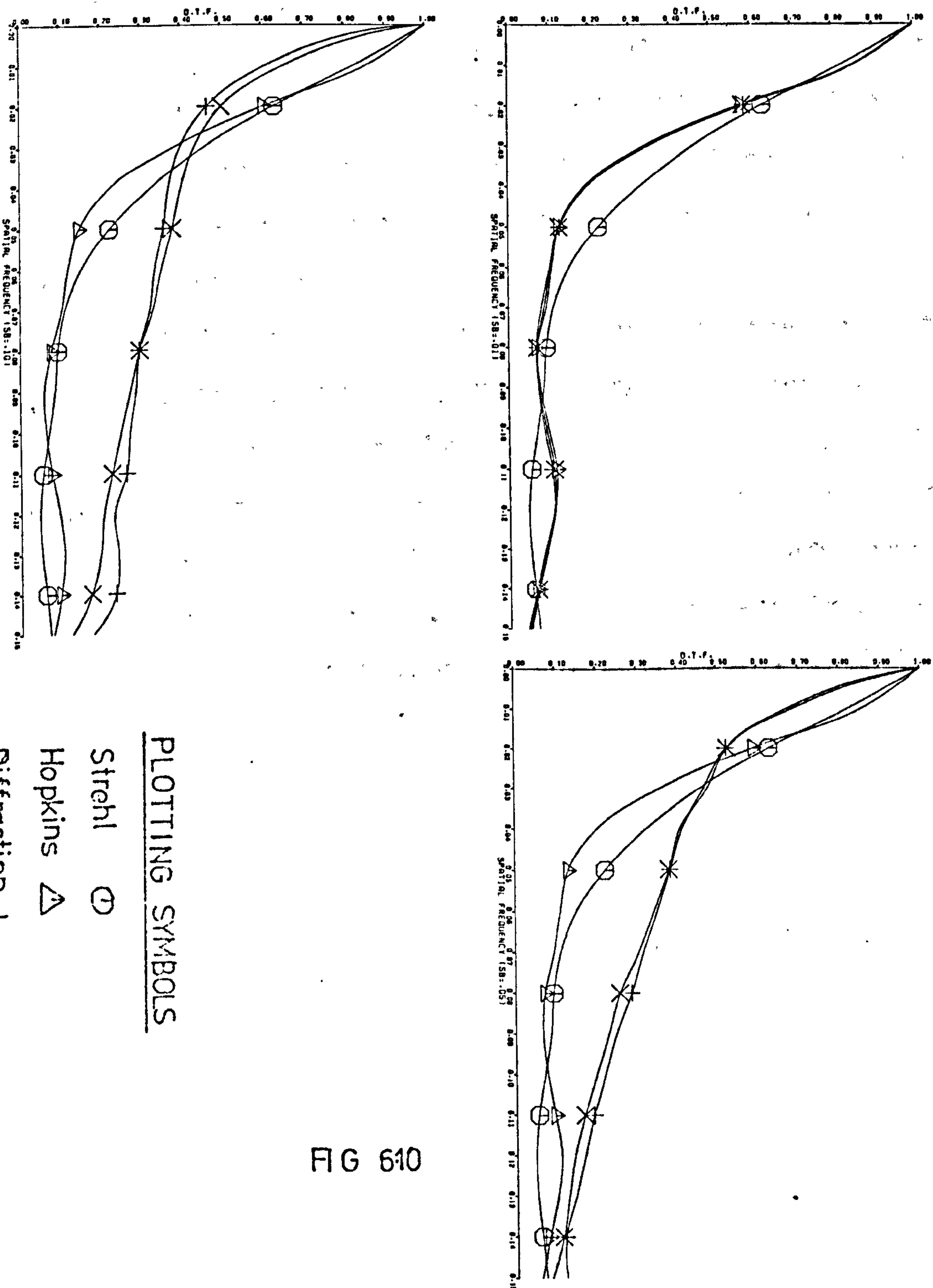
DIFFRACTION O.T.F. (COM) AT $M51 = -40$ 

FIG 610

PLOTTING SYMBOLSStrehl $\bigcirc$ Hopkins $\triangle$ Diffraction $+$ Geometric $\times$

be seen that as aberration size and spatial frequency increase, both diffraction and geometric wavefronts move from being over-corrected, $B_{35} < -1.2$ to being under-corrected, $B_{35} > -1.2$. The Hopkins wavefront is, as always, limited by being independent of aberration size, and remains highly over-corrected, although within the tolerance limits these balances are almost identical to the diffraction balances. The Strehl-Maréchal balances, as before, attempt to provide an optimum balance over all frequencies and are mid-way between the diffraction high and low frequency balances, as expected.

The o.t.f. curves produced by these balances show that, as with the axial case, the geometric balance produces o.t.f. values which are only marginally lower than the maximum diffraction values over the whole region of aberration size and frequency. The Hopkins criterion is, as before, only useful within its limited tolerance range, while the Strehl-Maréchal criterion remains a valid criterion for overall o.t.f. optimisation.

Chapter 7

High Frequency Aberration Balancing

The previous chapters amply justify the use of the geometrical o.t.f. method for use where an o.t.f. criterion of optimisation is required. Although the regions examined are wider than used in previous studies, a complete survey over all spatial frequencies, aberration sizes and combinations of aberrations would obviously be impractical. To extend the survey, many more o.t.f. values would need to be evaluated at high spatial frequencies and aberrations, and as stated previously the amount of computation required increases rapidly as the o.t.f. decreases. To present the results of such a survey in the amount of detail of the last two chapters would give a set of results both unmanageable to handle and difficult to evaluate.

It was decided, however, that for the sake of completeness, some investigation should be carried out at higher spatial frequencies. Accordingly the axial survey of chapter 5 was continued at frequencies above 0.10. This part of the study is intended to demonstrate more fully the regions of validity of the various criteria, and so a more compact form has been used to present the results. The set of graphs, figures 7.1 and 7.2, show the maximum diffraction o.t.f. values which can be obtained by optimising the four criteria under investigation. These show the results of balancing over the same aberration range as before but at spatial frequencies up to 0.3.

Figure 7.1 shows the aberration size plotted as abscissa, and the maximum o.t.f. has been plotted on lines of constant spatial frequency. The curves of figure 7.1c give the maximum o.t.f.

which can be obtained for any given aberration size.

Figure 7.2 reverses this situation, since the spatial frequency has been plotted as the abscissa, and lines of constant aberration size have been plotted. It must be remembered that each point on these graphs has been found by balancing at the particular spatial frequency, and so the curves of figure 7.2 must not be taken as true o.t.f. curves, except for the zero-aberration curve which has been plotted on each graph.

To interpret these graphs and to provide meaningful regions of validity for the criteria, a tolerance value of 5% has been set on each criterion. Thus, in this context, for a criterion to be valid the o.t.f. value produced by the aberration balance must not be more than 5% lower than the maximum diffraction o.t.f. at that aberration size and spatial frequency. The limit of 5% is arbitrary, and is only intended to serve as a reference for performance, although its significance will be shown later in this chapter.

In figures 7.1 and 7.2 a boundary line has been drawn on each set of curves to divide the region where the criterion falls outside the tolerance value. The shaded region shows the area where the criterion is invalid.

It can be seen immediately that for finite aberrations both the Strehl and Hopkins balances are generally valid only when the o.t.f. is above 60%. This agrees with the results found by King for the low-aberration case. The geometrical criterion remains valid over almost the whole range of aberrations and frequencies. The only region where the geometric criterion falls below the

CONSTANT FREQUENCY LINES

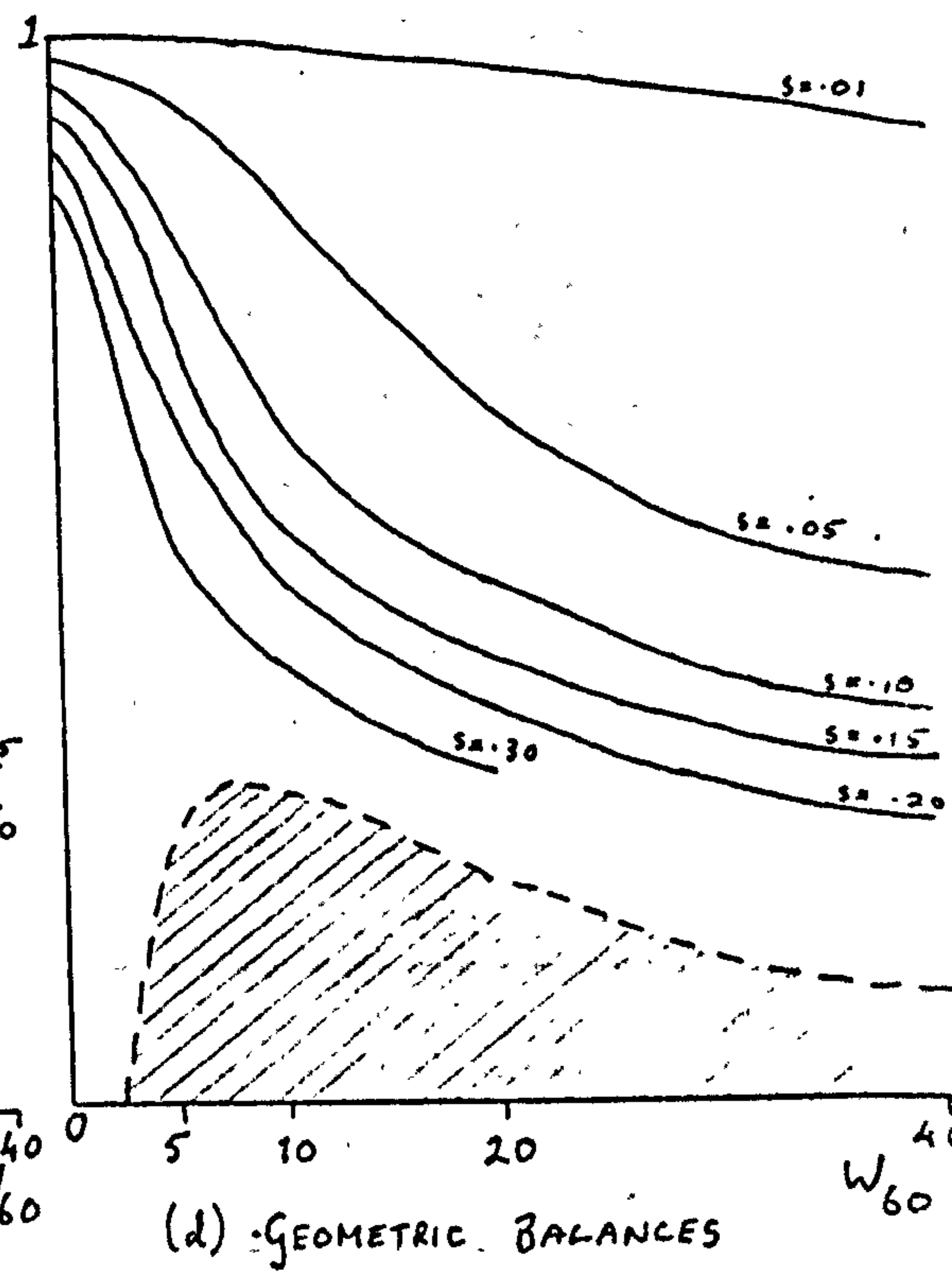
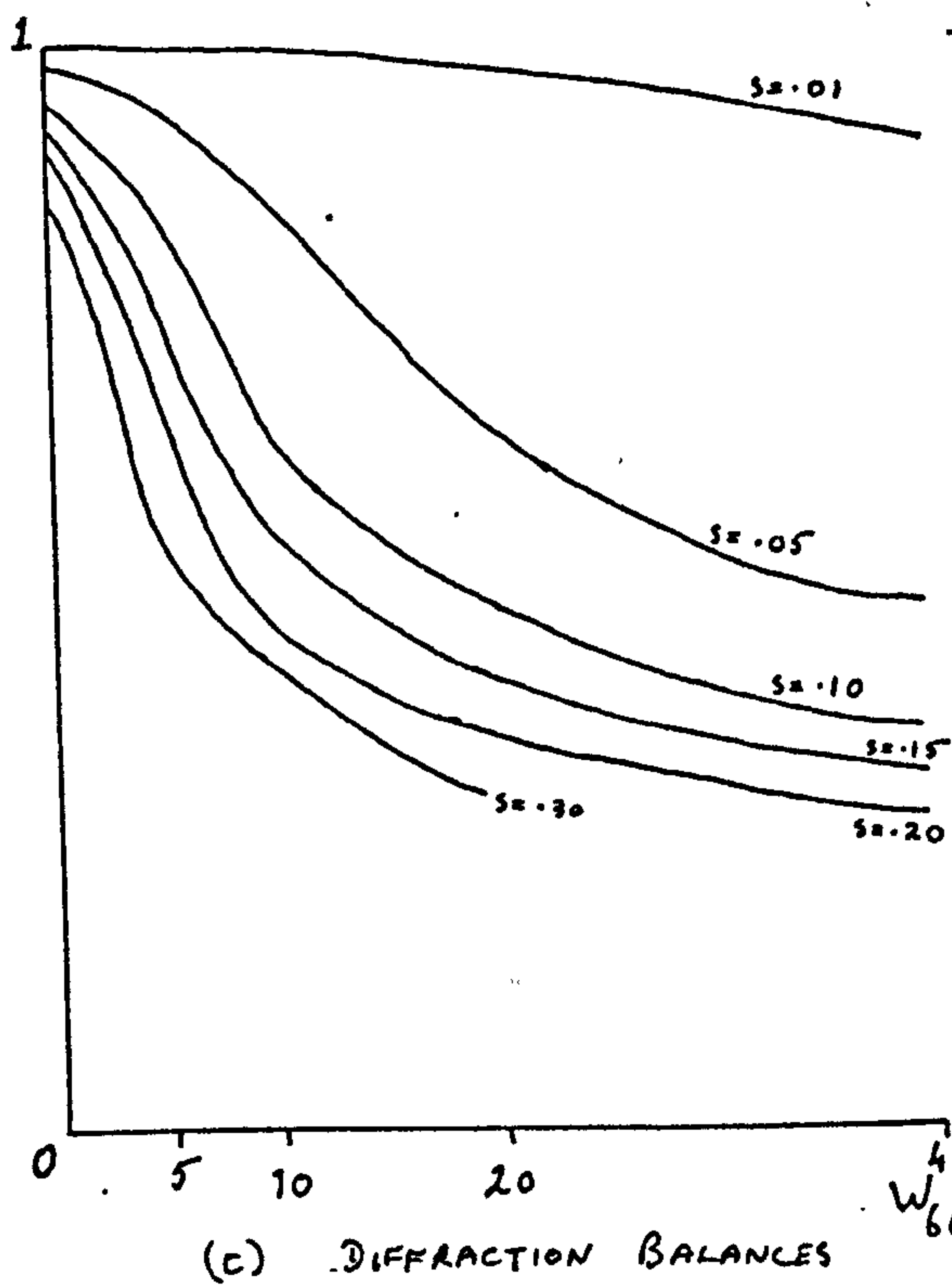
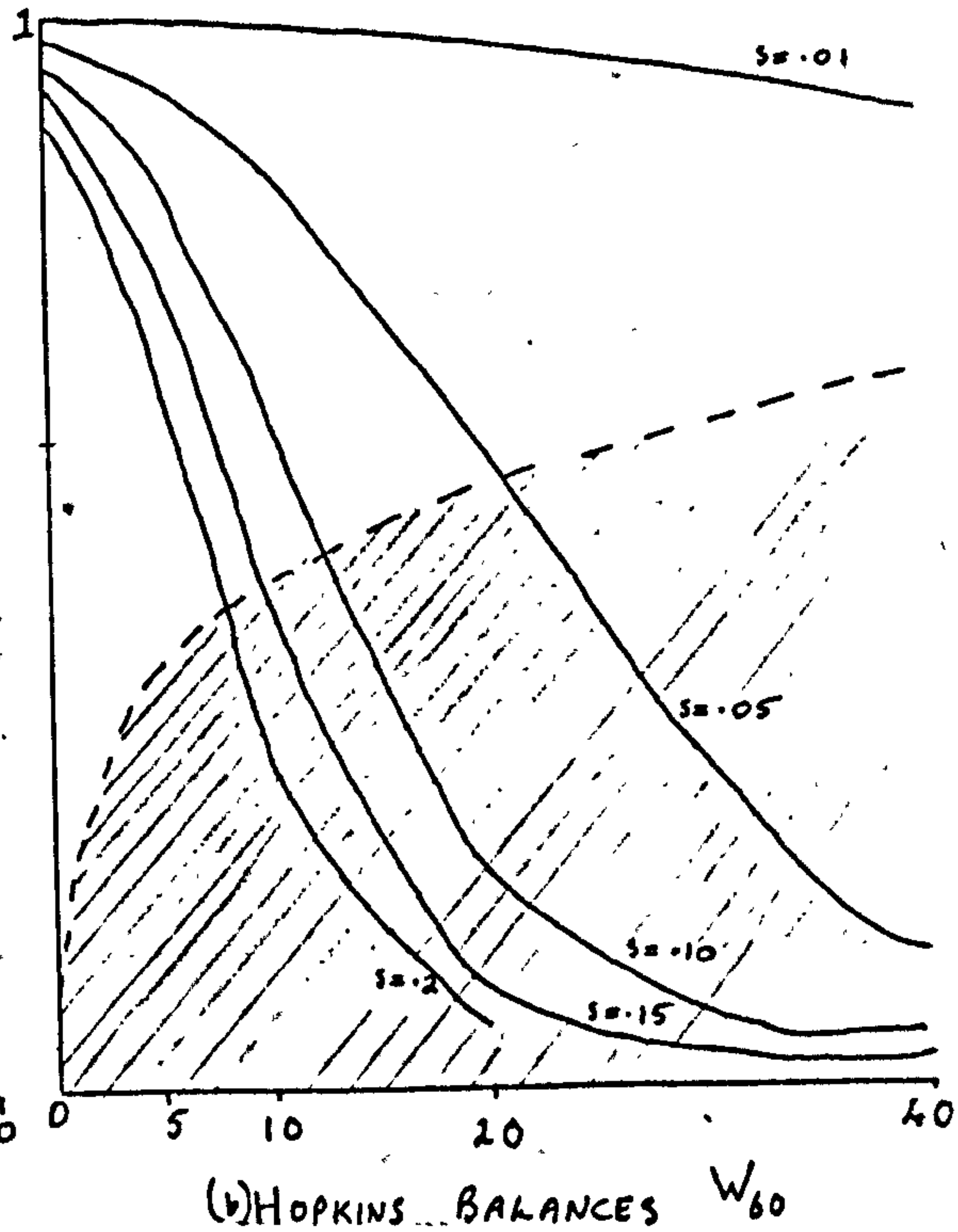
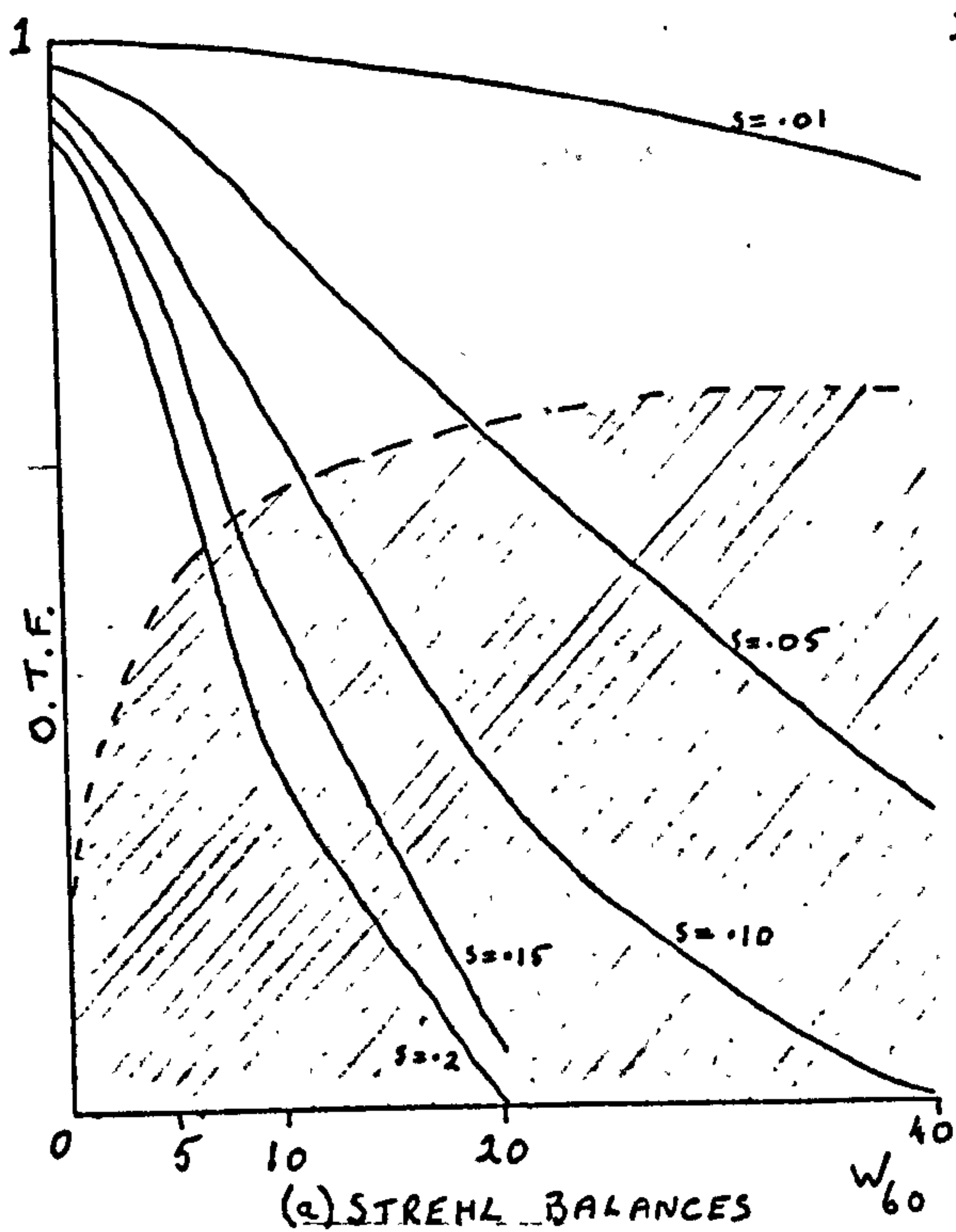


FIG 71

CONSTANT ABERRATION SIZE LINES

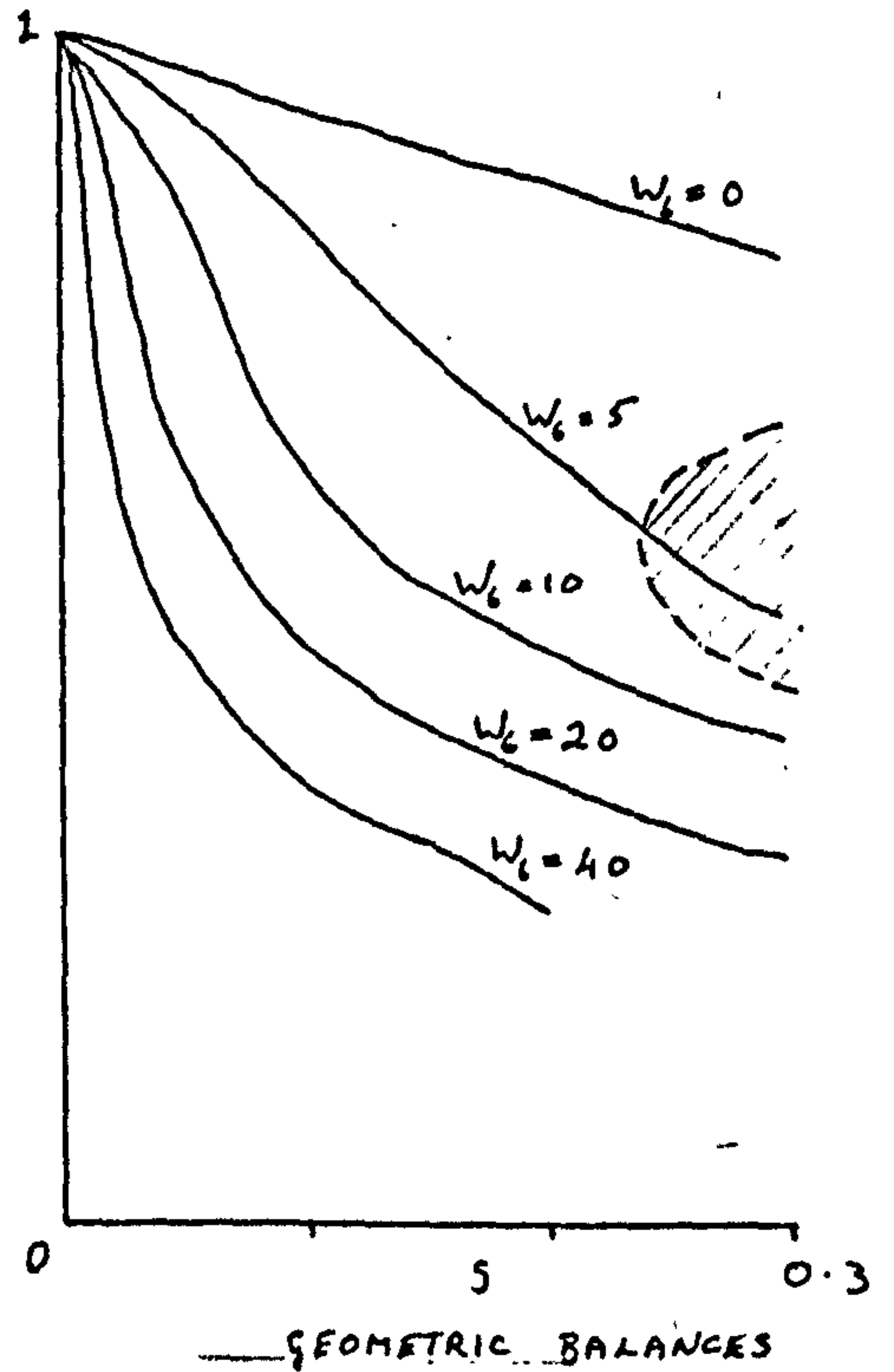
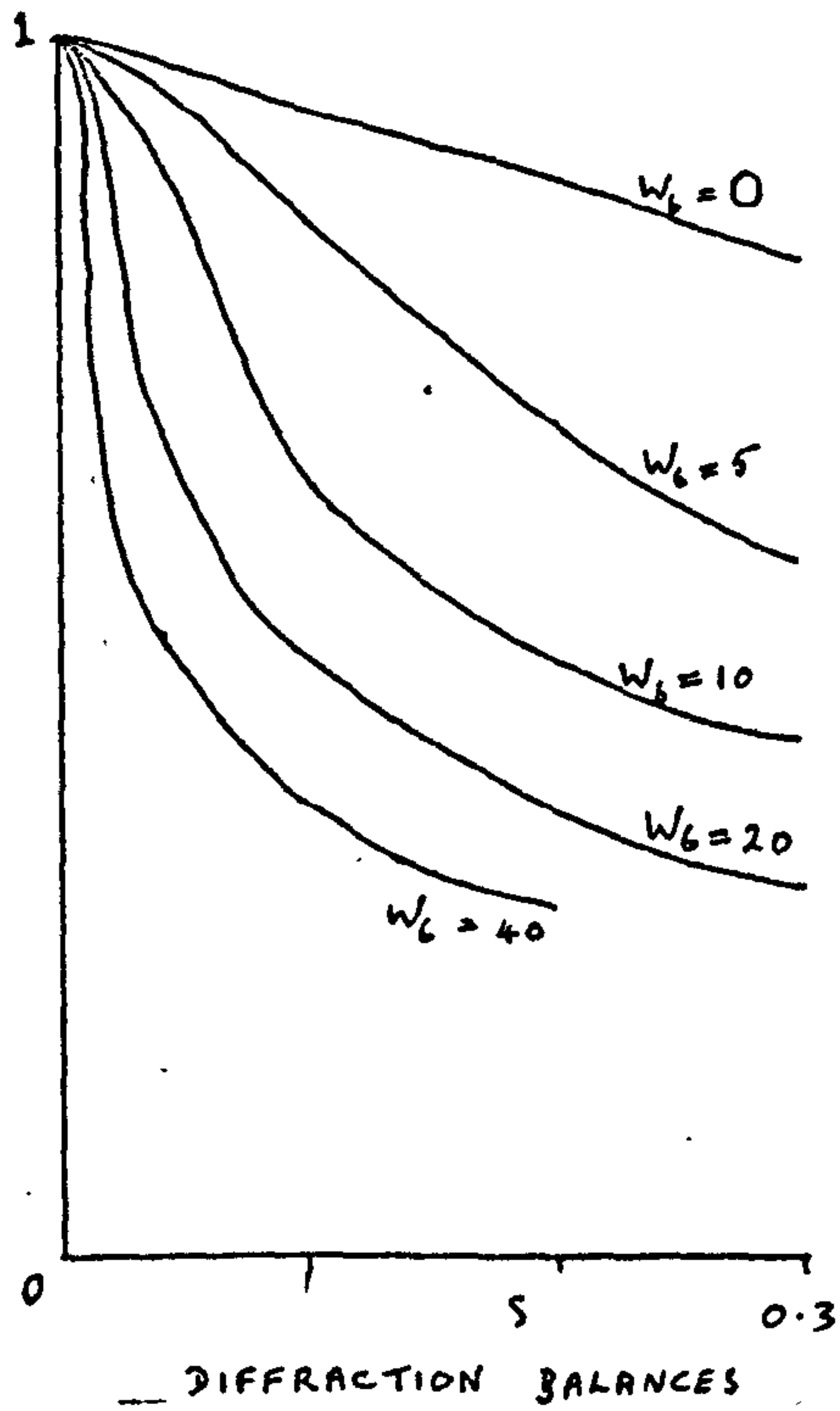
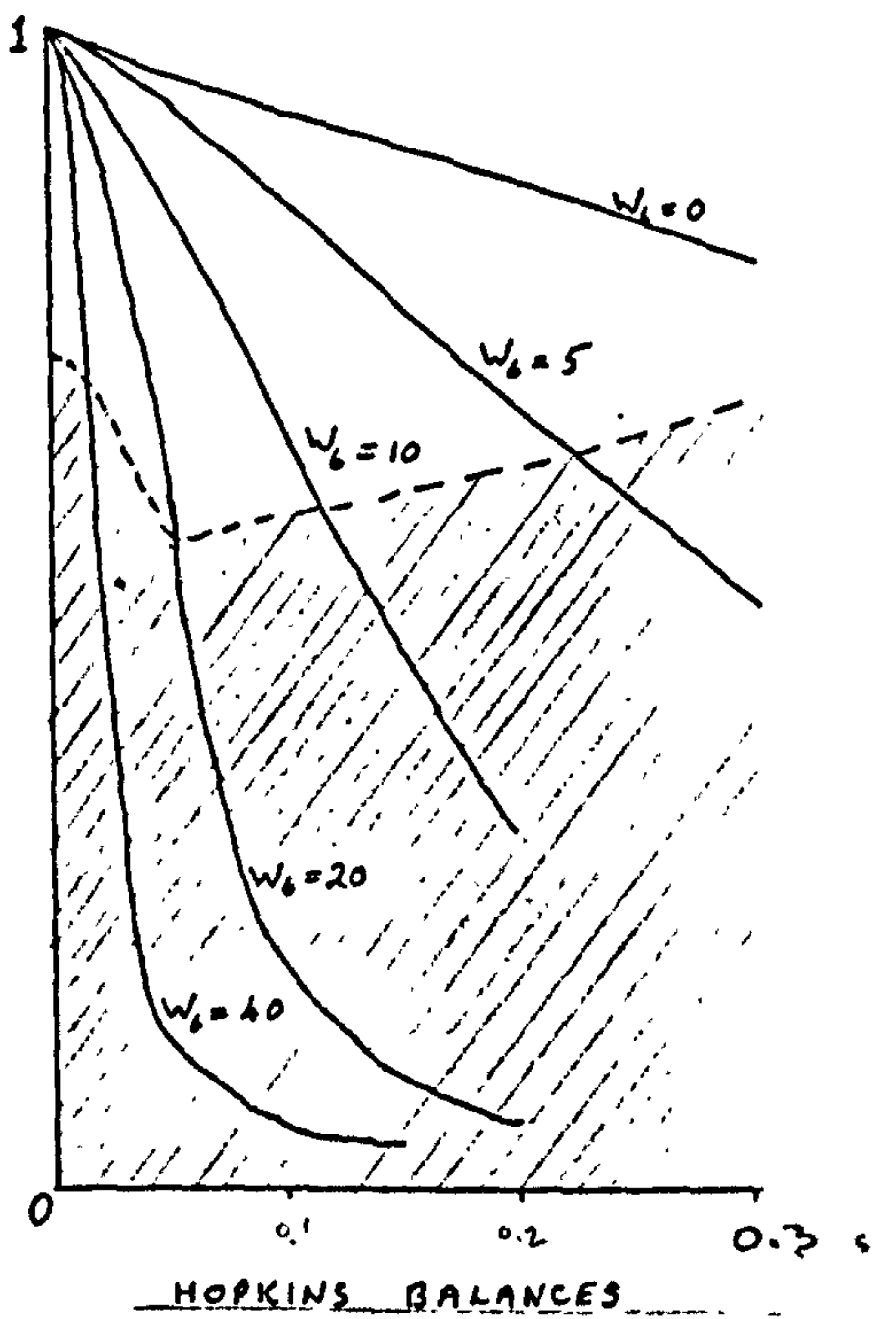
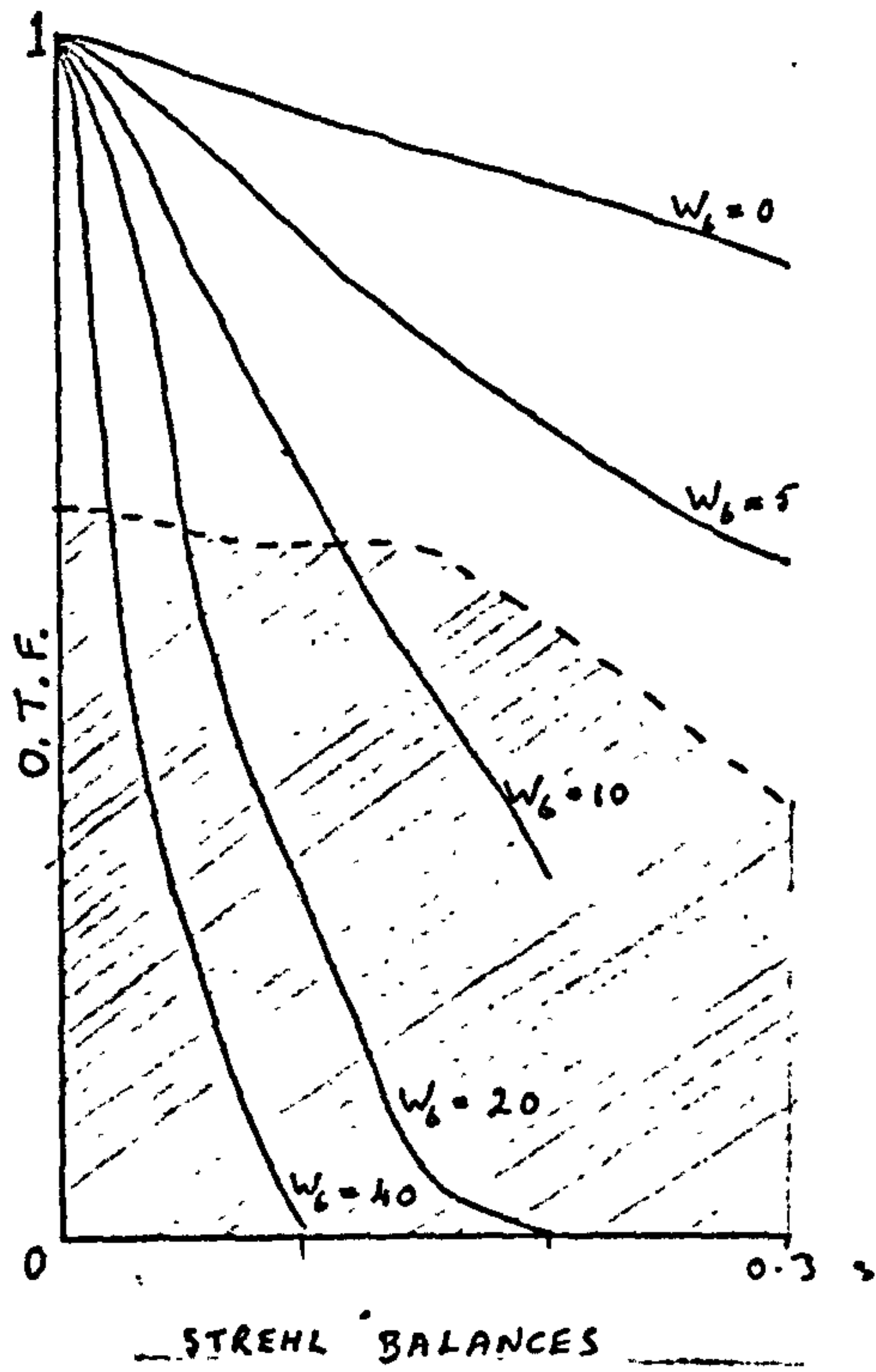


FIG 7.2

limit is at high spatial frequencies and low aberrations. This is expected on theoretical considerations, since as shown in chapter 3 the largest contribution to the error in the geometric integral at high frequencies comes from the difference in the areas of integration, since the area for the diffraction integral depends on the spatial frequency. This effect is only noticeable at low aberrations, since at high aberrations the o.t.f. is low and only low frequencies are of practical interest.

Although 5 wavelengths of aberration is not normally considered to be low with respect to the diffraction limit, it was shown in chapter 5 that with the appropriate balance this amount of second order aberration represents only about 0.7 wavelengths maximum aberration over the pupil. This lies within the limits 0.2λ to 2.0λ given by Miyamoto as the region of least accuracy for the geometric o.t.f.

The curves of figures 7.1 and 7.2 have only taken spatial frequencies up to 0.3, since above this value the maximum diffraction o.t.f. rapidly drops below a useful level for all but the lowest aberrations. It was not considered worthwhile to evaluate the balances in these low o.t.f. regions. To consider the effects of balancing at high frequencies and low aberrations the balances for 5 wavelengths of secondary spherical aberrations were calculated for spatial frequencies up to 1.0. These balances are plotted in terms of W_{20} and W_{40} in figure 7.3. The results show that, as predicted by the Hopkins formulae, the optimum diffraction balances follow an approximate parabolic curve, although the curves depart significantly from the Hopkins curve outside the Hopkins tolerance limits.

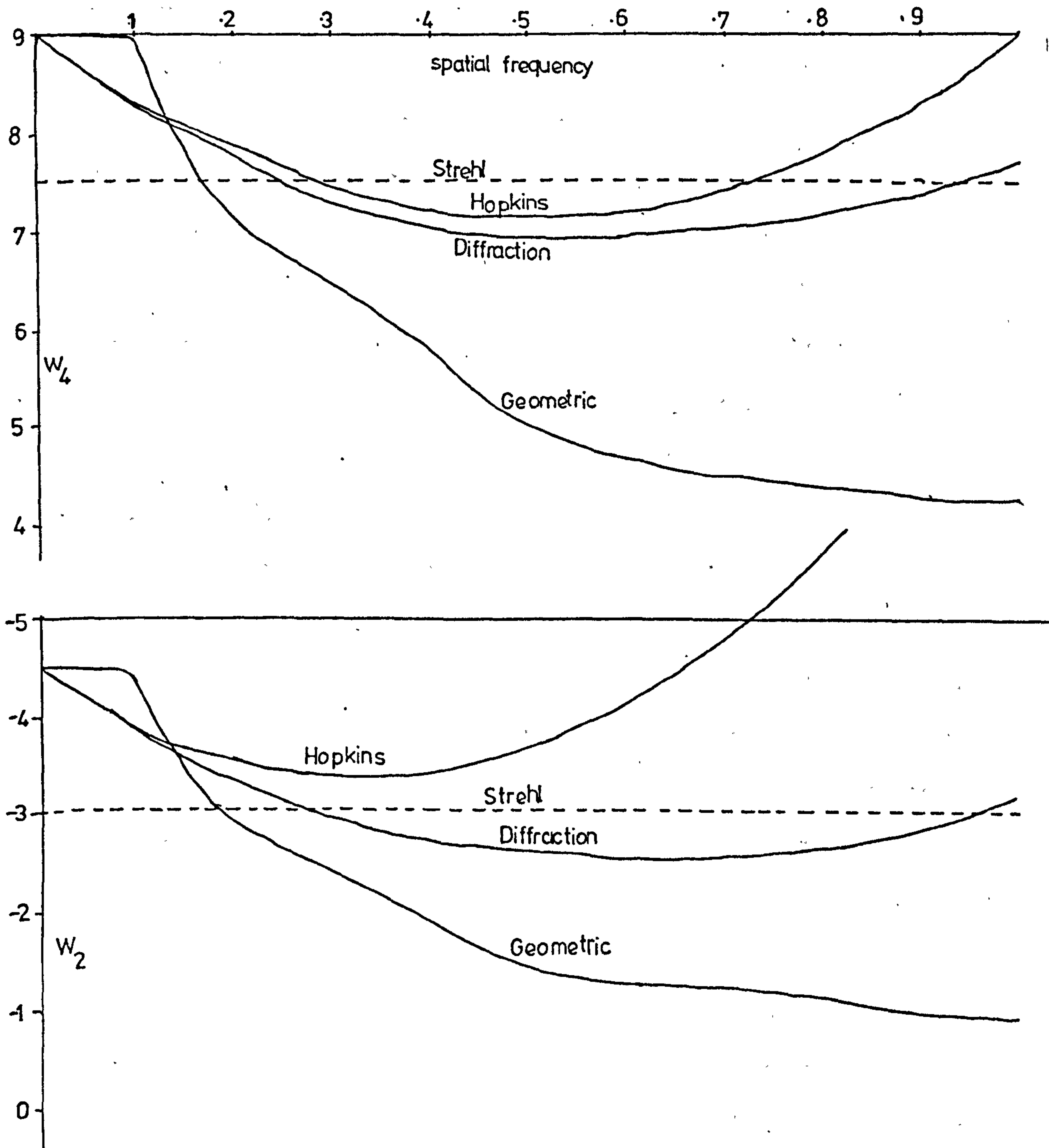


FIG 7.3

Outside the low frequency region both the geometric and Hopkins formulae depart from the optimum values, and it can be seen that the constant Strehl-Maréchal balance remains close to the optimum throughout the complete range. This is in agreement with the o.t.f. values of figure 7.2, where the Strehl balance is the only one to remain within the tolerance limit throughout the whole low aberration curve.

It should be noted, however, that due to the parabolic form of the curves it would be possible to produce acceptable high-frequency balances by using the geometric or Hopkins criteria at low frequencies in this low-aberration region.

Outside the low-aberration, high-frequency region the curves of figures 7.1 and 7.2 agree with the results of chapters 5 and 6 that the geometric criterion remains greatly superior to the other criteria investigated over the complete range of practical aberrations and spatial frequencies. There is one further result from the values produced in the previous chapters which concerns the Strehl intensity. It was mentioned in chapter 3 that the Maréchal variance is only an approximation to the true Strehl ratio, and thus it is expected that it would have only a limited useful range. The high aberration curves of chapters 5 and 6 show that for all but the lowest frequency balance the diffraction and geometric balances give o.t.f. curves which contain a much greater area, and hence have a higher Strehl ratio, than do those produced by the Strehl-Maréchal criterion itself. The geometric criterion was proposed to optimise systems at a specific frequency but these results show that this technique also produces better overall frequency balances than criteria based on an overall o.t.f. criterion. This result is found to hold extremely well in

the practical results which will be given in the second part of this thesis. It may be possible for a criterion to be derived which uses the exact Strehl intensity as the merit function, and if this were investigated the overall frequency response would probably be above that produced by the geometric criterion. It is not obvious, however, whether the shape of the o.t.f. curves produced by such a method would give such a good high-frequency response at high aberrations as the geometrical results presented here.

Phase and Optimisation Criteria

The previous sections have shown that, except at the low-aberration, high frequency region, the geometric approximation to the o.t.f. provides aberration balances which are only marginally inferior to the theoretical maximum for any given aberration and frequency. In deriving the merit function based on the geometric o.t.f., however, the imaginary term of the o.t.f. was disregarded. It can be simply shown that this does not reduce the value of the merit function.

The o.t.f. may be written either in terms of modulus and phase or in terms of real and imaginary parts as;

$$M(s)\exp(i\theta(s)) = R(s) + iI(s) \quad 7.1$$

where

$$M(s) = (R(s)^2 + I(s)^2)^{\frac{1}{2}} \quad 7.2$$

$$\theta(s) = \tan^{-1}(I(s)/R(s))$$

If the optimisation programme uses a technique which optimises the modulus of the o.t.f., $M(s)$, then the phase term $\theta(s)$ must

be controlled separately. If only the real part of the o.t.f. is optimised equation 7.2 shows that this will automatically increase the modulus and reduce the imaginary term $I(s)$. This will then reduce the phase term $\theta(s)$ by reducing the ratio I/R . Thus instead of being a drawback to the usefulness of the criterion the omission of the imaginary term is a benefit for the control of the phase of the o.t.f.

It is appropriate in this section to consider further the effects of phase in optical imagery.

A computer programme was constructed to investigate the effects of phase by simulating the effect on given object distributions of passage through optical systems with given phase terms in the o.t.f. The method, which is illustrated in figure 7.4, consists of taking the frequency spectrum of the object intensity, multiplying it by the o.t.f. and then applying the inverse Fourier transform to produce the image intensity. This image is then directly compared with the object to judge the effects of the o.t.f.

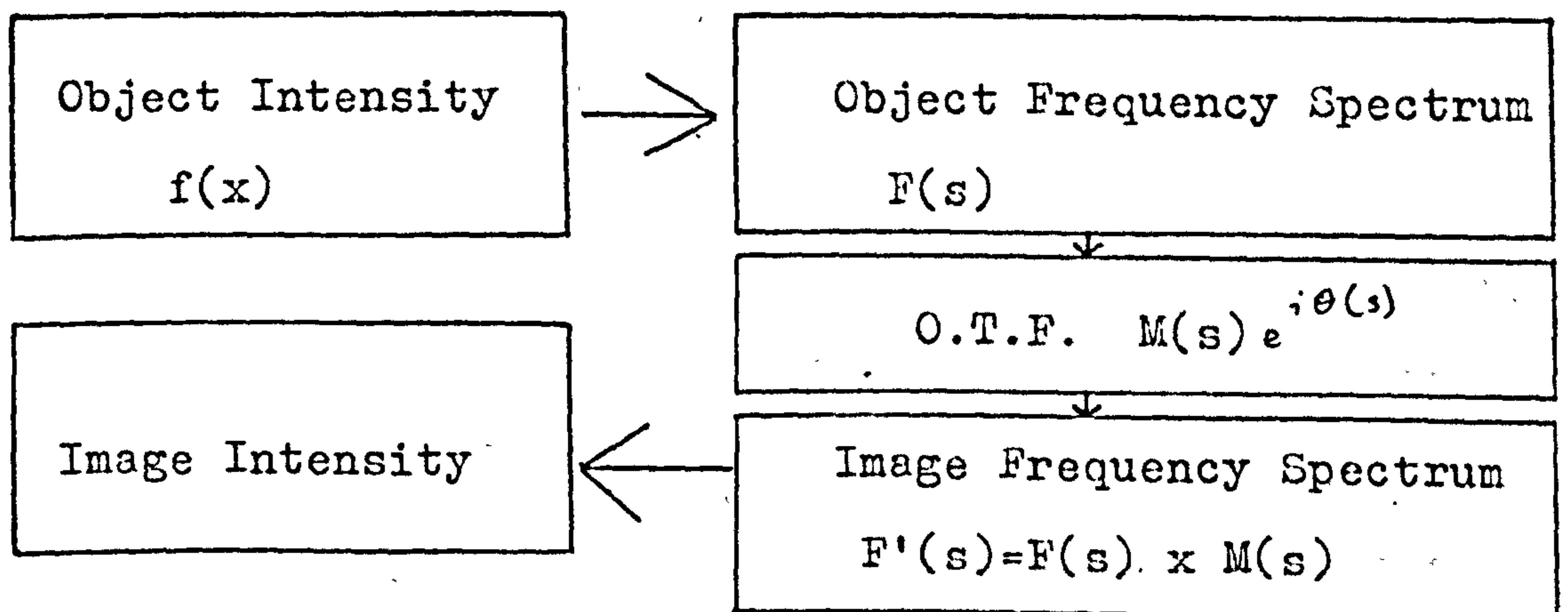


Figure 7.4 Numerical Simulation of Optical Imagery

The object distribution was taken as one of a series of simple distributions such as square waves, rectangular functions and Gaussian functions, all with varying characteristic parameters. These functions have the added advantage that their Fourier transforms reduce to simple analytic formulae. For most of the investigations the o.t.f. was taken to be the perfect diffraction o.t.f. with the addition of a phase term which had the form of a polynomial in the spatial frequency, s .

The results from this survey are too extensive to be included in this section of the thesis, and so one example has been chosen to exhibit most of the important features. The example chosen is a rectangular distribution of unit width and height, $\text{rect}(x)$, centred on $x = 0$. Using normalised units, this gives rise to an object frequency spectrum $\text{sinc}(s)$. Figure 7.5 shows the o.t.f. of the system with a phase term $0.2\pi s^2$. The s^2 term is the simplest non-trivial type of phase term, since a phase which is simply proportional to s merely gives a transverse focal shift. Figure 7.5b shows the object distribution together with the image distribution with the phase term in the o.t.f. and the image distribution produced by a perfect o.t.f., i.e. without the phase term.

It is easily seen that the effect of the phase term is small compared with the effect of the reduction in amplitude of the increasing frequency components. This result was found to hold for nearly all the cases in the survey. For objects having a continuous frequency spectrum, such as the Gaussian and rectangular functions the effects of phase are generally small. For periodic objects, having discrete spectra, the effects of phase become more marked, although the degree of distortion caused by the phase

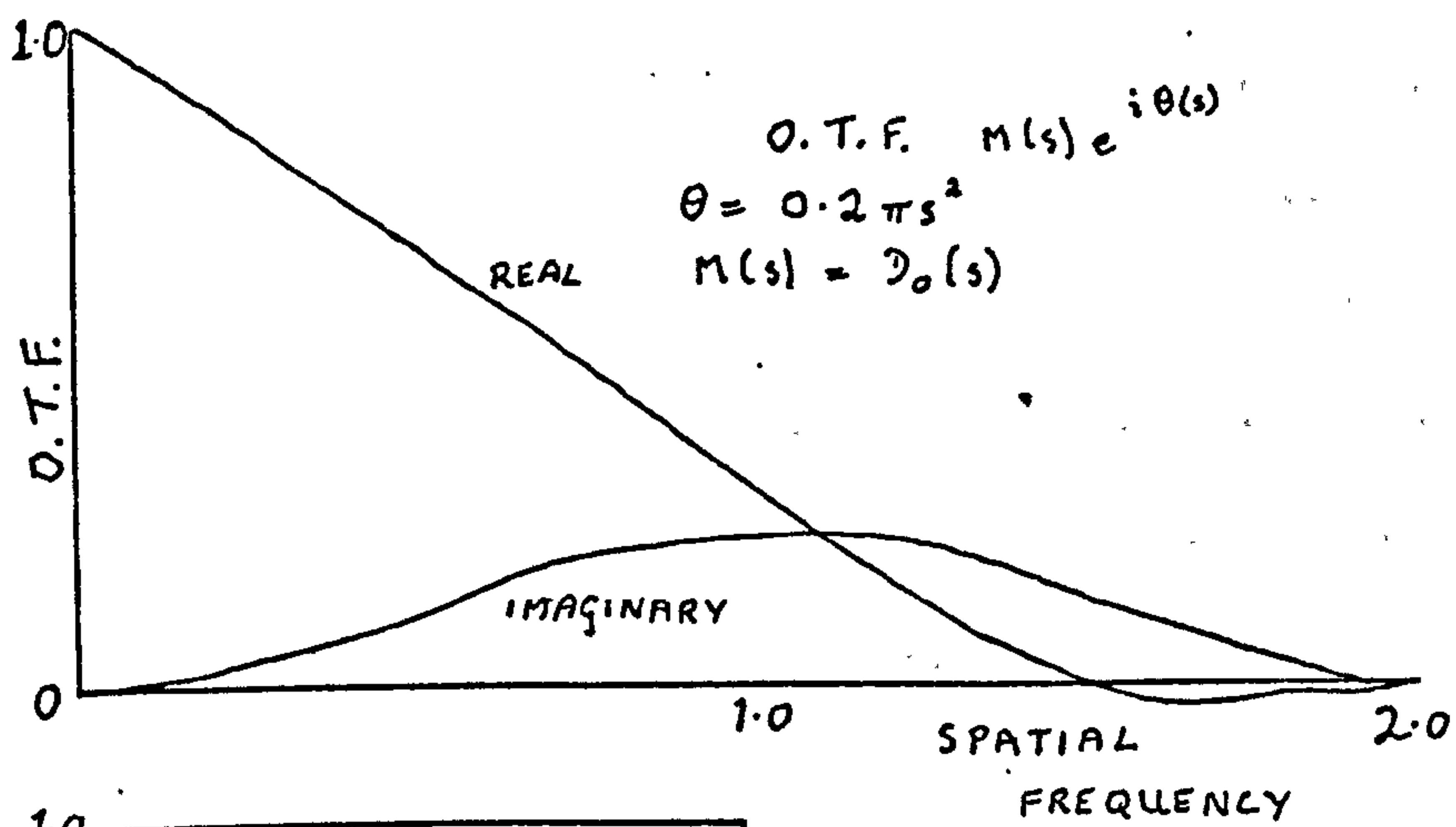


Fig 7.5(a)

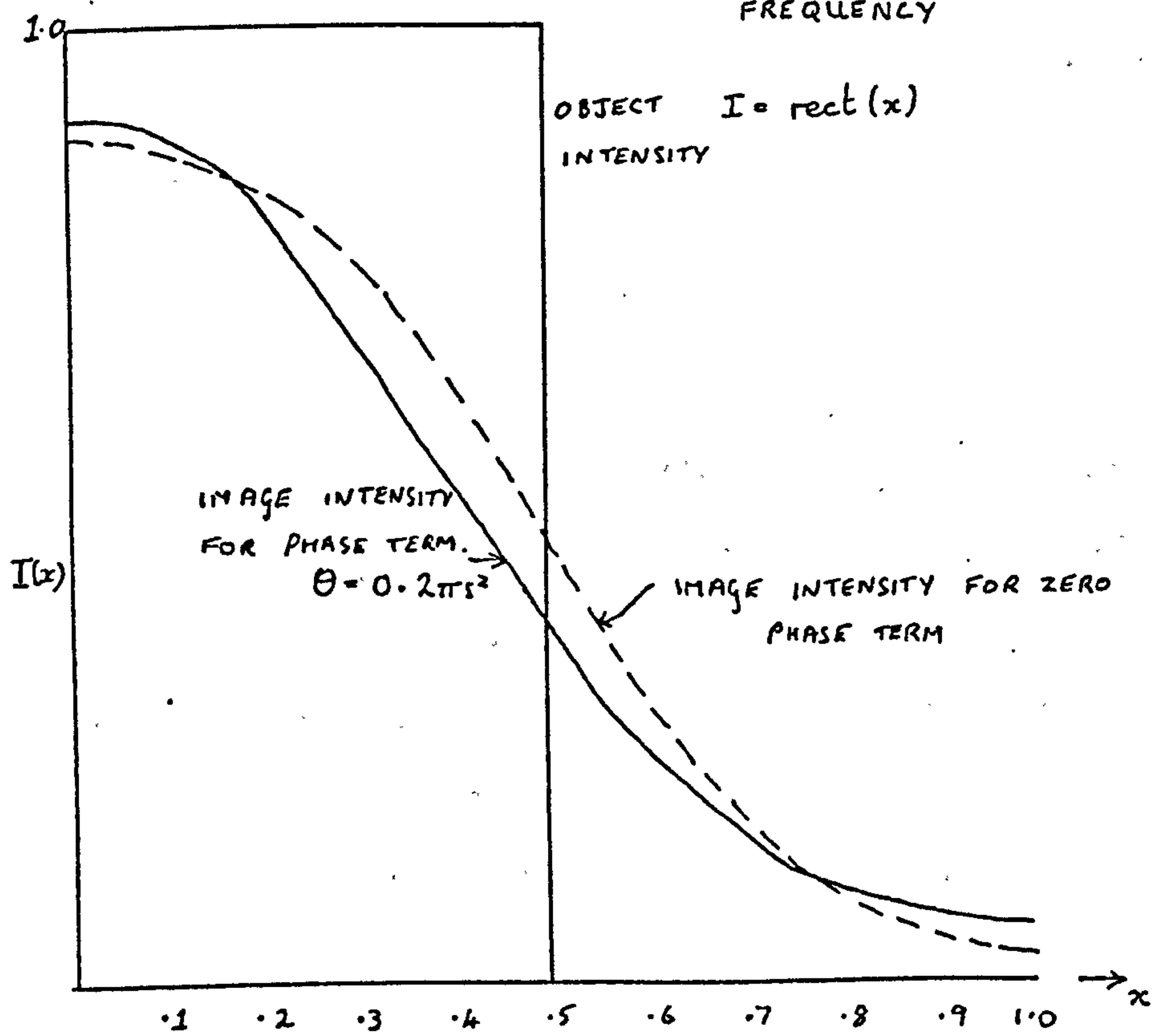


Fig 7.5(b)

effects depends on the spacing of the object. For objects with small spacing and hence a high fundamental spatial frequency the high frequency components will be greatly attenuated by the modulus of the o.t.f., and hence any phase shifting of these components relative to the fundamental will have little overall effect on the image. For serious distortion to occur sufficient components must pass through the system with high modulation and large phase shifts. Even in the most extreme cases investigated the distortions produced by phase terms of the magnitude usually apparent in optical systems did not impair the image to any great extent.

Physiological Optics and the O.T.F.

It has been assumed during this thesis that an o.t.f. criterion is appropriate for the type of system being optimised. Naturally the usefulness of an o.t.f. specification for a system depends upon the proposed use of the system. For photographic, and particularly photo-reconnaissance work a considerable amount of research has been carried out into the applicability of image quality criteria. Much of this work is performed under classified projects, and many reports have not been widely published, but this section summarises some of the most important results which have been reported on o.t.f. criteria.

One advantage of an o.t.f. performance specification is that in a compound optical system, for incoherent imagery, the o.t.f. of the total system is simply the product of the o.t.f.'s of the individual components. This means that, knowing the frequency response of a given film emulsion, the performance of the rest of

the system may be specified to give the required lens-film performance. In this following section the o.t.f. used is always the o.t.f. of the complete lens-film system. Much of this work may be found in the excellent book by Biberman (1974).

Several o.t.f.-based criteria have been proposed for system evaluation. The earliest of these were the one and two-dimensional Strehl intensities

$$I_1 = \int_0^\infty M(s) ds \quad 7.3$$

$$I_2 = \iint_0^\infty M(s,t) ds dt \quad 7.4$$

Shade (1952) suggested an equivalent passband N_e , defined by

$$N_e = \int_0^\infty M(s)^2 ds \quad 7.5$$

These criteria were analysed by methods of subjective:objective testing. Observers were asked to rank photographs according to certain image qualities, hence providing a subjective ranking of the photographs. The o.t.f. values of the system used to create the photographs were calculated, and these were combined by the proposed criteria and the resulting figures gave an objective ranking of the photographs. The correlation between the two sets of rankings gave a measure of the usefulness of the criteria for system evaluation.

Early studies concluded that the straightforward two-dimensional Strehl intensity gave the best performance in these ranking tests. It was suggested that the criteria should be modified to include the effects of film granularity and the response of the human observer. For subjective image evaluation it was found that the region of greatest importance for a human observer was

where the o.t.f. was between 10% and 40%. As shown in the results of chapters 5 and 6 it is in this region that the geometric otf criterion shows its greatest advantage over the other optimisation criteria. It was also shown from the ranking tests that the minimum detectable difference in o.t.f. values which could be distinguished when evaluating two photographs of the same scene was 5%, which justifies the use of the 5% tolerance used in the first section of this chapter.

Charman and Olin (1965) attempted to include the effects missing from the previous criteria. They determined that the effects of film granularity, characteristic curve and the human observer set a limit on the minimum resolvable contrast for any given spatial frequency, $M_t(s)$. Since, for most aerial reconnaissance situations, the average modulation present in the object is 20%, they proposed that any criterion should include only the area enclosed by 0.2 times the system o.t.f. and the threshold curve. This is illustrated in figure 7.5.

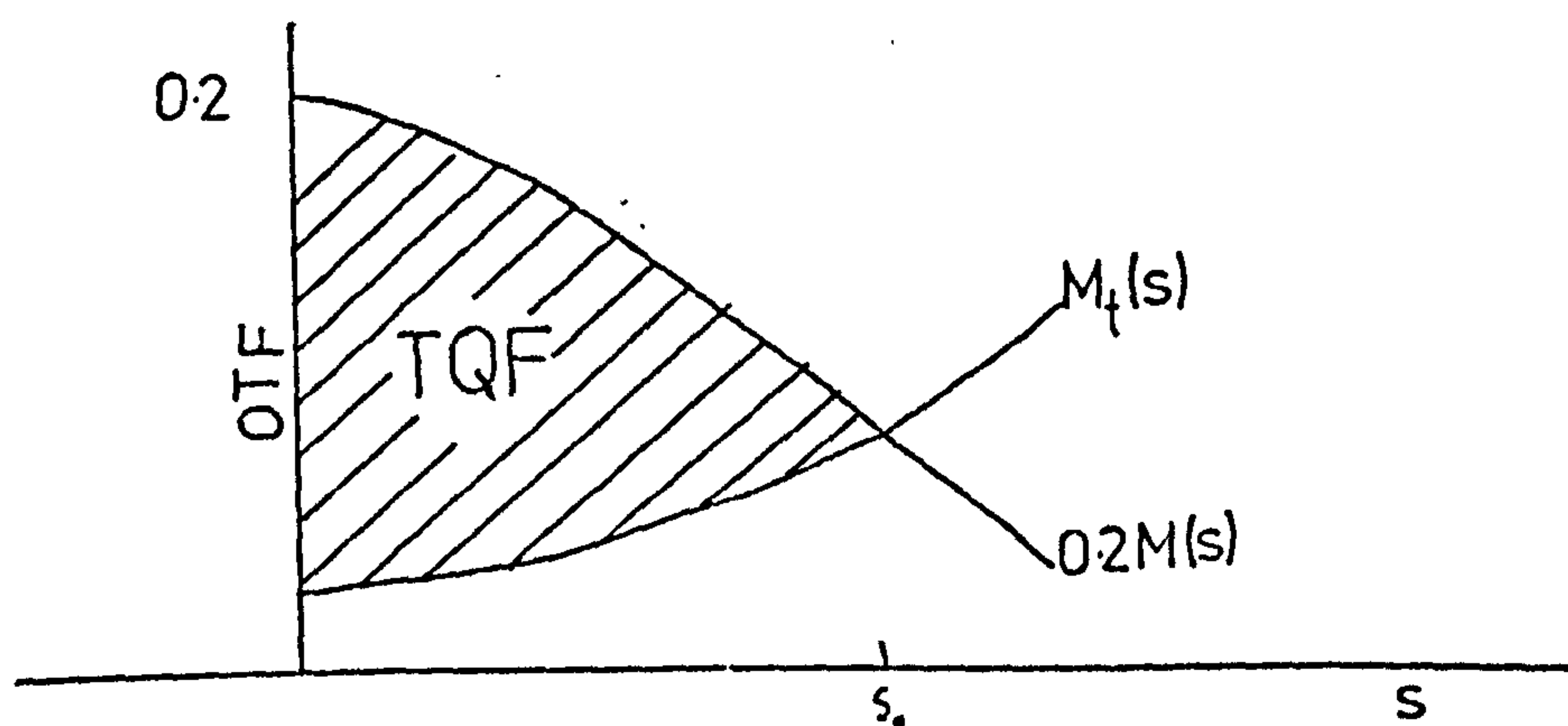


Fig. 7.5 The Threshold Quality Factor

The threshold quality factor (TQF) is thus defined as

$$TQF = \int_0^{s_0} (0.2 M(s) - M_t(s)) ds \quad 7.6$$

The TQF deals only with modulation values, and is thus not affected by phase shifting in the o.t.f.

This approach was followed by Borough et al (1967) who proposed and tested two similar criteria, based on the area-weighted m.t.f. the MTFA. The first of these criteria is the linear MTFA, defined as

$$\text{MTFA}(\text{lin}) = \int_0^{s_1} (M(s) - M_t(s)/M_o(s)) ds \quad 7.7$$

where $M_o(s)$ is the perfect o.t.f.

The second criterion defined the area on a logarithmic scale, to give the MTFA(log-log) as

$$\begin{aligned} \text{MTFA}(\text{log-log}) = & \int_{\log s_0}^{\log s_1} [\log(M(s))] d \log s \\ & - \int_{\log s_0}^{\log s_1} [\log(M_t(s)/M_o(s))] d \log s \end{aligned} \quad 7.8$$

For both these criteria s_1 represents a high frequency cut-off at the overlap of the threshold and m.t.f. curves. In the second criterion s_0 is introduced to prevent over-weighting the low-frequency components.

Borough reported correlations of 0.92 and 0.90 for the two criteria for objective:subjective ranking tests.

Granger and Cuperey (1967) also introduced a low-frequency cut-off, but this was introduced to take account of the band-pass of the human eye. Their criterion, the subjective quality factor SQF was defined on a semi-logarithmic scale as

$$\text{SQF} = \int_0^{s_1} (M(s) - M_t(s)) d(\log(s)) \quad 7.9$$

Granger and Cuperey reported the astonishingly high figure of 0.98 for the correlation of the SQF in ranking tests.

All these results suggest that for situations where human analysis of images is required, several o.t.f. based criteria provide accurate and efficient means of system evaluation, and hence o.t.f. should be used as the optimisation criterion for the design of the system. The type of o.t.f. curve produced by the geometric o.t.f. optimisation criterion appears to be the only available criterion which will automatically produce suitable balances for this type of application at high aberration sizes.

Practical Optimisation Techniques

1. The SLAMS Programme

The optical design programme used at Imperial College is a direct descendant of the least-squares programme originally described by Wynne (1959). This programme uses the technique of successive linear approximations at maximum step, SLAMS, which is described below, and which has been modified to use the new geometric o.t.f. criterion.

The general merit function, ψ , may be taken as a sum of squares of m weighted aberrations f_i as discussed in chapter 2.

Then

$$\psi = \sum_{i=1}^m f_i^2 \quad 8.1$$

The treatment of the least-squares theory is simplified by the use of matrix notation. The notation used in this chapter is that two-dimensional matrices are represented by capital letters and column or row vectors are represented by lower case letters. Subscripted letters represent individual elements of the matrices. With this notation the merit function may be written

$$\psi = |f|^2 \quad 8.2$$

The aberrations may be considered as functions of the n variable system parameters, x_j , which will be altered by the programme. At the start of each iteration the programme sets up a matrix of differentials A , where

$$A_{ij} = \partial f_i / \partial x_j \quad 8.3$$

If the origin is set to the point $x_j = 0$, and parameter changes x_j are applied, then within the limits of linearity

each aberration f_i becomes f'_i , where

$$f'_i = f_i + \sum_{j=1}^n A_{ij} x_j$$

$$f' = f + A \cdot x$$
8.4

If the vector f' is set to zero, then assuming the number of aberrations m is equal to the number of parameters n equations 8.4 give the parameter changes required to simultaneously reduce all the aberrations, and hence the merit function, to zero. This assumes that the aberrations are linear functions of the parameters, and that the equations have a solution i.e. the matrix A is not singular. In nearly all practical situations the functions are not linear, and the matrix A is often ill-conditioned. For this reason the number of aberrations is taken to be larger than the number of parameters, and a least-squares method is used to minimise the sum of squares of the residual aberrations f' .

Differentiating the merit function ψ with respect to the parameter x_k gives

$$\frac{\partial \psi}{\partial x_k} = 2 \sum_{i=1}^m f_i \frac{\partial f_i}{\partial x_k}$$
8.5

We may define a vector g as

$$g_k = \frac{1}{2} \frac{\partial \psi}{\partial x_k} = \sum_{i=1}^m f_i \frac{\partial f_i}{\partial x_k} = \sum_{i=1}^m A_{ik} \cdot f_i$$
8.6

$$g = \tilde{A} \cdot f$$
8.7

At the minimum of ψ we may set the gradient, and hence the vector g , to zero. At this point the residual aberrations will be f' , and hence from equations 8.4 and 8.7 we may write

$$g = \tilde{A}.A.x + \tilde{A}.f = 0 \quad 8.8$$

In terms of the individual aberrations these equations become

$$g_k = \sum_{i=1}^m A_{ik} \left(\sum_{j=1}^n A_{ij} \cdot x_j \right) + \sum_{i=1}^m A_{ik} \cdot f_i = 0 \quad 8.9$$

where k ranges over the n parameters of the system. This gives n equations in the n unknowns, x_j , and hence, if the equations are not singular, may be solved exactly.

Since the functions are non-linear, the derivatives will only be valid over small regions. It is necessary, therefore, to restrict the parameter changes to remain within the bounds of linearity of the system. Levenberg (1944) suggested restricting the magnitude of the change vector x , which is known as the step length, by including in the merit function a term ϕ , where

$$\phi = p^2 R = p^2 \tilde{x}.x \quad 8.10$$

The scalar p governs the magnitude of the step length R , as will be shown. Differentiating ϕ with respect to x_k gives

$$\frac{\partial \phi}{\partial x_k} = 2 p^2 x_k \quad 8.11$$

This term is added to the gradient of ψ from equation 8.5 to give

$$\begin{aligned} g_k &= \frac{1}{2} \frac{\partial \psi}{\partial x_k} + \frac{1}{2} \frac{\partial \phi}{\partial x_k} \\ g_k &= \sum_{i=1}^m A_{ik} \cdot f_i + p^2 x_k \end{aligned} \quad 8.12$$

or
$$g = \tilde{A}.f + p^2 x$$

The least-squares equations, 8.8 and 8.9, become

$$\begin{aligned} g &= \tilde{A}.A.x + p^2 x + \tilde{A}.f = 0 \\ g_k &= \sum_{i=1}^m A_{ik} \left(\sum_{j=1}^n A_{ij} \cdot x_j + p^2 x_k \right) + \sum_{i=1}^m A_{ik} \cdot f_i = 0 \end{aligned} \quad 8.13$$

This latter equation may be written as

$$g = (\tilde{A}.A + p^2.I).x + \tilde{A}.f = 0 \quad 8.14$$

$$g_k = \sum_{j=1}^n \left(\sum_{i=1}^m A_{ik}A_{ij} + p^2 I_{kj} \right) x_j + \sum_{j=1}^m A_{jk}f_j$$

A constant term p^2 has been added to each diagonal element of the square matrix $\tilde{A}.A$, and the effect of this damping factor p is to limit the change in each of the parameters x_j such that as p increases the step length R decreases, and vice-versa. By suitable choice of p the step length can be controlled to lie just within the range of linearity of the equations. The control of the step length in the original SLAMS programme has been discussed by Wynne (1963) and consists of successive halving or doubling of p . More details of this technique, and the later modifications are given below.

The SLAMS programme has been continuously developed over the years, and a detailed description of its use was given by Kidger (1971). The current version, version 14, has been largely written by Mrs Prudence Wormell of the Optical Design Group, and has a modular structure which enables users to simply construct additional factors for the merit function. The basic programme has facilities for optimisation using transverse and wavefront aberrations, including provision for a shift of the reference sphere radius to give the Maréchal variance, and angular aberrations for afocal systems. Chromatic aberrations are treated either by the Conrady d-D formulae, or by raytracing in several wavelengths. Glass changing is performed by allowing the glass parameters, refractive index and dispersion, to vary within an area on the standard glass chart. After initial optimisation the glasses are usually set to the

nearest values for manufactured glasses before a final balancing run.

The standard programme requires that the designer gives the full set of rays to be traced together with the individual weights for each of the aberrations to be included in the merit function. This means that the designer must provide a considerable amount of input data, since the programme can accept up to 50 variable system parameters and 100 terms in the merit function. The limit on the number of terms in the merit function sets a limit on the number of rays which can be traced dependant on the number of aberrations per ray, usually between three and eight.

The programme considers, for efficiency, five types of rays, set by an index on the data cards. These rays are;

1. Paraxial ray from an axial object
2. Finite ray from an axial object
3. Principal ray from an off-axis object
4. Meridian ray from an off-axis object
5. Skew ray from an off-axis object.

The rays were split in this way since for each of these groups different values are required from the raytrace. The types of aberrations calculated for each ray depend upon the type of ray and the nature of the merit function. For simplicity only the case of raytracing in several wavelengths will be discussed below.

As well as the individual ray aberrations the programme must also control certain system values. These may typically be the required focal length, overall length, throw and total

glass thickness. These may either be given target values, in which case their variance from the target is taken as an aberration, or they may be given limits, in which case they are only considered to be aberrations when they move outside these limits.

Most of these overall values are calculated from the paraxial ray, type 1, for the chosen mean wavelength. The variations of the back focal lengths in the other wavelengths are calculated with respect to the mean value, to control the longitudinal chromatic aberration.

The off-axis principal ray, type 3, is used to give the image height for each wavelength, and the astigmatism, distortion and magnification for the mean wavelength. The image point in the principal wavelength acts as the reference point for the aberrations for the meridian and skew rays from the same object point.

The finite meridian and skew rays, types 2, 4 and 5, give the transverse and wavefront aberrations in all wavelengths and, for the axial meridian rays, the value of OSC, the offence against the sine condition.

For the new programme using the o.t.f. criterion the individual rays and weights are selected by the programme, and the work of the designer is greatly reduced. The programme is given the paraxial marginal ray to determine focal length and the other system values, and the coordinates of the required off-axis principal rays. These rays give the same aberration types as in the standard programme, and hence weights may be included to control distortion, astigmatism and other standard variables. There is also provision for the designer to specify standard rays and weights to be used in addition to the automatically

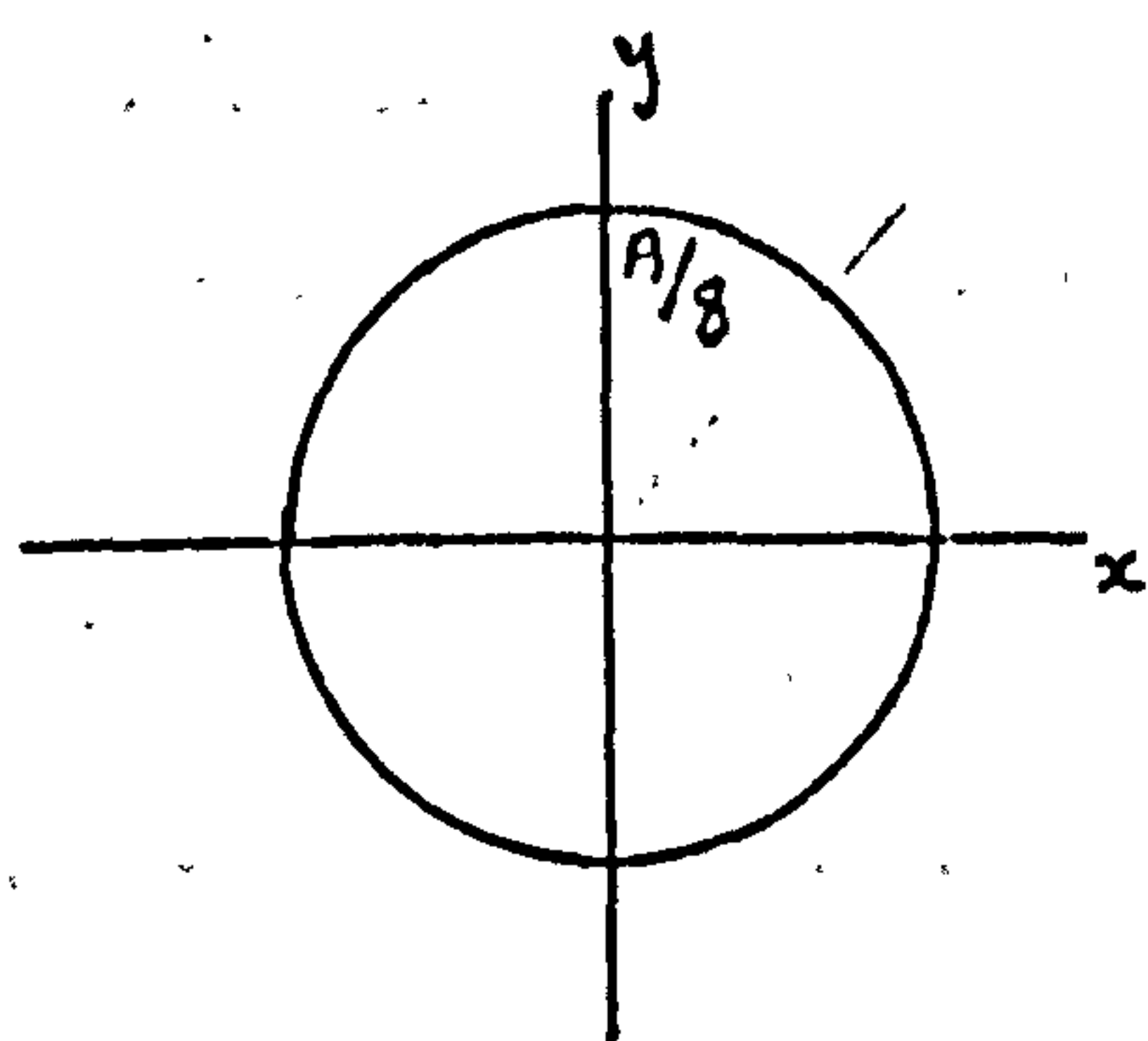
selected ones if required.

The programme is given the values for the spatial frequencies to be optimised - this may be one or more values - and the designer sets relative weights for the o.t.f. in the various frequencies, wavelengths, field angles and azimuths. It is usual to give chromatic weights to correspond to the spectral sensitivity of the final recording medium of the system, such as a film or phosphor.

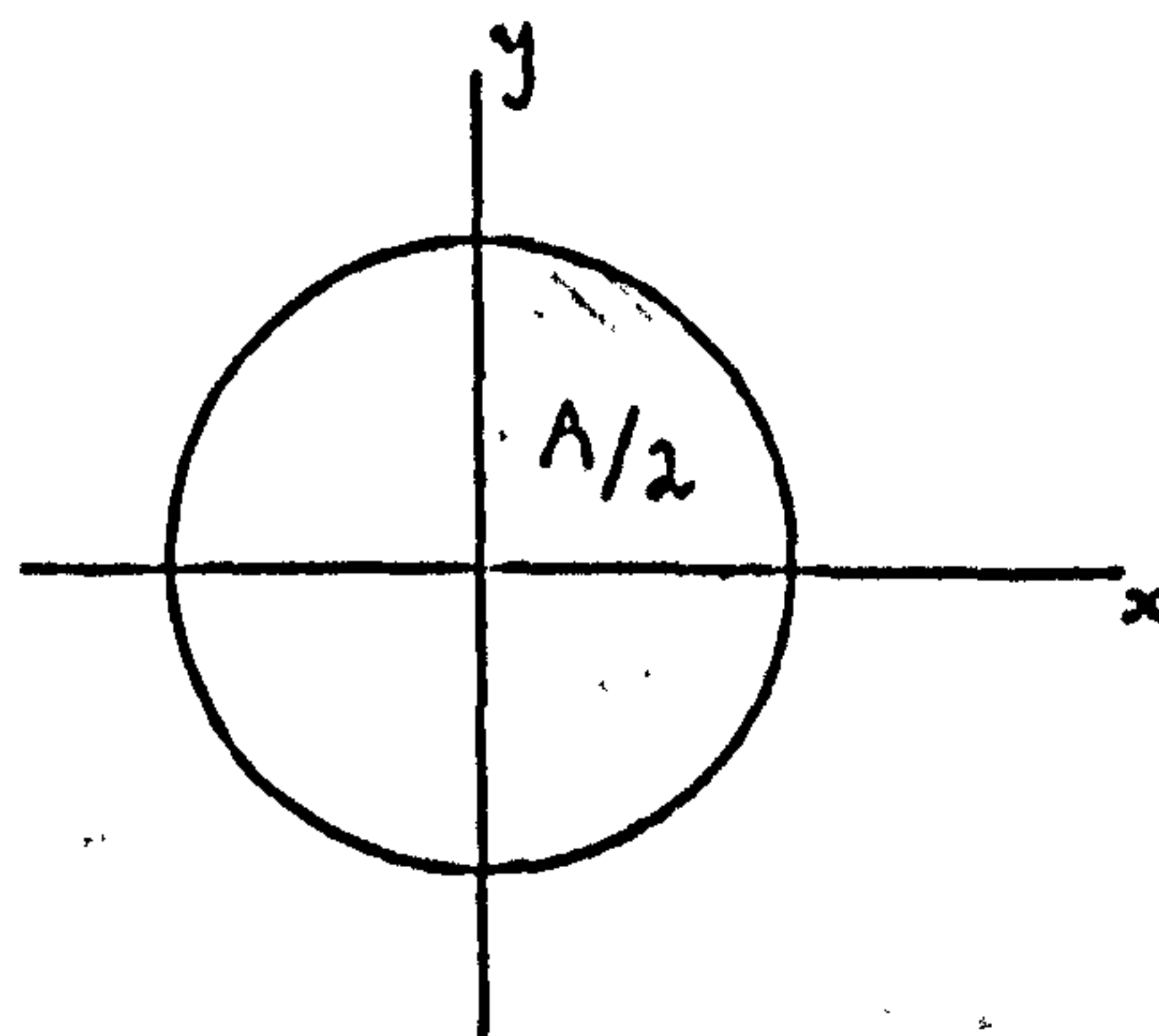
The programme has been enlarged to accept a maximum of 300 aberrations, but the designer may choose to reduce the amount of computation required by reducing the maximum number of aberrations and hence the number of rays generated. The designer gives the programme a set of clear apertures and the programme starts the ray generation process.

Ray Patterns for o.t.f. Optimisation

It is generally known that for o.t.f. calculations symmetry considerations reduce the amount of computation needed. For the axial case the area of integration may be reduced to one octant of the pupil, as shown in figure 8.1a



(a)



(b)

Fig 8.1

Areas of integration for axial (a) and off-axis (b) o.t.f.

The geometric o.t.f. integral on axis may be written

$$\begin{aligned}
 D(S) &= 1 - \iint_A 2\sin^2(\pi S \zeta\eta) dx dy \\
 &= 1 - \iint_A 2\sin^2(\pi S \zeta\eta) dx dy
 \end{aligned}
 \tag{8.15}$$

Since there is symmetry about both the x- and y-axis we may write

$$\zeta\eta(x,y) = \zeta\eta(-x,y) = \zeta\eta(x,-y) = \zeta\eta(-x,-y)$$

Then

$$D(S) = 1 - 4 \iint_{A/4} 2\sin^2(\pi S \zeta\eta) dx dy \tag{8.16}$$

However, since the aberrations are radially symmetric

$$\zeta\eta(x,y) = \zeta\eta(y,x) \tag{8.17}$$

and 8.16 reduces to

$$D(S) = 1 - 8 \iint_{A/8} 2\sin^2(\pi S \zeta\eta) dx dy \tag{8.18}$$

When the object moves away from the axis, the only symmetry remaining is about the meridian plane, the y-z plane, and so the area of integration may be halved as shown in figure 8.1b.

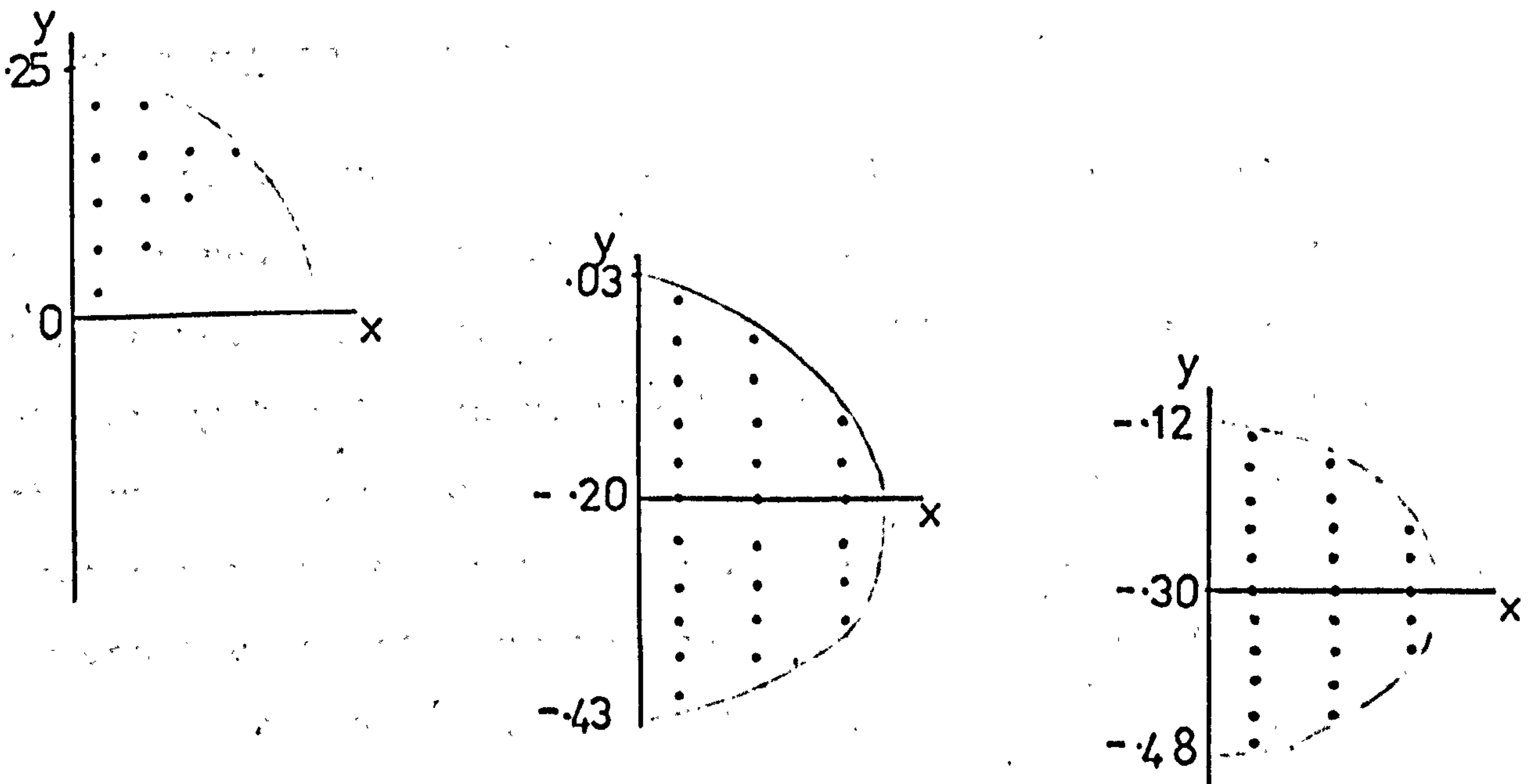
The programme sets up rays on a rectangular grid inside these areas. On axis the ray coordinates are calculated from the object position and the numerical aperture of the system, but for off-axis objects the vignetting means that pupil exploration techniques must be used to determine the entrance pupil limits. Starting from the principal ray in each field angle, meridian rays are traced using a simple iterative technique to find the upper and lower limits of the pupil in the meridian section. These values are used to set up a rectangular mesh across the entrance pupil, and only those rays which pass within the specified clear apertures are accepted for the optimisation.

RAYS GENERATED FOR DOUBLE-GAUSS SYSTEM

AXIS

-13°

-17°



x, y - pupil co-ordinates

FIG 8.2

A typical set of rays, generated for a double-Gauss system, is shown in figure 8.2.

It is possible that the shape of the entrance pupil will vary during the optimisation process. If the clear apertures are maintained by edge thickness minimum constraints and new rays are generated to fill the pupil at the start of each optimisation run, then the pupil cannot reduce in area, and hence vignetting is effectively controlled.

Optimisation Process

One great advantage of the new optimisation criterion is that it required relatively minor modifications to the existing programme. All rays, apart from the principal rays and rays chosen by the designer, are treated as skew rays of type 5 in the raytrace section, and hence only one of the five raytrace routines needs to be modified. At each iteration the programme traces the set of rays to produce the values of the current weighted aberrations, the vector f . For the skew ray transverse aberrations the standard programme takes the terms

$$f_{i_s} = a_i \delta \xi_i \quad 8.19$$

and $f_{i_t} = a_i \delta \eta_i$

As shown in chapter 3 the only change needed is to replace the the transverse aberrations $\delta \xi, \delta \eta$ by the terms $\sin(\pi S \delta \xi)$ and $\sin(\pi T \delta \eta)$ respectively. This gives the aberrations as

$$\begin{aligned} f_{i_s} &= a_i \sin(\pi S \delta \xi) \\ f_{i_t} &= a_i \sin(\pi T \delta \eta) \end{aligned} \quad 8.20$$

The terms $S \delta \xi, T \delta \eta$ will be referred to as aberration products. When the first raytrace has been finished the programme uses

a finite difference method to set up the matrix A . Each parameter x_j in turn is given an increment δx_j and the raytrace is repeated to give the new weighted aberration vector f' . The matrix A is then constructed as

$$A_{ij} = (f'_i - f_i) \quad 8.21$$

The increments, δx_j , are set by the programmer at the start of the optimisation. As a compromise between having separate increments for each parameter or one increment for all parameters, the programme uses separate increments for the four classes of parameters, the curvatures, separations and the two glass parameters.

Once the matrix A has been constructed, the programme enters the p-selection routine. Early versions of the programme used a technique described by Wynne and Wormell (1963). For each value of the damping factor p the new value of the merit function ψ' is compared with the value predicted by substituting the changes into equations 8.4. This linear forecast is denoted by ψ_{LIN} . If ψ' is much larger than ψ_{LIN} , the step length must be reduced, and hence p must be increased, while if there is close agreement between the two values the programme is well within the linearity region and the step length could safely be increased. In this way the step length is adjusted so that the solution is always near the limit of linearity.

Version 14 of the programme uses a more sophisticated method of p-selection. Starting with a given p -value, the programme solves the least-squares equations for values of $2p$, p and $0.5p$. These produce merit function values of ψ , ψ_1 , ψ_2 . If these

three values span a minimum value of Ψ then a curve fitting technique is used to select the value of p which should give this minimum, and the equations are re-solved with this value. If the three values do not span a minimum, the value of p giving the highest Ψ value is rejected and a new value of p is used at the other end of the series. It is possible, especially in the later stages of optimisation, that repeated trials fail to span a minimum, and in this case the old method of p -selection is used.

When the p -selection is complete the programme sets all the parameters to their new values and prints the results of a new raytrace, if required, before the next iteration.

It is in the process of differentiation and p -selection that problems were first found with the new programme. The problems stem, in part, from the accuracy of the method of differentiation with finite differences. The accuracy of the difference equation 8.21 will depend on the nature of the function f , and the size of the increment δx . To avoid problems of non-linearity the increment should be as small as possible, but since all digital computers have a finite word length the effect of rounding errors must not be neglected. The choice of the increment size must be a compromise between rounding error considerations and accuracy.

When the terms in the merit function are taken as the sines of the aberration products, the sine will introduce a further non-linearity on top of the non-linearity of the aberrations themselves. This means that the increments must be chosen with greater care to ensure that accurate derivatives A_{ij} are found.

There will also be an increased discrepancy between Ψ' and Ψ_{LIN} for a given value of p .

To investigate these effects a series of optimisations was performed on one of the lenses described later in this thesis. The lens was a modified double-Gauss P.E.P.R. lens, of focal length 100 mm. The curvatures ranged between plane, $c = 0$, and $c = 0.1 \text{ mm}^{-1}$. For this curvature range the standard programme would use an increment 10^{-4} , and trials were made with curvature increments δc from 10^{-4} to 10^{-15} . All trials started with the same system and the same merit function value of 577. The CDC 6400 computer has a precision of 15 significant figures with its 64-bit word length and so, as expected, the trial with an increment of 10^{-15} produced virtually no improvement. Figure 8.3 shows the final merit function produced by each of the runs. These results show that there is a strong dependence on increment size, with a clear minimum at $\delta c = 10^{-9}$, although there is a reasonable region between 10^{-7} and 10^{-11} . The numerical value chosen will depend on the curvature values of the system, and in future optimisations the curvature increment was taken to be 10^{-8} times the largest curvature value, with similar choices for the other three increments.

The increased non-linearity due to the sine term becomes apparent when the aberration product becomes greater than 0.5. At this point the term $\sin(\pi \delta \xi)$ moves into the next quadrant. When the aberration product approaches the critical value 0.5 the derivative with respect to any of the parameters becomes zero. The programme assumes, therefore, that in the linear case no parameter change will produce an improvement in this aberration, whereas, in fact, any parameter change will produce an

FINAL MERIT FUNCTION VALUE, ψ , against CURVATURE
INCREMENT δc

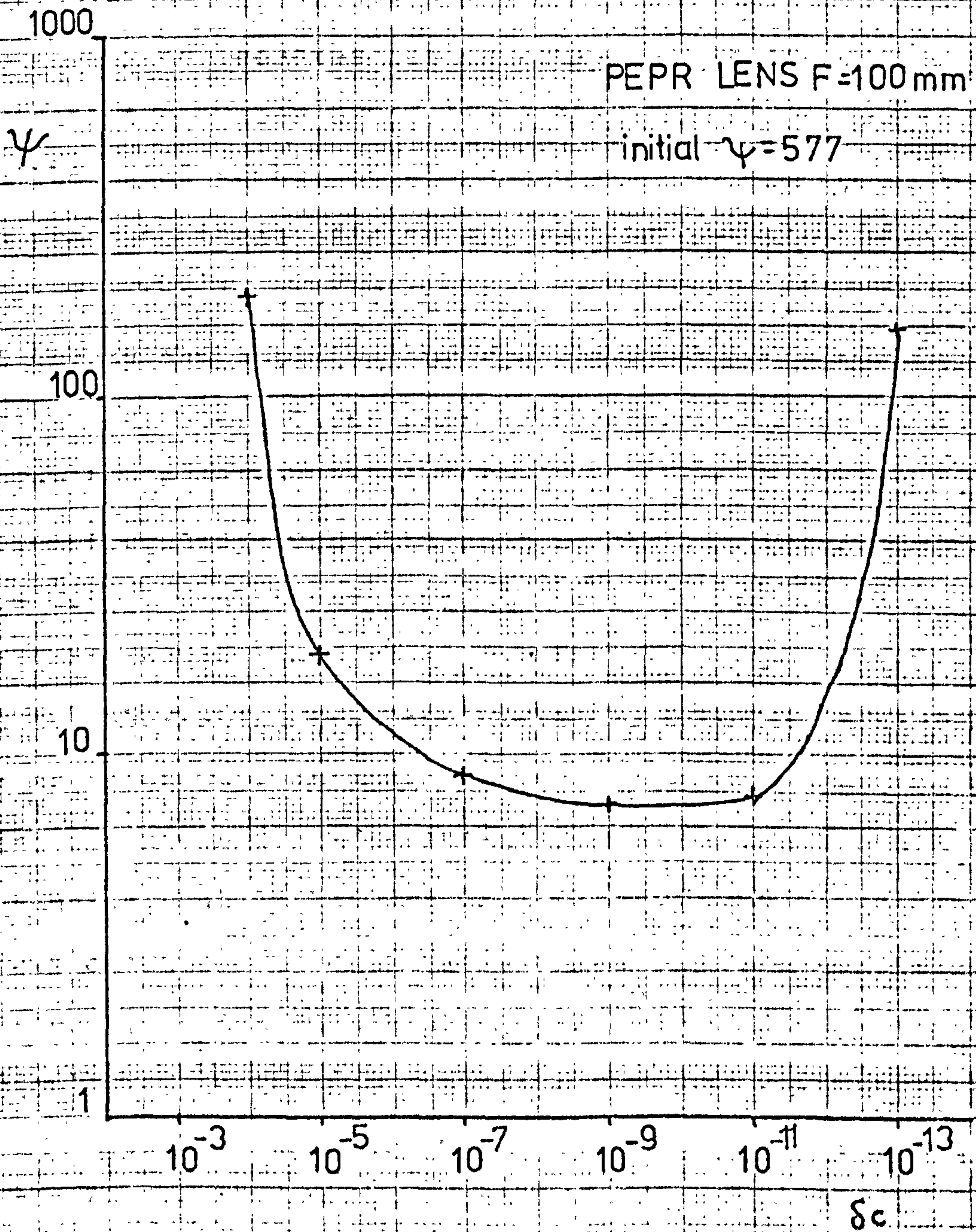


FIG 8.3

improvement.

The improved p-selection technique described above has particular benefit for this non-linear situation, since it considers real values of the merit function, and does not rely on the comparison between Ψ' and Ψ_{LIN} , which can often be wildly misleading for the new merit function. This p-selection effectively compensates for inaccuracies in the differentials, providing the number of critical aberrations is small compared with the total number of aberrations.

In practice these effects are reduced by using appropriate values for the spatial frequency of optimisation. It is usual that any system optimised by the standard programme will contain many aberrations whose aberration products would exceed 0.5 for the required spatial frequency. By choosing a low spatial frequency such that most of the aberration products are below 0.5, a preliminary optimisation may be performed. This optimisation has the effect of reducing the aberrations so that the aberration products at the required spatial frequency will be below 0.5. It would be possible, if needed, to use a series of intermediate optimisations at increasing spatial frequencies, but it has not been found necessary in any of the systems optimised to date to use more than one preliminary case.

Although, as has been stated, the choice of weighting factors is simplified in the new programme, thought must still be given to relative weights for the required spectral sensitivity and for the system values. For each wavelength λ the o.t.f. terms in the merit function are of the form

$$\Psi_{\lambda} = \sum_i a_{\lambda i}^2 \sin(\pi S \delta \xi_i)^2 \quad 8.22$$

As shown in chapter 3 this corresponds to a merit function

$$\Psi_{\lambda} = a_{\lambda}^2 \left(1 - \sum_i \cos(\pi \lambda s_i \xi_i) \right),$$

$$\Psi_{\lambda} = a_{\lambda}^2 \left(1 - \text{O.T.F.}(S)_1 \right) \quad 8.23$$

Thus the relative weights a_1 must be chosen as the square roots of the required relative chromatic response.

Although weighting the aberrations equally over the azimuths for each given wavelength does produce the maximum average o.t.f. it is usual to require that the minimum o.t.f. value is as large as possible, to ensure a satisfactory performance over the whole field. For this reason the weights are chosen to give highest weights to those azimuths and fields with the lowest o.t.f. at the start of the run.

For control of other system values the normal aberration terms from the standard programme have been retained. Some of these aberrations have traditionally been used in optical design, and are of particular value in first order design. For final o.t.f. optimisation, however, only those values which describe overall characteristics, such as distortion and magnification, need be retained. Other terms, such as astigmatism, describe point imagery, and since the aberration product terms are being used to control o.t.f. rather than point imagery, these other terms should be neglected. It is true that extreme values of astigmatism would not be consistent with high o.t.f. performance in both azimuths in one focal plane, but the astigmatism should be controlled indirectly, by the aberration product terms and be free to take whatever

values are determined by the o.t.f. optimisation. This is shown very clearly in the case of the microfilm lens given in chapter 9.

Thus, as well as providing direct o.t.f. optimisation the new programme requires considerably less work on the part of the designer than the traditional optimisation programme.

Practical Lens OptimisationIntroduction

This chapter gives details of some of the optimisations performed with the new o.t.f.-based programme, and shows the type of improvement to be expected from its use.

To provide a meaningful test of the new programme the systems to be optimised had to fulfil certain conditions. Firstly, the systems had to be ones which required a high o.t.f. performance, and secondly they must have been designed with a conventional optimisation programme using a merit function which has been constructed to give, as far as possible, the optimum o.t.f.

The systems to be rebalanced with the new programme had all been optimised using the SIAMS programme, and all were systems where the o.t.f. performance was the main quality criterion. Some of these systems, however, did not fulfil the second condition above. These were double-Gauss objectives which had been designed by Michael Kidger of the Optical Design Group, and are extensively discussed in his thesis (1972). Some of the early designs, although requiring high o.t.f. performance, had been designed purely on spot size considerations. When these systems were rebalanced by the new programme, the high frequency o.t.f. showed a rapid improvement, in one case from the original 18% to 56%. This showed the ease with which the new programme could improve conventionally-balanced systems, but did not serve as a good test of the programme since the original systems could undoubtedly be greatly improved by adjusting the merit function in the standard programme.

It was decided, therefore, to use the new programme on systems which had either been recently developed by the Optical Design group, or were currently under development. These designs would represent the best performance which could be

obtained from a conventional programme, using all the techniques available to the designer.

All the conventionally-optimised systems have been designed using version 12 of the standard SLAMS programme, and these will be designated 'V12' on the diagrams in this chapter. All systems optimised on the new programme will be designated 'OTF'. All figures in this chapter are in centimetres, unless explicitly labelled.

1. Padua P.E.P.R Lens

This system, designed for the Instituto Nazionale de Fisica Nucleare at Padua, was the first production lens to be designed on the new programme. The design was a modification of a previous PEPR lens designed by the Optical Design Group, and the specifications for the new system were

Image-Object Distance	98 -100 cm
Object Diameter	21 cm
Axial Magnification	0.7
Edge-of-field Distortion	1.3%
Numerical Aperture	0.1

The system was designed to image onto a photocathode, with the spectral response shown in figure 9.1. The basic lens was used in conjunction with a glass faceplate next to the object, and a condenser near the image plane.

The specification required an o.t.f. performance not less than 40% at 400 cycles/cm anywhere in the field, and so was an appropriate system to design on the new programme.

The conventional programme was used for the initial optimisations, working at three field positions, on axis, full field ($\eta = 10.5$ cm) and 0.7 field ($\eta = 7.35$ cm). This gave the 7 component system shown in figure 9.2. The full design details

of this system, and all other systems, are given in appendix A.

Since this was a well-corrected system, nearly all the aberration-products for a spatial frequency 400 cycles/cm were below the 0.5 limit, and thus the system could be optimised directly at 400 cycles in the new programme, without the need for an intermediate run. The previous optimisation had left the system with 1.7% distortion at the edge of the field, and the new weights were arranged to reduce this figure. The system was optimised at four wavelengths within the required spectral range at 400, 420, 390 and 440 nanometres. The weights for the optimisation run were chosen, according to the theory of chapter 8, to be proportional to the square roots of the relative responses at these wavelengths given from figure 9.1.

After one optimisation at 400 cycles the system shown in figure 9.3 was produced.

As expected, the programme has made relatively minor changes to the system, but, as will be shown by the o.t.f. performance, these small changes are sufficient to produce significant improvements. The optimisation programmes produce the geometric o.t.f. values at the end of each run, and the curves for the two systems are shown in figure 9.4. The curves are given for the three field positions used, for spatial frequencies up to 750 cycles/cm.

These curves clearly show a general rise of the o.t.f. for the new system, but, as always, it is the performance at the specified spatial frequency which is important. The o.t.f. values for the two systems at 400 cycles are given below.

Geometric o.t.f. at 400 cycles/cm

	Version 12		'OTF'	
	sag	tan	sag	tan
Axis	.543		.566	
$\eta = 7.35$	.515	.502	.519	.569
$\eta = 10.5$	.425	.418	.476	.542
Average		.480		.534

Although these figures show a clear 5.4% rise in average o.t.f. and a 4.8% rise in the worst o.t.f. value, it must be determined whether the diffraction o.t.f. figures show similar improvements, in accordance with the previous theory.

The geometric o.t.f. curves were produced by weighting the o.t.f. values produced at the four wavelengths used in the optimisation. It is found that more wavelengths are usually needed to give an accurate forecast of the spectral response for the diffraction o.t.f. A diffraction o.t.f. programme was used to evaluate the diffraction o.t.f. at eight wavelengths within the spectral band, and these values were weighted according to the relative response of figure 9.1. The resulting diffraction o.t.f. curves for the two systems are plotted in figure 9.5.

These curves show a similar rise to the geometric o.t.f. curves between the two systems, and this rise can be clearly seen by considering, as before, the values at 400 cycles, given below.

Diffraction o.t.f. at 400 cycles/cm

Version 12			'OTF'	
	sag	tan	sag	tan
Axis	.507		.594	
$\eta = 7.35$	.471	.546	.558	.620
$\eta = 10.5$	.570	.533	.603	.605
Average	.525		.596	

For these results the new system shows an overall rise of 7% in average o.t.f., but the worst o.t.f. figure, which is the most important criterion has risen from 47% to 56%, which is a significant improvement.

It should be noted, by comparison of the two sets of o.t.f. values, that although there is no direct relationship between the geometric and diffraction values, for each field angle and azimuth both the geometric and diffraction values in the new system are higher than the corresponding values in the version 12 system. This shows that, as predicted, optimisation of the geometric o.t.f. has lead to optimisation of the diffraction o.t.f., even though individual values do not always closely agree.

To illustrate the spectral response of the two systems, the diffraction o.t.f. values calculated at 400 cycles have been plotted as functions of wavelength in figure 9.6. As a check on the accuracy of these curves, the second set of curves, for the 'OTF' system, have been calculated with 17 points along the curve, and these extra points do not show any large fluctuations. The shapes of the two sets of curves are similar, and show the usual type of response for systems which are limited by secondary spectrum.

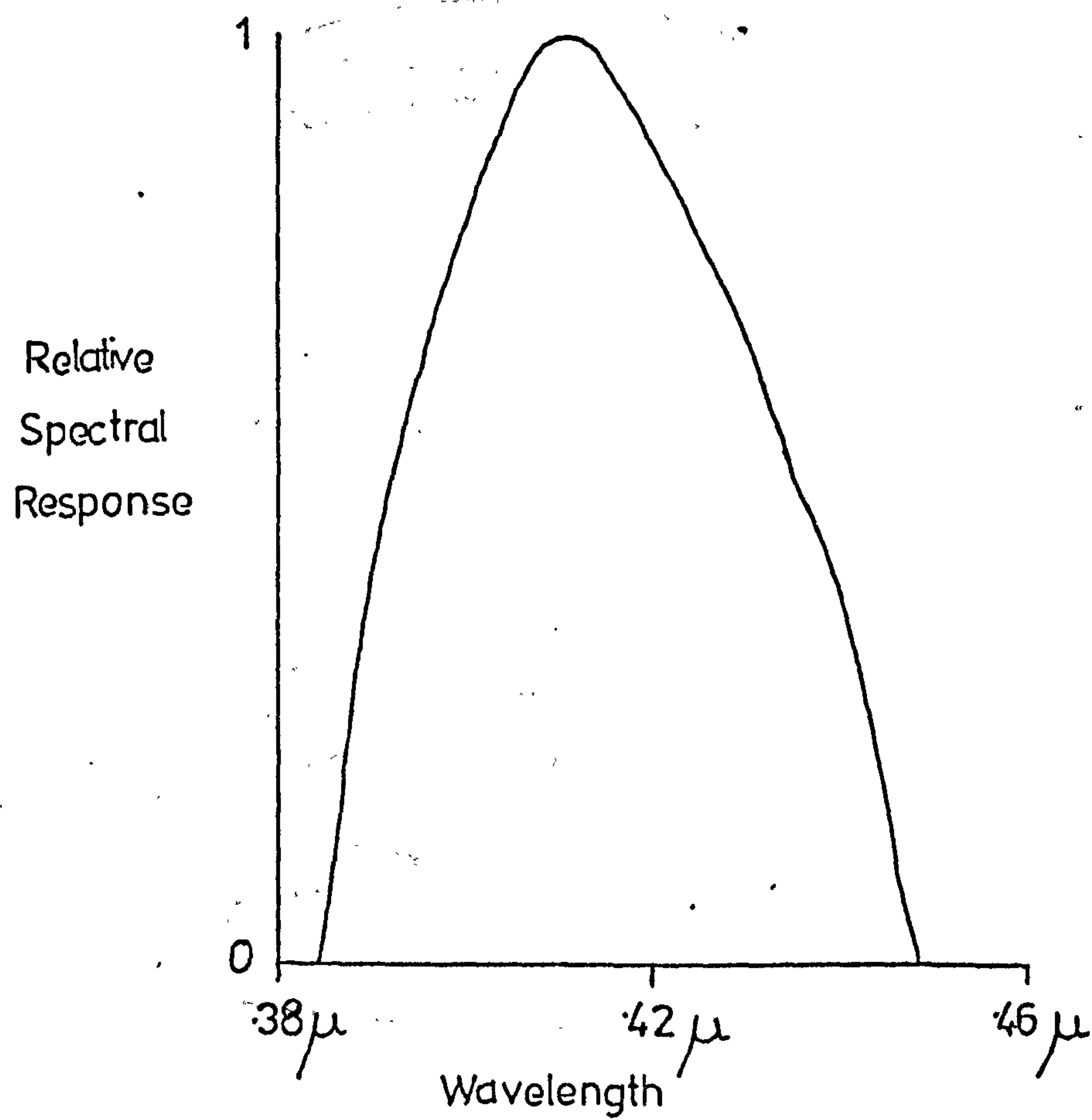
PADUA SYSTEMPHOTO-CATHODE SPECTRAL RESPONSE

FIG 9.1

PADUA LENS 'Y12'

2cm

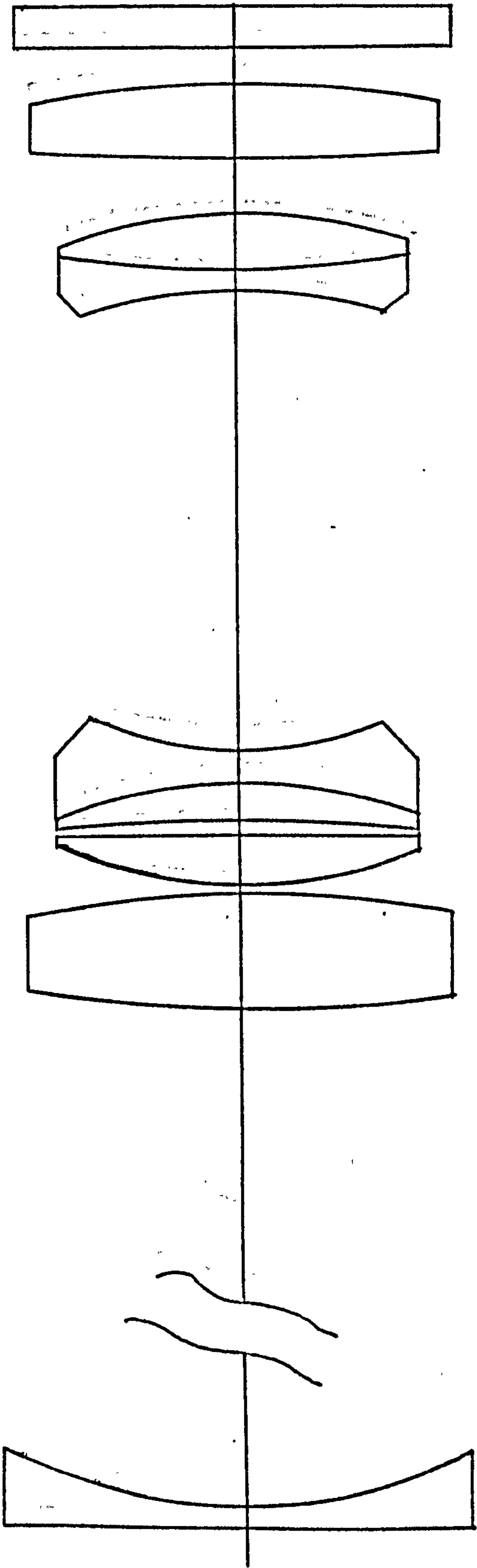


FIG 9.2

PADUA LENS

OTF

2cm

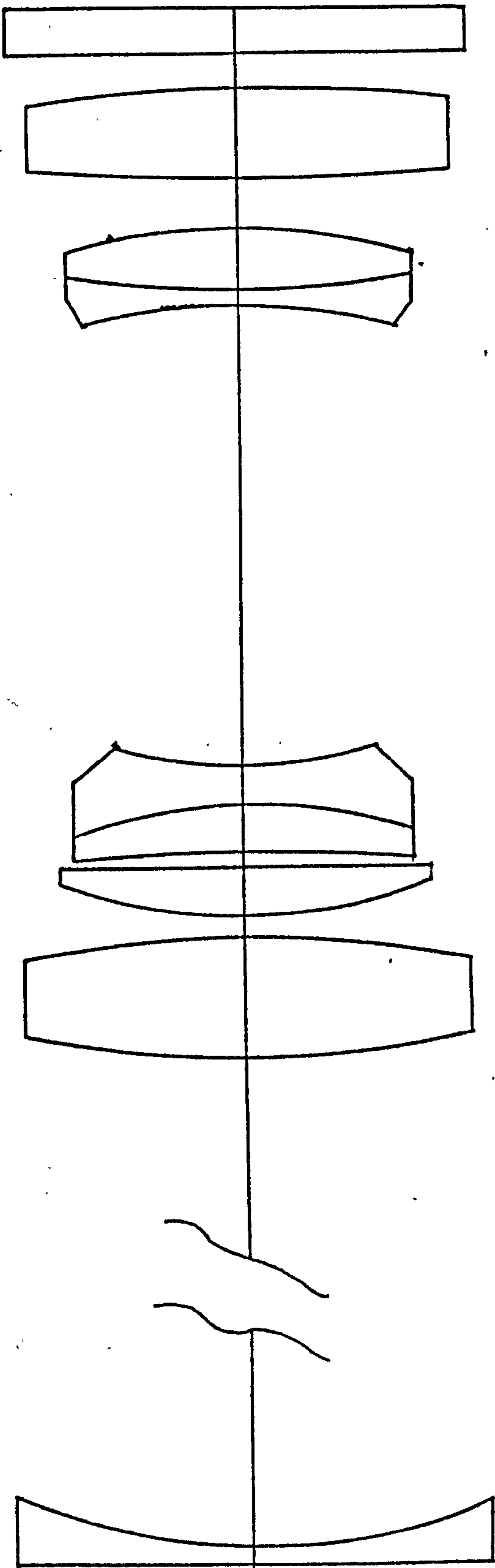


FIG 9.3

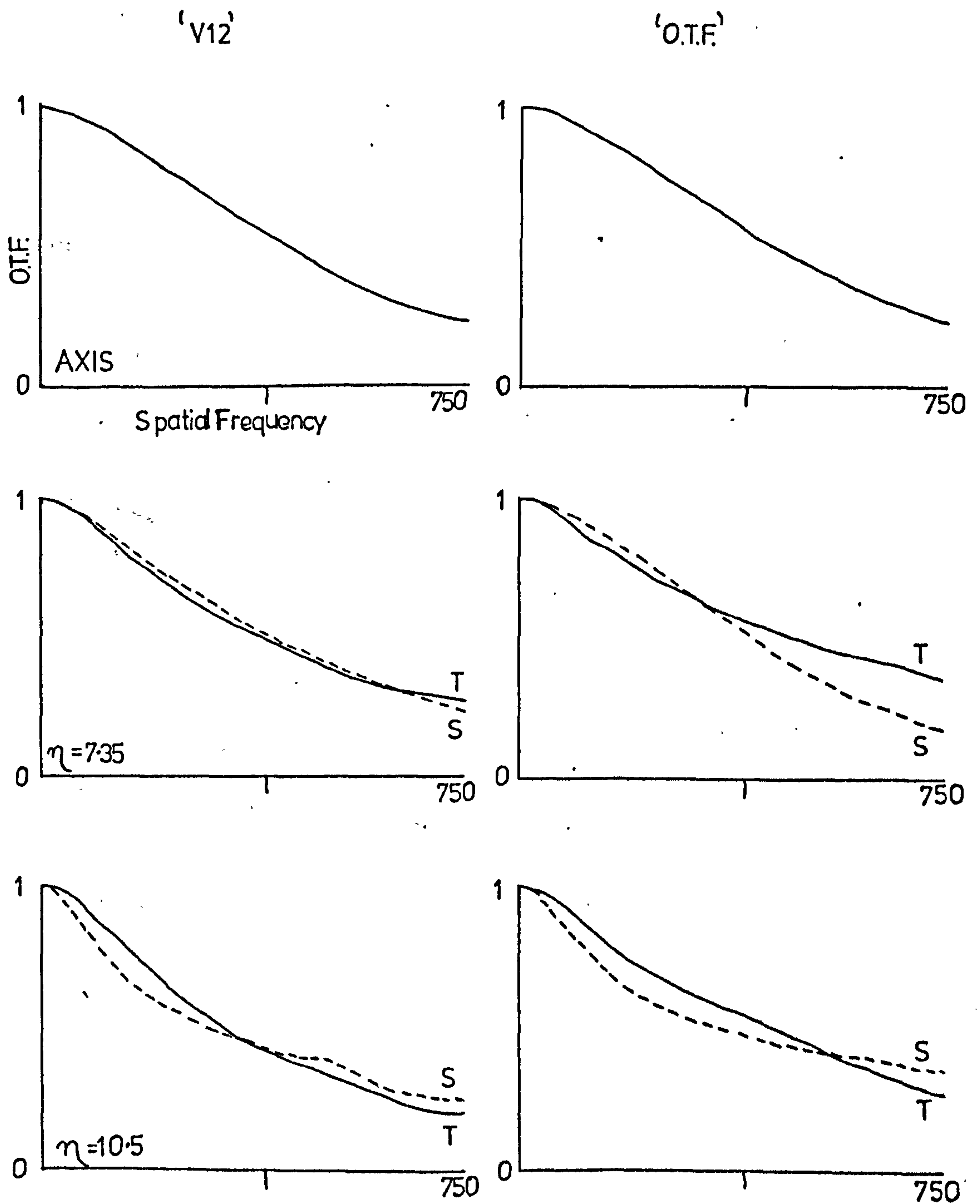


FIG 94

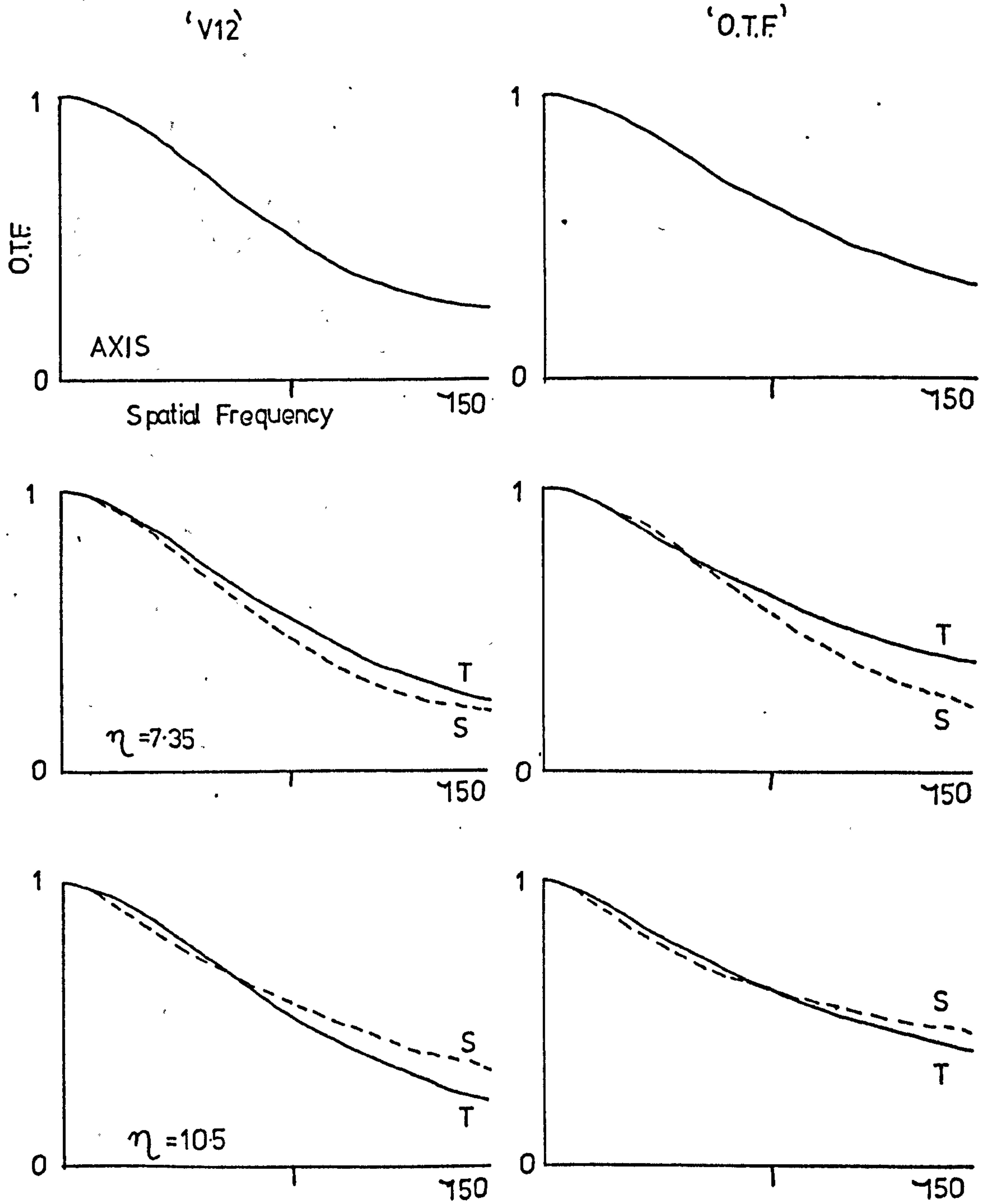
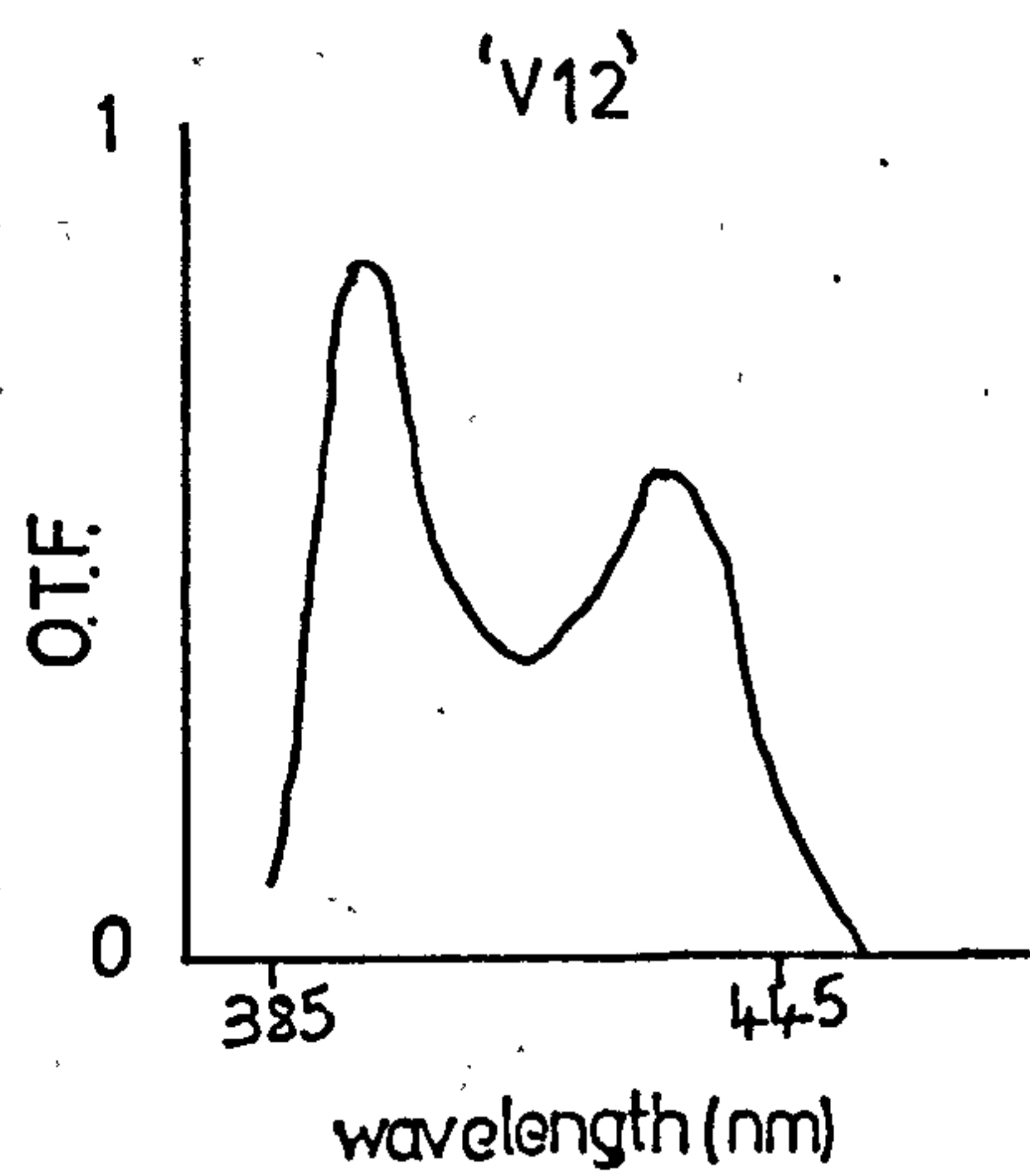
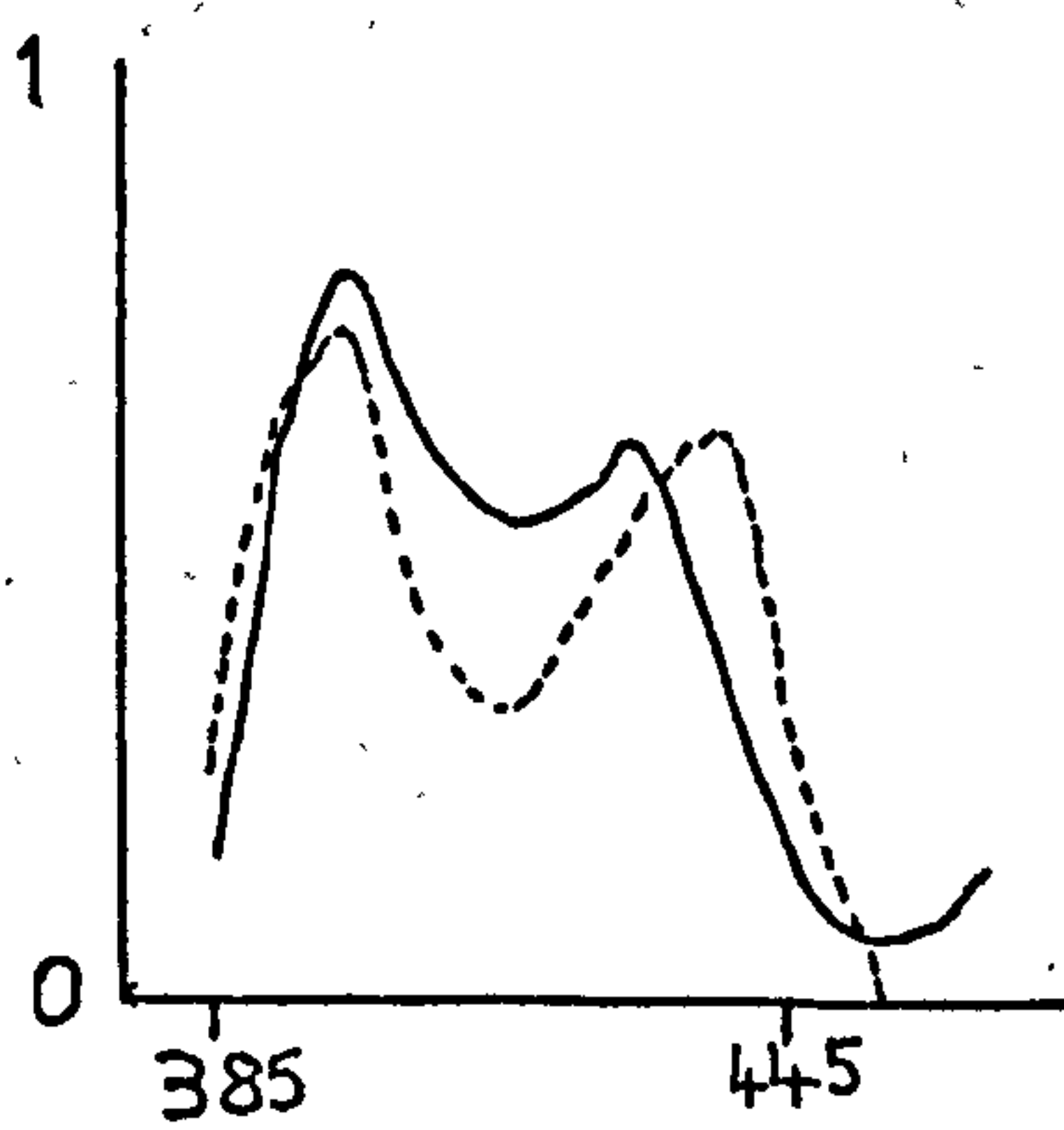
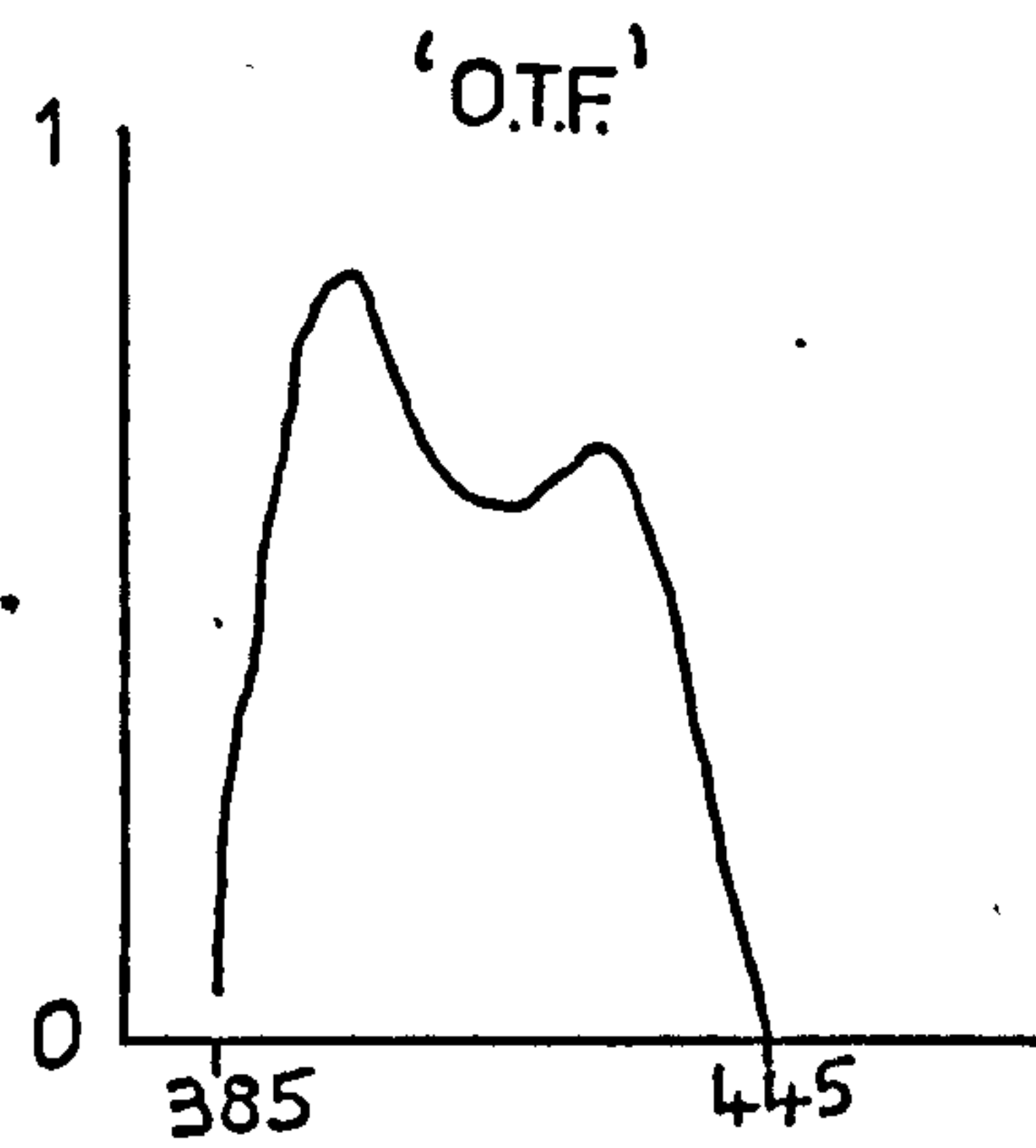


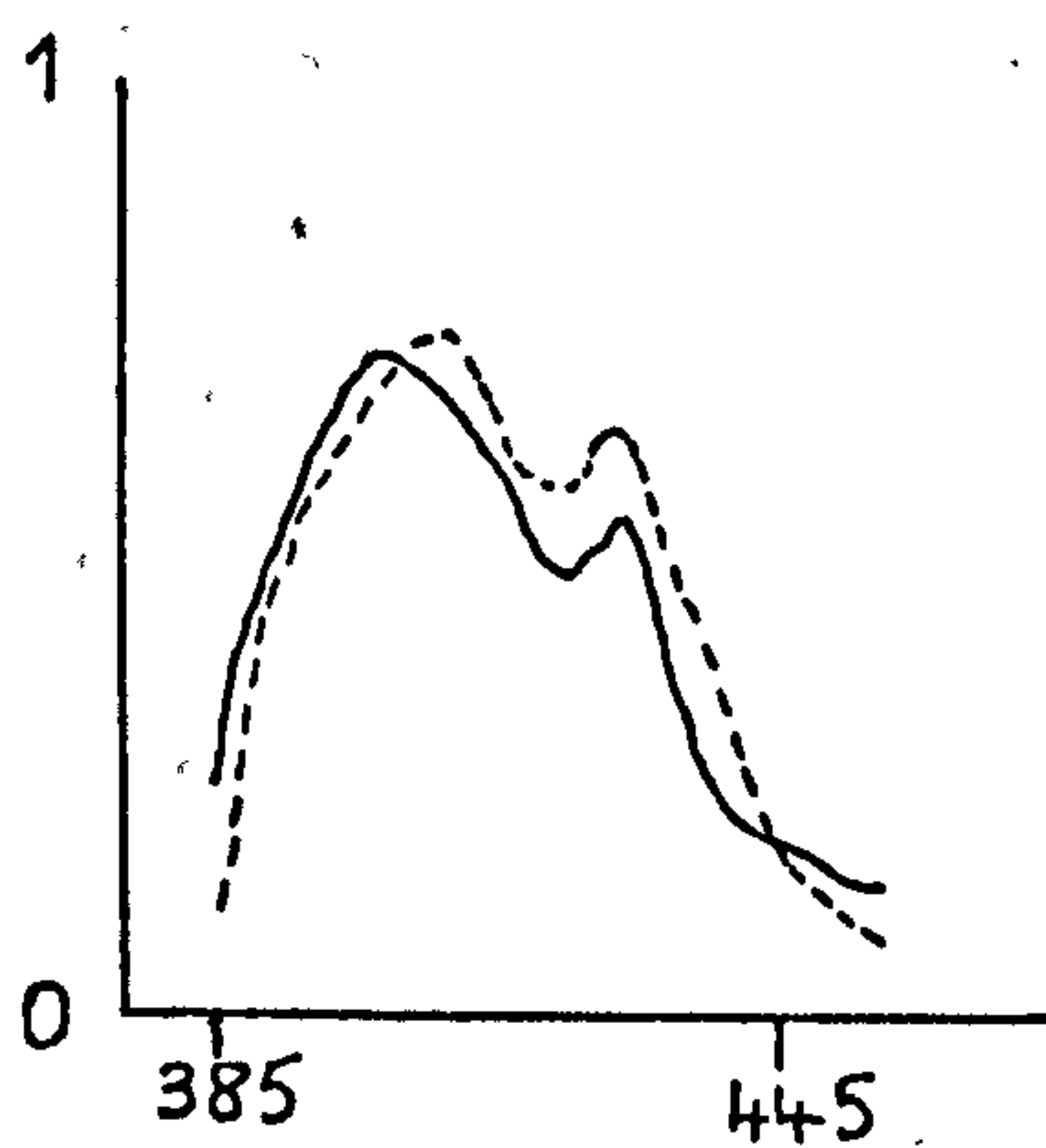
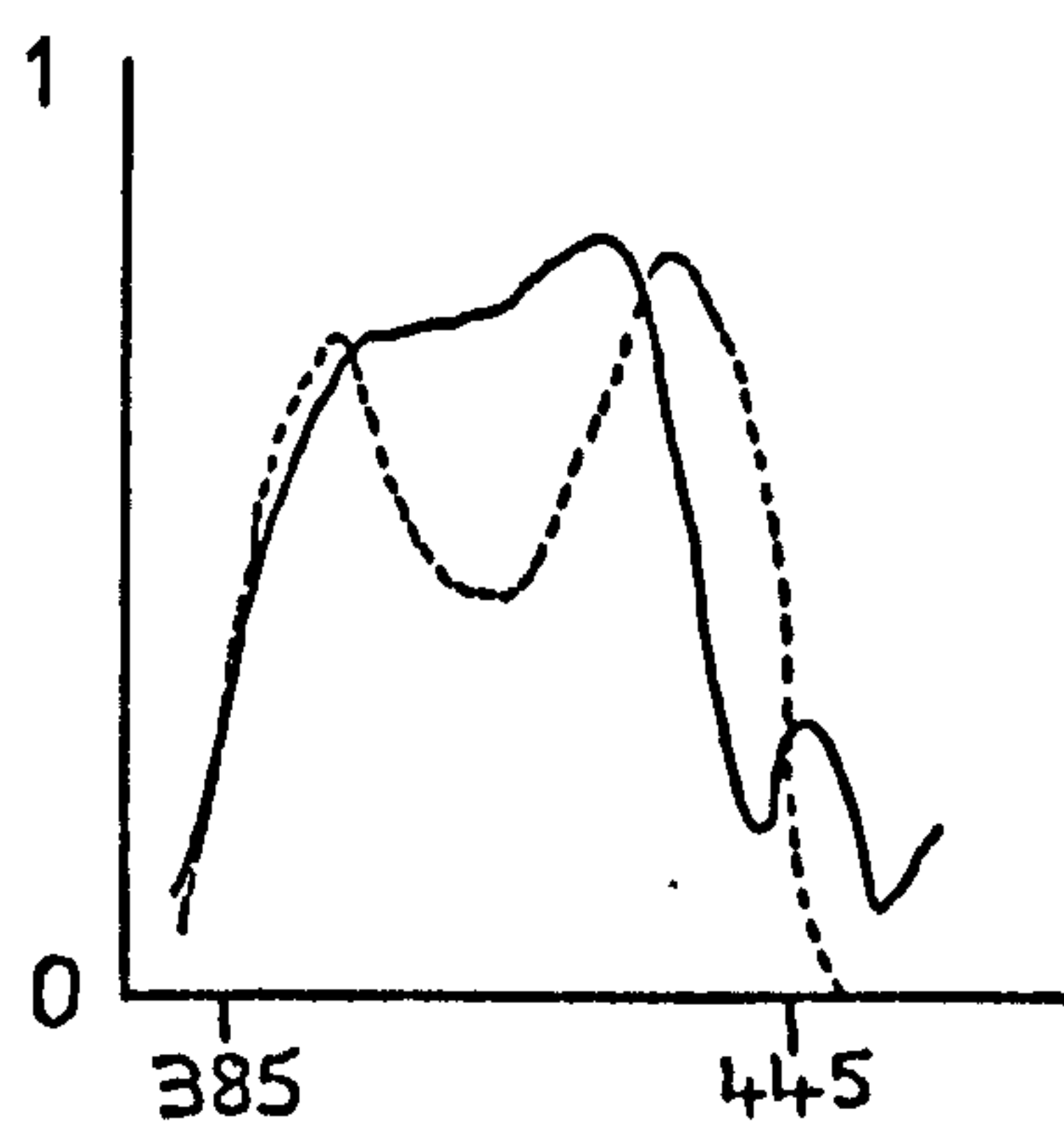
FIG 95



AXIS



$\eta = 7.35$



$\eta = 10.5$

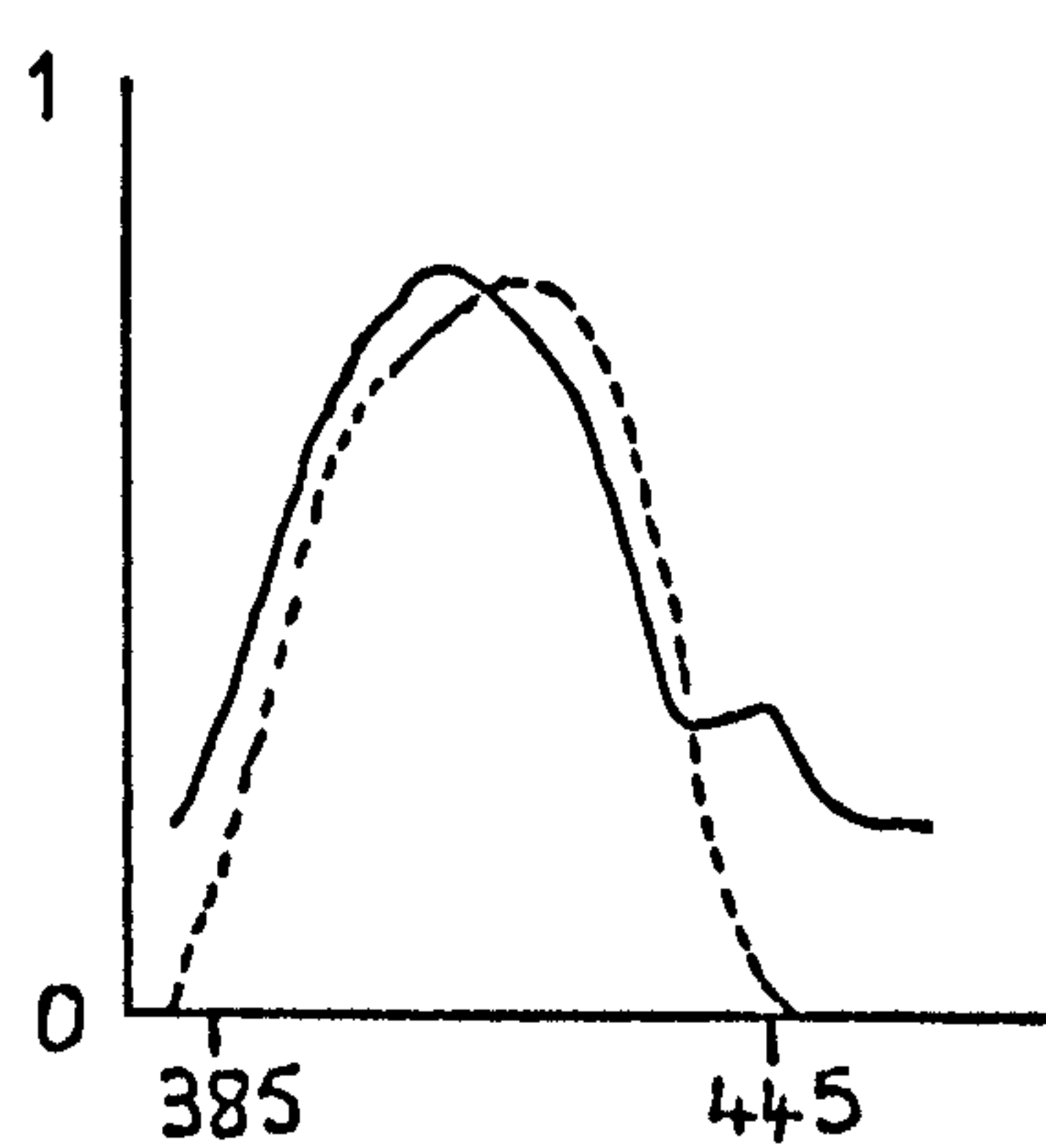


FIG 96

The distortion was reduced by the new programme to 1.5%, which was considered to be acceptably close to the desired figure. A later run was performed to see whether this figure could be reduced still further with a larger distortion weight. This run reduced the distortion to 1.3% without any deterioration of the o.t.f. performance.

2. 3 cm Microfilm Lens.

Since it has been shown in the last section that an improvement in the geometric o.t.f. leads to improvement in the diffraction o.t.f. of a system, only the geometric o.t.f. curves will be given for the next two systems.

This system was designed for microfilm imaging, where high o.t.f. is again required. The system described here was designed for an object at infinity, although the system was later rebalanced for finite conjugates. The system had a focal length of 3 cm, and worked at a numerical aperture of 0.125. Optimum o.t.f. performance was required at 1000 cycles/cm.

The optimisations used three wavelengths, 580, 610 and 550 nanometres, and three field angles of zero, 14 and 20 degrees.

The basic lens was a modified double-Gauss design, with a split rear element giving a 5 component, 7 glass system.

The system was optimised on the conventional programme, and after several trials produced the system shown in figure 9.7. When this system was examined it was seen that many of the transverse aberrations were outside the 0.5 limit for aberration products at 1000 cycles. An intermediate optimisation was carried out at 500 cycles, and this had the desired effect of reducing the aberrations to within the limit.

The o.t.f. optimisation programme was run at 1000 cycles,

and this produced the system shown in figure 9.8. As with the Padua lens, the programme has made only minor modifications to the system, but as before the increase in performance is clear.

The geometric o.t.f. curves for the two systems are shown in figures 9.9 and 9.10. As well as the o.t.f. in the plane of focus chosen by the programmes, through-focus o.t.f. curves at 1000 cycles/cm are given. For the o.t.f.-optimised system the chosen focal plane is obviously the optimum plane both for the average o.t.f. and for the optimum worst o.t.f. For the version 12 system the o.t.f. curves do not all peak at the chosen focal plane, and so the system may possibly be improved by further defocussing. The o.t.f. figures at 1000 cycles in the chosen focal planes are given below.

<u>Geometric o.t.f. at 1000 cycles/cm</u>				
	Version 12		'OTF'	
	sag	tan	sag	tan
Axis	.438		.612	
14°	.580	.495	.614	.662
20°	.327	.621	.518	.531
Average	.492		.587	

As with the Padua lens it is the worst performance figure over the field which is important. For the o.t.f.-balanced system the worst figure, from the values above, is 52%. The worst figure for the version 12 system is 32% at the chosen focal plane, but it can be seen from the through-focus curves that this figure can be improved to 36% by a defocus of .001cm. Even after this defocus the new system is 16% better, which is an unexpectedly high improvement.

Since there is such a large improvement in the o.t.f. performance it is instructive to consider the aberration curves for both systems.

The wavefront and transverse aberrations for the two systems in all three wavelengths are given in figures 9.11 to 9.14. The longitudinal aberrations for both systems in the mean wavelength, 580 nm, are given in figure 9.15.

The spatial frequency, 1000 cycles/cm, corresponds to a reduced spatial frequency 0.464, which is a relatively high spatial frequency, according to the previous theory. Naturally the aberration balancing in this practical situation is more complex than the simple models considered earlier, and the wavefront aberrations cannot be easily interpreted. The transverse aberration curves show that both systems have good chromatic balances, but that except for the far-field case the aberrations in the o.t.f.-balanced system are smaller than those from the conventional programme. This is due to the intermediate optimisation, which reduced the transverse aberrations so that the aberration products were sufficiently small for the high frequency optimisation. For the far-field case, the aberrations at the edge of the pupil are above the limit of 5 microns for aberration products of 0.5 at 1000 cycles/cm. However, it can be seen that the shape of these large aberration curves corresponds to the transverse aberrations which would be given by the high-frequency, high aberration balanced wavefronts shown in the first part of this thesis.

The longitudinal aberration curves provide more obvious signs of the difference between the two optimisations. The most noticeable effect is due to the lack of astigmatism control

MICROFILM LENS 'V12'

0.2 cm

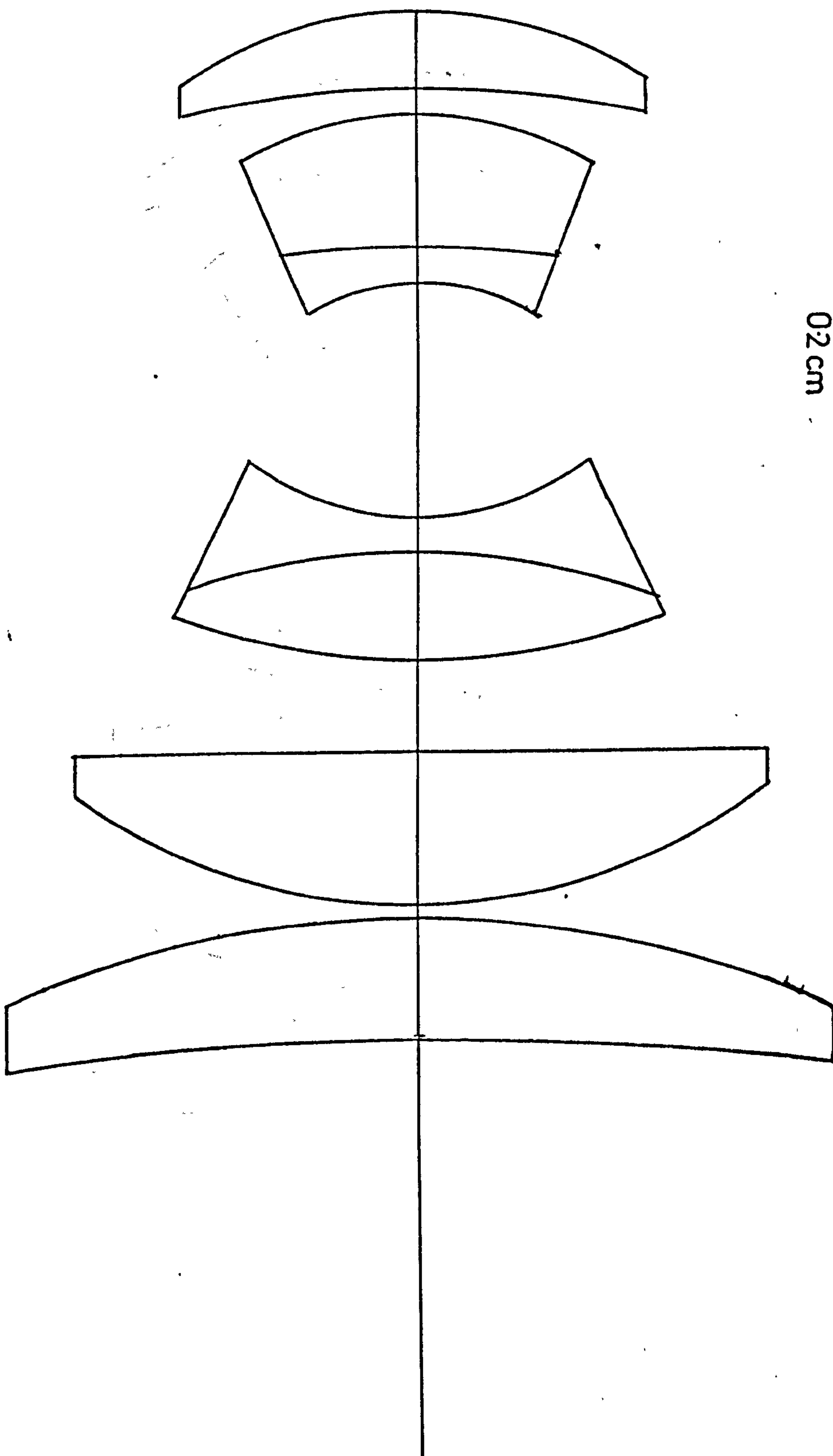
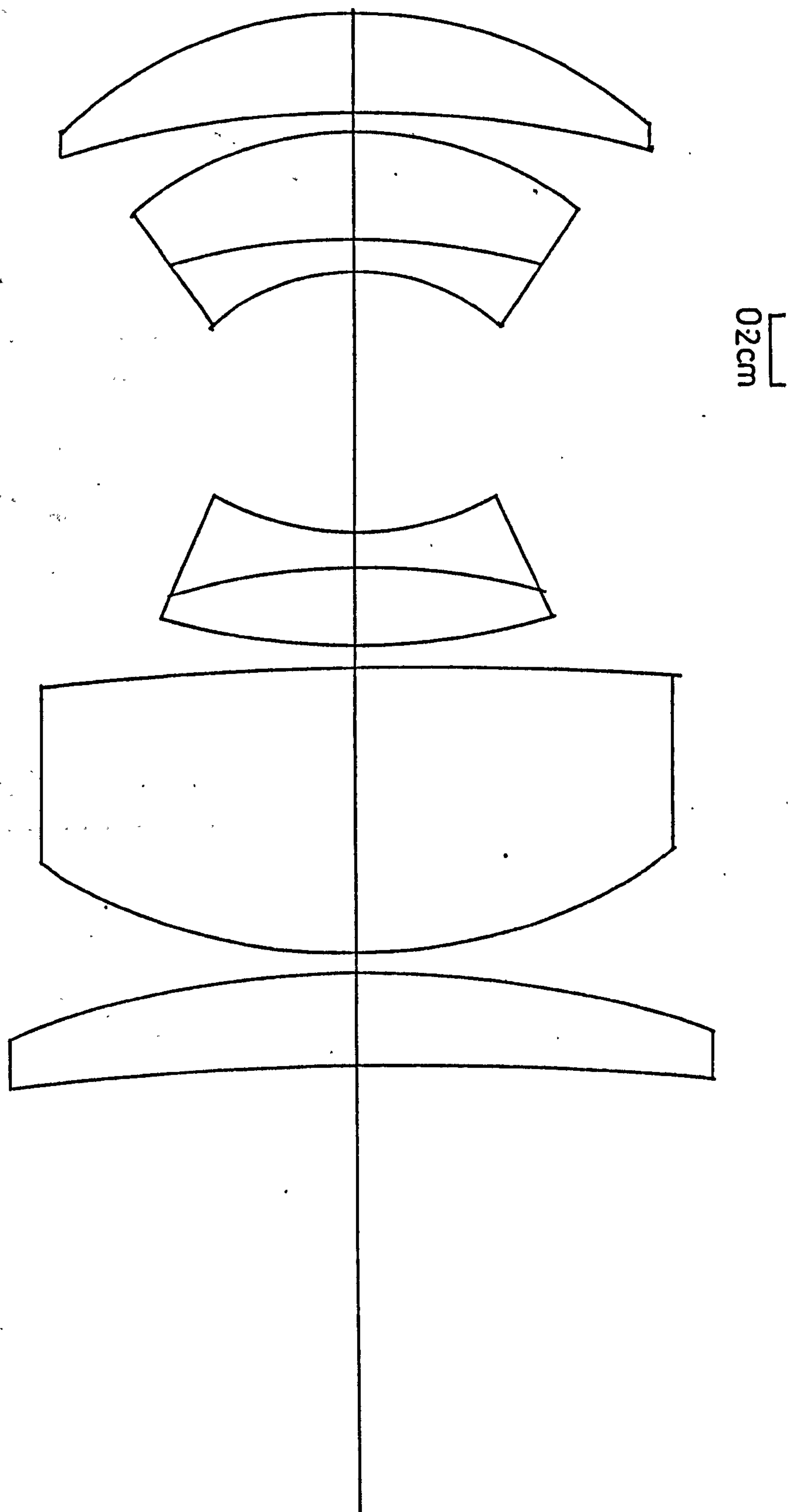


FIG 9.7



MICROFILM LENS

'O.T.F.'

FIG 9.8

O.T.F.
MICROFILM LENS 'V12'

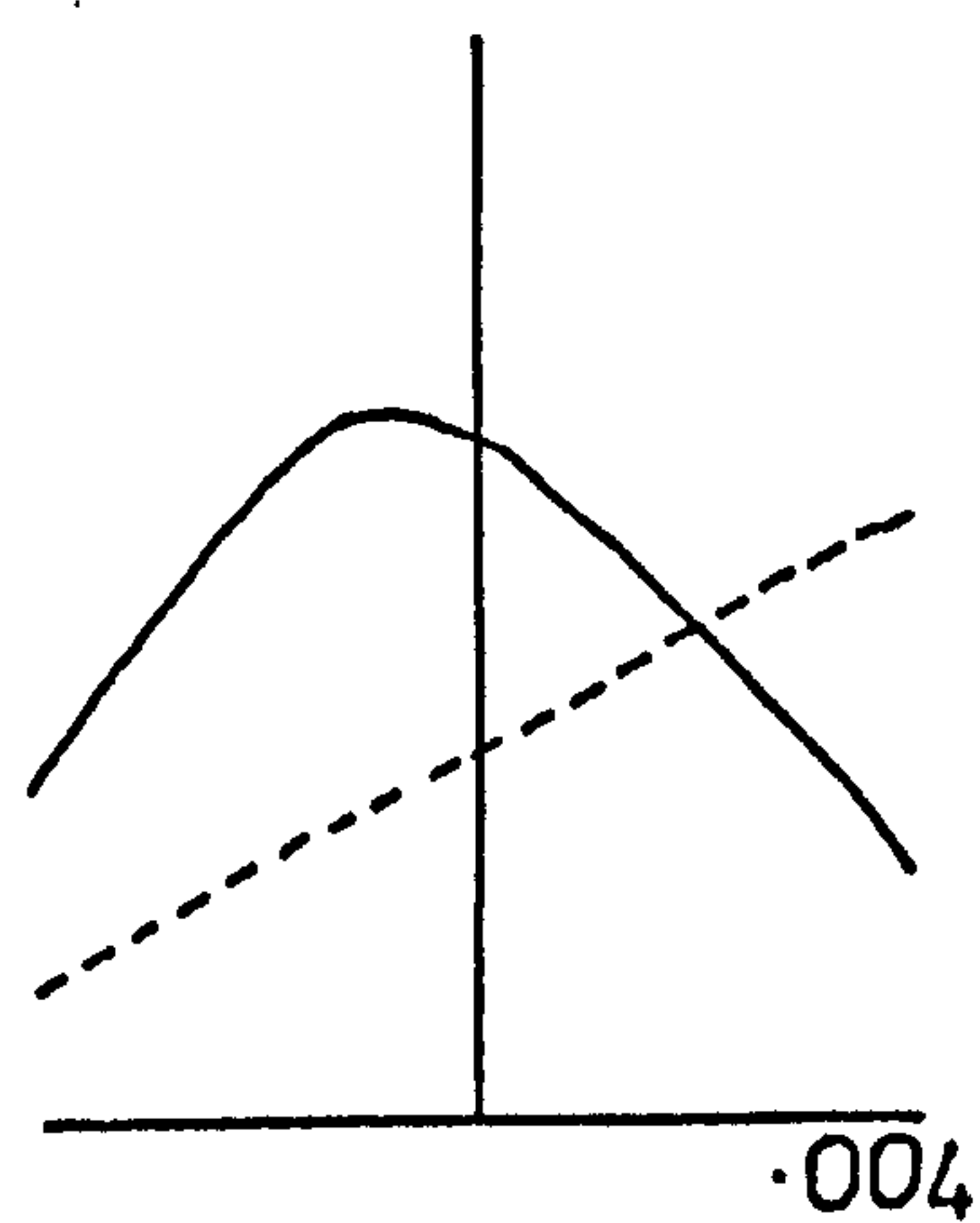
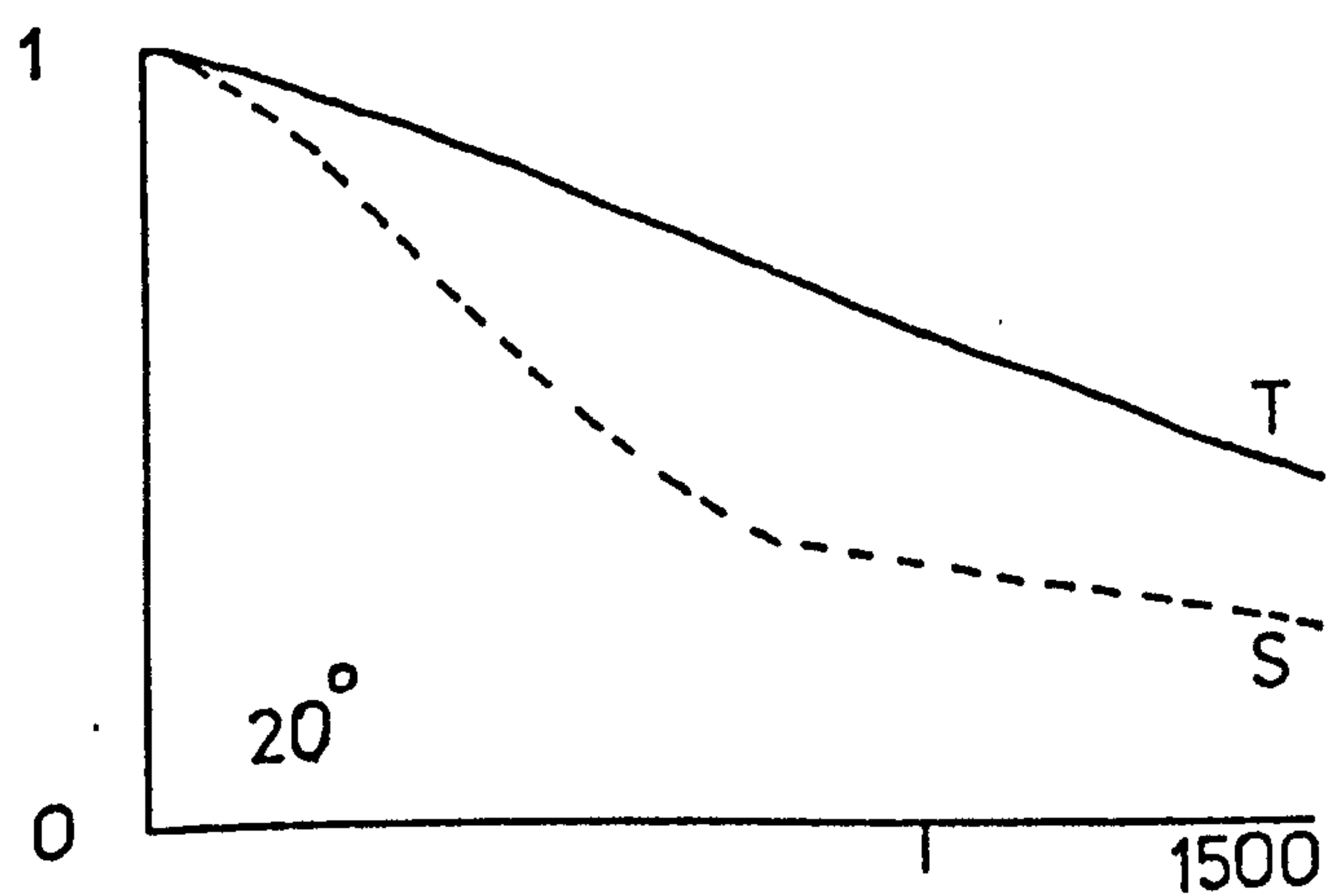
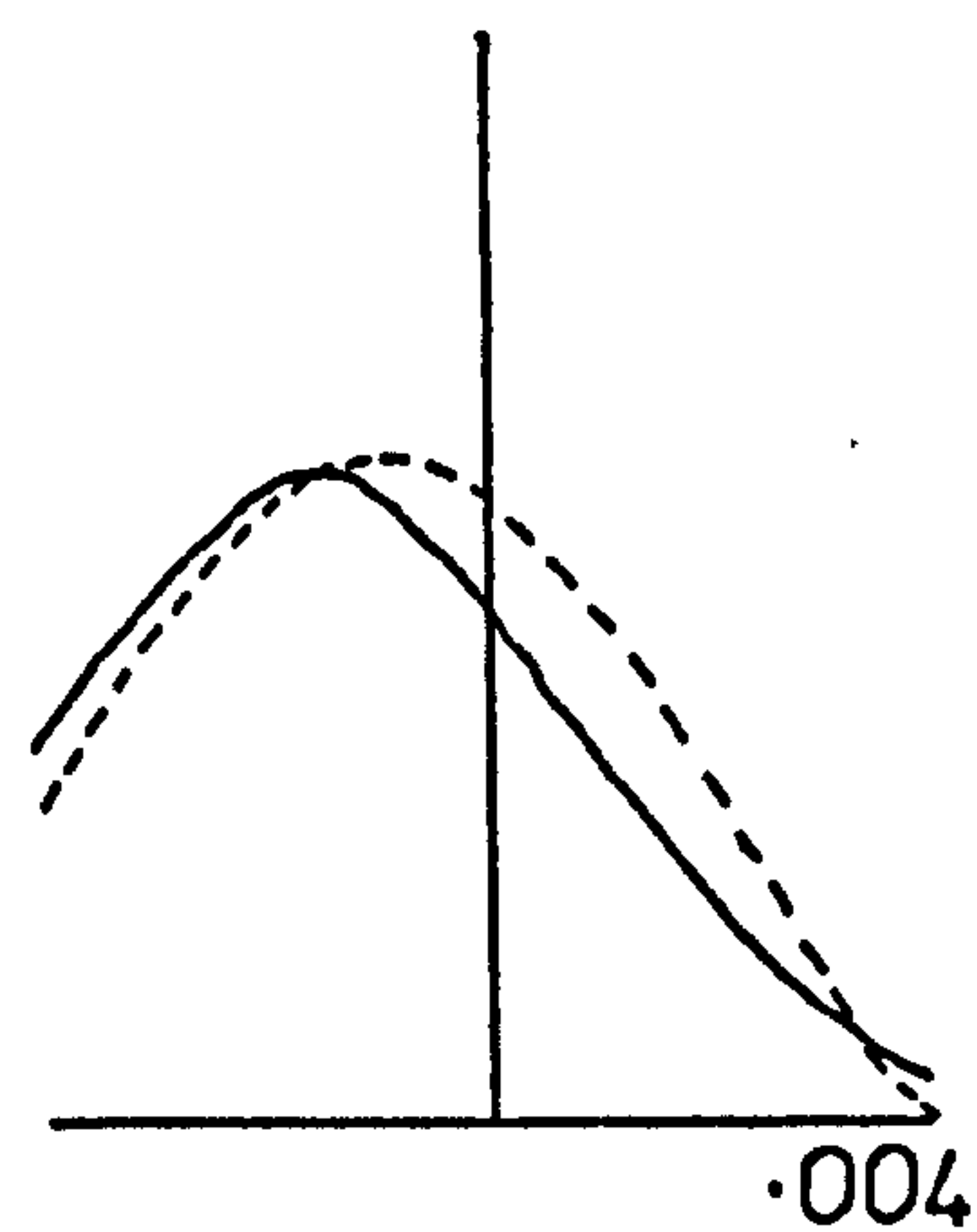
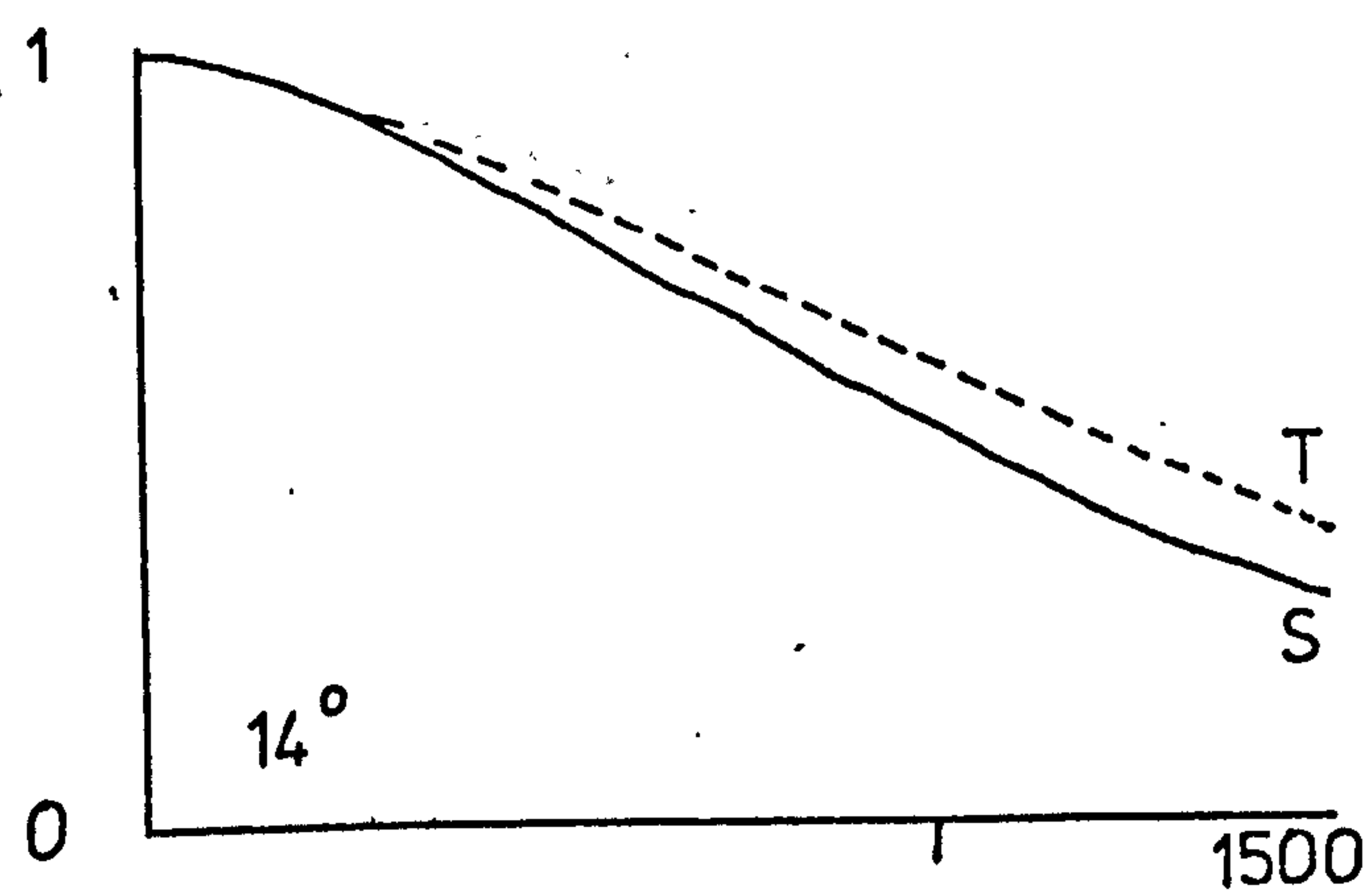
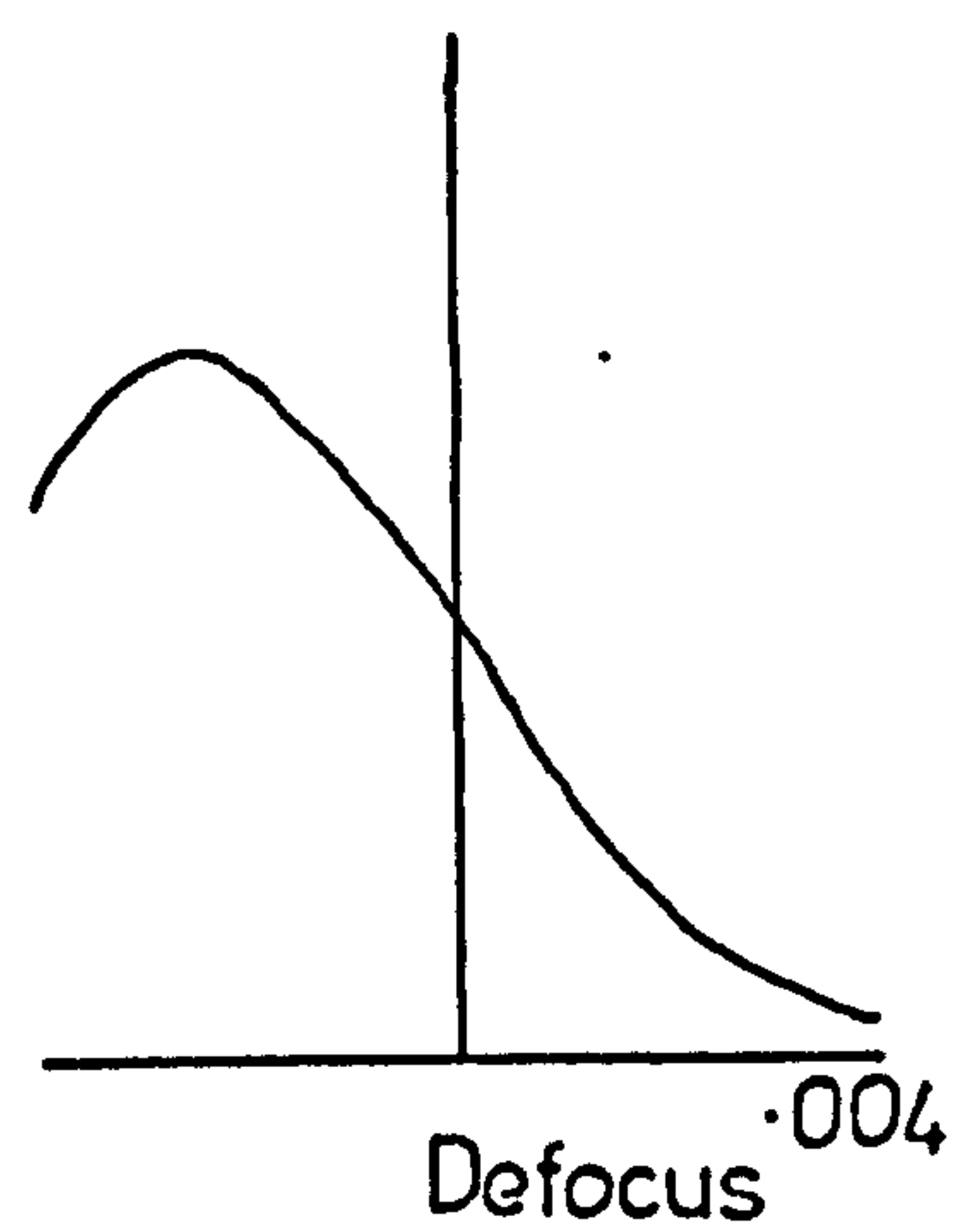
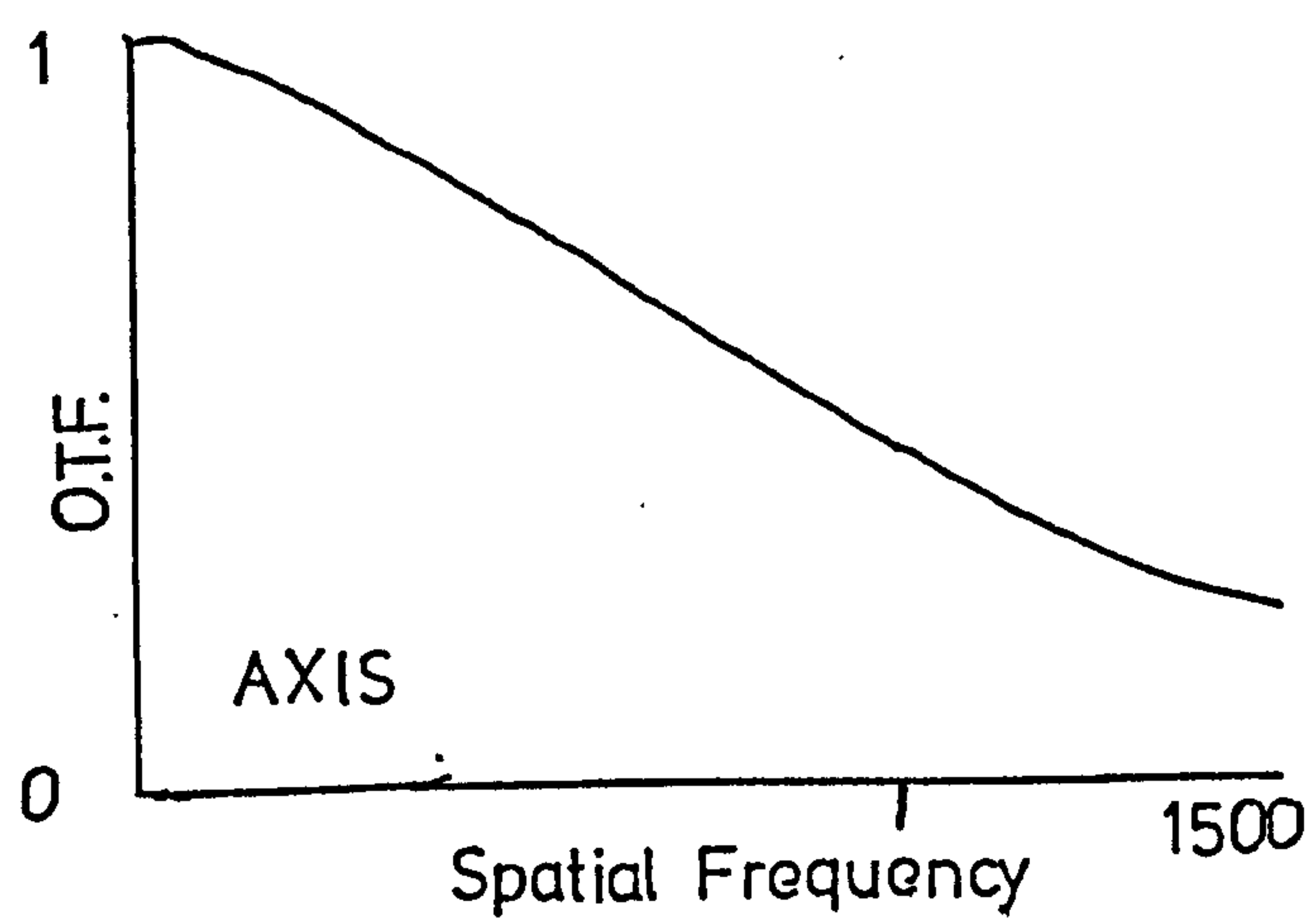


FIG 9.9

O.T.F.

9.18.

MICROFILM LENS

'O.T.F.'

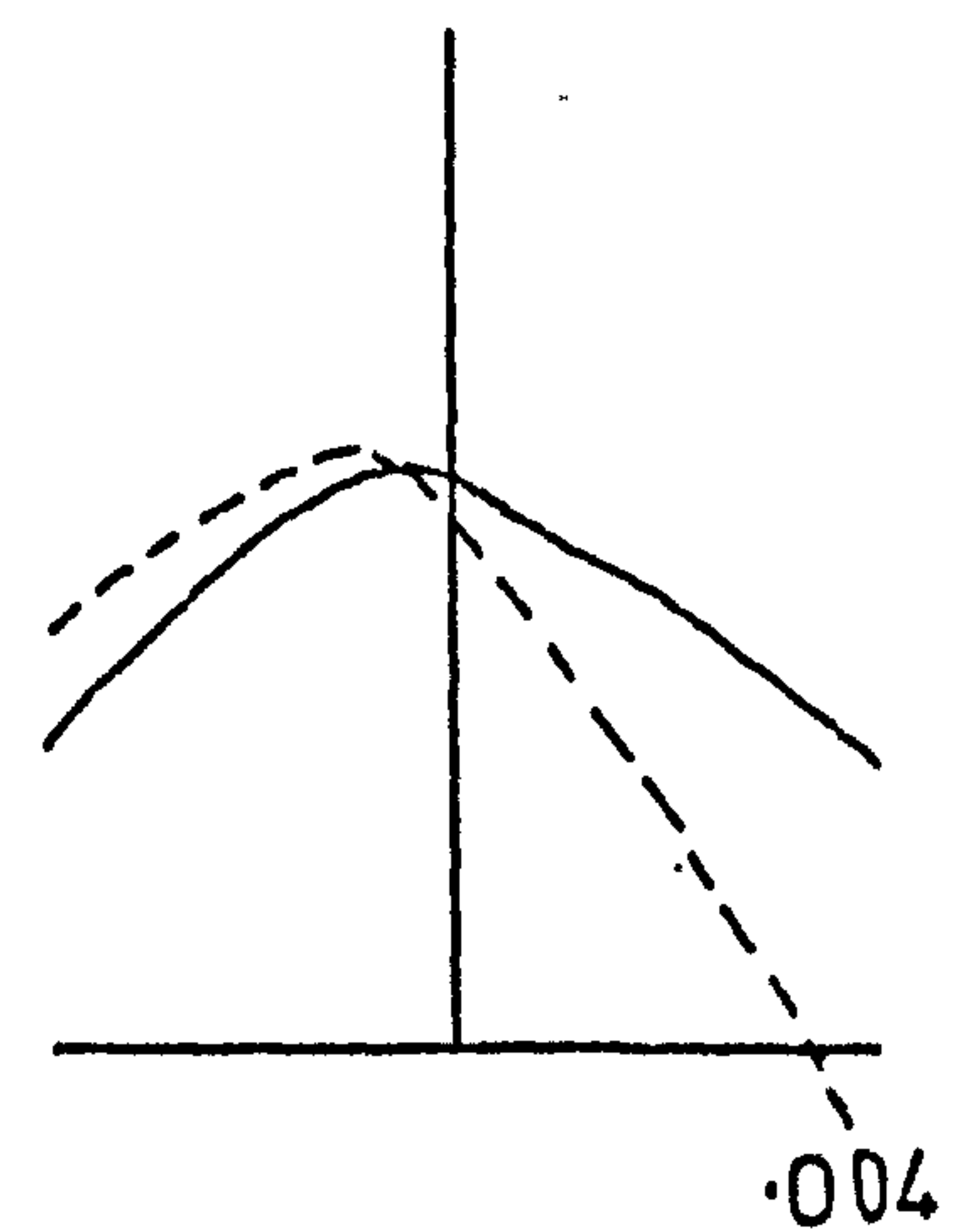
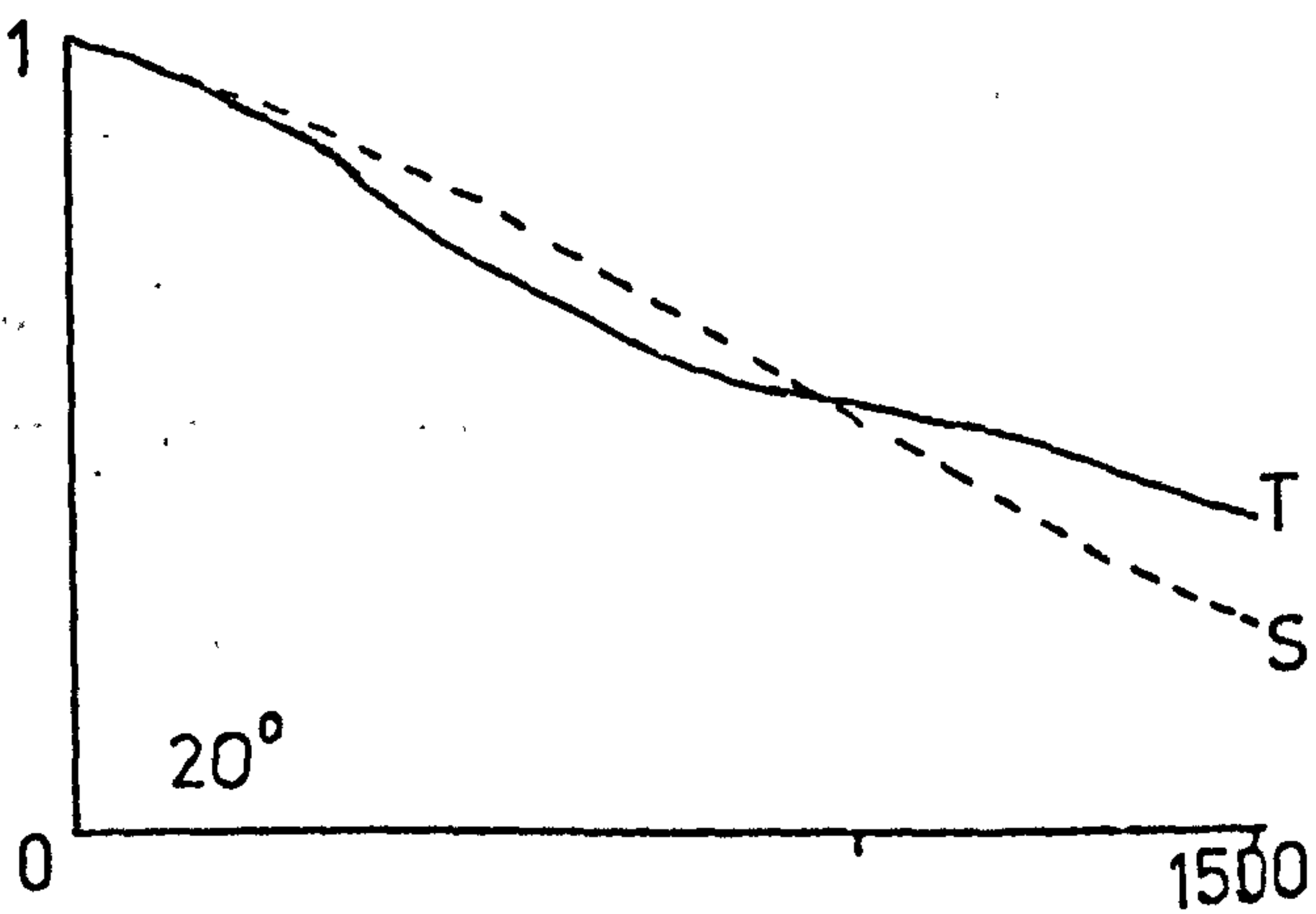
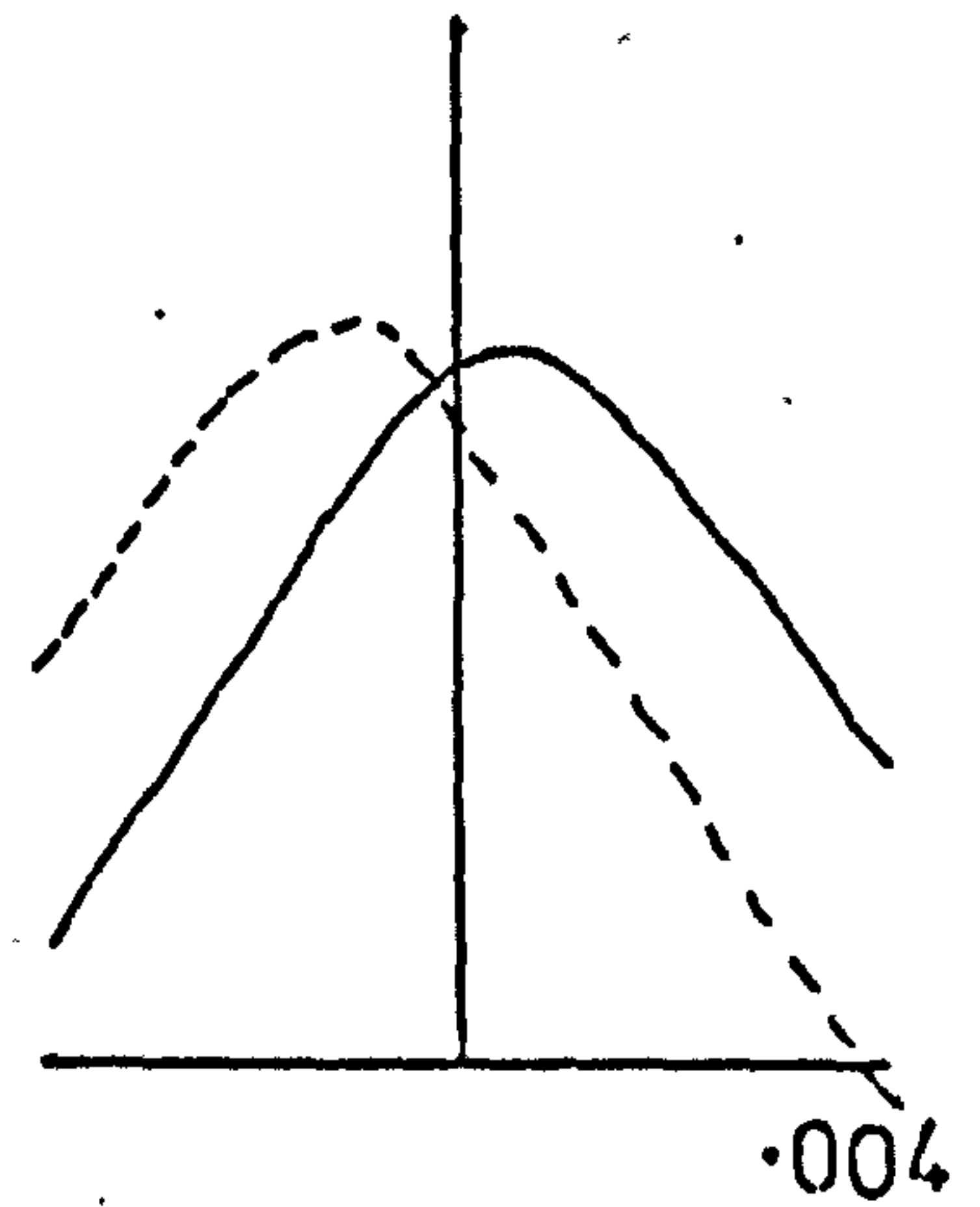
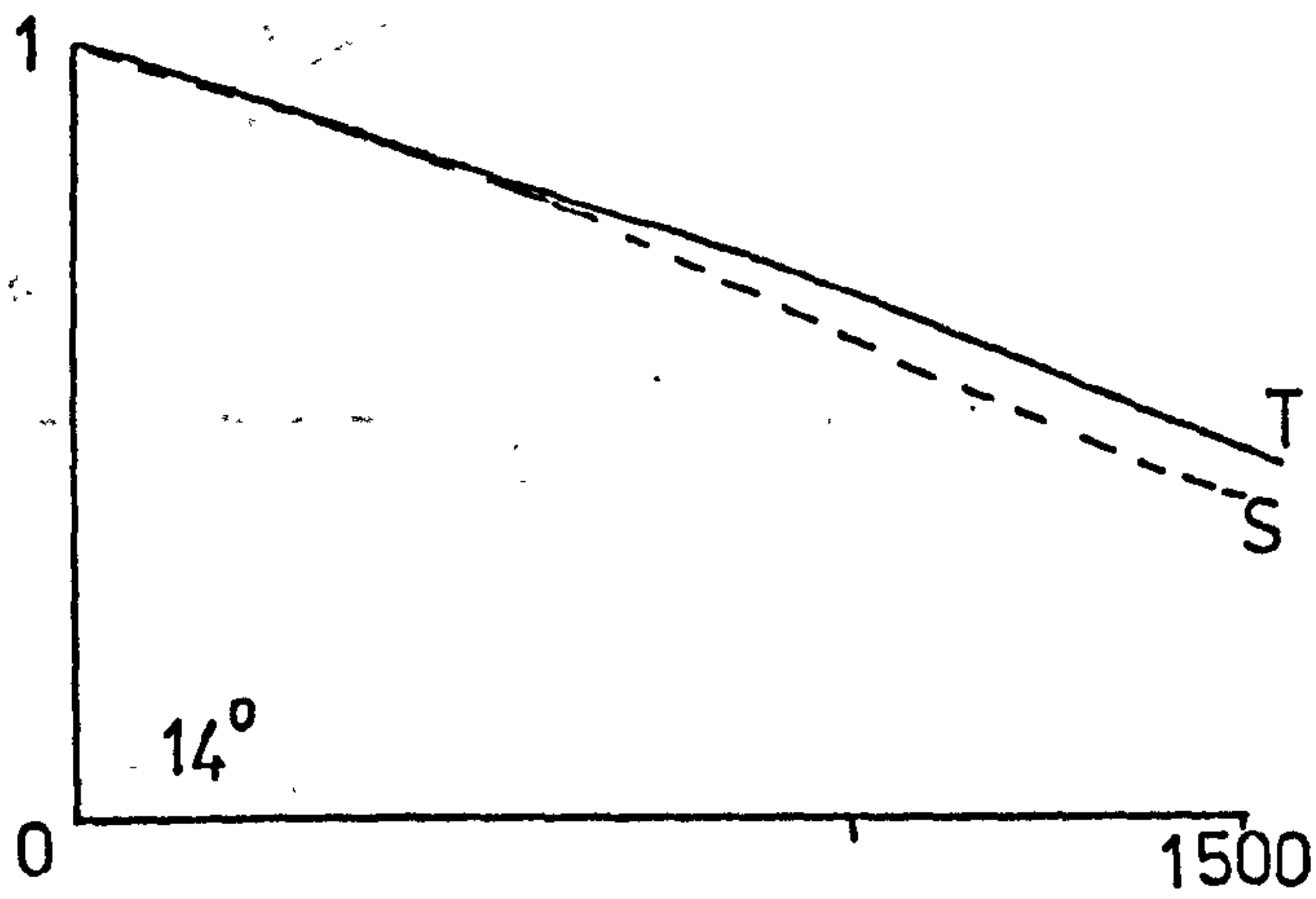
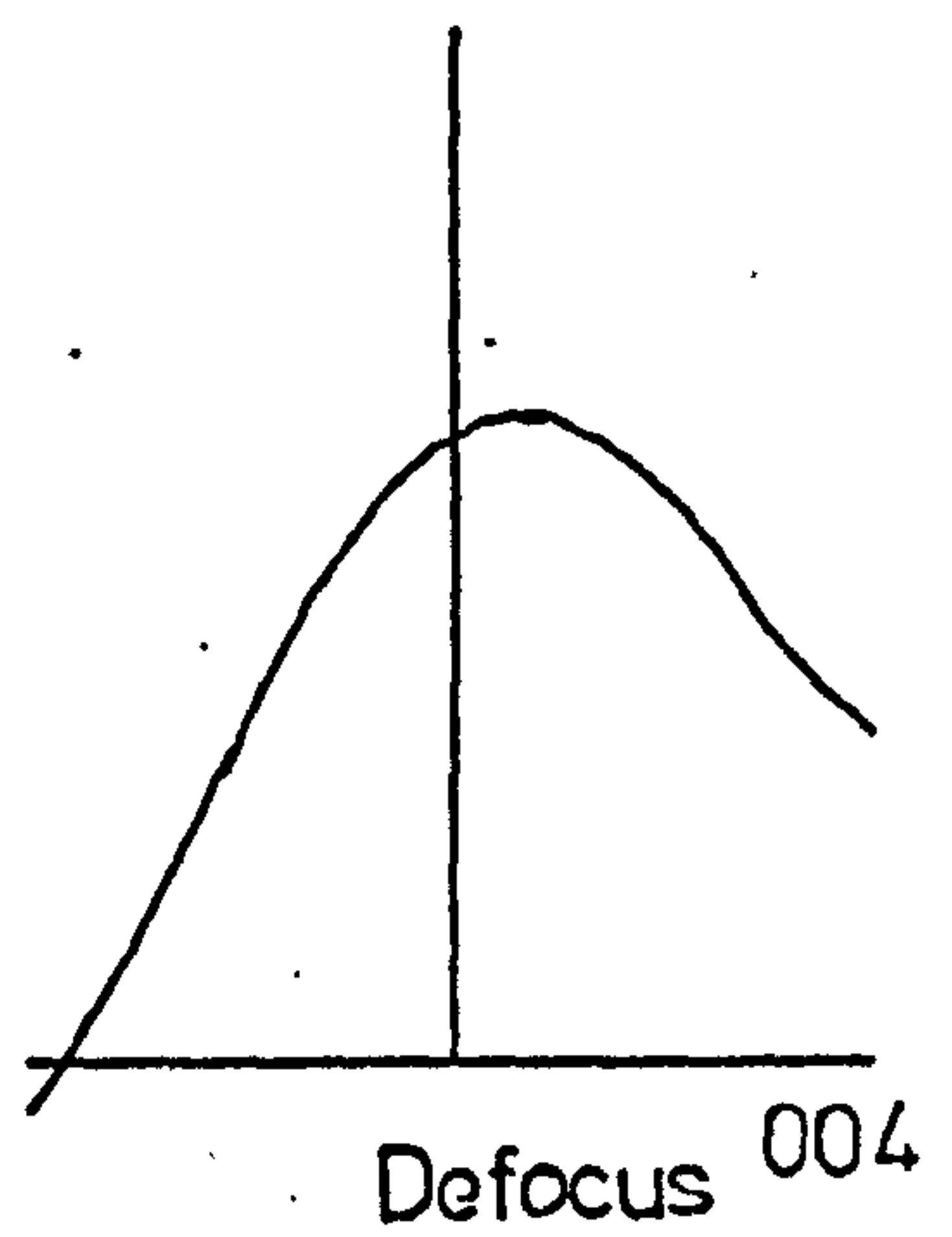
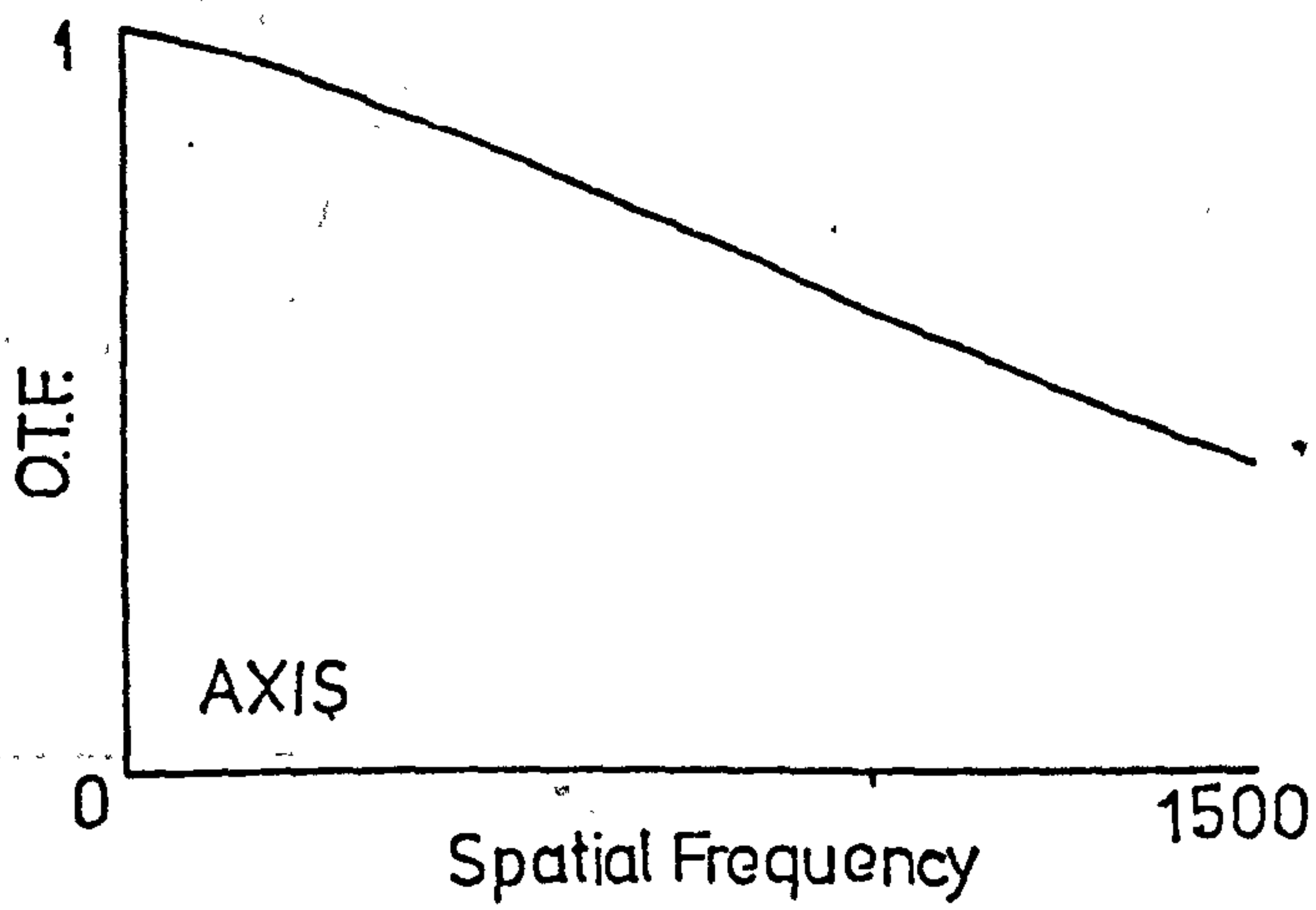


FIG 9.10

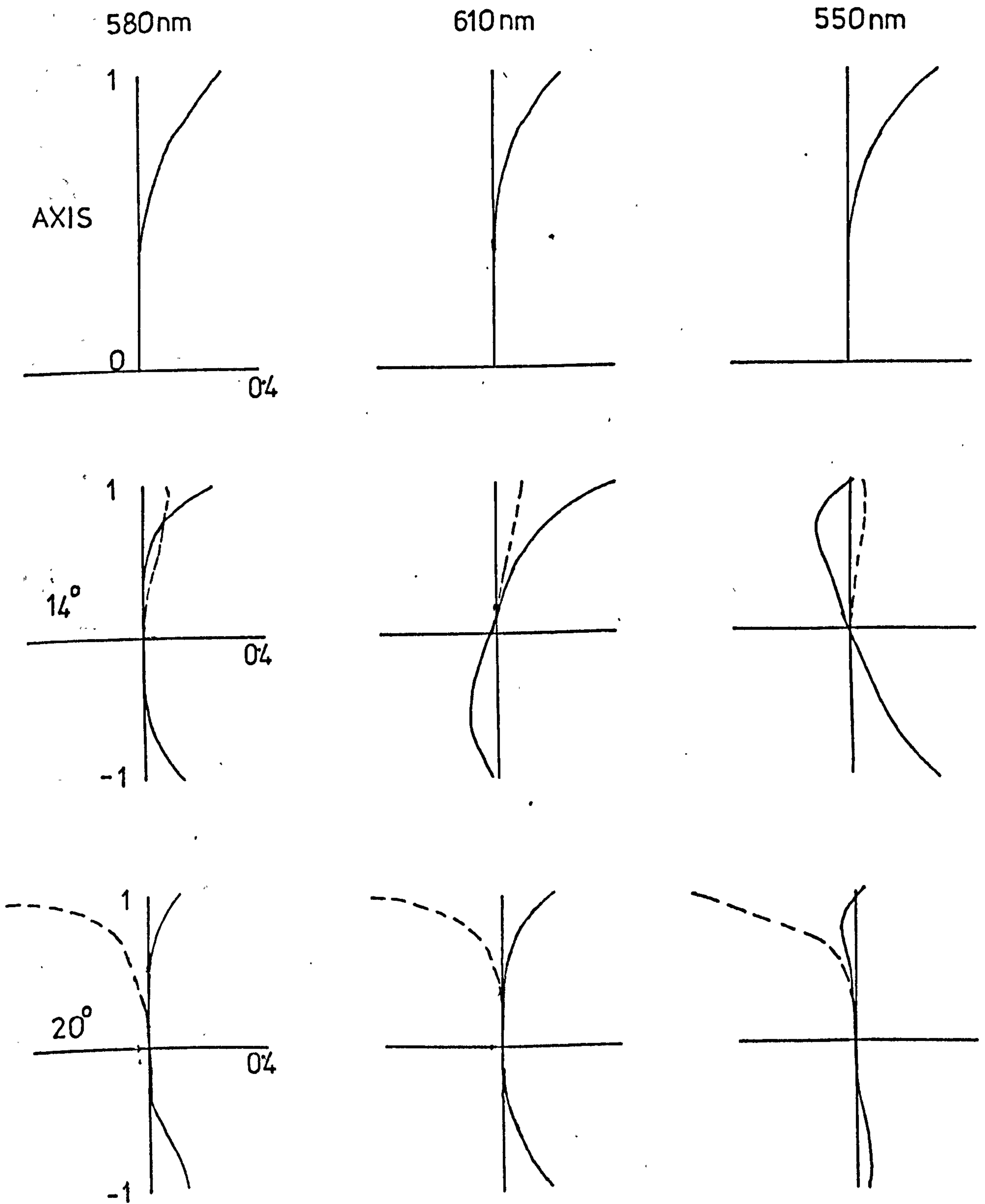


FIG 9.11

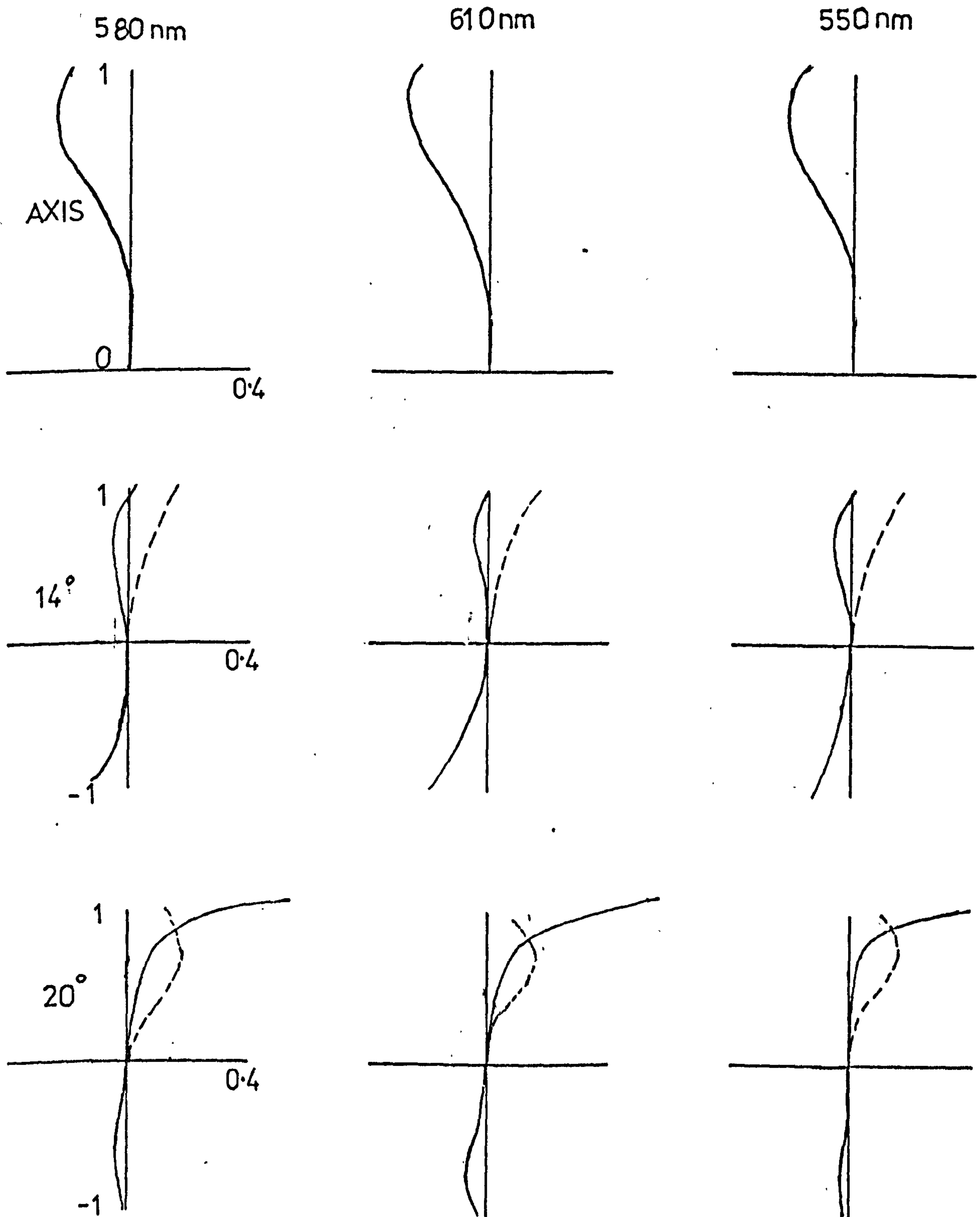


FIG 9.12

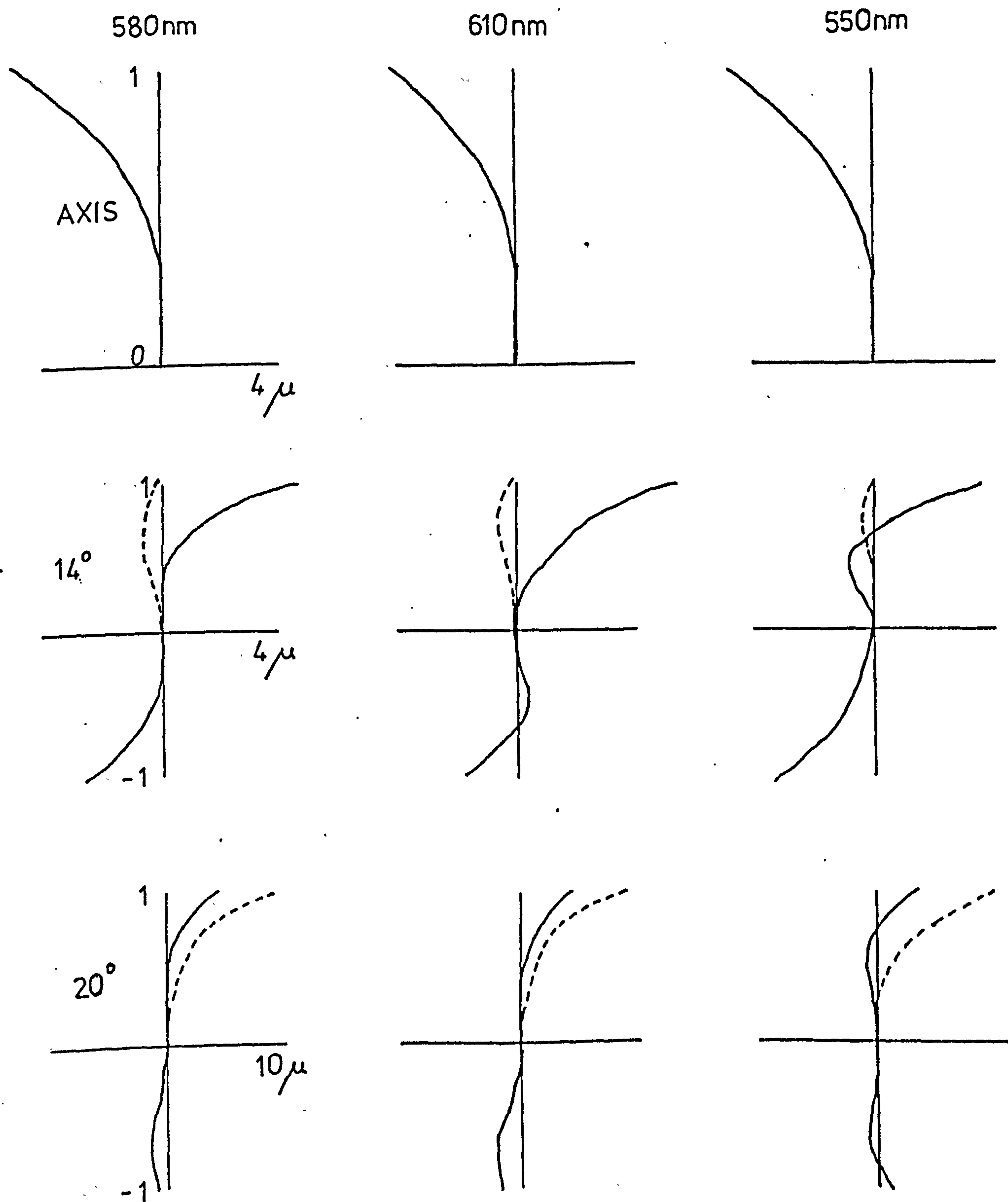


FIG 9.13

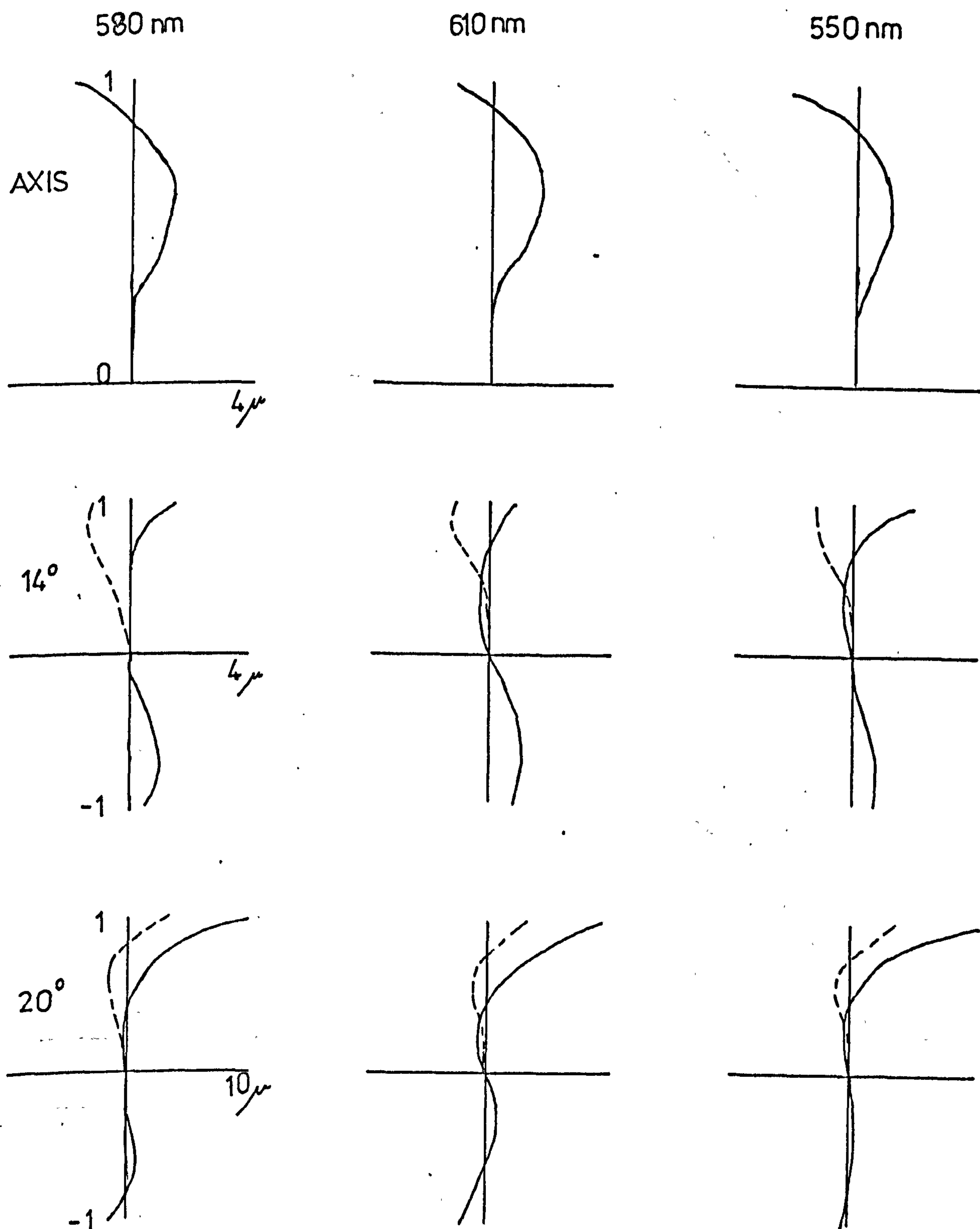
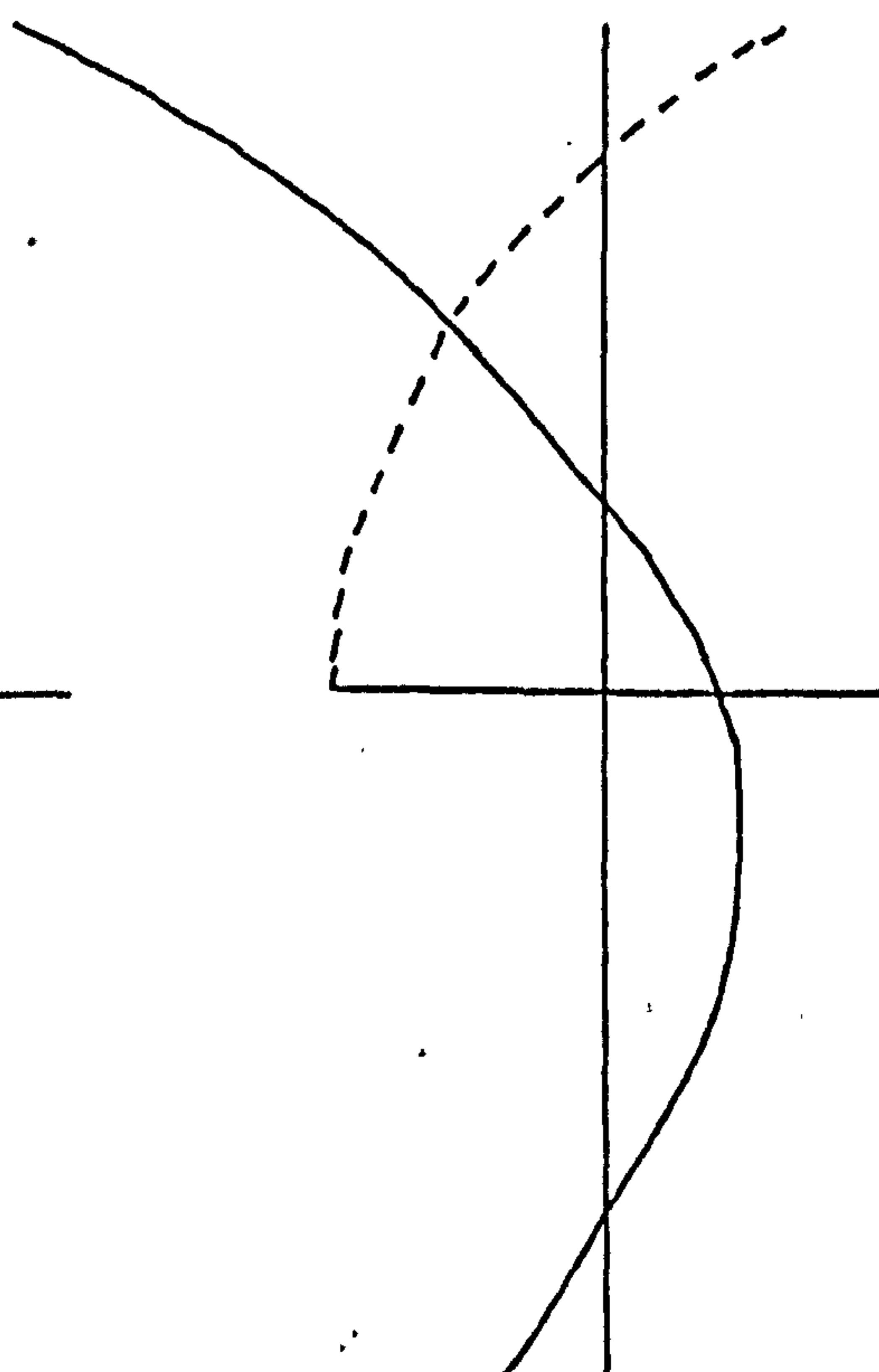
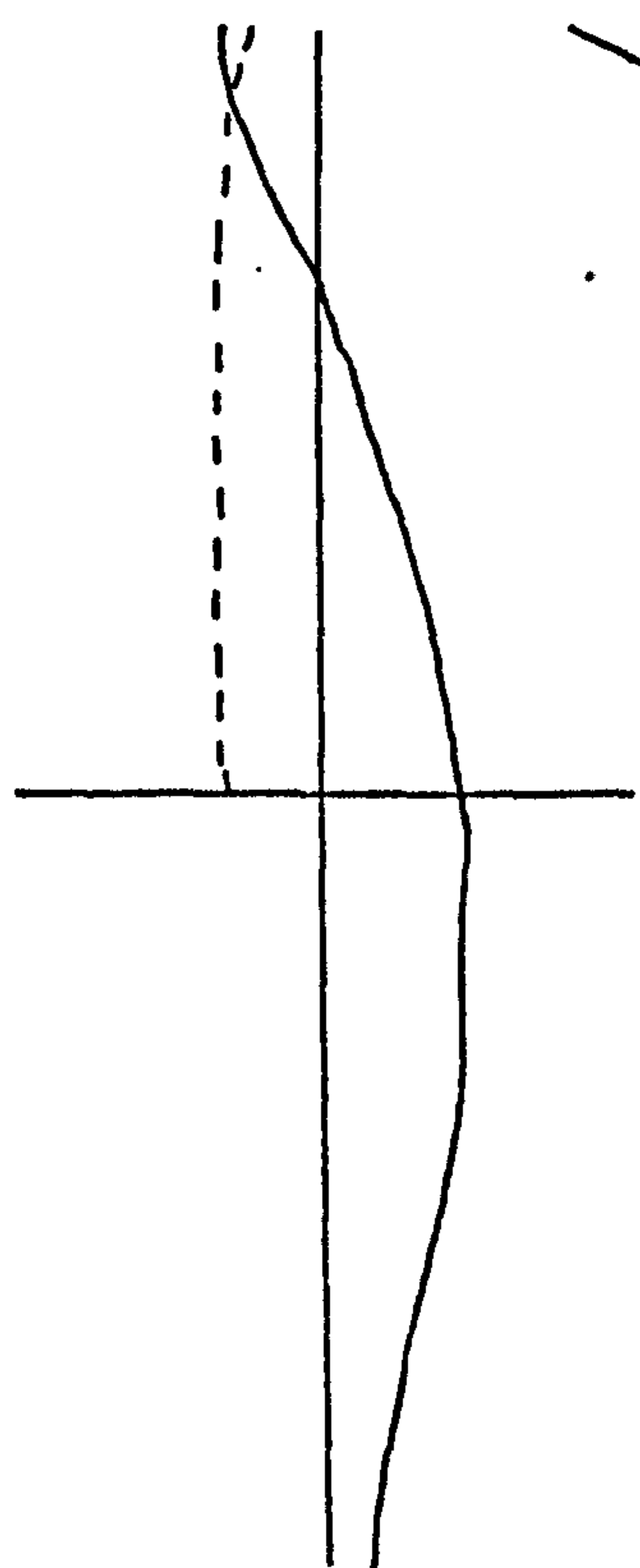
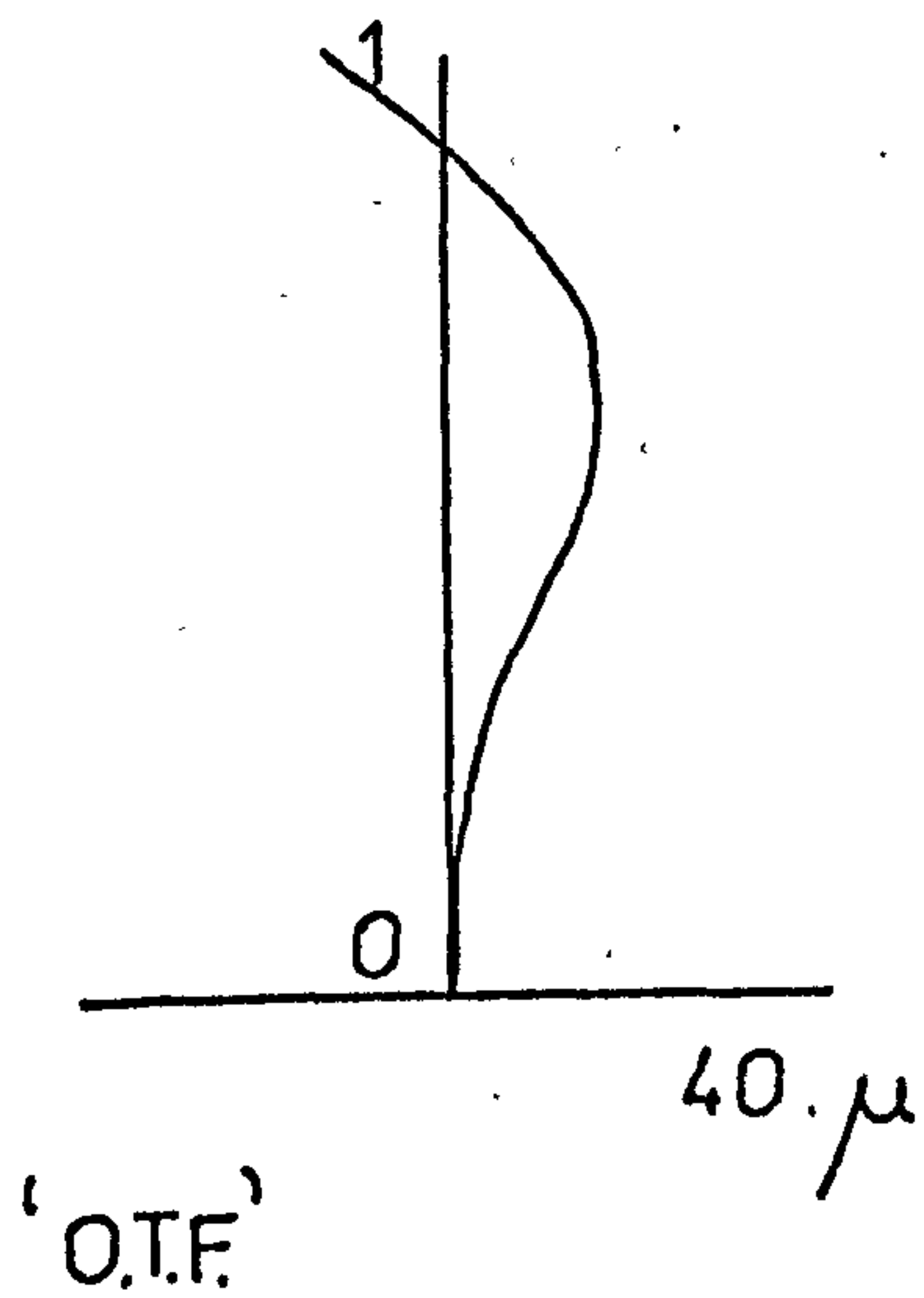
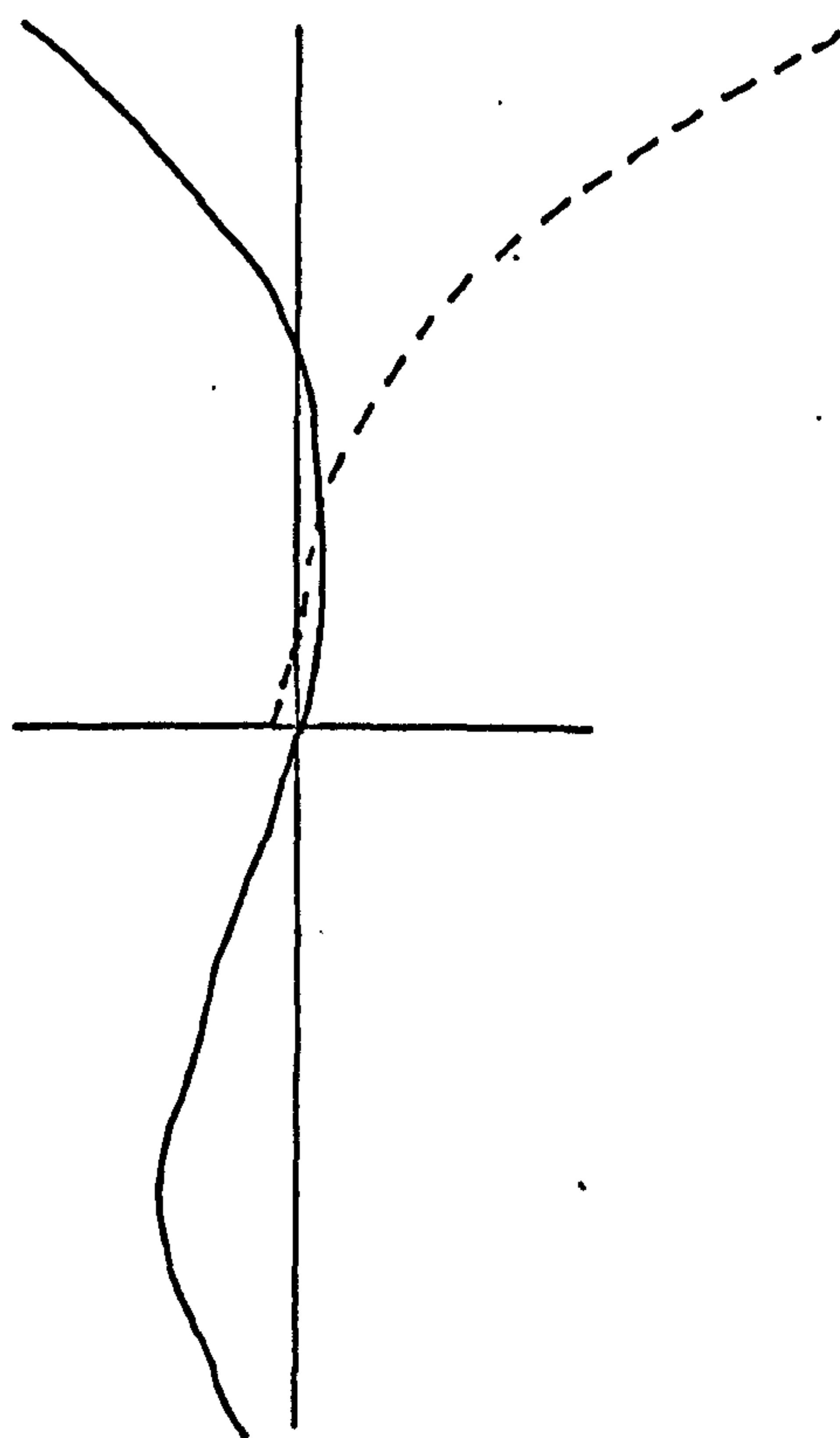
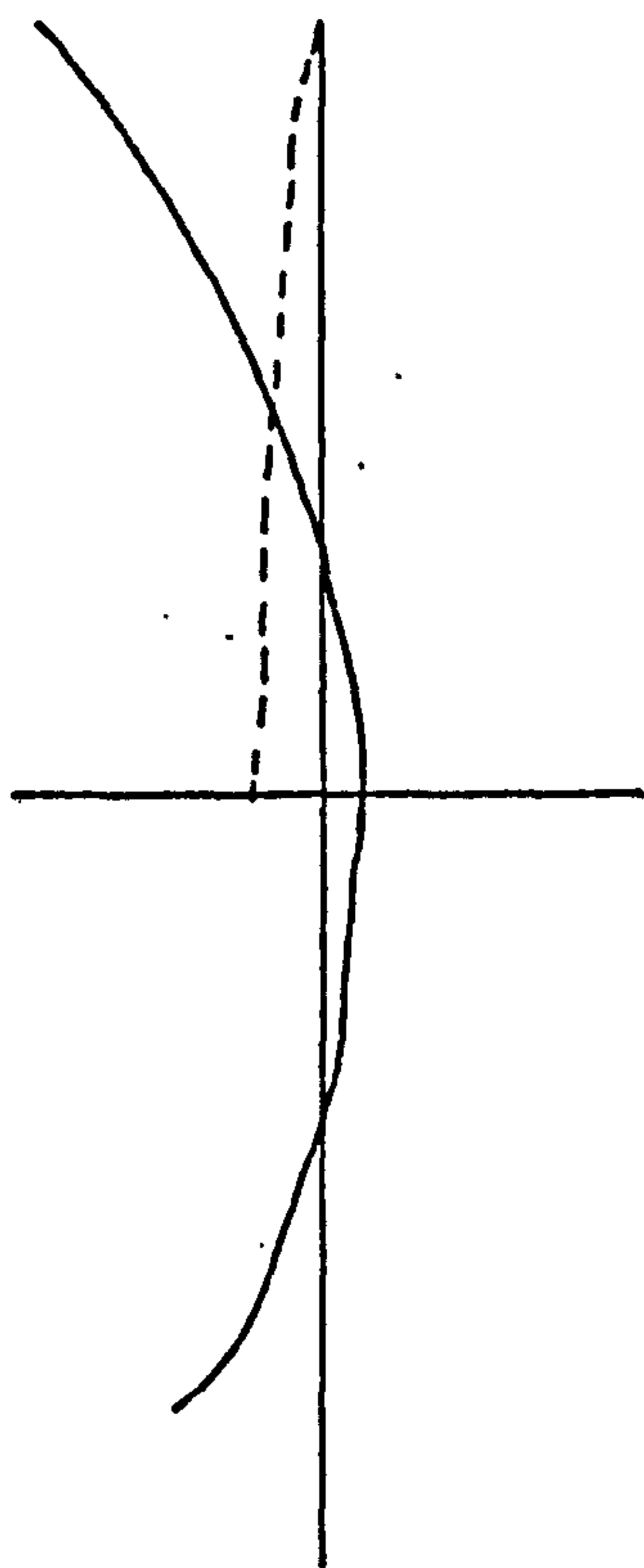
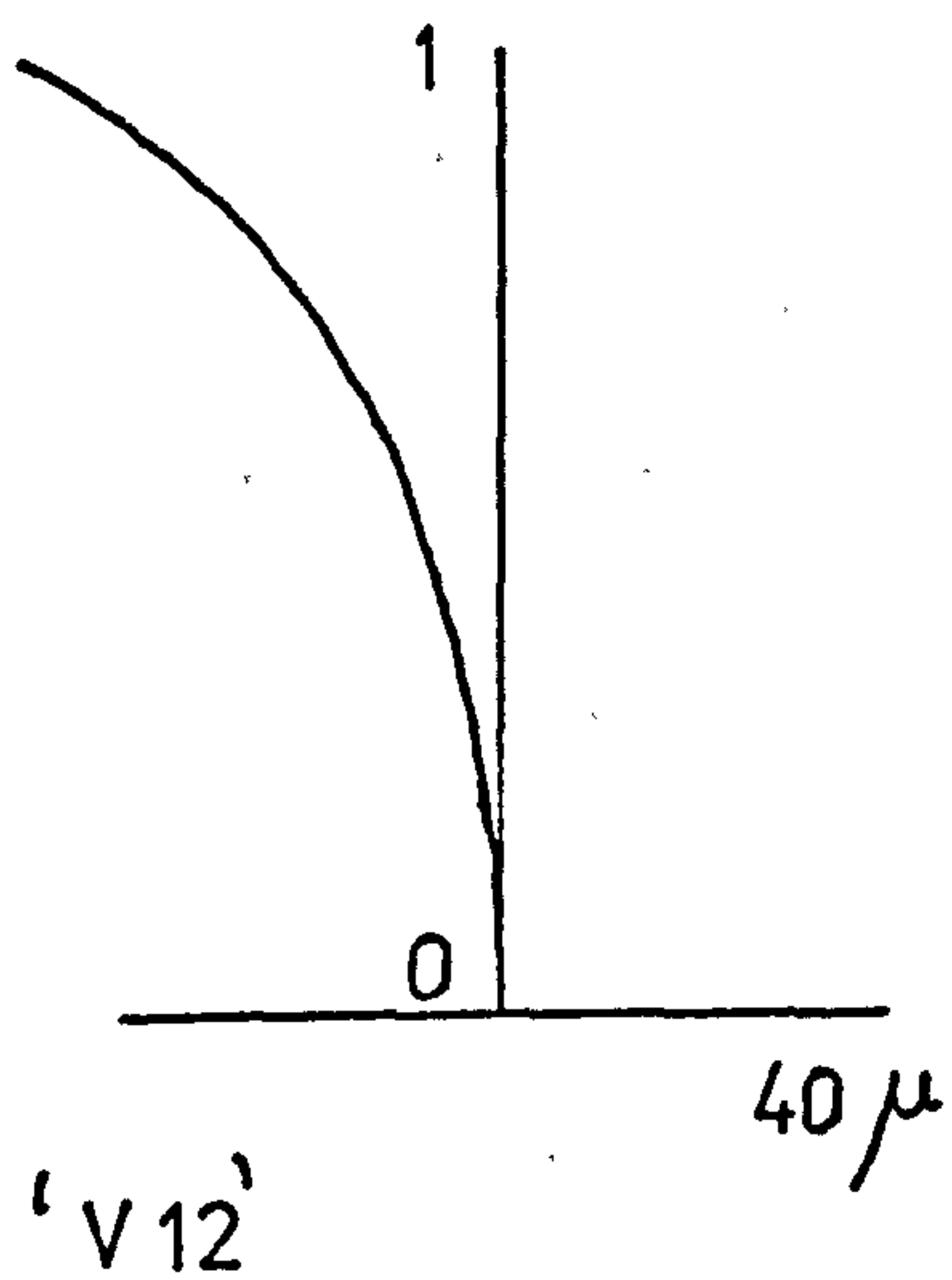


FIG 9.14

580nm



AXIS

14°

20°

FIG 9.15

in the weights for the new programme. The astigmatism at full-field is 0.0065 cm, compared with 0.0002 cm for the version 12 system. The high o.t.f. performance with this large astigmatism shows that it may be possible to relax the astigmatism weights for the standard programme with some benefit.

3. Petiau P.E.P.R. lens

This system, designed for a similar application to the Fadua system, had been designed by the Optical Design Group, and was already being manufactured when the new programme was ready for use. The system required a final balancing run when the actual melt glasses had been polished, and so the programme was allowed to vary only the separations of the components.

The system was optimised on the version 12 programme, and produced the system shown in figure 9.16. The system was designed to work with a prism after the final component, but this has been omitted on the diagram.

After a preliminary optimisation at 100 cycles/cm the new programme was run at the required frequency of 450 cycles/cm. As expected, since the programme had only a few separations to vary it was unable to make large changes to the system. The largest adjustment was an increase of 0.3 cm in one separation, and since this was only just resolvable on the scale of the lens diagram, the diagram has not been re-drawn for this system. Full details of both systems can, as usual, be found in appendix A. Even though the programme had only a few parameters to vary it was still able to produce an improvement in the geometric o.t.f. The two sets of o.t.f. curves are shown in figure 9.17; and these clearly show the increase due to the new programme. As before it is the performance of the systems at the specified

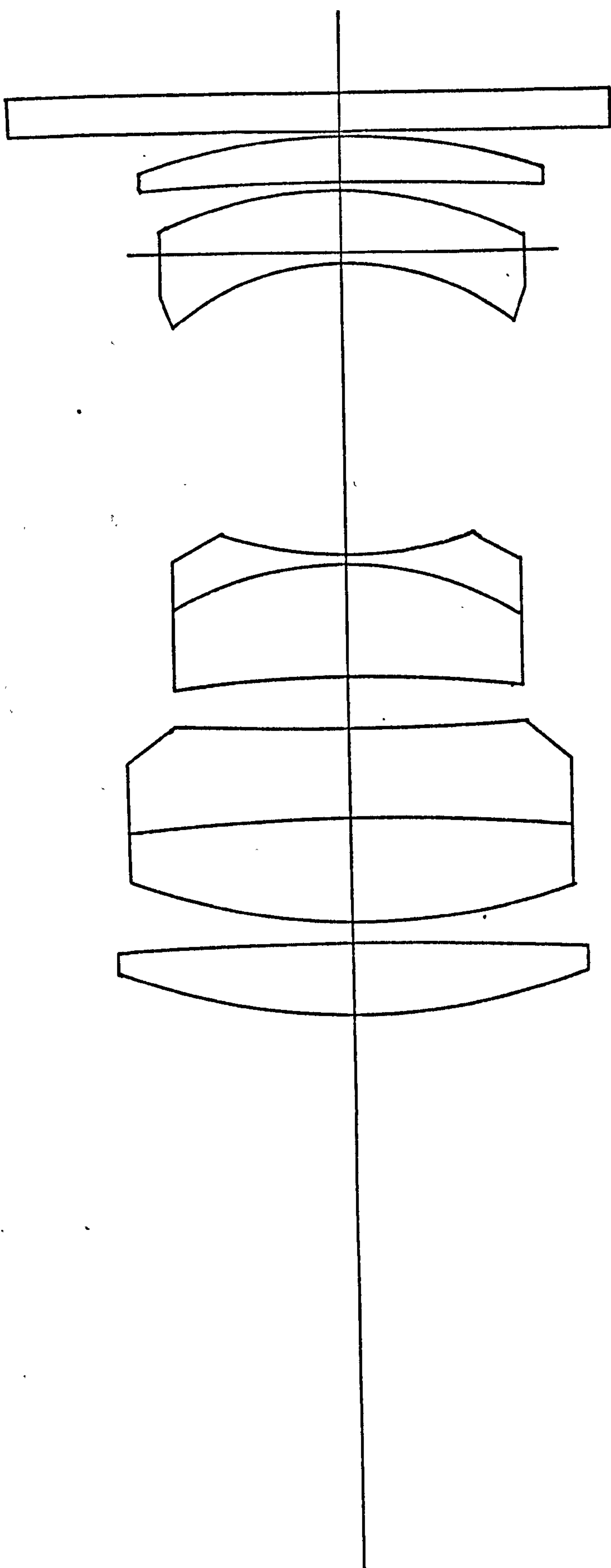
spatial frequency which is important, and these figures are given below.

<u>Geometric O.T.F. at 450 cycles/cm</u>				
Version 12			'CTF'	
Ax	sag	tan	sag	tan
Axis	.312		.474	
$\eta = 7.35$	.348	.460	.427	.496
$\eta = 10.5$	.322	.427	.330	.450
Average	.374		.435	

In this case, although there is a marked increase in the average o.t.f., there is no corresponding increase in the worst o.t.f. figure. This can be attributed in the second system to the low full-field sagittal o.t.f. In an attempt to improve this figure, the new programme was used with an increased full-field sagittal weight, and this had the effect of increasing the o.t.f. to 36% from 33%. A further run, with a still higher weight increased the sagittal o.t.f., but severely reduced the o.t.f. at the other fields, leading to a worse overall performance.

These effects may be seen more clearly by considering the through-focus o.t.f. curves for the three systems, which are given in figure 9.18. These curves show that by defocussing the version 12 system the worst o.t.f. can only be increased to 32%. The first o.t.f.-optimised system can also benefit by defocussing, and the worst o.t.f. figure can be increased to 37%. The second o.t.f.-optimised system does not gain any improvement by defocussing, since the rebalance has already effectively moved the full-field sagittal curve.

It should be noticed that for the first o.t.f.-optimised



2cm

PETIAU LENS

'V12'

FIG 9.16

PETIAU LENS

GEOMETRIC OTF

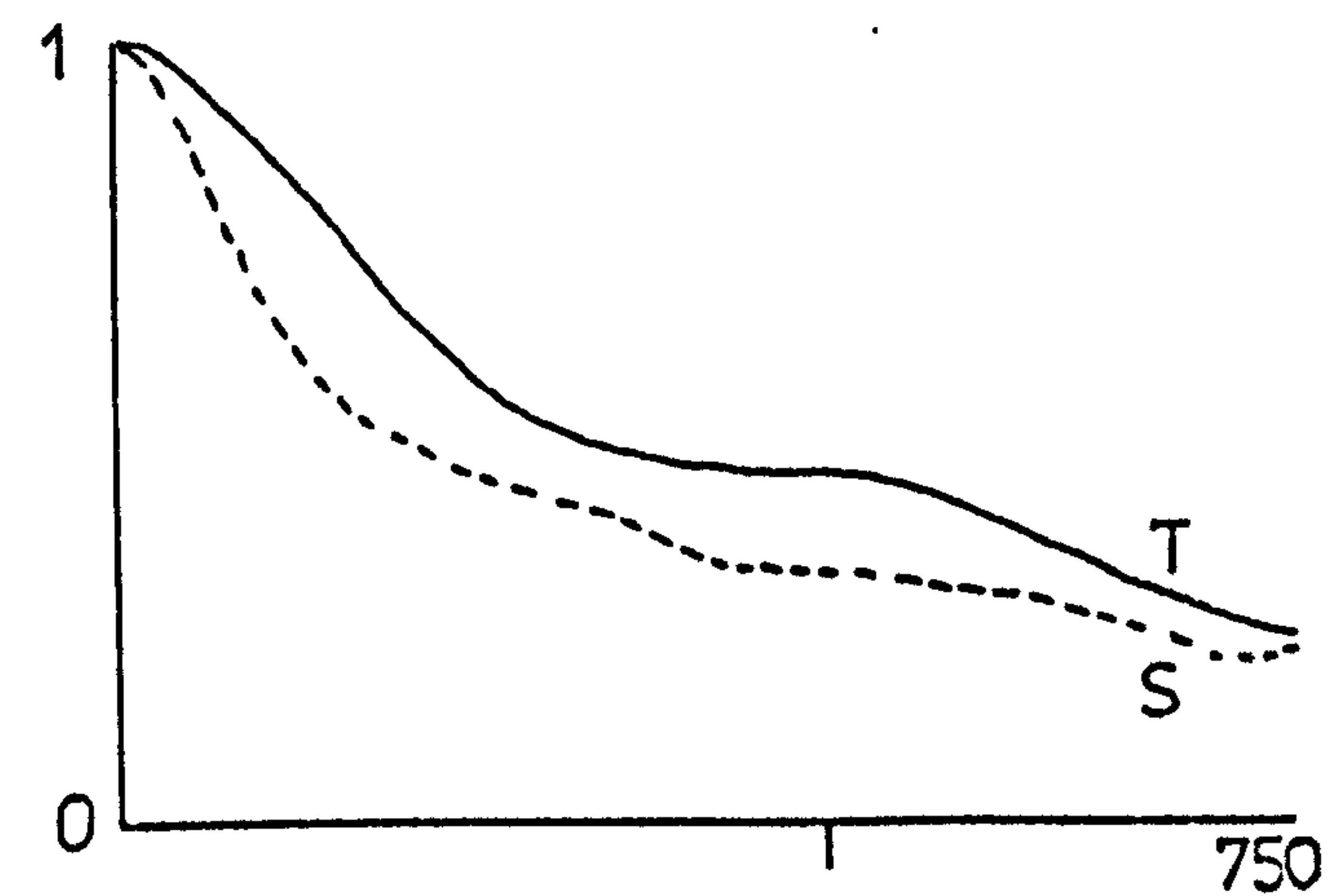
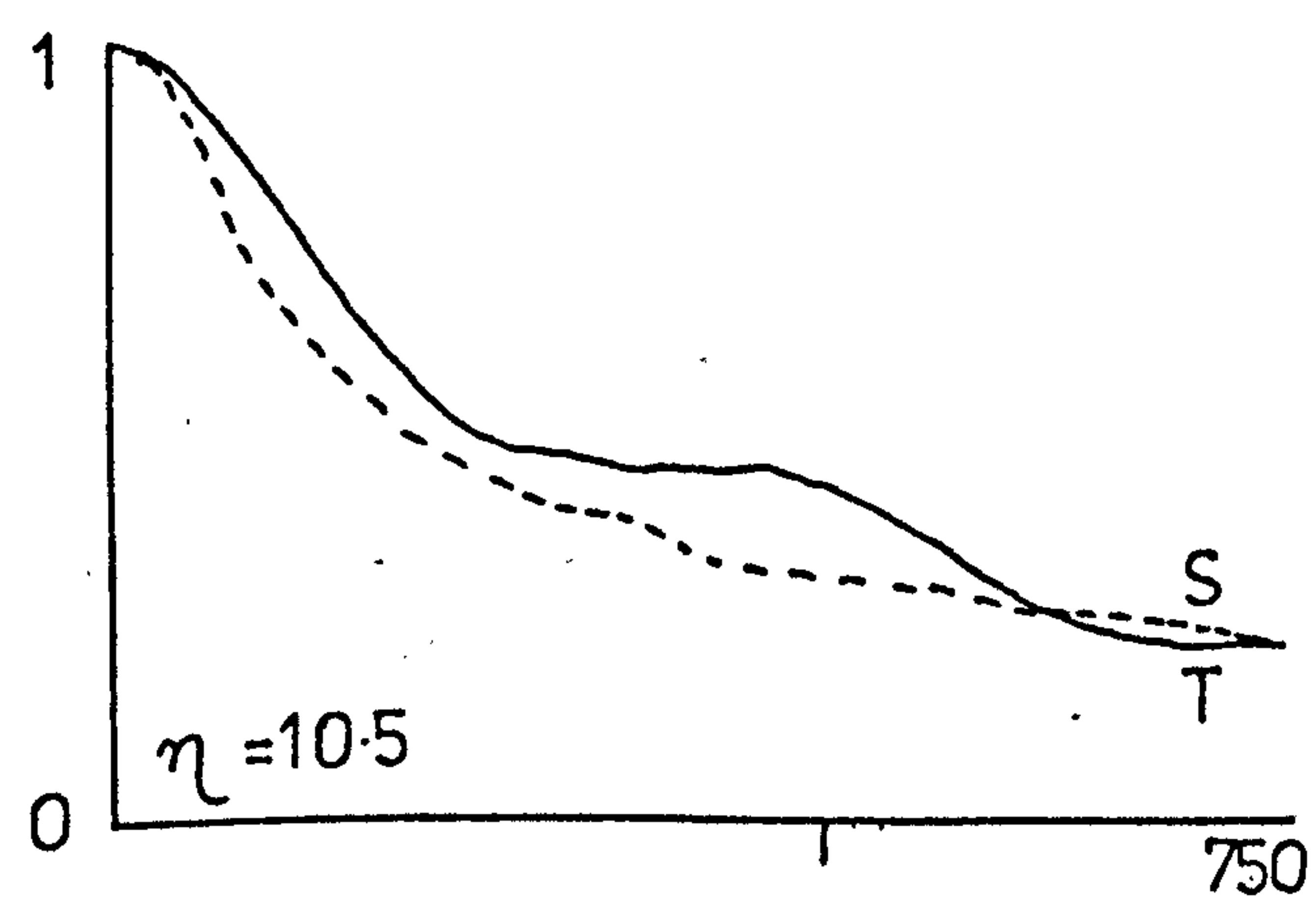
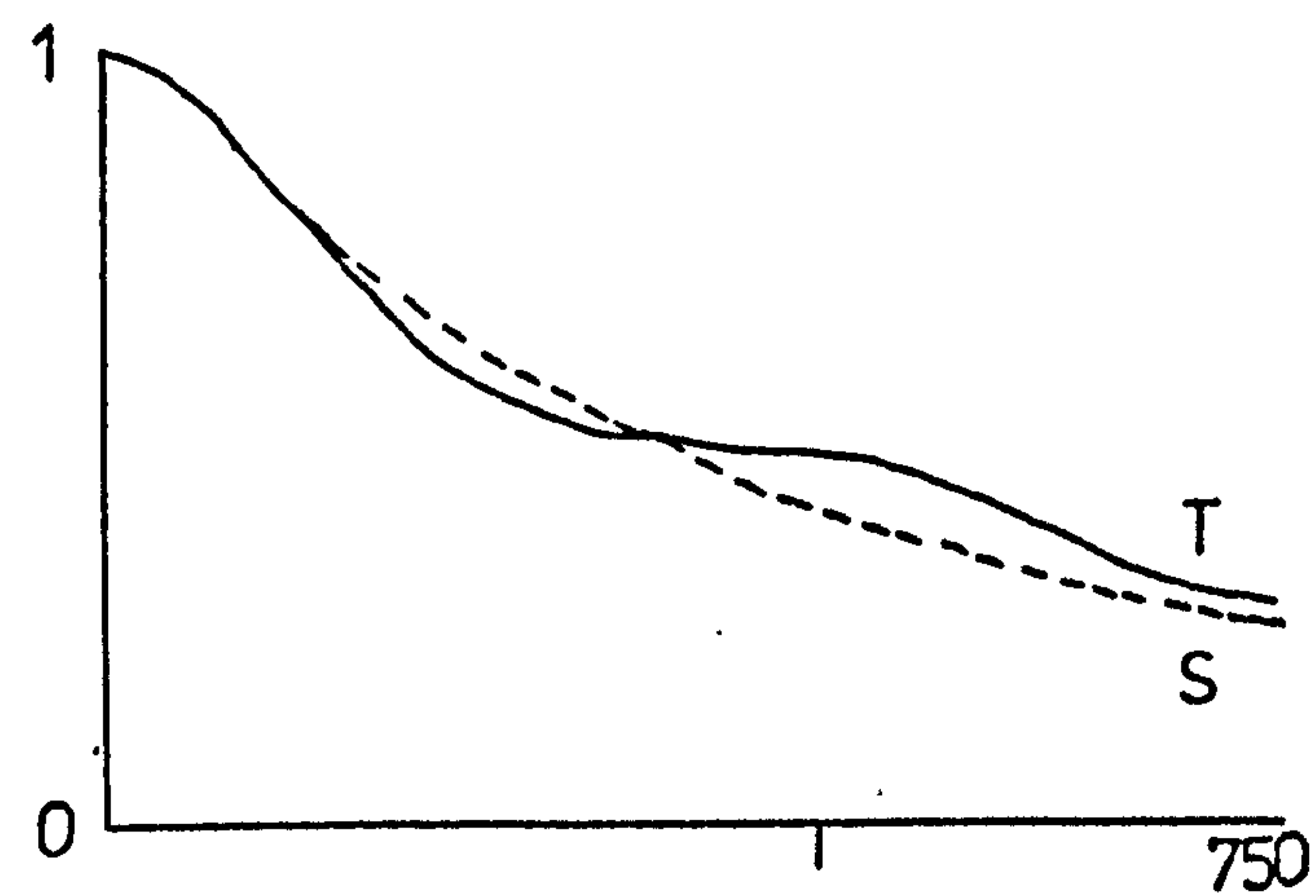
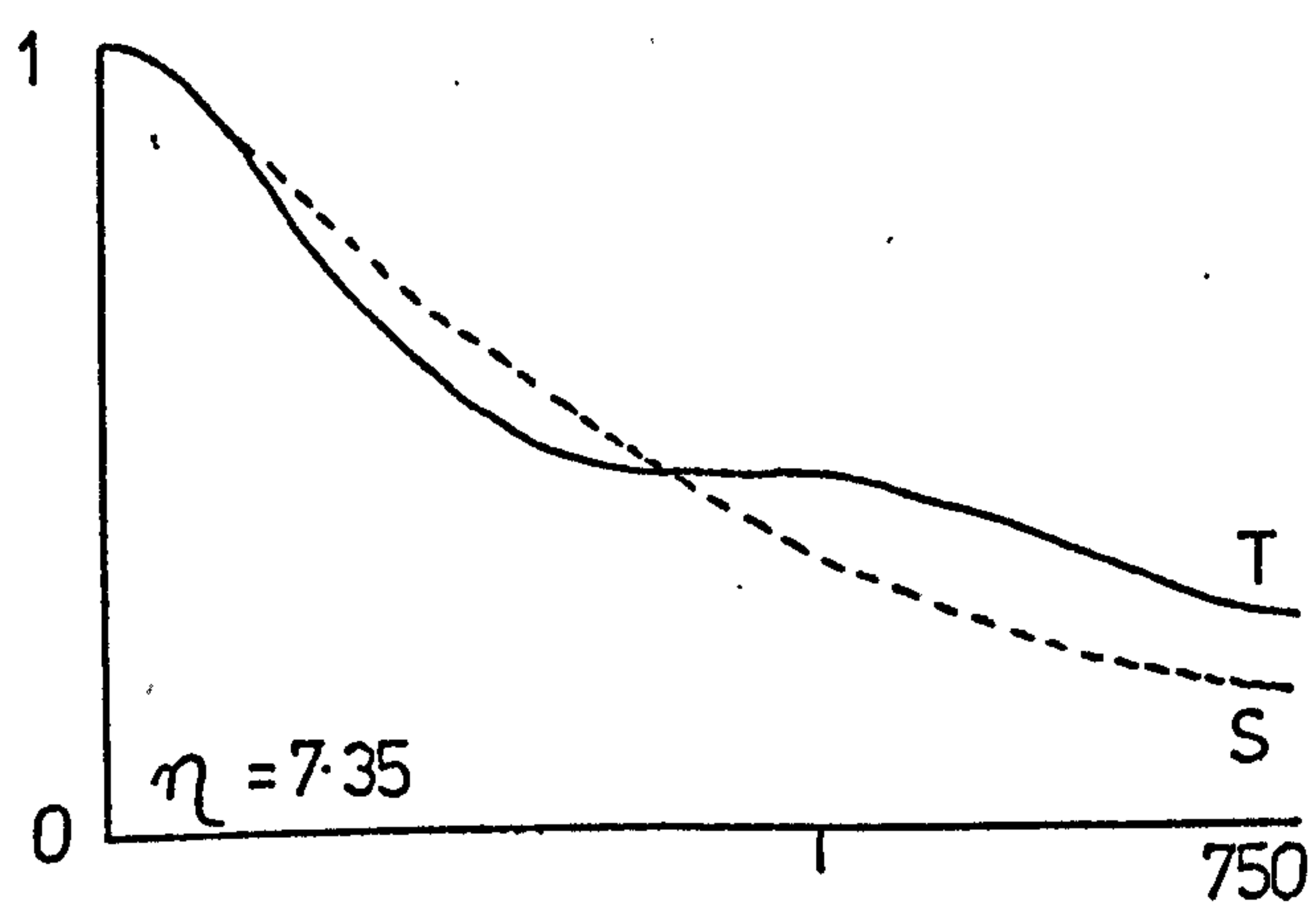
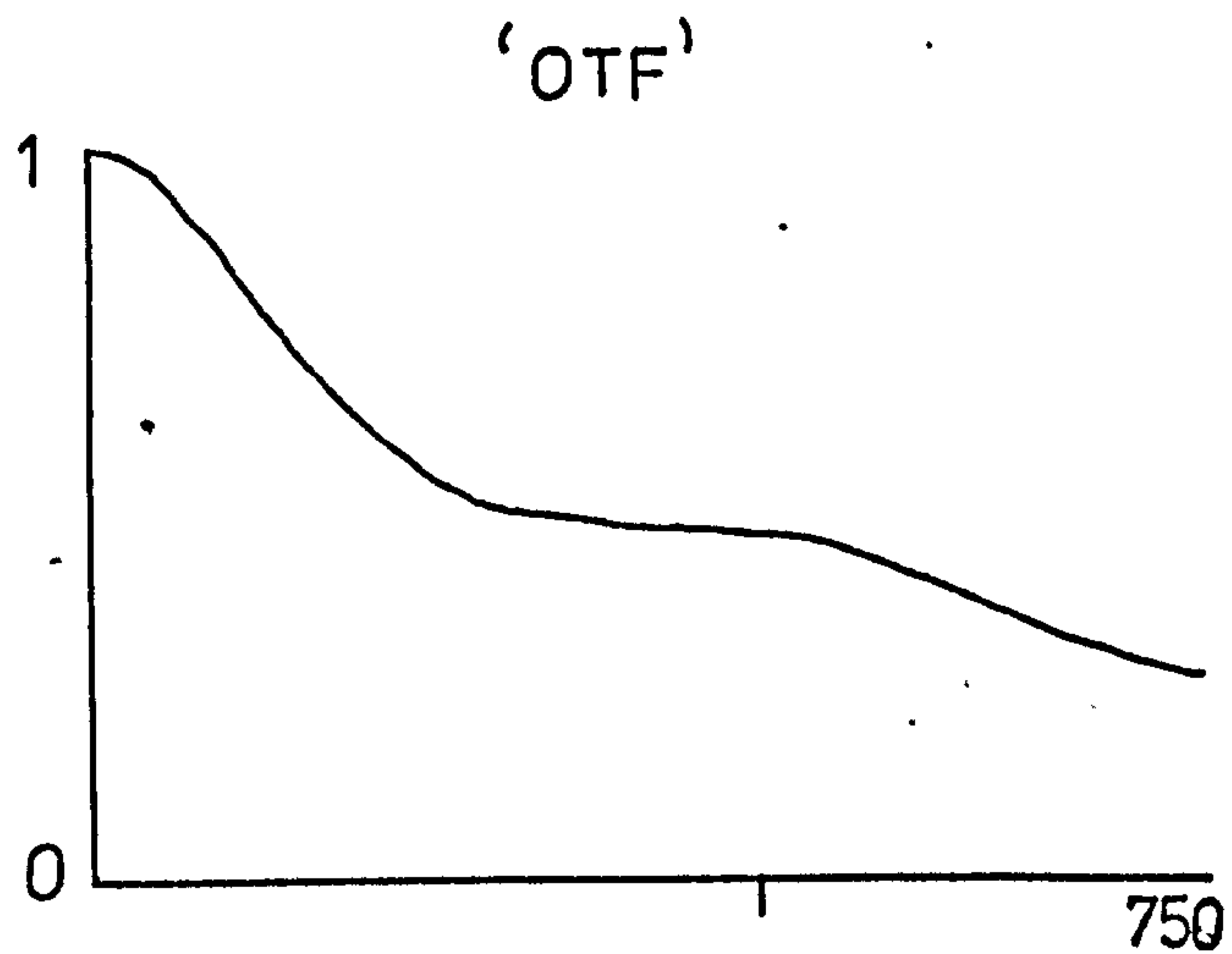
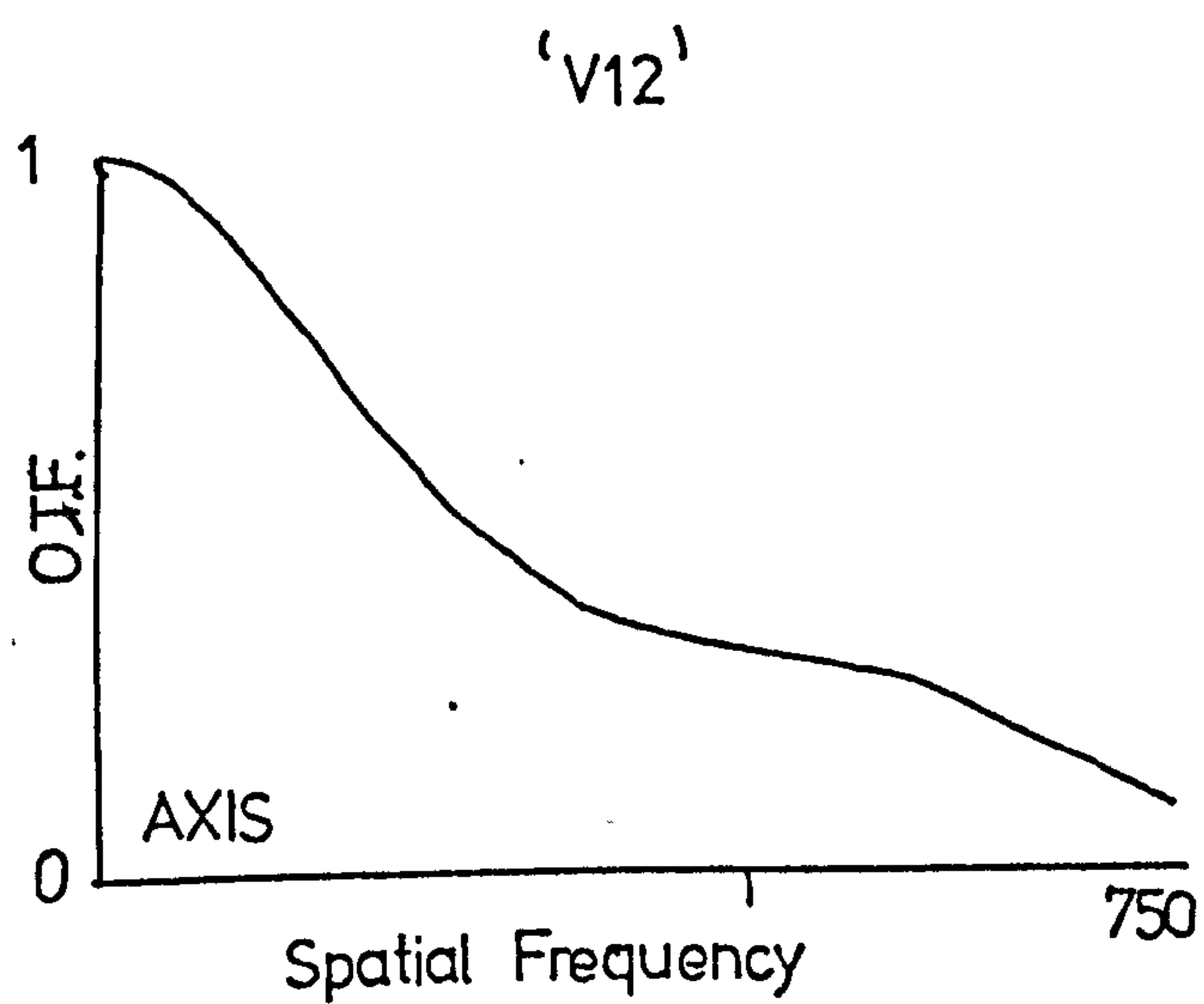


FIG 9.17

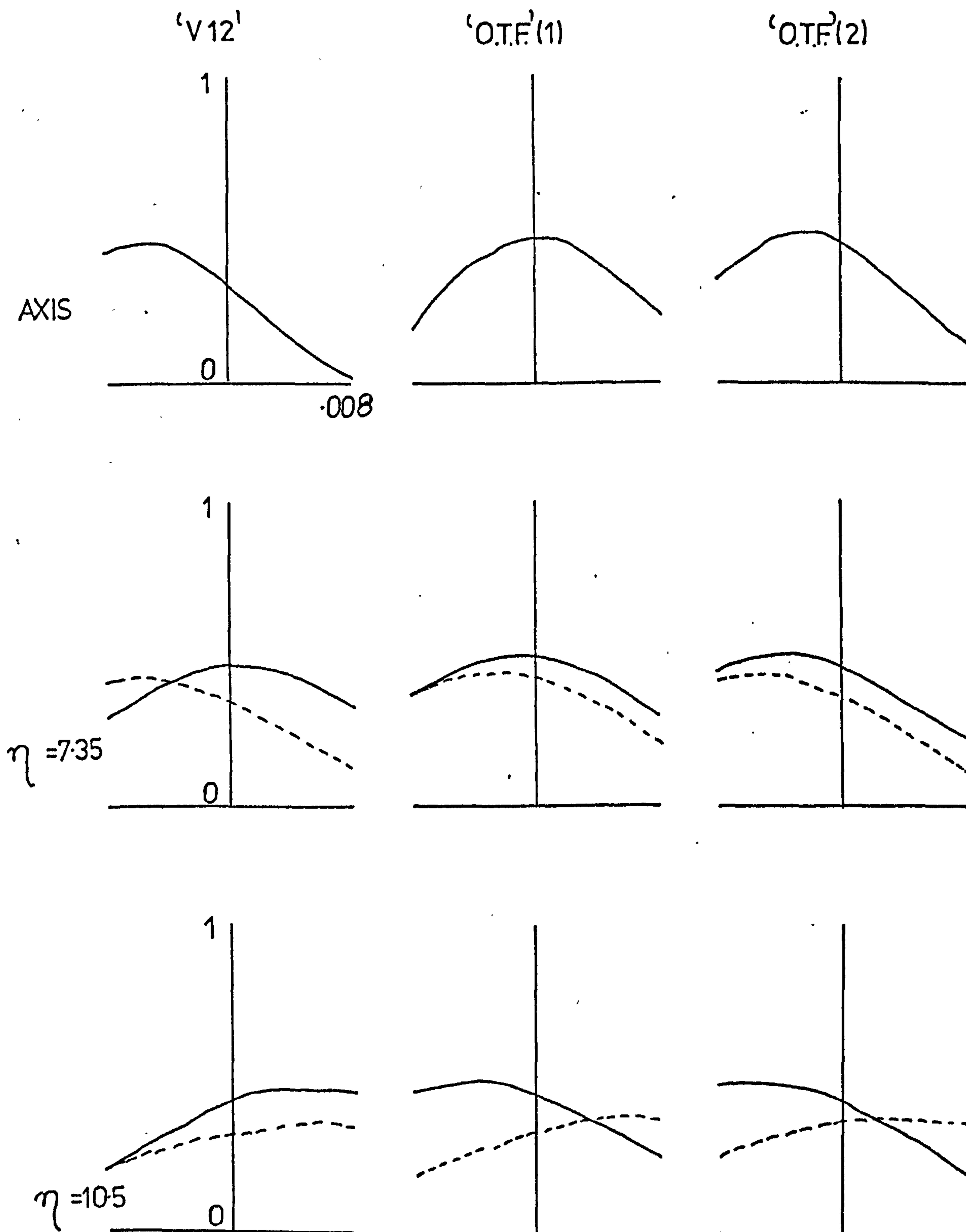
PETIAU LENSTHROUGH-FOCUS O.T.F.

FIG 9.18

system the average o.t.f. is higher (43.5%) at the chosen plane of focus than at the defocussed plane (41.5%), which shows that the programme is correctly optimising the average o.t.f., as required.

Conclusions

All three systems show that considerable improvements are possible in geometrical o.t.f. performance by the use of the new programme, and, as predicted by the previous theory, the results of the first system show that this improvement is accompanied by an increase in the diffraction o.t.f. of the system.

Appendix ALens Systems after Version 12 and 'OTF' Optimisations

All the systems referenced in chapter 9 are given in this appendix. All curvatures are given in cm^{-1} , all separations are given in cm, and all wavelengths are given in nm.

For ease of comparison, only the first two wavelengths, and hence the first two refractive indices are given for each system.

Padua P.E.P.R. LensVersion 12

No	Curvature	Separation	400nm	420nm
			Index	Index
1	Plane	1.27	1.530847	1.528386
2	Plane	1.0	air	
3	.031178	2.22772	1.764271	1.755197
4	-.007285	1.558391	air	
5	.071338	1.606774	1.609579	1.60581
6	-.031396	0.5	1.764271	1.755197
7	.064422	6.671135	air	
8	Plane	6.231847	air	
9	-.093934	0.950459	1.652148	1.646089
10	.071281	1.083342	1.692759	1.685478
11	.017522	0.363135	air	
12	-.006205	1.467709	1.699774	1.695878
13	-.080315	0.222332	air	
14	.03987	3.244349	1.699774	1.695878
15	-.023061	25.217354	air	
16	-.070112	.5	1.590183	1.587038
17	-.001001			
	Defocus	-.013056		

<u>Padua P.E.P.R. Lens</u>			<u>'OTF'</u>	
No	Curvature	Separation	400nm Index	420nm Index
1	Plane	1.27	1.530847	1.528386
2	Plane	1.0	air	
3	.026057	2.303291	1.764271	1.755197
4	-.013103	1.507113	air	
5	.070868	1.673225	1.609579	1.60581
6	-.035695	0.5	1.764271	1.755197
7	.06171	6.580578	air	
8	stop	5.698176	air	
9	-.086811	1.118775	1.652148	1.646089
10	.073038	1.338752	1.692759	1.685478
11	.021049	0.397892	air	
12	-.005094	1.327201	1.699774	1.695878
13	-.072515	0.578479	air	
14	.035988	3.227796	1.699774	1.695878
15	-.030213	24.877106	air	
16	-.064535	0.5	1.590183	1.587038
17	.006559			
	Defocus	-.013616		

3cm Microfilm LensVersion 12

No	Curvature	Separation	580nm	610nm
			Index	Index
1	.750643	0.209531	1.6915	1.68962
2	.279	0.080399	air	
3	.946466	0.370099	1.6915	1.68962
4	.252541	0.03	1.62657	1.624
5	1.600109	0.114169	air	
6	stop	0.588931	air	
7	-1.234576	0.09	1.69986	1.69648
8	.524327	0.327667	1.6915	1.68962
9	-.48081	0.25706	air	
10	-.003112	0.454176	1.6915	1.68962
11	-.575661	0.05	air	
12	.399415	0.358011	1.73716	1.73383
13	.089796			
	Defocus	-.000021		

3cm Microfilm Lens'OTF'

No	Curvature	Separation	580nm	610nm
			Index	Index
1	.757548	0.287735	1.6915	1.68962
2	.315325	0.05	air	
3	.991683	0.322772	1.6915	1.68962
4	.394814	0.09	1.62657	1.624
5	1.631026	0.365998	air	
6	Plane	0.376395	air	
7	-1.153566	0.09	1.69986	1.696948
8	.482056	0.227363	1.6915	1.68962
9	-.472015	0.05	air	
10	.020491	0.827801	1.6915	1.68962
11	-.593306	0.05	air	
12	.347202	0.262205	1.73716	1.73383
13	.064437			
	Defocus	-.004219		

Petiau P.E.P.R. LensVersion 12

No.	Curvature	Separation	400nm	420nm
			Index	Index
1	Plane	1.27	1.53139	1.52892
2	Plane	0.1	air	
3.	.054413	1.373264	1.71737	1.713211
4	.00663	0.1	air	
5	.084095	1.780685	1.69901	1.695136
6	Plane	0.5	1.6387	1.634092
7	.115056	6.059282	air	
8	stop	2.576991	air	
9	-.084095	0.5	1.60814	1.603195
10	.102041	3.360498	1.64395	1.638096
11	.027092	1.589603	air	
12	-.009965	2.5	1.76523	1.756116
13	.016393	3.5	1.700204	1.696312
14	-.041623	.433747	air	
15	.009965	2.1	1.69901	1.695136
16	-.040154	0.1	air	
17	Plane	11.0	1.53139	1.52892
18	Plane			
	Defocus	-.0347		

Petiau P.E.F.R. Lens'OTF'

400nm

420nm

No	Curvature	Separation	Index	Index
1	Plane	1.27	1.53139	1.528920
2	Plane	0.1	air	
3	.054413	1.373264	1.71737	1.713211
4	.00663	0.193926	air	
5	.084095	1.780685	1.69901	1.695136
6	Plane	0.5	1.6387	1.634092
7	.115056	5.976958	air	
8	stop	2.681422	air	
9	-.084095	.05	1.60814	1.60395
10	.102041	3.360498	1.64395	1.638096
11	.027092	1.379473	air	
12	-.009965	2.5	1.76523	1.756116
13	.016393	3.574434	1.700204	1.696312
14	-.041623	.731644	air	
15	.009965	2.1	1.69901	1.695136
16	-.040154	0.1	air	
17	Plane	11.0	1.53139	1.52892
18	Plane			
	Defocus	-.038463		

Acknowledgements

I would like to thank Professor C.G.Wynne for his invaluable help and guidance throughout these studies, and Dr.M.J.Kidger for many helpful discussions.

I would also like to thank Mrs P.M.J.H.Wormell for her guidance through the intricacies of the SLAMS programme.

I am indebted to Professor Wynne and Dr.Kidger for the original designs of the systems described.

I am grateful to Dr.K.G.Birch for much of the material on physiology and o.t.f.

Finally, I wish to acknowledge the financial support from the Science Research Council.

References

1

- Barakat R. 1964 J.O.S.A. 54 1 page 38
- Borough et al 1967 Boeing Company D2-114058-1
- Biberman L.M. 1974 'Image Evaluation Techniques'
Plenum Press
- Charman W.N. 1965 Photo Sci Engnrg 9 6 385
and Olin A.
- Goodbody A.M. 1958 Proc Phys Soc 72 411
- Hopkins H.H. 1955 Proc Roy Soc A 91 231
1957 Proc Phys Soc B70 449
1957 ibid 1162
1962 Proc Phys Soc 79 889
1966 Optica Acta 13 343
- Kidger M.J. 1971 Thesis University of London
1974 To be published
- King W.B. 1968 J.O.S.A. 58 655
1968 Applied Optics 7 489
1968 ibid 1193
1969 J.O.S.A. 59 692
1972 J.O.S.A. 62 230
- Levenberg K. 1944 Quart.Appl.Math 2 164
- Lukosz W. 1958 Thesis Tech Hochschule Braunschweig
- Maréchal 1947 Rev Opt 26 257
- Meiron J. 1968 Applied Optics 7 667
- Miyamoto K. 1958 J.O.S.A. 48 57-63
- Offner 1970 Perkin-Elmer Report
- Rosenhauer K. 1966 Applied Optics 5 415
- Strehl K. 1902 Z InstrumKde 22 213
- Wynne C.G. 1959 Proc Phys Soc 73 777
- Wynne C.G. and 1963 Applied Optics 2 164
- Wormell P.M.J.H.