AN EXPERIMENTAL METHOD OF PRESENTING THE PRINCIPLES DETERMINING THE GENERAL PROPERTIES OF OPTICAL GRATINGS.

By A. L. McAulay, B.Sc., B.A., Ph. D.

(Read 6th August, 1923.)

In a laboratory which is not well equipped with modern optical instruments, the actual appearance and meaning of many of the phenomena whose theory is studied is not in the least appreciated. This seems especially the case in connection with the elements of the more advanced theory of optical gratings, and with the theory of interference spectroscopes.

The following presentation is an attempt to provide a set of experiments, made with fairly cheap and simple apparatus, which shall illustrate the various effects that produce the phenomena exhibited by an optical grating, and includes such simple theoretical discussion as is necessary to understand the experiments.

In the course of the experiments the effects that determine such things as the resolving power of a grating, the intensity of the spectrum of a given order, and so on, are directly observed. At the end of the paper the results obtained in the discussion are applied to the cases of the ordinary plane grating and the Michleson Echelon.

SECTION 1. APPARATUS.

The arrangement of apparatus is shown diagrammatically in Figure 1. G is an ordinary metallic filament electric lamp of about 60 candle power, enclosed in a box which is fairly light tight; a kerosene tin lying on its side, with a black cloth over one end, and a window cut in the other serves the purpose very well. A is a vertical slit placed against the window. It should have good straight jaws, and be adjustable. Considerable use may be made of this adjustment, for some purposes it is convenient to have the slit very much wider than for others. It is best to have a large diameter short focus lens in the box to diffuse the light falling on the slit. B is a short focus lens. It must be well corrected for chromatic aberration, as it must work with a fairly large diaphragm. A symmetrical camera lens by Beck of about 28 cms. focal length, working at F8, proved very satisfactory.

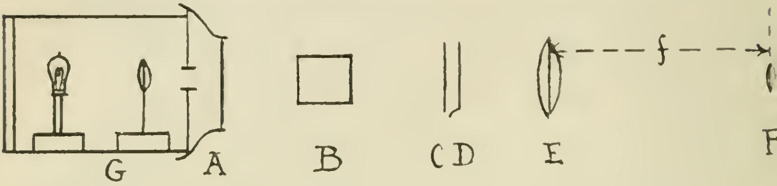


Figure I

C is another vertical slit. Its jaws must be good and it should be fitted with a micrometer adjustment enabling its width to be measured. Otherwise, some such device as closing it on a piece of metal foil whose thickness is afterwards measured by a micrometer must be employed to determine its width. D is a wire grid, made by soldering together a metal rectangle, two sides of which are screws of fine pitch. Twenty-two gauge copper wire is wound round the rectangle over the screws, thus forming a double grid of the same pitch as the screws. The wire is next fastened to the rectangle with sealing wax, and the strands on one side cut away. In this way a grating of 22 lines, of pitch .76 mm., was made, and proved satisfactory. E is a telescope lens, corrected for chromatic aberration, of about a metre focal length. F is a micremeter eyepiece in whose field the phenomena to be described are observed, and by means of which distances on the diffraction patterns are measured. An ordinary eyepiece and scale could probably be made to serve.

SECTION 2. ADJUSTMENT.

A is opened to about a third of a millimetre, and B adjusted to render light from A parellel. C and D are removed, and E and F so arranged that A is sharply focussed along the vertical crosswire of F. B, E, and F should be sufficiently well corrected to give a brilliant image of A without striking aberrations.

SECTION 3. DIFFRACTION PATTERN DUE TO SINGLE SLIT. VISIBILITY CURVE.

The slit C is first completely closed, and then slowly opened. When its width is about .1 mm. the field of the eyepiece is seen to be slightly illuminated. On increasing its width the illumination increases, and at a certain stage dark vertical bands appear, one from each side of the field, and

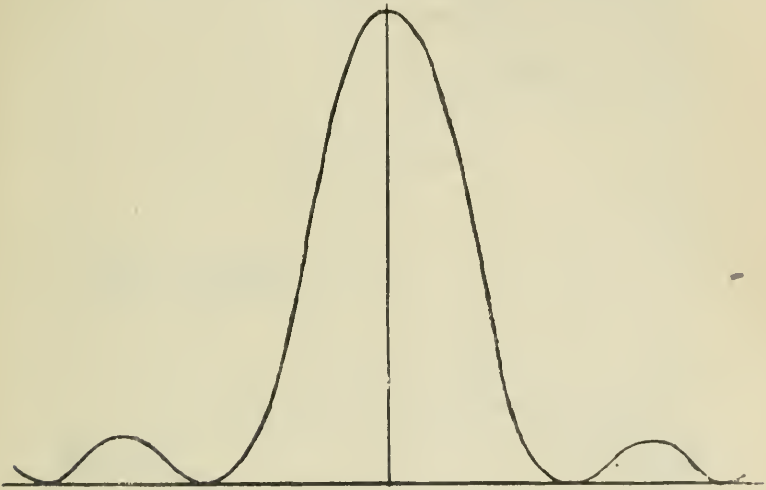


Figure II

move towards the centre. The slit is opened till it completely uncovers one aperture of the grid D, i.e., until its width is about .4 mm. The appearance is now as indicated in Fig. II., where abscissæ represent distances from the cross-wire of F, and ordinates the intensity of the light. This curve in what follows will be referred to as the visibility curve.

Measure the distance, f, from E to F, the width of the slit, e, and with the micrometer eyepiece the corresponding distance, 2d, between the first minima on the visibility curve. Repeat with three different slit widths, ending by opening the slit till one whole aperture of the grid is exposed. Show that in each case $d = \lambda f/e$, where λ is the mean wave length of the light used (about 6×10^{-5} cms.).

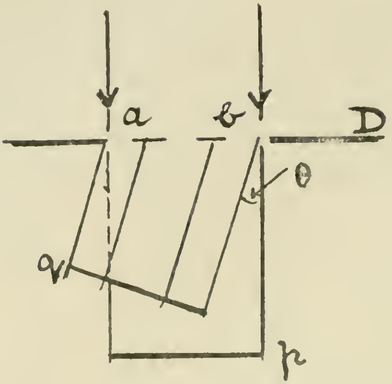
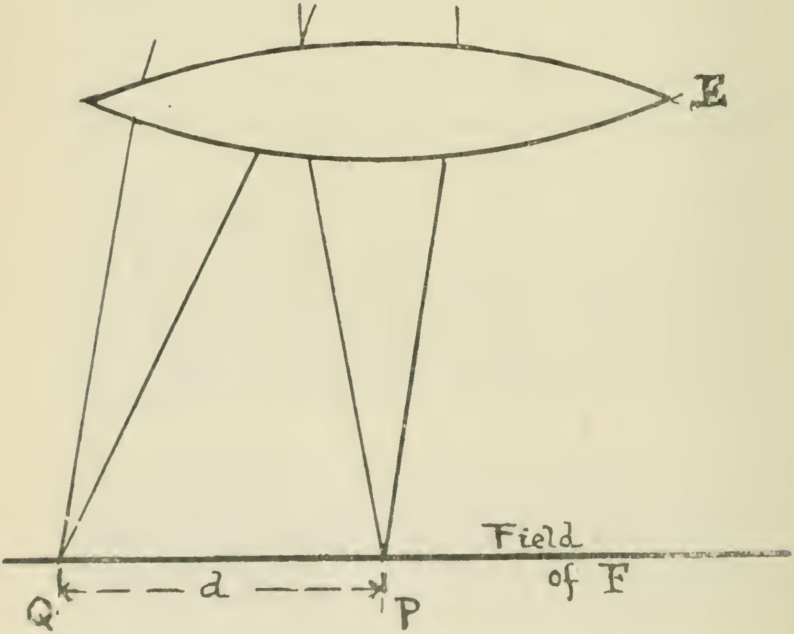


Figure III



SECTION 4. THEORY OF SECTION 3.

Figure III. shows diagrammatically the slit D and the lens E. Light falls on D normally, and therefore leaves every part of it in the same phase. Consider a beam leaving the slit at an angle θ with the direction of the incident light. The disturbances over a plane such as q perpendicular to this direction will be brought together by E at Q, the point at which a wave front at q would be focussed. The disturbances over q at any instant will, however, not be in exactly the same phase, and it is the combined effect of a set of out of phase waves that will produce the illumination at Q. Let p, P be the wave front, and focus for $\theta = 0$. It is obvious that this is the position of the image of the slit, that is, it is on the vertical crosswire as adjusted in section 3.

It is required to investigate the illumination at Q when Q takes up different positions. For this purpose the slit D will be thought of as made up of a large number of elements, and the combined effect of the wave trains from each will be considered. The waves arriving at Q will be represented as vectors in the usual way, and, as there are an equal number of wave lengths between each element of q and Q, a vector drawn for q will equally well stand for the effect of the same wave train at Q. a and b are two adjacent elements of the slit. Then obviously the path difference at q of the light coming from a and b is $ab \sin \theta$, and the phase difference of the wave trains at q (or Q) is $\frac{2\pi}{\lambda} ab \sin \theta$. Now $PQ = f \sin \theta = d$. Therefore, the phase difference between the wave trains is $\frac{2\pi}{\lambda} ab \frac{PQ}{f}$ i.e., it is proportional to the distance of Q from P.

Consider the vectors representing disturbances from successive elements as short rods hinged to each other at the ends (see Figure IV.). Then the line joining the ends of the composite rod will be the vector representing the resultant effect of all the elements. At P the waves are all in the same phase, the jointed rod or chain lies stretched out in a straight line (Figure IV. 1), and the resultant is the arithmetic sum of the components. At Q, from what has been said above, wave trains from adjacent elements will have a phase difference of $\frac{2\pi}{\lambda} ab \frac{d}{f} = d\alpha$ say, i.e., each section of the chain will make an angle of $d\alpha$ with the one next to it. Obviously,

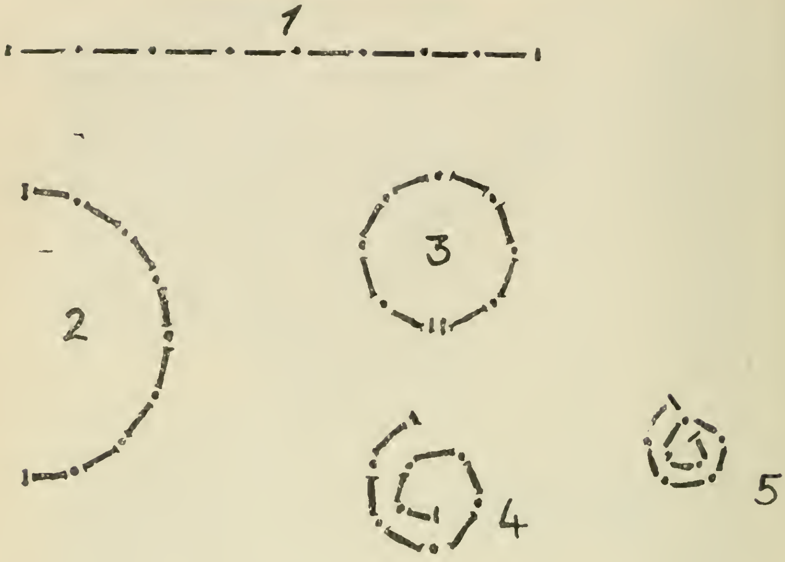


Figure IV

if there are n elements, the angle between the end vectors will be $\frac{2\pi n\, ab}{\lambda}\frac{d}{f} = n d\alpha = \alpha$ and as n ab=e= the width of the slit $\alpha = \frac{2\pi e}{\lambda}\frac{d}{f}$. Figure IV., 1, 2, 3, 4, and 5, show the chain of vectors for $\alpha = o, \pi, 2\pi, 3\pi, 4\pi$. Remembering that the resultant is the line joining the ends of the chain, it is easy to see that the visibility curve must have the general form found for it in section 3, Figure II. It must not be forgotten that the intensity of the illumination is proportional to the square of the amplitude of the wave, while the vector represents the amplitude. It is evident from a consideration of Figure IV. that the first minimum of the curve will occur when $\alpha = 2\pi$ which gives as found experimentally $d = \frac{f\lambda}{e}$ or $e \sin\theta = \lambda$. A more complete theory (see Schuster, Theory of Optics, p. 99 *et seq.*) shows that the expression giving the form of the visibility curve is $I = I_0\left[\frac{\sin^2(\alpha/2)}{(\alpha/2)^2}\right]$ where I is the intensity, I_0, a constant and $\alpha = \frac{2\pi}{\lambda} e \sin\theta$.

SECTION 5. THE TWO LINE GRATING. POSITION OF SPECTRA FORMED BY A GRATING.

The slit C is widened till a second aperture of the grid begins to be uncovered. When a very narrow strip is uncovered the pattern of section 3 is still clearly seen in the eyepiece F, but it is furrowed by dark lines. A typical curve connecting illumination with distance from the crosswire is shown in Figure V. 1. On further widening the slit till the whole of the second aperture is uncovered, the dark bands deepen, and the pattern splits into bright strips, all but the centre one being coloured at the edges. The colours are dispersed more the further the strips are from the crosswire. These bright strips with coloured edges correspond to the spectra formed by gratings, and are, in fact, the spectra formed by a grating of two lines. Figure V. 2 shows the way in which V. 1 would develop on widening the slit. As will presently be shown, these spectra should be evenly spaced. Measure the distances of their centres from the centre of the central uncoloured strip with the micrometer eyepiece, and calculate the corresponding values of $\sin \theta$, θ having the same meaning as in the last section. It will be noticed that some gaps are double others, indicating that certain spectra are missing. This is due to the fact that they should appear where there is a minimum of the visibility curve, and where consequently there is no light available for their production. c Figure V. 2 represents such a case.

Draw and dimension curves similar to Figures V. 1 and V. 2 from your observations, and compare them with the dimensioned visibility curve you obtained as in section 3. Make a special note of any spectra that are missing. Compare the values you obtain for $\sin \theta$ for the centres of the spectra with $\frac{m\lambda}{l}$ where l is the distance between two successive apertures of the grid, m is an integer, and λ is a mean value for the wave length of white light, say 6×10^{-5} cms.

SECTION 6. THEORY OF SECTION 5.

The case taken for consideration is that observed when the two apertures are equal in width, i.e., when two complete apertures of the grid D are uncovered by the slit C. Consider the illumination at a point on the field of F corresponding to an angle of diffraction θ. Then from what has gone before, the illumination due to each slit separately is that

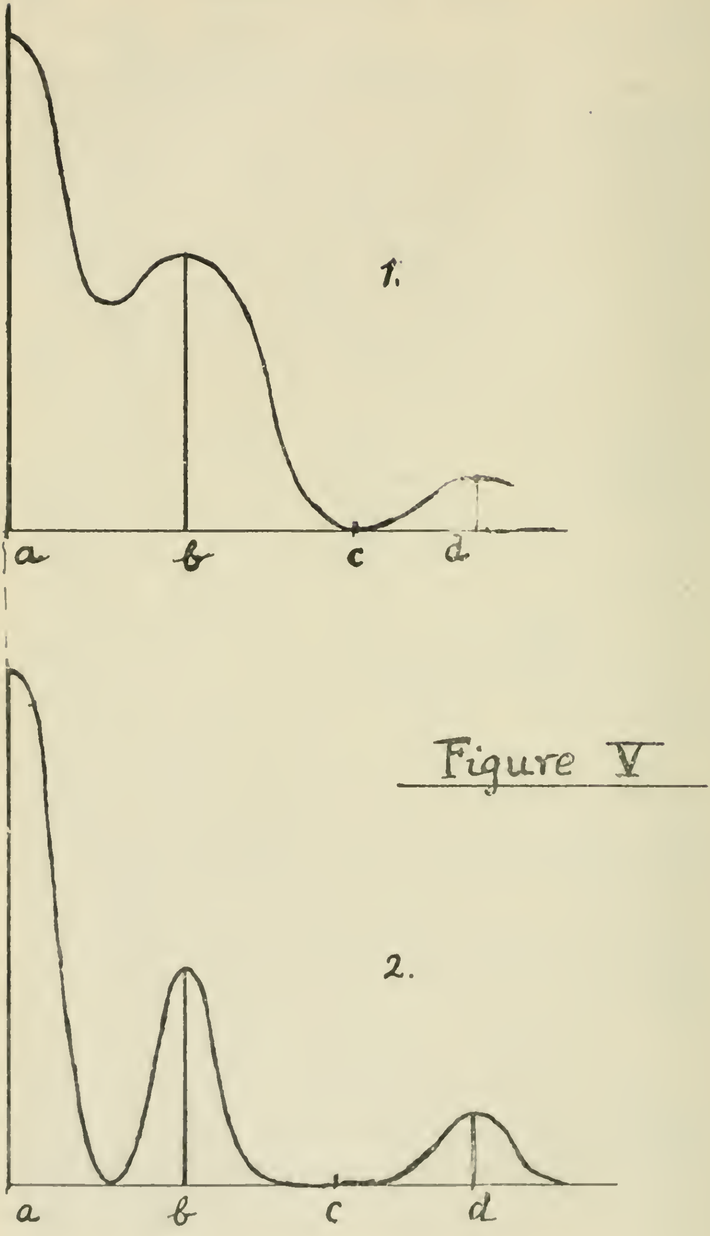


Figure V

given by the ordinate of the visibility curve, and these two beams will combine to produce illumination on the field.

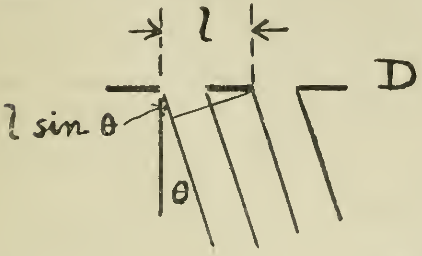


Figure VI

If the distance between corresponding points of the two apertures is l, the phase difference between the two beams is evidently $\frac{2\pi}{\lambda} l \sin\theta$ (see Figure VI.). Thus the two beams will completely interfere when $\frac{2\pi}{\lambda} l \sin\theta = \pi (2m-1)$ where m is an integer, i.e., where $\sin\theta = \frac{\lambda(2m-1)}{2l}$. The centres of the bright bands or spectra evidently come where $2\pi/\lambda \; l \sin\theta = 2m\pi$, i.e., where $\sin\theta = \frac{\lambda}{2l} 2m = \frac{m\lambda}{l}$. The spectra formed for m=1, 2, 3, etc., are called first, second, third, etc., order spectra. If the value of $\sin\theta = \frac{m\lambda}{l}$ is such that it coincides with the value for a minimum of the visibility curve, the spectrum will have zero intensity. Obviously in general, the intensity of the spectra will depend on the position they happen to occupy in the visibility curve, and this depends on the ratio of the pitch of the grating, l, to the width of the aperture, e.

The above discussion has been concerned with homogeneous light. In the case of white light, each colour forms its spectra in a different place, determined by the value of λ; hence the colour effects at the edges of the spectra.

SECTION 7. GRATING WITH MORE THAN TWO LINES. SECONDARY MAXIMA, ETC.

Widen the slit C till three apertures of D are opened and observe the appearance of the pattern in the field of F. Next uncover four apertures, then five and six. Finally, remove the slit C so that the whole of D is effective. D should consist of about twenty apertures. As more slits are uncovered, it will be seen that the spectra, although their centres remain unchanged in position, become narrower and sharper, and the colours purer, and no longer confined to the edges. Also, in the dark regions between the spectra, now wider than before, secondary maxima appear. These are much narrower and fainter than the spectra themselves. It will be found that if N is the number of apertures exposed, the number of secondary maxima is N−2. They can readily be counted, for N=3, 4 and 5, but after that, become rather too faint and close together. On replacing the white light of the lamp by a sodium flame, it will be seen that the spectra are now quite narrow lines. This was, of course, indicated by the purity of the colours in the white light spectra.

Draw curves as in section 5, Figure V., and dimension them to show the positions of the centres of the spectra, the width of the dark bands separating them, and the positions of the secondary maxima. Verify that the positions of the spectra have not changed on increasing the number of apertures.

SECTION 8. THEORY OF SECTION 7.

For simplicity, the discussion will be confined in the first instance to illumination by homogeneous light. As in section 6, the visibility curves from the different apertures are superimposed, and the illumination at any point is limited by the ordinate of the visibility curve.

Consider first a grating with a large number, N, of lines. Let E be the total width of the grating. It may be looked on as an aperture of width E divided into N elements. Figure III, of section 4, represents this case as well as the case of a slit. The scale is, however, different. The whole discussion of section 4 will also be seen to hold, but instead of there being an infinite number of elements supplying light, there is a large finite number. E is much larger than e, and therefore the distances from P of corresponding parts of the diffraction pattern are much smaller, and instead of a broad central band, there will be a fairly narrow central line flanked by secondary maxima, the whole phenomenon being confined to a very narrow region near P. The first minimum occurs where $d = f\lambda/E$ or calling d/f (the very small angle which separates the maximum from the minimum) $d\theta$ it occurs where $d\theta = \frac{\lambda}{E}$

The essential difference between the two cases lies in the fact that on further increasing θ in the case of the grating, a point is reached where the resultant disturbance from one element differs from that from the one next it by exactly one wave length, i.e., the disturbances over q (Figure III.) are all in phase again. This is evidently where $l \sin \theta = \lambda$ (see Figure VI., section 6), and results in the first order spectrum. This can never occur for the slit, because the elements are infinitely close together, and $l=0$.

Refer to Figure IV., section 4, and consider the illumination at a point Q, as represented by the distance between the ends of the jointed rod or chain. As Q moves away from P, the chain passes through the stages 1, 2, 3, 4, and 5, and then continues coiling up on itself, the successive maxima becoming smaller and smaller. In the case of the grating, there comes a point where one section is turned back completely over the one next it, and a further rotation of the sections begins to uncoil the chain once more, the uncoiling continuing till one section has turned through 360 degrees relatively to the one next it, and the chain is again stretched straight out. It is here that the first order spectrum is found. In the case of the slit the sections being infinitesimal, the chain continues to coil up for ever, degenerating to a point when Q reaches infinity.

The above shows the reason for the existence of the spectra and the secondary maxima near the central line, but a little further consideration is needed to extend the discussion to the region in the neighbourhood of the spectra, and to find the position of the first minimum associated with a spectral line.

Referring to Figure VII., p is a wave front of the beam of rays which goes to form the central maximum of a spectrum, and q is the plane normal to the bundle which is united at the first minimum flanking the spectral line. Then the disturbances at C on both p and q are in the same phase, and the disturbances at C and B are in the same phase. Therefore, the phase difference between extreme rays of q is the same as the phase difference between B and D which is evidently $\frac{2\pi}{\lambda}[E \sin(\theta + d\theta) - E \sin \theta]$. $d\theta$ is very small. Identifying $\cos d\theta$ with 1 and $\sin d\theta$ with $d\theta$, this expression becomes $\frac{2\pi}{\lambda} E \cos \theta \, d\theta$. From section 4, Figure IV., the first minimum occurs when the phase difference between extreme rays of q is 2π, i.e., when $\frac{2\pi}{\lambda} E \cos \theta \, d\theta = 2\pi$ or when $d\theta = \frac{\lambda}{E \cos \theta}$.

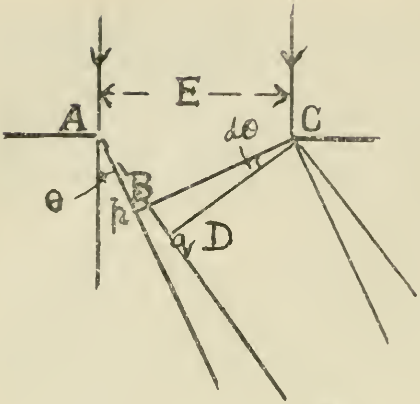


Figure VII

SECTION 9. COMPARISON OF THE EXPERIMENTAL RESULTS OF SECTION 7 WITH THOSE FOUND THEORETICALLY IN SECTION 8.

Using methods similar to those of section 4, and diagrams of the nature of Figure IV., find the number of secondary maxima that should appear between spectra given by gratings of 3, 4, and 5 lines. Roughly estimate their positions with respect to the spectra, and compare your results with those obtained experimentally in section 7.

SECTION 10. APPLICATION OF THE FOREGOING TO THE CASE OF THE ORDINARY TRANSMISSION AND REFLECTION GRATINGS.

The plane transmission grating consists of a transparent surface closely ruled with lines; 15,000 lines to the inch is an average spacing. In such a case, a transparent aperture will have a width of roughly 10^{-4} cms. The light from these apertures passes through a lens, and is focussed, all the individual visibility curves being superimposed. The angle through which the central maximum of the visibility curve extends on one side of the centre is given by $\sin\theta = \frac{\lambda}{e}$ (see sections 3 and 4). Here therefore $\sin\theta = 6\times10^{-5}/(1\times10^{-4})$, i.e., it extends over more than 30 degrees on either side of the direction of the incident light. The form of the curve will not be exactly that deduced for smaller angles owing to the influence of obliquity, but the error introduced will not be large.

The spectra will appear at positions given by $\sin\theta = \frac{m\lambda}{l}$ (section 6), or if $\frac{N}{E}$ is the number of lines per cm. $\sin\theta = \frac{mN\lambda}{E}$ and their intensities will be as the ordinates of the visibility curve for corresponding values of θ.

The resolving power of a grating is defined as the reciprocal of the fraction of a wave length that separates two spectral lines which the grating can just exhibit as distinct, that is, if the lines have wave lengths λ and $\lambda + d\lambda$ the resolving power is $\frac{\lambda}{d\lambda}$. Experience indicates that if the maximum of one line falls on the first minimum of a line adjacent to it, the two lines can just be recognised as distinct This, therefore, is taken as the criterion of resolution. Let two lines that a grating just resolves have wave lengths λ and $\lambda + d\lambda$. Then as $\sin\theta = \frac{mN\lambda}{E}$ (section 6) $\cos\theta\, d\theta = \frac{mN\,d\lambda}{E}$, and $d\theta$, the angle by which they are separated is $mN\,d\lambda / E\cos\theta$. Now the angle between the maximum of a line and its first minimum is $d\theta = \frac{\lambda}{E\cos\theta}$ (section 8). Then, as the two lines under consideration are just resolved, these two values of $d\theta$ must be equal, and $\frac{\lambda}{E\cos\theta} = \frac{mN\,d\lambda}{E\cos\theta}$ or $\frac{\lambda}{d\lambda} = mN$, i.e., the resolving power is the product of the order of the spectrum, and the total number of lines in the grating.

The foregoing discussion applies equally well to reflection gratings, and with slight modifications to the case of oblique illumination.

References.

Baly, Spectroscopy, Chap. VI.
Houston, Treatise on Light, p. 171 to p. 180.

Questions.

In a certain transmission grating the transparent spaces are the same width as the opaque spaces. Where on the visibility curve do the second and third order spectra lie?

Why cannot a two line grating resolve two spectral lines for which $\frac{\lambda}{d\lambda}$ is less than 6, although the expression obtained above would give 6 as the resolving power in its third order spectrum?

SECTION 11. APPLICATION TO THE ECHELON.

In this instrument some twenty plates of optically plane parallel slabs of glass, all accurately of the same thickness (about 1 cm.), are piled above one another, each one overlapping the one beneath it by about a millimetre. Light is passed normally through the pile, emerging from the overlapping ledges, which behave as the clear spaces of a grating, the beam coming from one ledge being retarded many thousands of wave lengths behind that from the ledge next it, owing to its passage through a greater thickness of glass. The spectra observed are thus of about the ten thousandth order.

The following brief discussion should be supplemented by reading. Treatments will be found in Schuster, Theory of Optics, p. 116; Baly, Spectroscopy, p. 190; Wood, Physical Optics, p. 274; and in other text books.

Figure VIII. shows two plates of an echelon. HL is a wave front for wave length λ in the m th order.

The Visibility Curve.—The width of the slit is about 1 mm. Thus (sections 3 and 4) the breadth of the central maximum of the visibility curve is given by $2\,\theta$ where $\sin\theta = \lambda/e = \frac{6 \cdot 10^{-5}}{1 \times 10^{-1}}$ Therefore $2\,\theta =$ about 4 minutes. Outside this narrow range the spectra have not sufficient intensity to be observed, consequently the echelon is only useful for examining the fine structure of a spectral line or determining the separation of two lines very close together.

Separation of the Orders.—Let HK (Figure VIII.) be a wave front for wave length λ in the $m+1$ th order, e is about 1 mm. Therefore, $d\,\theta = LK/HL = \lambda/e =$ the angle between the two orders, is about $\frac{6 \times 10^{-5}}{\cdot 1}$ i.e., is half the breadth of the central maximum of the visibility curve. The result is that there are in general two orders of every wave length in view in the field, and that the different lines are piled on one another in an inextricable jumble. It is, therefore, necessary to use the echelon in conjunction with an ordinary spectrometer or other device for selecting a narrow range of wave length.

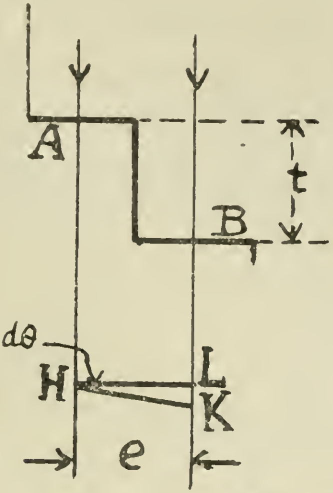


Figure VIII

Dispersion.—Let HK (Figure VIII.) be a wave front of wave length $\lambda + d\lambda$ of the m th order. From general considerations, it is obvious that the dispersion will be large, for if the refractive index of glass were a constant, LK would equal $m\,d\lambda$ [the ray BL would have an optical path $m\lambda$ and the ray BK a path $m(\lambda + d\lambda)$ longer than the ray AH], and thus with m large $d\theta$ must be relatively large. Actually the conditions are complicated by the fact that μ is a function of λ, μ being the refractive index of glass. It can very easily be shown, however (see references above), that the dispersion $d\theta/d\lambda = \dfrac{m - t\,d\mu/d\lambda}{e}$

Resolving Power.—Evidently to find the angle, $d\theta$, between the maximum of a spectral line and its first minimum, the discussion of section 8 holds without alteration, but with the restriction that θ is always very small, and therefore that $\cos\theta = 1$, very nearly. Therefore, $d\theta = \dfrac{\lambda}{E\cos\theta} = \dfrac{\lambda}{E}$

This is also the angle between the maximum of the λ line and the maximum of one of wave length $\lambda + d\lambda$ which is just resolved from it (see section 10), and this is given by the expression for the dispersion obtained in the last paragraph. Equating these two values of $d\theta$, $\frac{m\,d\lambda - t\,d\mu}{e} = \frac{\lambda}{E}$ and E = Ne where N is the number of apertures. Therefore, $\frac{\lambda}{d\lambda} = N\left(m - \frac{d\mu}{d\lambda}t\right)$ Now $t\,d\mu/d\lambda$ is almost always less than .1 of m. Thus the resolving power is given very nearly by Nm, the same expression that holds for the ruled grating.