International Journal of Theoretical Physics, Vol. 15, No. 3 (1976), pp. 169-180

The Interpretation of Diffraction and Interference in Terms of Energy Flow

R. D. PROSSER

Physics Department, University of Stirling, Stirling, Scotland

Received: 20 *December* 1974

Abstract

Solutions to Maxwell's equations at a semi-infinite plane and a double slit are used to construct lines of constant amplitude, constant phase and energy flow. The lines of energy flow show how the electromagnetic boundary conditions necessitate a particular undulation in the path of the light energy and that the consequent redistribution of energy corresponds with a diffraction or interference pattern. This interpretation complements the interpretation in terms of the interaction of secondary wavelets due to Huygens.

1. *Introduction*

The classical theory of diffraction and interference has its roots in the principle proposed nearly three hundred years ago by Christian Huygens. According to this principle every point on a wave front (i.e. a surface of constant phase) is regarded as a source of secondary spherical wavelets and the subsequent development of the wave front can be derived from the superposition of these wavelets. Fresnel assumed that the amplitude of the secondary wavelets varied as $\cos\theta$, where θ is the angle between the direction of the incident light and the direction of propagation of the wavelets. He was thus able to account for the intensity distribution of diffraction patterns, although the corresponding phase distribution was not correct. A more refined expression of Huygens' principle followed from the work of Kirchhoff who showed that it was necessary to assume that the amplitude of the wavelets varied as $1 + \cos\theta$. This modification gives results which are in good agreement with experiment at distances greater than several wavelengths from the diffraction edge.

Kirchhoff showed that the light wave at any point in space can be expressed as an integral over a closed surface surrounding that point. By allowing the surface to coincide with the diffraction screen and evaluating the integral with suitable boundary conditions the solution to any scalar diffraction problem can in principle be obtained. However this requires a knowledge of the amplitude and phase together with their derivatives with respect to the normal over the whole surface. In practice these are never known exactly so that only approximate solutions can be obtained with the help of simplifying assumptions. In Kirchhoff's theory, Huygens' wavelets appear as contributions to the integral, but their amplitude is dependent upon the boundary conditions and cannot be precisely specified so that the wavelet concept begins to lose its simple physical significance.

The complete solution to diffraction problems can be obtained from Maxwell's equations. This approach allows the polarization of the light as well as the electrical properties of the diffraction screen to be taken into account, although exact solutions have been obtained in only few cases of physical interest because of the mathematical difficulties involved. Maxwell's equations also suggest an alternative interpretation of diffraction and interference phenomena, in terms of an undulation of the light path defined by the Poynting vector, which is complementary to the wavelet concept of Huygens. In this paper we examine the solution to Maxwell's equations at a semi-infinite plane and at a double slit and show how the resulting diffraction and interference patterns arise from the nonlinear trajectory of the energy flow.

2. *The Exact Solution to Maxwell's Equations at a Semi-Infinite Plane*

The exact solution to Maxwell's equations for a uniform plane wave incident upon an infinitesimally thin, perfectly conducting sheet of semi-infinite extent and bounded by a straight edge was originally obtained by Sommerfeld (1954). A further step was taken by Braunbek and Laukien (1952) who showed that Sommerfeld's solution can be expressed in the form of intensity and phase distributions, and they plotted this solution, together with lines of energy flow, over a region extending to one wavelength from the diffraction edge. In Figure 1 similar plot has been extended to several wavelengths from the edge so that certain features of the solution can be more clearly seen.

Following Braunbek and Laukien, we consider a situation in which the incident wave travels in the positive y direction, and the semi-infinite plane lies in the xz plane with its edge along the z axis. We deal only with the case in which the magnetic vector lies in the z direction (parallel to the diffraction edge). The electric vector then lies always in the xy plane.

It follows from Sommerfeld's solution, for the particular case of normal incidence that

$$H_z = A\,[F(\sigma)\exp(-i\gamma) + F(\sigma')\exp(i\gamma)] \tag{2.1}$$

$$\left.\begin{aligned}
\gamma &= \frac{2\pi r}{\lambda}\sin\phi \\
\sigma &= 2\left(\frac{r}{\lambda}\right)^{\frac{1}{2}}\left(\sin\frac{\phi}{2} - \cos\frac{\phi}{2}\right) \\
\sigma' &= -2\left(\frac{r}{\lambda}\right)^{\frac{1}{2}}\left(\sin\frac{\phi}{2} + \cos\frac{\phi}{2}\right) \\
F(\sigma) &= \int_{-\infty}^{\sigma} \exp(-i\pi\tau^2/2)\, d\tau
\end{aligned}\right\} \quad (2.2)$$

$$\begin{aligned}
E_x &= \frac{\lambda}{2\pi i}\left(\frac{\mu_0}{\epsilon_0}\right)^{\frac{1}{2}}\frac{\partial H_z}{\partial y} \\
E_y &= \frac{-\lambda}{2\pi i}\left(\frac{\mu_0}{\epsilon_0}\right)^{\frac{1}{2}}\frac{\partial H_z}{\partial x}
\end{aligned} \quad (2.3)$$

where λ is the wavelength, and r and ϕ are polar coordinates in the xy plane. The time factor, $\exp i\omega t$, has been omitted.

Equation (2.1) can be written in the form

$$H_z = H(r, \phi) \exp[i\Psi(r, \phi)] \quad (2.4)$$

which enables one to preserve the amplitude H and phase Ψ of the magnetic field. The reduction is somewhat cumbersome and is not reproduced here, but it follows directly from the preceding equations.

The components of the Poynting energy flow vector which represents the energy flow averaged over one cycle are

$$\begin{aligned}
S_x &= \tfrac{1}{2}\, Re\,(H_z \cdot E_y^*) \\
S_y &= -\tfrac{1}{2}\, Re\,(H_z \cdot E_x^*)
\end{aligned} \quad (2.5)$$

where Re refers to the real component and the * superscript indicates the complex conjugate.

The differential equation for the energy flow lines which indicate the direction of the energy flow is

$$\frac{dr}{d\phi} = \frac{r(S_x \cos\phi + S_y \sin\phi)}{(S_y \cos\phi - S_x \sin\phi)} \quad (2.6)$$

which can be integrated after inserting functions from equations (2.1) (2.3) and (2.5).

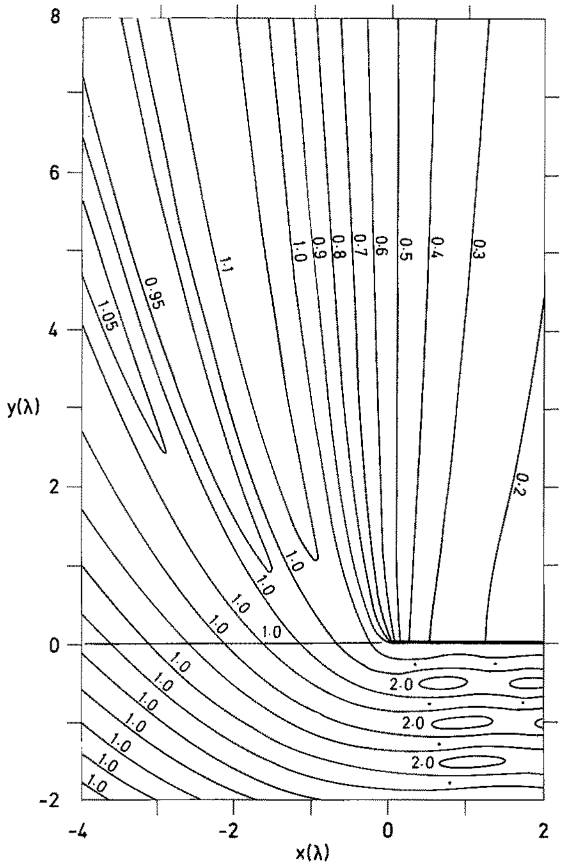
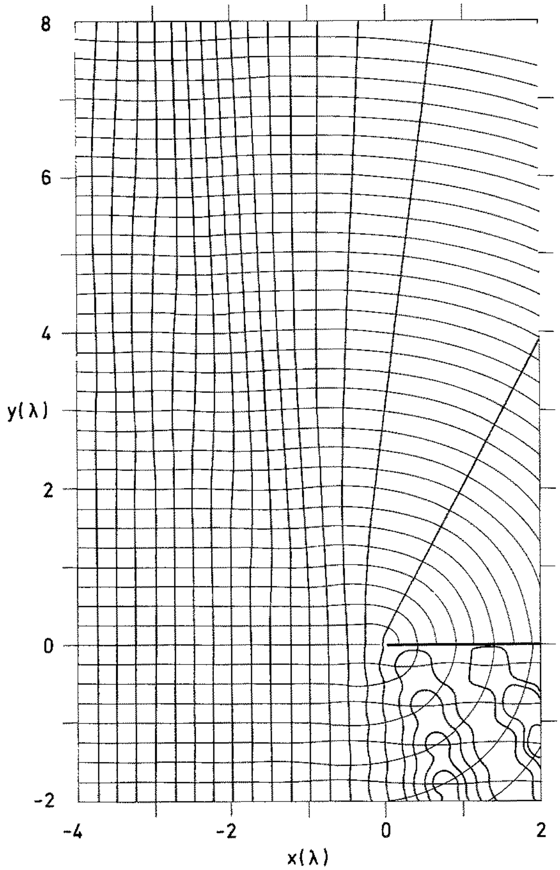


Figure 1—Solution for the semi-infinite plane: (a, above) lines of constant amplitude; (b, facing page) lines of constant phase (——) and lines of energy flow (——).

The lines of constant amplitude are shown in Figure 1a. The lines of constant phase are drawn as thin lines in Figure 1b. The interval between neighbouring lines is $\pi/2$ radians. At the points of intersection on the bright side of the screen the phase contours undergo a discontinuous change of π radians. These points correspond with the points of zero amplitude in Figure 1a. For full details of the absolute magnitude of the phase the large scale plots of Braunbek and Laukien (1952) should be consulted. The lines showing the direction of energy flow are drawn as thick lines in Figure 1b. It can be seen that these lines are everywhere orthogonal to the lines of constant phase, as follows from equations (2.3) and (2.5).

Figures 1a and 1b were obtained with the aid of a computer and an associated graph plotter. The expressions for H and Ψ, representing the amplitude and phase of the magnetic field in equation (2.4) were first



obtained in terms of x and y coordinates by appropriate substitutions from equations (2.1) (2.2) (2.3). The Fresnel integrals in these expressions arising from equations (2.2) were calculated by means of a subroutine taken from the program library at CERN and accurate to 8 significant figures. The line of constant phase passsing through a given point was constructed by calculating the gradients $\delta\Psi/\delta x$ and $\delta\Psi/\delta y$. Since the phase increment corresponding to the position increments δx and δy is given by

$$\delta\Psi = \frac{\delta\Psi}{\delta x}\,\delta x + \frac{\delta\Psi}{\delta y}\,\delta y \tag{2.7}$$

the gradient of the line of constant phase, for which $\delta\Psi = 0$, is $-(\delta\Psi/\delta x)/(\delta\Psi/\delta y)$. A step was taken in this direction and a new point established. The process was repeated starting from the new point and thus an array of points on the constant phase contour was derived. Any error due to the finite size of the step was corrected by adjustment of the subsequent step. The step size was varied to suit the curvature of the contour, and was normally set to 0.05λ. The automatic graph plotter was used to draw a smooth curve through the set of points. The constant amplitude contours were constructed in a similar manner using the amplitude gradients $\delta H/\delta x$ and $\delta H/\delta y$. The lines of energy flow were constructed so as to be orthogonal to the lines of constant phase. Errors in the plotted figures are not greater than 0.02λ.

3. *Solution for the Single Slit*

In order to deal with the double slit, we shall first construct the solution for the single slit as a superposition of solutions for the half plane and establish the necessary notation.

Equations (2.3) and (2.4) represent the complete solution in polar coordinates for a half plane which extends from $x = 0$ to $x = +\infty$. This solution is represented by ${}_{+}\Phi_0$ (see Figure 2).

The solution for the case in which the half plane extends from $x = a$ to $x = +\infty$ is represented by ${}_{+}\Phi_a$; this solution is obtained by substituting $x - a$ for x in ${}_{+}\Phi_0$. The solution for the case in which the half plane extends from $x = b$ to $x = -\infty$ is represented by ${}_{-}\Phi_b$; this solution is obtained by substituting $b - x$ for x in ${}_{+}\Phi_0$. The solution for free space is represented by Φ_f; this solution is the limiting form of ${}_{+}\Phi_0$ as $x \to -\infty$. The solution for the case of a perfectly conducting and infinite plane is represented by Φ_r; this is the limiting form of ${}_{+}\Phi_0$ as $x \to +\infty$.

The field components relating to the single slit are denoted by a single prime superscript, e.g. E_x', where distinction is required. For the double slit, a double prime is used.

It is well known that any sum of solutions to Maxwell's equations is itself another solution. We now consider the solution

$$ {}_{+}\Phi' = {}_{+}\Phi_a + {}_{-}\Phi_b - \Phi_f \tag{3.1}$$

and show that it satisfies the electromagnetic boundary conditions for a slit in an infinite plane. The plane is infinitesimally thin, perfectly conducting and perpendicular to the y axis. The slit edges are at $x = a,b$ and run parallel to the z axis. As in the case of the half plane we consider the solution in

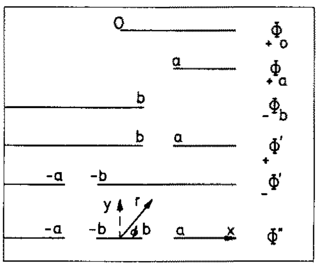


Figure 2—Physical situations corresponding to the solutions.

which the magnetic vector H_z' lies only in the z direction and the electric vectors E_x', E_y' lie always in the xy plane. The boundary conditions require that E_x' should be zero at the conducting surfaces; elsewhere all the field components should be continuous.

It can be seen from the graphical solution for the half plane that if $x < b$ and $y = 0$, then $\partial H_z/\partial y$ in ${}_{+}\Phi_a$ is approximately equal to $\partial H_z/\partial y$ in Φ_f. This is because ${}_{+}\Phi_a \rightarrow \Phi_f$ as $x \rightarrow -\infty$. From equations (2.3) and (3.1) it follows that E_x' is approximately equal to E_x in ${}_{-}\Phi_b$. Now E_x in ${}_{-}\Phi_b$ is zero for $x < b, y = 0$, since this is the region which corresponds with the conducting surface of the half plane. Hence we have that for $x < b, y = 0, E_x'$ is approximately zero. By a similar argument it follows that for $x > a, y = 0, E_x'$ is again approximately zero. Thus we find that at the conducting surfaces forming the slit the boundary condition for E_x' is approximately satisfied. The continuity requirement for the field components H_z', E_x' and E_y' is also satisfied in ${}_{+}\Phi'$ since this requirement is satisfied by each of the terms on the right hand side of equation (3.1), and we can conclude that ${}_{+}\Phi'$ represents the solution for the single slit.

It is possible to obtain an exact solution for the single slit in terms of Mathieu functions. The advantage of the approximate solution ${}_{+}\Phi'$ is that it contains only trigonometric functions and Fresnel integrals which can be rapidly and accurately calculated by computer. We shall leave the physical justification for the use of the approximate solution until later.

4. *Solution for the Double Slit*

Form of the Solution.

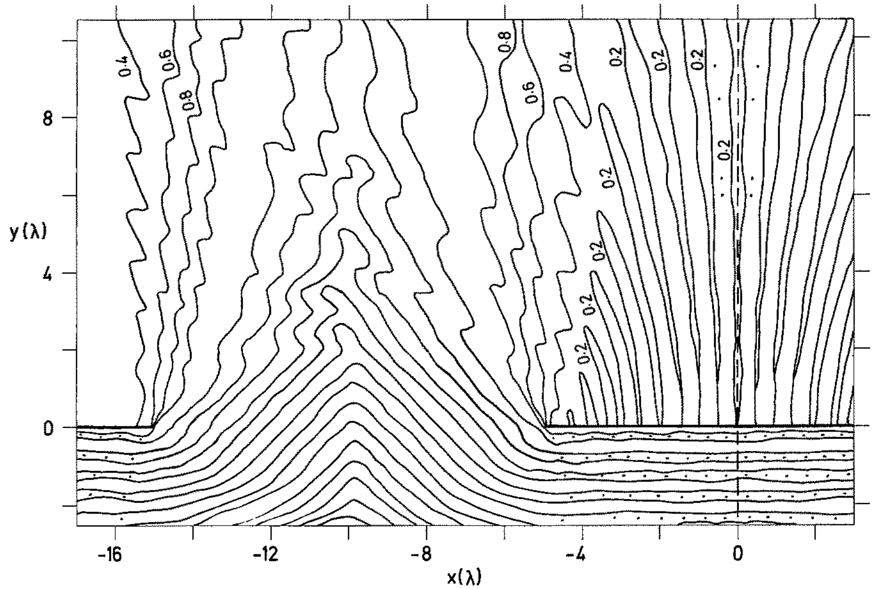
Let ${}_{-}\Phi'$ represent the solution for a single slit with edges at $x = -a, -b$ (Figure 2). Then the solution

$$\Phi'' = {}_{+}\Phi' + {}_{-}\Phi' - \Phi_r \tag{4.1}$$

is the exact solution for the double slit provided that Φ' represents the exact solution for the single slit. In ${}_{-}\Phi', E_x'$ is zero everywhere in the plane of the slit except in the aperture where $-a < x < -b$. Similarly in ${}_{+}\Phi', E_x'$ is zero everywhere in the plane of the slit except where $b < x < a$. Hence in the summations represented by equation (4.1), E_x'' is zero everywhere in the plane $y = 0$ except where $-a < x < -b$ and where $b < x < a$, i.e., E_x'' is zero everywhere on the conducting surfaces of the double slit arrangement shown in Figure 2. In order that the field components elsewhere should be continuous the term in Φ_r is required; otherwise there would be a discontinuity in the apertures of the slits.

Application to a Particular Case.

As in the case of the half plane the solution for the double slit can be expressed in the form of contours of constant amplitude and phase. To illustrate the application of the solution, we consider the case of two slits



(a)

Figure 3—Solution for a double slit. (a) Lines of constant amplitude. Unlabeled lines have unit amplitude. Points of zero amplitude are shown by a dot; (b, facing page) Lines of constant phase. (c, facing page) Lines of energy flow. The axis of symmetry is represented by the dashed line.

each having a width of 10 wavelengths and a separation of 20 wavelengths between their respective centres.

The lines of constant amplitude are shown in Figure 3a. The pattern consists of the rapid fluctuations associated with interference upon which are superimposed the more slowly varying fluctuations which characterize diffraction. It can be shown that at an infinite distance from the slits the amplitude distribution corresponds with the usual classical formula for interference. An interesting feature of Figure 3a is that the points of zero amplitude, which are found only on the bright side of the conducting surface in the case of the half plane, are now found on the shadow side also.

The lines of constant phase are shown in Figure 3b at intervals of π radians. The discontinuities in the lines, which occur at the intersection with points of zero amplitude, indicate a discontinuous change in phase of π radians. As in the case of the half plane, there are again undulations in the phase lines and these become very marked in the region between the slits.

The lines of energy flow are determined by a differential equation exactly analogous to that used for the half plane. These lines indicate the direction of the Poynting vector, corresponding with the direction of energy flow at any point, and are shown in Figure 3c.

The method used for constructing the lines of constant phase and energy flow was similar to that used previously for the half plane. A computer was

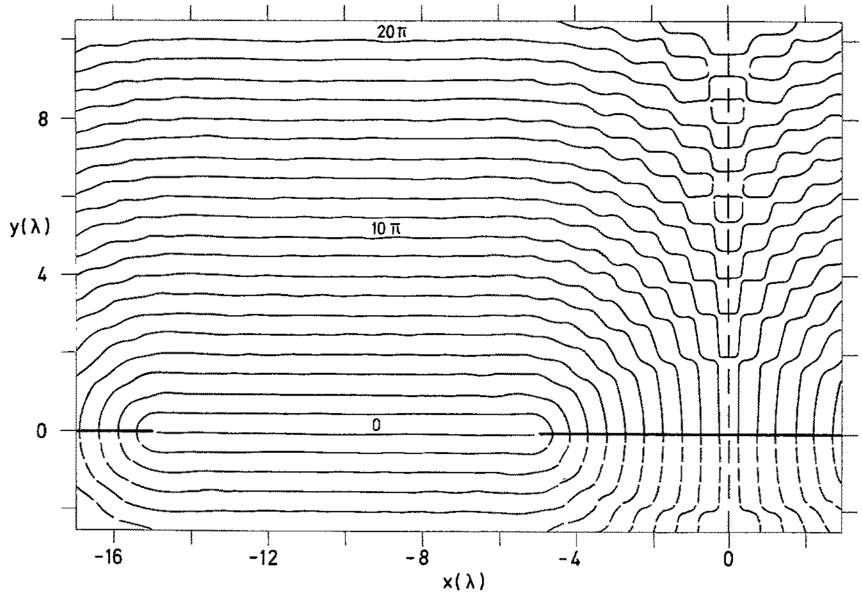


Figure 3b.

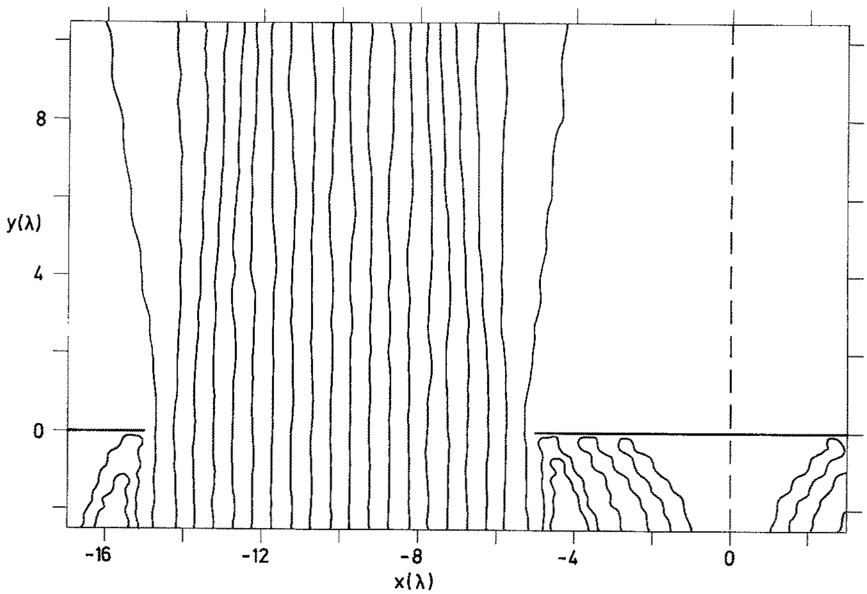


Figure 3c.

used to construct an array of points on the contour and this information was fed to an automatic graph plotter so that a continuous curve was drawn between the points of the array. The interval between the points was varied to suit the curvature of the contour and was normally set to 0.05λ. Plotting errors in these curves are not greater than 0.02λ. The lines of constant amplitude were constructed manually by interpolation from a tabular form of the solution, since the features of this diagram do not merit very precise reproduction. It is accurate to about 0.1λ.

Finally we return to the question of the error in the solution itself due to the use of an approximate solution for the single slit. We have already noted that the boundary conditions require that E_x'' should be zero at the conducting surfaces. Figure 4 shows the value of E_x'' derived from the solution for the shadow side of the conducting surface in the region between the slit edge and the axis of symmetry where the error will be greatest. It can be seen that at the conducting surface E_x'' falls sharply to a value of about 10^{-3} of that in the centre of the slit aperture. A similar result is obtained for the bright side of the conducting surface. In practice it is not possible to realize the physical conditions of the solution exactly since any real material must possess finite conductivity and thickness. For a real conducting material E_x'' will consequently not be zero at the surface. In the case of an infinite reflecting plane made of copper it can be shown by the usual methods that, for visible wavelengths, the value of E_x at the surface is about 2×10^{-2} of that in the incident wave. Hence the errors in the solution are unlikely to be greater than those which would inevitably arise from the application of the exact solution for the idealized case of a perfectly conducting and infinitely

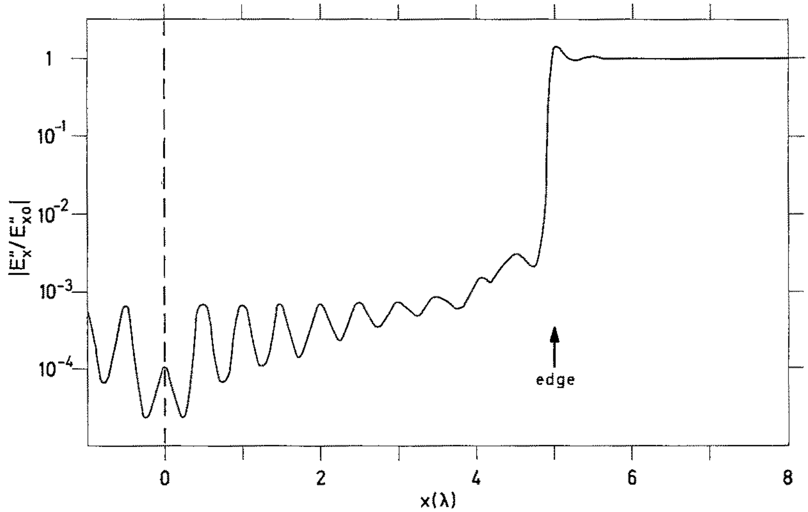


Figure 4—Tangential field E_x'' in the plane of the slits (shadow side). E_{x0}'' is the value of E_x'' in the center of the slit aperture.

thin material to a practical situation. The approximate solution described should generally be satisfactory in cases where the slit widths are greater than one wavelength.

5. *Discussion of the Solutions with Respect to Diffraction and Interference Phenomena*

The lines of energy flow which delineate the path of the light rays in the vicinity of a diffraction edge are shown in Figure 1b. In the unobstructed region beyond the half plane, where x is negative and diffraction fringes occur, it can be seen that the lines weave alternately to left and right. The departure from a straight path is only a fraction of a wavelength, and we see that the idea that light travels in a straight line is very nearly true, yet it is just this weaving trajectory which enables the light energy to be redistributed in the form of a diffraction pattern. Comparison with Figure 1a shows that the flow lines bunch together at the diffraction maxima, while at the minima they are most widely dispersed. In the shadow region beyond the plane the flow lines become straight, corresponding with the absence of fringes in this area. In the reflection region on the bright side of the plane the lines undergo a reversal, rotating about points of zero amplitude. Thus all the features of the diffraction and reflection brought about by the plane can be understood in terms of the trajectory of energy flow.

A similar situation obtains in the case of interference at a double slit, as can be seen in Figure 3. The weaving of the flow lines is more pronounced than for the half plane, corresponding with the greater variations in amplitude. There appears to be no obvious systematic pattern in this case, although as before the lines have the property that they are most densely gathered together at regions of maximum amplitude, while at the minima they are most widely dispersed. By reason of slight undulations in their trajectories, the flow lines converge and diverge in such a manner that the energy is distributed according to the characteristic pattern of interference.

The solution in Figure 3 covers only the Fresnel region close to the slits whereas most interference experiments are conducted in the Fraunhofer region far from the slits. The extension of the plots into this region would show that as the distance from the slits increases the bunching of the flow lines becomes more marked and the amplitude of the fringes increases until it corresponds with the Fraunhofer formula for interference at infinity.

An interesting feature of Figure 3c is that no flow lines cross the axis of symmetry between the slits. This is related to the fact that the phase lines always cross the axis at right angles, and hence the flow lines which are orthogonal to the phase lines cannot cross it. Therefore the energy which illuminates the region to the left of the axis of symmetry passes entirely through the left hand slit, and no energy passes into this region from the right hand slit.

These studies of the solutions to Maxwell's equations show that it is possible to interpret diffraction and interference in terms of the pattern of energy flow. We may summarize this interpretation in the following way. In the absence of obstacles the rays of light energy follow a linear trajectory. The introduction of an obstacle, such as a plane which may contain slits, alters the electromagnetic boundary conditions in that plane and this causes the light rays to deviate from their straight path. This deviation may take the form of a change in the direction of linear propagation, as when light enters the region of shadow behind a diffracting screen, or it may take the form of undulations in the trajectory. In this latter case there will be a resultant concentration of energy in places where the light rays converge and a corresponding depletion where they diverge. This redistribution of energy due to the nonlinear trajectory is what is observed as a diffraction or interference pattern.

References

Braunbek, W. and Laukien, G. (1952). *Optik*, 9, 174.

Sommerfeld, A. (1954). *Lectures on Theoretical Physics*, Vol. 4, Ch. 5, pp. 247–72. Academic Press, New York. (Note: Sommerfeld's solution for the semi-infinite plane was originally published in German in 1896 in *Math. Ann.* 47, 317.)