The diffraction theory of waves yields cumbersome mathematical expressions (typically integrals of the complex amplitudes) that in only a few cases can be solved analytically (as for the Fraunhofer or far-field conditions). This fact is an important drawback to using the theory in many practical cases. Moreover, the limits of validity of several widespread approximations (as the Fresnel expressions) are not well established or have inherent limitations for points near the shadow of the aperture borders (as the geometrical theory of diffraction or the stationary phase approach). In the computer era, an approach based on the exact computation of the relevant integrals could be conceived, but up to now direct numerical evaluation of the integrals involved in diffraction theory has not been considered as a practical operating method for a general case. That this is so is due to (i) the highly oscillatory nature of the function to be integrated inside (ii) the two-dimensional aperture. Our present objective is to set up a general procedure for the evaluation of Eq. (1) by use of only elementary numerical integration techniques. The basis of our method is geometrical: determination of the zones in which the integrand does not oscillate. Subsequently the integration is performed at the intersection of these zones with the aperture. Moreover, our approach (since it computes the exact case of the exact diffraction formula, i.e., Eq. (1)) can also be used to test the validity of any approximation. First, the exact point spread function for the pinhole camera is calculated. Second, the computation of the diffraction pattern at the focusing plane of a monochromatic converging spherical wave limited by a centered circular diaphragm is analyzed (and the validity of the Airy pattern is checked).

As mentioned above, the key point in our approach is to isolate the oscillations of the integrand in space. Owing to the linearity of Eq. (1) with respect to $U(x,y)$, we first note that any monochromatic disturbance $U(x,y)$ that reaches the aperture, $\Sigma$, may be represented by a sum of (diverging or converging) spherical waves, and our problem is reduced to

$$
U(P) = \frac{1}{2\pi} \iint_{\Sigma} U(x,y) \frac{\exp(ikr_P)}{r_P} \times \left(\frac{1}{r_P} - ik\right) \cos(\hat{n},r_P) \, dx \, dy,
$$

where $\Sigma$ is in the plane $z = 0$, $k = 2\pi/\lambda$, $r_P$ is the vector from $P$ to $(x,y)$, and $\hat{n}$ is the normal to the aperture at $(x,y)$ away from the volume enclosing $P$. The calculation of Eq. (1) (both analytically or numerically) is very difficult because of its inherent complexity. Usually the relation between the dimensions of $\Sigma$ and $A$ and the distance from $S$ and $P$ is such that the integrand is highly oscillatory. Consequently, approximations to Eq. (1) are mandatory. Some of these simplifications, such as neglecting the obliquity factor $\cos(\hat{n},r_P)$, are conceptually simple, but usually the limits of validity of the approximation procedures are unclear and (or) the resulting integrand is still highly oscillatory within $\Sigma$.

To our knowledge, no attempts at numerical computation of Eq. (1) for a general configuration have been reported. Our previous work was a limited development along this line and was illustrated for a rectangular aperture. Our present objective is to set up a general procedure for the evaluation of Eq. (1) by use of only elementary numerical integration techniques. The basis of our method is geometrical: determination of the zones in which the integrand does not oscillate. Subsequently the integration is performed at the intersection of these zones with the aperture. Moreover, our approach (since it computes the exact case of the exact diffraction formula, i.e., Eq. (1)) can also be used to test the validity of any approximation. First, the exact point spread function for the pinhole camera is calculated. Second, the computation of the diffraction pattern at the focusing plane of a monochromatic converging spherical wave limited by a centered circular diaphragm is analyzed (and the validity of the Airy pattern is checked).

As mentioned above, the key point in our approach is to isolate the oscillations of the integrand in space. Owing to the linearity of Eq. (1) with respect to $U(x,y)$, we first note that any monochromatic disturbance $U(x,y)$ that reaches the aperture, $\Sigma$, may be represented by a sum of (diverging or converging) spherical waves, and our problem is reduced to

$$
U(P) = \frac{1}{2\pi} \iint_{\Sigma} U(x,y) \frac{\exp(ikr_P)}{r_P} \times \left(\frac{1}{r_P} - ik\right) \cos(\hat{n},r_P) \, dx \, dy,
$$

where $\Sigma$ is in the plane $z = 0$, $k = 2\pi/\lambda$, $r_P$ is the vector from $P$ to $(x,y)$, and $\hat{n}$ is the normal to the aperture at $(x,y)$ away from the volume enclosing $P$. The calculation of Eq. (1) (both analytically or numerically) is very difficult because of its inherent complexity. Usually the relation between the dimensions of $\Sigma$ and $A$ and the distance from $S$ and $P$ is such that the integrand is highly oscillatory. Consequently, approximations to Eq. (1) are mandatory. Some of these simplifications, such as neglecting the obliquity factor $\cos(\hat{n},r_P)$, are conceptually simple, but usually the limits of validity of the approximation procedures are unclear and (or) the resulting integrand is still highly oscillatory within $\Sigma$.
isolating the oscillations of the integrand in Eq. (1) for a component spherical wave. This question was already formulated in Ref. 9 in terms of a source point, S, a calculation point, P, and generalized semiperiodic zones in which the relevant function did not oscillate. The semiperiodic zones are the intersections between the aperture plane and the surfaces,

$$r_p \pm r_S = \overline{SP} \pm n\lambda/2,$$

(2)

where \(n\) is the index of the zone, \(r_p\) and \(r_S\) are the distances to \(P\) and \(S\), respectively, and \(\overline{SP}\) is the distance from \(S\) to \(P\). The plus sign corresponds to a diverging wave (intersections between aperture plane and ellipsoids); the minus sign, to a converging wave (intersections between aperture plane and hyperboloids). In this Letter we complete our previous developments by use of the geometrical approach described above to propose a computation scheme that is valid for any diffraction geometry. This scheme is illustrated through two examples, a pinhole camera and a converging wave through a circle. It is worth mentioning that computing the second case will require extension of the procedures presented in Ref. 9 to the important case (not included there) in which \(P\) and \(S\) lie in a plane parallel to the aperture.

The pinhole camera is the simplest image-forming optical instrument: a small circular hole in front of a screen. Given wavelength \(\lambda\) and distance \(d\) between the hole and the screen, only one parameter can be optimized to enhance image quality: the diameter of the hole. Image quality can be defined by the point spread function of the instrument, i.e., the image of a point object. In the case of a distant point, there is a point object. 2

Given wavelength \(\lambda\) and distance \(d\) between the hole and the screen, only one parameter can be optimized to enhance image quality: the diameter of the hole. Image quality can be defined by the point spread function of the instrument, i.e., the image of a point object. 2 In the case of a distant point, there is a point object. 2

In the focusing plane of a wave converging through a circular aperture. As mentioned above, our previous approach did not include this case, since the line joining source \(S\) and calculation point \(P\) does not intersect the aperture plane. If the distance from the aperture to the image plane is \(d\) and the aperture diameter is \(\phi\), one crucial parameter here is the f-number, \(f = d/\phi\). Figure 2 illustrates the situation for \(f = 1\). Given a focus point \(S\) and a calculation point \(P\), the solid curves show the cross section (in a plane perpendicular to the aperture) of the branches of the hyperboloids defined by \(r_S - r_P = \overline{SP} - n\lambda/2\). The intersections of these hyperboloids and the aperture plane define the semiperiodic zones. When analyzing the figure, we note that \(d \gg \lambda\) (required even for the validity of the scalar theory of diffraction), whereas \(\overline{SP} \approx \lambda\).

The numerical integration in the intersection of any single semiperiodic zone and the diffracting aperture is made by use of a two-dimensional Romberg algorithm. This allows one to control the fractional accuracy within each zone. 8 Finally, the results for the zones are simply added; the diamonds in Fig. 1 represent the exact FWHM. For radius \(r\) in the range 0.18–0.34 mm, the FWHM decreases smoothly from 0.18 to 0.1 mm (the points under the hyperbola). A further increase of the size of the aperture leads to non-monotonically decreasing peaks (with rings and strong oscillations) that make the definition of the FWHM useless (the FWHM’s for aperture radii of 0.4, 0.45, 0.5, and 0.55 mm are shown in Fig. 1 anyway), indicating that the point image quality would decrease dramatically. Accordingly, \(r_2 = \sqrt{\lambda d}\) can be considered the optimum size, and values \(r > r_2 = \sqrt{\lambda d}\) and \(r < r_1 = 0.245\lambda d\) are inadequate. These results are in agreement with the images shown in Ref. 12.

One of the most common configurations in optics and, broadly speaking, in diffraction theory, is imaging in the focusing plane of a wave converging through a circular aperture. As mentioned above, our previous approach did not include this case, since the line joining source \(S\) and calculation point \(P\) does not intersect the aperture plane. If the distance from the aperture to the image plane is \(d\) and the aperture diameter is \(\phi\), one crucial parameter here is the f-number, \(f = d/\phi\). Figure 2 illustrates the situation for \(f = 1\). Given a focus point \(S\) and a calculation point \(P\), the solid curves show the cross section (in a plane perpendicular to the aperture) of the branches of the hyperboloids defined by \(r_S - r_P = \overline{SP} - n\lambda/2\). The intersections of these hyperboloids and the aperture plane define the semiperiodic zones. When analyzing the figure, we note that \(d \gg \lambda\) (required even for the validity of the scalar theory of diffraction), whereas \(\overline{SP} \approx \lambda\).

![Fig. 1. Exact calculation of the optimal aperture radius in a pinhole camera.](image-url)
Thus (i) the semiperiodic zones on the aperture are virtually defined by the asymptotes of the hyperbolas, and (ii) the most relevant parameter will be the \( f \)-number. The dotted lines in Fig. 2 indicate a cone with its vertex in \( S \) defined by any circular aperture with \( f = 1 \), and a diaphragm at \( d = 10 \) is also represented.

To calculate the diffracted amplitude in the present configuration requires finding the intersections of each branch of the hyperboloid [for each allowed \( n \) in Eq. (2) with the minus sign] and the aperture, integrating inside each semiperiodic zone (using the numerical algorithms cited above) and adding the results. Thus the final precision attained is defined by the fractional accuracy within each zone times the number of semiperiodic zones involved.

For \( f \)-numbers 32, 16, 8, 4, 2, and 1, we computed the diffracted amplitudes according to Eq. (1) and neglected the obliquity factor. Comparing our results with the well-known theoretical result for the far-field (Fraunhofer) approximation (Airy pattern, \( J_1(\rho)/\rho \), where \( J_1 \) is the Bessel function of the first kind and \( \rho \) the normalized distance to \( S \)), no visual differences are apparent for \( f = 32, \ldots, 4 \); for \( f = 2 \) and \( f = 1 \) some differences are evident (mainly for \( f = 1 \)). Figure 3 shows the comparison in the vicinity of the first zero for \( f = 1 \). The solid curve in Fig. 3 shows the Fraunhofer limit, the circles correspond to the Rayleigh–Sommerfeld integral (1), and the diamonds are computed without the obliquity factor \( \cos(\hat{n},\mathbf{r}_P) \). The intensity was normalized so that the value of \( \rho = 0 \) mm is 1 in all cases. Thus we are able to show (since our method uses no approximation) that the exact Eq. (1), when we neglect its obliquity factor, and the Fraunhofer formula lead to appreciably different results. A careful experiment could reveal which of these is the most appropriate theoretical approach.

In summary, the procedures that we have presented constitute a new approach to the study of diffraction phenomena and make feasible the numerical integration of the relevant formulas in diffraction theory, such as Eq. (1), by the definition of the semiperiodic zones for any diffraction configuration. The method was applied to a pinhole camera (and compared with experimental results) and was also used to show the limits of validity of the classical Fraunhofer diffraction formula for a circle.

S. Bosch’s e-mail address is sbp@optica.ub.es.

References