

Article

The Fresnel Approximation and Diffraction of Focused Waves

Colin J. R. Sheppard 

Department of Optometry and Vision Science, UNSW Medicine, Sydney, NSW 2052, Australia;
colinjrsheppard@gmail.com

Abstract: In this paper, diffraction of scalar waves by a screen with a circular aperture is explored, considering the incidence of either a collimated beam or a focused wave, a historical review of the development of the theory is presented, and the introduction of the Fresnel approximation is described. For diffraction by a focused wave, the general case is considered for both high numerical aperture and for finite values of the Fresnel number. One aim is to develop a theory based on the use of dimensionless optical coordinates that can help to determine the general behaviour and trends of different system parameters. An important phenomenon, the focal shift effect, is discussed as well. Explicit expressions are provided for focal shift and the peak intensity for different numerical apertures and Fresnel numbers. This is one application where the Rayleigh–Sommerfeld diffraction integrals provide inaccurate results.

Keywords: diffraction theory; focusing; Fresnel approximation; focal shift; Fresnel number; beam propagation; laser beam systems; micro-optics; diffractive optics; planar optics; teraHertz optics; microwave lenses

1. Introduction

Diffraction by a circular aperture of a uniform plane wave, or focused wave, is an important fundamental topic in optics with numerous applications. Hundreds of papers have been written about this area. Many of these papers have been concerned with applications in acoustics, and a book *Scalar Diffraction from a Circular Aperture* by Daly and Rao was even published, concentrating especially on the time domain behavior [1]. Another classic source in the optics discipline is the book *Waves in Focal Regions* by Stamnes, which stresses asymptotic approaches to evaluating diffraction integrals [2].

The aim of the present paper is to reconsider diffraction of monochromatic plane waves, and in particular focused spherical waves, by a circular aperture. We limit our treatment to the scalar case for simplicity, as it will be seen that even the scalar case is quite complicated. Osterberg and Smith calculated the axial intensity for the focused case using the first Rayleigh–Sommerfeld diffraction formula (RSI) and found that under some circumstances the maximum intensity could be situated further from the aperture than the geometrical focus [3]. Li pointed out that because experimentally the maximum should be closer to the aperture, RSI may not always be accurate, as the assumed field in the aperture will not be correct [4]. He instead calculated the focal shift using a boundary diffraction wave theory. In this paper, we consider different forms of the Fresnel approximation applied to Kirchhoff diffraction theory for diffraction of focused spherical waves. Optical systems with different geometry, dimensions, and wavelength can be compared by introducing normalized optical coordinates.

These results have potential applications in the areas of laser beam systems, micro-optics, diffractive optics, planar optics, teraHertz optics, and microwave lenses.

The obvious further work is to extend the treatment to the vectorial case. The topic of diffraction is very diverse, and any paper cannot cover all areas. Other modern research which could be related to the present study includes diffraction by disks, rather than apertures [5]; diffraction of vortex beams [6,7]; and diffractive optics and meta-lenses [8–10].



Citation: Sheppard, C.J.R. The Fresnel Approximation and Diffraction of Focused Waves. *Photonics* **2024**, *11*, 346. <https://doi.org/10.3390/photonics11040346>

Received: 29 February 2024

Revised: 5 April 2024

Accepted: 7 April 2024

Published: 9 April 2024



Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

2. Historical Background

Fresnel diffraction theory is taught in most optics courses and optics text books. Fresnel proposed the principle of half-period zones (~ 1818) as a mathematical model for Huygens' principle of secondary waves (~ 1678). The Fresnel number is the number of half-period zones of which an even number tend to cancel by destructive interference. As many papers have discussed Fresnel diffraction, in this section we identify a selection of some key papers in its historical development.

We mention first the work of Young, who proposed a boundary diffraction wave theory (BDW) of diffraction in which the diffracted field can be considered as a summation of a direct beam and a wave scattered from the boundary of an aperture [11].

Airy derived the diffracted intensity for a circular aperture, the Airy disk [12]; his result was in the form of a power series expansion, as this predated the general use of Bessel functions.

Huygens' principle was placed on a rigorous mathematical basis by Kirchhoff, who applied Green's theorem over a closed surface to provide the Kirchhoff integral theorem (~ 1882), although Helmholtz had previously derived a related result in the frequency rather than the time domain (~ 1859). Kirchhoff went on to propose that for diffraction by an aperture in a screen, the field and its normal derivative within an aperture can be assumed equal to their values in the absence of the screen, while the field on the screen and elsewhere on the closed surface of integration can be taken as zero. These are known as the Kirchhoff (or physical optics) boundary conditions. The combination of the Kirchhoff integral theorem and Kirchhoff boundary conditions is called the Kirchhoff diffraction theory. This theory can be applied over a variety of different surfaces, including a planar surface within the aperture (which we call Kp) or over a spherical wavefront (Ks).

In the conventional Fresnel approximation, a square root in an exponent in the diffraction integral is approximated using a power series expansion. Fresnel theory for diffraction by a circular aperture results in integrals that can be evaluated analytically in terms of Lommel functions of two variables, $U_n(u, v)$, $V_n(u, v)$ [13,14]. The functions $V_n(u, v)$ are used in the illuminated region, as they converge quickly, while the functions $U_n(u, v)$ are used in the shadow region. The two variables (transverse (v) and axial (u), respectively) are so-called optical coordinates, dimensionless quantities that allow the resulting intensity to be expressed in a way that is independent of the wavelength and the system geometry and dimensions. This property is quite useful, as even if the field can be computed for specific cases it is difficult to appreciate the overall performance and trends. The same functions can be used to investigate the important cases of diffraction by a circular aperture of a collimated wave or a focused wave in the paraxial approximation. For illumination with a collimated plane wave, the axial optical coordinate u is related to the Fresnel number N_z (the number of half-period zones) $u/2\pi = N_z = a^2/\lambda z$, where a is the radius of the circular aperture, z is axial distance, and λ is the wavelength. The transverse optical coordinate is $v = kaR/z$, where $k = 2\pi/\lambda$ and R is the cylindrical radius, meaning that $v/u = R/a$. Nijboer developed an alternative expansion for the focal field of a lens with circular aperture [15]. Boersma proposed that this expansion can be used for efficient numerical calculation of the Lommel functions [16].

Lord Rayleigh (1891) conducted an analysis of the pinhole camera (camera obscura) [17]. He found that light diffracted by a circular aperture illuminated with collimated light comes to a focus at a distance corresponding to about $N_z = 1$, which corresponds to the situation where the phase difference between marginal and axial rays is equal to $\lambda/2$, meaning that there is a single Fresnel zone. Note that an aperture illuminated with collimated light is analogous to a lens with infinite focal length, and the focus is closer to the lens than infinity, meaning that this can be considered as a simple example of focal shift. Rayleigh went on to say that if the distance is increased such that $N_z = 0.5$, then the phase difference is reduced to $\lambda/4$, which reduces the wavefront aberration (defocus) but also reduces the brightness.

Rayleigh (1897) derived the Rayleigh diffraction formulae, in which the integrals are performed over an infinite planar surface [18]. Rayleigh considered an aperture illuminated

with a plane wave; the more general case where Kirchhoff-like boundary conditions are applied results in the Rayleigh–Sommerfeld diffraction formulae [19]. There are two Rayleigh–Sommerfeld diffraction formulae, which we call RSI and RSII. Note that the literature does not seem to be consistent in naming the two Rayleigh–Sommerfeld integrals, which seems to result from the fact that Rayleigh considered RSII first, as it is physically relevant for acoustic waves. Our notation (the most common) is that RSI uses a known field on the integration boundary surface and RSII uses a known field derivative. The predicted field from Kp is the mean of RSI and RSII: $U_{Kp} = (U_{RSI} + U_{RSII})/2$. The difference between RSI and RSII lies in the resulting obliquity factors. Under many practical conditions, they provide identical results; however, in the far field of a small simple planar source, the field predicted by RSI has a factor $\cos \theta$, where θ is the angle of diffraction, equivalent to the planar source behaving as an axial dipole.

Debye (1909) proposed his diffraction theory of focusing, which is basically an angular spectrum theory in which the aperture is assumed to be in the far field relative to the focus [20]. The theory is often further approximated into a paraxial Debye theory [14].

In the acoustics discipline, several papers on the diffraction of convergent beams were published that do not seem to have become well known by the optics community. Williams investigated the acoustic field excited by a concave piezoelectric crystal [21]. He found that the peak in intensity did not coincide with the geometrical focal point, but was displaced towards the crystal. This effect is now known in optics as focal shift. Later papers in acoustics were by Fein, who mentioned that [22]

“Focusing measurements show agreement with A. O. Williams’ prediction that the point of maximum acoustic intensity in the radiation pattern is not necessarily at the centre of curvature of the crystal,”

by O’Neil, who analyzed the system and found that [23]

“The point of greatest intensity is not at the centre of curvature . . . ,”

and by Lucas and Muir [24]. All of these papers found a focal shift towards the aperture, i.e., a negative distance from the geometrical focus.

Andrews studied diffraction of a plane wave by a circular aperture using microwave experiments and Kirchhoff diffraction theory [25]. He found that the positions of the intensity maxima and minima were satisfactorily predicted by the Fresnel zone theory even if the field within the aperture varies in amplitude.

Based on spheroidal wave functions, Bouwkamp developed a rigorous theory of diffraction of an acoustic plane wave by a small circular hole [26]. He then developed the analogous theory for diffraction of electromagnetic waves by a hole in a perfect conductor [27]. Numerical calculation of these results becomes more difficult as the diameter of the hole becomes larger than the wavelength. These results were discussed in more detail in a paper which reprinted his PhD dissertation [28]. As far as I know, the treatment of diffraction of a converging spherical wave has not been studied by a rigorous treatment. Bouwkamp published a review paper on diffraction of both scalar and electromagnetic waves [29].

Linfoot and Wolf investigated the field near the focal region for a circular aperture (as well as for annular apertures) based on the paraxial Debye approximation, and plotted contours of constant intensity in the focal region [30]. The plot was later reproduced in Born and Wolf [14]. The corresponding optical coordinates are $v = k\rho \sin \alpha$ and $u = kz \sin^2 \alpha$, where α is the semi-angular aperture and ρ and z are cylindrical coordinates. Thus, for a contour plot of the intensity in a meridional section, contours of constant v, u are parallel to the z and ρ coordinates, respectively, and the intensity in the focal region is symmetrical about the focal plane, meaning that the intensity maximum is at the focal point [14].

Farnell calculated the field distribution of a microwave lens in the scalar approximation [31]. He found that the axial maximum in intensity was found to occur closer to the aperture than the geometrical focus, and confirmed these results experimentally using microwaves [32].

Richards and Wolf investigated focusing by a high numerical aperture (NA) system in the vectorial nonparaxial Debye (angular spectrum of plane waves) case [33]. However, they used a paraxial axial (defocus) optical coordinate $u = kz \sin^2 \alpha$, which is not the most appropriate choice for high NA systems.

Based on the first Rayleigh-Sommerfeld diffraction integral (RSI), Osterberg and Smith [3] derived an exact expression for the amplitude distribution along the optical axis for a focused scalar wave. They found that under certain circumstances the focal shift could be in the positive direction, i.e., further from the aperture than the geometrical focus. For a paraxial system in the Fresnel approximation, the focal shift is found to depend on the value of the Fresnel number.

McCutchen showed that the field in the focal region of a lens can be calculated as a three-dimensional (3D) Fourier transform of a 3D generalization of the lens aperture (related to what is now often called the 3D coherent transfer function, or CTF) [34]. This approach avoids the need for a paraxial approximation.

Kogelnik and Li described the theory, in the paraxial approximation, for propagation of Gaussian laser beams [35]. The focal shift effect is a noticeable property of this theory. In laser optics, the Rayleigh distance z_0 ($2z_0$ is called the confocal parameter) is provided by $z_0 = \pi w_0^2 / \lambda$, where w_0 is the radius of the beam waist; thus, $z/z_0 = z\lambda / \pi w_0^2 = 1 / \pi N_w$, where N_w is a Fresnel number associated with the waist of the beam.

Sherman showed the equivalence of the formalism for angular spectrum of plane waves and the RSI diffraction integral [36]. Note that this equivalence holds for the complete RSI diffraction integral, including a factor $(1 - ikr)/r$, and including evanescent waves. Analogous angular spectrum theories also hold for RSII and the Kirchhoff diffraction theory with integration over a planar surface (Kp).

Dainty extended the contour plot of intensity for diffraction of a focused wave by a circular aperture for the paraxial and large Fresnel number case to larger values of u [37]. He suggested that intensity zeros occur only in the focal plane and along the optical axis.

Zemanek calculated the diffracted field from a planar circular acoustic transducer, and confirmed Rayleigh's result that the best focus was when $N_z = 1$ [38]. He started from the Huygens-Fresnel diffraction formula (without an obliquity factor) and calculated the integral exactly, i.e., with no Fresnel approximation. He presented contour plots of intensity, and remarked that

"Two surprising features become evident from examining these figures . . . The first feature is that the -3 dB contour has a minimum diameter or spot size of less than one-fourth the transducer diameter."

This is reminiscent of the reduction in cross-section of a water spout when emanating from a circular orifice.

Heurtley calculated analytic expressions for the axial intensity for diffraction of a focused wave by a circular aperture using RSII and Kp [39], and presented plots showing the focal shift.

Welford showed that a diffractive lens on a spherical surface can be free from spherical aberration [40]. The curvature of the surface modifies the radii of the Fresnel half-period zones.

Papoulis developed a theory of Fresnel diffraction and Fourier optics for an aperture that involved coordinate transformations of the independent variables of the ambiguity function [41]. He derived a simple expression for the variation in width of a diffracted beam in the paraxial approximation in terms of moments of the amplitude of the incident wave.

Carter calculated the field of a scalar spherically-symmetric source numerically, excluding evanescent waves [42]. This is equivalent to the problem of focusing a complete uniform hemispherical converging wave, i.e., $\alpha = 90^\circ$. He also presented a plot of the intensity for a semi-angle of convergence of $\alpha = 30^\circ$.

Arimoto applied the Fresnel approximation to a lens with circular aperture. He found there was a factor $f/(f + z)$ in the axial amplitude, and showed that the point spread function of an optical system of low NA (for a constant focal length) is not symmetrical

about the focal plane [43]. He stressed that this effect becomes particularly important for the focusing of laser beams:

“The large f -number optics will become more important as the laser becomes more popular.”

Harvey showed that the breakdown in the Fresnel approximation close to the aperture is equivalent to the introduction of aberrations [44].

Southwell discussed the connections between the Fresnel approximation and the paraxial approximation [45].

The focal shift for optical systems of finite Fresnel number became well known in the traditional optics community only in 1981, when three papers were independently published [46–48].

Sheppard and Wilson studied high NA focusing of scalar waves, and introduced a redefined axial (or defocus) optical coordinate $u = 4kz \sin^2(\alpha/2)$, which is a better approximation for nonparaxial systems than the paraxial form $u = kz \sin^2 \alpha$ [49].

Li and Wolf showed that the optical coordinate of the paraxial Debye theory scales nonlinearly with the axial distance according to the value of the Fresnel number $N = a^2/\lambda f$, where f is the focal length and a is the radius of the aperture [47,50]. They found that

$$\begin{aligned} v &= \left(\frac{f}{f+z} \right) k\rho \sin \alpha = \frac{ak\rho}{f+z} = \frac{2\pi N\rho f}{a(f+z)}, \\ u &= \left(\frac{f}{f+z} \right) kz \sin^2 \alpha = \frac{a^2 kz}{f(f+z)} = \frac{2\pi Nz}{f+z}. \end{aligned} \quad (1)$$

The Debye theory is valid only for systems with large Fresnel number values, rather than a small f -number as proposed by Arimoto [43]. For the case of a finite Fresnel number, the paraxial diffraction integral undergoes a coordinate transformation such that the diffracted intensity is not symmetrical about the focal plane. Li and Wolf plotted the intensity in the focal region [50]. Lines of constant u are straight lines perpendicular to the z axis that get further apart with increasing u , while lines of constant v are straight lines that radiate from a point at the centre of the aperture. Born and Wolf's *Principles of Optics* was revised (from the 6th corrected edition, in 1986) to provide conditions of validity for the paraxial Debye approximation [51]:

“However if $f \gg a \gg \lambda$, and $a^2/\lambda f \gg 1$, the Debye integral can be expected to give a good approximation to the light distribution in the region of focus.”

Li and Wolf derived an expression for the fractional focal shift for the paraxial case as [47]

$$\frac{\Delta f}{f} = -\frac{1}{1 + \pi^2 N^2/12}. \quad (2)$$

Wilson and Sheppard analyzed diffraction of a focused wave by a circular aperture using Fresnel diffraction theory in the paraxial approximation, and showed that there is a paraboloidal phase factor $\exp(iv^2/4\pi N)$ in the focal plane [52]. In addition, they showed that the axial optical coordinate for a point at a distance z from the focus is

$$u = ka^2 \left(\frac{1}{f} - \frac{1}{f+z} \right) = \frac{ka^2 z}{f(f+z)} \approx \frac{ka^2 z}{f^2} \quad (3)$$

only if $z \ll f$. Thus, it can be seen that the paraxial Debye form for the axial optical coordinate u makes the assumption that $z \ll f$.

Li [53] calculated the focal shift for a nonparaxial scalar system based on the BDW theory. He found that the focal shift effect could be expressed as a function of two parameters, namely, the Fresnel number N and the f -number. Experimental confirmation of the focal shift effect for visible laser light was demonstrated by Li and Platzer [54]. In particular, it was found that the focal shift relative to the geometrical focal point was towards the aperture.

Sheppard described how the focal shift effect for a low NA system could be explained by a compromise between an increase in spatial frequency cut-off for observation points closer to the aperture and blurring caused by defocus [55]. According to this theory, the axial optical coordinate is modified to become $u = kz \sin \alpha \sin \beta$, where the effective semi-angular aperture at the observation point β is a function of axial position.

Li extended the results of his previous paper on the coordinate transformation resulting from finite Fresnel number in a Fresnel approximation calculation to small Fresnel numbers, including $N = 0$ [56].

Sheppard and Matthews proposed the pseudo-paraxial diffraction theory of focusing [57]. According to this theory, the focused field for a system of moderate NA is better approximated by the paraxial approximation if the axial optical coordinate is redefined as $u = 4kz \sin^2(\alpha/2)$ [49] such that $kz - u/2 = kz \cos \alpha$ (the axial component of the phase change along a marginal ray). They also showed that the amplitude of a full uniform spherical focused wave with its maximum intensity normalized to unity is $U = \sin(kr)/(kr)$, where r is the spherical radius.

English and George extended the treatment of Osterberg and Smith for the on-axis amplitude for illumination of a circular aperture by a plane wave to provide an analytic expression for the vectorial regime [58].

Steane and Rutt derived an expression for the transverse derivative of the amplitude for illumination of a circular aperture by a plane wave in the Fresnel approximation [59]. They showed that the accuracy of the Fresnel approximation can be improved by redefining the Fresnel number, and thereby the axial optical coordinate.

Bertilone calculated an analytic expression for the focused field at the focus of a uniform hemispherically-focused scalar wave in terms of Lommel functions [60,61].

Kraus proposed calculating the Huygens–Fresnel diffraction integral for diffraction of a diverging spherical wave by a circular aperture by integrating over the wave front rather than the plane of the aperture [62]. He claimed at the time that this approach was more accurate, but later showed that if the full Kirchhoff diffraction integral is used (rather than Huygens–Fresnel) there is in fact no advantage from the point of view of accuracy [63].

Sheppard and Hrynevych [64] provided expressions for the optical coordinates in the nonparaxial finite Fresnel number regime based on a generalization of Fresnel diffraction. In this theory, the optical coordinates are redefined taking into account the critical points in a stationary phase treatment. However, this is not an asymptotic evaluation, and as such is valid even at the focal point. They went on to consider oblique illumination, and derived expressions for the optical coordinates in the meridional plane. The validity of this approach is supported by Young's BDW theory [11] and the results of Andrews [25]. The accuracy of the generalized Fresnel diffraction theory for diffraction of a plane wave by a circular aperture was investigated in detail by Hrynevych [65]. He found that the results for intensity were accurate to 1% for $ka^2/z \leq 1500$, as compared with $ka^2/z \leq 100$ for the conventional Fresnel theory, while for the phase they were accurate to 1% for $ka^2/z \leq 200$ as compared with $ka^2/z \leq 150$ for the conventional Fresnel theory.

Andrés et al. studied the off-axis focal shift for non-rotationally symmetric apertures [66].

Hsu and Barakat [67] calculated the focal distribution for a finite Fresnel number system of high NA. They assumed, however, that the distribution could be expressed in terms of Li and Wolf's optical coordinates, which are not the most appropriate choice for high NA systems.

Wang et al. investigated the far-field behaviour of focused fields for different Fresnel numbers [68,69]. Their numerical results were restricted to the paraxial domain.

Forbes investigated the accuracy of the Fresnel approximation for diffraction of a plane wave by a circular aperture [70].

Sheppard and Török [71] reinterpreted the numerical results of Li [53] using different definitions for the focal length and Fresnel number, which are more appropriate for high NA.

Sheppard and Török considered different power series expansions of the square root in the diffraction integral and derived different optical coordinates corresponding to these expansions [72]. They also considered the vectorial focused case.

Sheppard stressed the requirement for satisfaction of the Debye approximation, i.e., that the axial displacement z must also satisfy $z \ll f$ in addition to the conditions mentioned by Born and Wolf ($f \gg a \gg \lambda$ and $a^2/\lambda f \gg 1$) [14,73]. The condition $f \gg a$ is necessary in order for the *paraxial* Debye approximation to be valid.

Sheppard and Török calculated the axial field, axial optical coordinate, and focal shift for systems with a high NA and finite Fresnel number based on various different scalar diffraction formulae [74]. For RSI, they confirmed the conclusion of Osterberg and Smith [3], namely, that the focal shift can be in the direction further from the lens than the geometric focus. They considered five different diffraction theories: RSI, RSII, and Kirchhoff (Kp) diffraction integrals performed over the plane of the aperture, a Kirchhoff diffraction integral over the spherical illuminating wavefront (Ks), and an approximate form of Ks valid for small displacements (which we call aKs). Of these, only RSI resulted in positive focal shifts.

Teng et al. showed how the amplitude on the axis and along the shadow edge for diffraction of a plane wave by a circular aperture can be calculated analytically using series of Bessel functions, as these are equivalent to Lommel functions [75].

Li reproduced Osterberg and Smith's results for RSI; as the occurrence of positive focal shift does not agree with experimental observations, he deduced that RSI is not an accurate diffraction model for this application [4]. Although RSI is rigorous for a scalar approximation, the assumed boundary condition, that is, that the field in the aperture is unperturbed by the presence of the screen, may not be valid. Using BDW theory to correct for the field in the aperture, they were able to obtain negative focal shifts.

Lin et al. used McCutchen's approach involving 3D Fourier transformation to calculate a focused field of finite Fresnel number in the paraxial approximation [76]. Kou et al. extended this approach to the nonparaxial regime using a 3D RSI model [77].

Glückstad and Madsen described how the predictions of Fraunhofer diffraction by a circular aperture can be extended into the Fresnel regime by renormalizing to the axial amplitude, which can be very simply calculated by the Fresnel diffraction expression [78].

A recent paper by Li considered the complete field, from the aperture to infinity, for diffraction of a focused wave in the paraxial case [79]. He introduced a mapping from the axial coordinate z to a new coordinate $Z = z/(f + z)$ (where $u = kZ \sin^2 \alpha$) in order to plot the behaviour in the far field. He showed that there are a finite number of intensity zeros between the geometrical focus and infinity.

Several papers have extended the treatment for finite Fresnel number and high NA to the case of oblique illumination [2,80–85]. These results have applications in digital deconvolution with objective lenses of finite tube length, although this utility has been reduced by the current adoption of infinite tube length objectives.

It is apparent from this historical review that the traditional optical community was rather slow to appreciate several features of diffraction that were well known in other disciplines, including laser beam optics, microwave optics, and acoustics. It is interesting to recall the rebirth of optics in the 1960s, which resulted from the introduction of optics and photonics into the electrical engineering curriculum through developments in lasers and quantum electronics, electromagnetic theory and microwaves, and information and Fourier optics.

3. The Fresnel Approximation

3.1. The Two Traditional Forms for the Fresnel Approximation (FrA1 and FrA2)

We first review the case of diffraction of a uniform plane wave by a circular aperture in order to introduce the Fresnel approximation. For illumination of a circular aperture

by a uniform plane wave, the Huygens–Fresnel (HF) diffraction integral (neglecting an obliquity factor) becomes

$$U_2 = -\frac{i}{\lambda} \iint \frac{e^{ikr}}{r} dS, \quad (4)$$

where r is the distance from a point in the aperture to the observation point, with

$$r = \left[(x_2 - x_1)^2 + (y_2 - y_1)^2 + z^2 \right]^{1/2}. \quad (5)$$

Here, e^{ikr}/r represents a spherical wave emanating from a point in the aperture and the factor $-i/\lambda$ comes from the fact that the source is driven in resonance, meaning that the far field is $\pi/2$ out of phase with the forcing.

Although it sometimes might not be appreciated, there are two forms of the Fresnel approximation [44,82]. In the first (FrA1), we replace r in the exponent by expanding the square root by a binomial expansion:

$$r \approx z \left[1 + \frac{R_1^2}{2z^2} + \frac{R_2^2}{2z^2} - \frac{(x_1x_2 + y_1y_2)}{z^2} \right], \quad (6)$$

where R_1, R_2 are cylindrical radii in the aperture and observation planes, respectively. In the second form (FrA2), we divide by the distance $r_0 = \sqrt{z^2 + R_2^2}$ from the centre of the aperture to the observation point:

$$r \approx r_0 \left[1 + \frac{R_1^2}{2r_0^2} - \frac{(x_1x_2 + y_1y_2)}{r_0^2} \right]. \quad (7)$$

The difference between these approximations is that if R_2 is large, R_2^2/z^2 cannot be assumed to be small, as in Equation (6).

In the denominator, r is replaced by the first term of the expansion Equations (6) and (7). Then, the diffracted amplitude (for FrA1) is

$$U = -\frac{ik}{z} e^{ikz} \exp\left(\frac{ikR_2^2}{2z}\right) \int_0^a \exp\left(\frac{ikR_1^2}{2z}\right) J_0\left(\frac{kR_1R_2}{z}\right) R_1 dR_1, \quad (8)$$

or

$$U = -ik \frac{\exp(ikr_0)}{r_0} \int_0^a \exp\left(\frac{ikR_1^2}{2r_0}\right) J_0\left(\frac{kR_1R_2}{r_0}\right) R_1 dR_1, \quad (9)$$

for FrA2, where J_n is a Bessel function of the first kind of order n . For FrA1, the exponential before the integral represents a paraboloidal wave, while for FrA2 it represents a spherical wave.

Setting $\rho_1 = R_1/a, \rho_2 = R_2/a$, we can introduce optical coordinates $v = ka^2\rho_2/z, u = ka^2/z$ (FrA1) or $v = ka^2\rho_2/r_0, u = ka^2/r_0$ (FrA2) to obtain

$$U = -\frac{i u}{2} e^{i v} I_L(v, u), \quad (10)$$

where

$$I_L(v, u) = \int_0^1 2 J_0(v\rho_1) \exp\left(\frac{i u \rho_1^2}{2}\right) \rho_1 d\rho_1. \quad (11)$$

A third optical coordinate w has now been introduced:

$$w = kz + \frac{u\rho_2^2}{2} = kz + \frac{v^2}{2u}, \quad (\text{FrA1}),$$

$$w = kr_0, \quad (\text{FrA2}), \quad (12)$$

where I_L can be expressed in terms of Lommel functions of two variables u, v . The factor of 2 in Equation (11) ensures that $I_L = 1$ for $v = 0, u = 0$.

Contours of u and v are plotted against z/a and ρ in Figures 1 and 2 for FrA1 and FrA2, respectively, where the variable $u = 2\pi N_z$, the shadow edge corresponds to $u = v$, and small N_z (small u) corresponds to the far field. Lines of constant v are straight lines in both cases. They are very similar for small values of v but become more different as v increases. Lines of constant u are straight lines for FrA1 and circles for FrA2. They agree quite well for small ρ and z/a larger than about three.

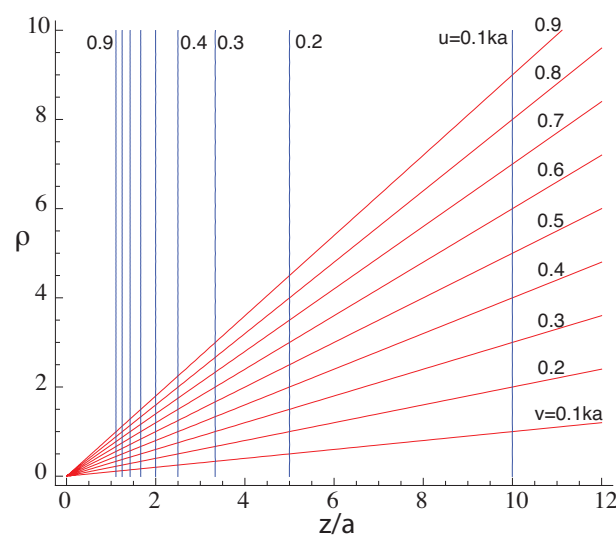


Figure 1. Contours of u (blue) and v (red) for FrA1 for diffraction of a plane wave by a circular aperture.

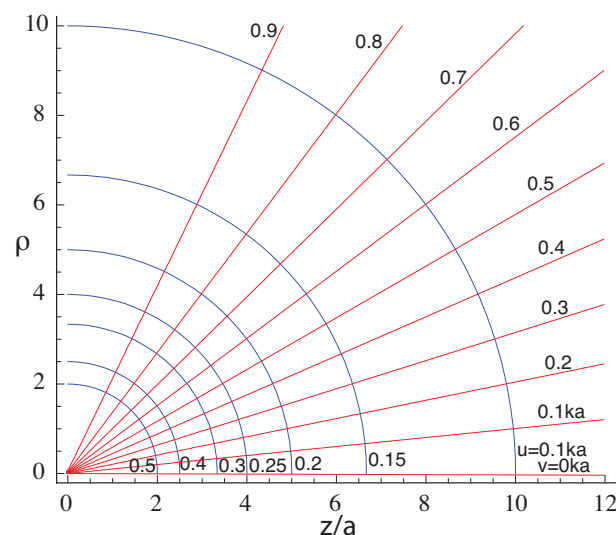


Figure 2. Contours of u (blue) and v (red) for FrA2 for diffraction of a plane wave by a circular aperture.

The advantage of FrA1 is that it allows the field to be propagated from a plane to a plane. However, FrA2 is more accurate than FrA1 in the far field for large values of ρ_2 , when the field strength can be significant for small apertures. It should be appreciated that

although the intensity can be expressed in a way that is independent of wavelength and geometry, the phase cannot because of the factor $\exp(ikz)$.

3.2. The Generalized Fresnel Approximation (gFrA)

The Fresnel approximation uses a parabolic approximation of r evaluated by a power series expansion that becomes worse as R_1 increases relative to z or r_0 . This is unfortunate, as we know from the stationary phase method that the edge of the aperture makes a significant contribution to the integral. Thus, alternative ways of matching can be expected to provide an improved result [59]. In the generalized Fresnel approximation (gFrA), we still use a parabolic approximation, but choose to match up at the critical points of r in the stationary phase evaluation of the diffraction integral [64]. However, we do not actually use a stationary phase expansion, and only use it to define the optical coordinates, meaning that the approach is valid even in the focal region. Here, we have

$$kr \approx w - v\rho \cos \phi + \frac{1}{2}u\rho^2. \quad (13)$$

In the illuminated region, there are three critical points, two of the second kind and one of the first kind, which provide three simultaneous equations for u, v, w . We introduce g_1, g_2 , respectively the maximum and minimum distances from the observation point to the aperture edge, as follows:

$$g_{1,2} = \left[z^2 + a^2(1 \pm \rho)^2 \right]^{1/2}. \quad (14)$$

Then, the two critical points of the second kind provide the two equations

$$kg_{1,2} = w \pm v + \frac{u}{2}, \quad (15)$$

where the $+$ and $-$ signs refer to g_1 and g_2 , respectively. Subtracting the two parts of Equation (15), we obtain the transverse optical coordinate

$$v = \frac{k}{2}(g_1 - g_2) \quad (16)$$

such that surfaces of constant v are hyperboloids of one sheet, with the foci located on the aperture edge. Contours of constant v are shown in Figure 3. If ρ and z/a are large enough, then lines of constant v tend asymptotically to straight lines, the slopes of which agree better with FrA2 than FrA1.

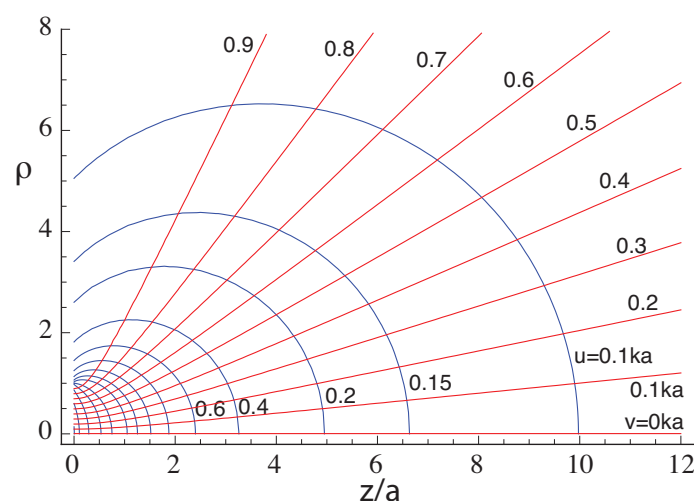


Figure 3. Contours of u (blue) and v (red) for the generalized Fresnel approximation for diffraction of a plane wave by a circular aperture.

Additionally, adding the two parts of Equation (15), we have

$$2w + u = k(g_1 + g_2). \quad (17)$$

The third simultaneous equation is provided by the critical point of the first kind:

$$w - \frac{v^2}{2u} = kz. \quad (18)$$

Eliminating w , we have

$$k(g_{1,2} - z) = \frac{(u \pm v)^2}{2u}, \quad (19)$$

which are positive, and which we can set equal to p^2, q^2 , respectively. $(u - v)^2$ results in slow intensity variations, while $(u + v)^2$ results in fine structure. We introduce the variables p, q , defined as [64]

$$\begin{aligned} p &= \sqrt{g_1 - z}, \\ q &= \sqrt{g_2 - z}; \quad \rho \leq 1 \\ &= -\sqrt{g_2 - z}; \quad \rho > 1, \end{aligned} \quad (20)$$

resulting in $v = (k/2)(p^2 - q^2)$. It is difficult to appreciate the behaviour of p, q for large z from Equation (20); thus, we rewrite them here as

$$\begin{aligned} p &= \frac{a(1 + \rho)}{\sqrt{g_1 + z}}, \\ q &= \pm \frac{a(1 - \rho)}{\sqrt{g_2 + z}}; \quad \rho \leq 1, \end{aligned} \quad (21)$$

where we have used $\sqrt{g_{1,2}^2 - z^2} = a(1 \pm \rho)$.

Then, the other two optical coordinates are

$$u = \frac{k}{2}(p + q)^2; \quad w = kz + \frac{v^2}{2u} = kz + \frac{k}{4}(p - q)^2. \quad (22)$$

Contours of constant u are shown in Figure 3. These are plotted over a larger domain than our previous plot [64], and are more accurate as a result of improvements in the contour plotting routines. Lines of constant u become close to circles for small values of u . For small ρ , they match FrA2 quite well for values of z/a greater than about four. For larger values of u (greater than about $1ka$), lines of constant u become more elongated and the shape of the lines for FrA1 agree better than those for FrA2, although neither of the standard Fresnel approximations provides the correct relationship between u and z/a . Contours of constant w are shown in Figure 4. For large w these are approximately circular, becoming more elongated for smaller values of w . For large z/a , it can be seen from Equation (17) that as u becomes small, the equation for constant values of w provide a family of ellipses with the centre at the origin, tending towards spheres $w = k\sqrt{z^2 + a^2}\rho^2$ for $w \gg ka$.

Along the axis, $\rho = 0$ and $p = q$; thus, $v = 0$ and $w = kz$, and we have

$$\begin{aligned} u &= 2k(\sqrt{z^2 + a^2} - z) = \frac{2ka^2}{z + \sqrt{z^2 + a^2}}, \\ z &= \frac{ka^2}{u} \left[1 - \frac{u^2}{16\pi(a/\lambda)^2} \right]. \end{aligned} \quad (23)$$

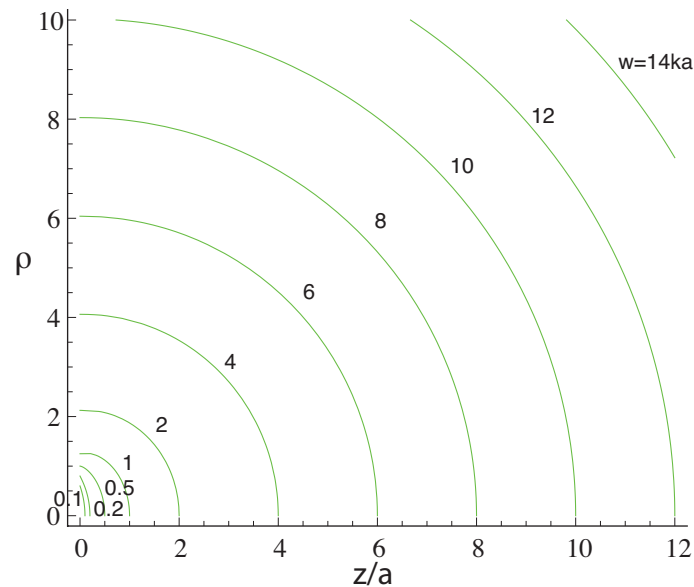


Figure 4. Contours of $w = kz + v^2/2u$ for the generalized Fresnel approximation for diffraction of a plane wave by a circular aperture.

Values of $u = 4\pi n$, where n is a positive integer, determine the positions of the intensity zeros along the axis. The axial intensity zero furthest from the aperture occurs at

$$z = \frac{a^2 - \lambda^2}{2\lambda}. \quad (24)$$

Thus, the number of axial intensity zeros is $\lfloor a/\lambda \rfloor$, indicating the axial behaviour depends on the value of the dimensionless parameter a/λ . The axial field agrees well with direct evaluation of the diffraction integral, but only agrees with FrA1 or FrA2 if $z > 3a$.

Along the shadow edge, $\rho = 1$, $q = 0$, $u = v = kp^2/2 = k(g_1 - z)/2$, and $w = kz + u/2$.

In the plane of the aperture, $z = 0$, $p^2 = a(1 + \rho)$, and $q^2 = a(1 - \rho)$; thus, $v = ka\rho$, and within the aperture, $u = ka(1 + \sqrt{1 - \rho^2})$ and $w = k[z + a(1 - \sqrt{1 - \rho^2})/2]$.

For small ρ , we have the following for v, u :

$$\begin{aligned} v &= \frac{ka^2\rho}{\sqrt{z^2 + a^2}} \left[1 - \frac{z^2 a^2 \rho^2}{2(z^2 + a^2)^2} \right], \\ u &= \frac{2ka^2}{z + \sqrt{z^2 + a^2}} \left[1 - \frac{(\sqrt{z^2 + a^2} + 2z)a^2 \rho^2}{4(z^2 + a^2)^{3/2}} \right]. \end{aligned} \quad (25)$$

If z/a is larger than about 5 [64], then we have

$$\begin{aligned} v &\approx \frac{ka^2\rho}{z} \left[1 - \frac{a^2}{2z^2}(1 + \rho^2) \right], \\ u &\approx \frac{ka^2}{z} \left[1 - \frac{a^2}{4z^2}(1 + 3\rho^2) \right], \\ w &\approx kz + \frac{ka^2\rho^2}{2z} \left[1 - \frac{a^2}{4z^2}(3 - \rho^2) \right]. \end{aligned} \quad (26)$$

The radius of curvature of contours of constant u at $\rho = 0$ tends to $2z/3$ for large z , i.e., the centre of curvature is at $z/3$, as can be deduced from Figure 3.

For $z \gg a$, $z \gg a\rho$, we obtain

$$\begin{aligned} p^2 &= \frac{a(1+\rho)}{\sqrt{2z}}, & q^2 &= \frac{a(1-\rho)}{\sqrt{2z}}, \\ u &= \frac{ka^2}{z}, & v &= \frac{ka^2\rho}{z}, \end{aligned} \quad (27)$$

which indicates agreement with FrA1.

The generalized coordinates work well for the illuminated region. In the shadow region, there are two critical points of the second kind, corresponding to the rim of the aperture. Then, the transverse optical coordinate v is

$$v = \frac{2ka^2\rho}{\sqrt{z^2 + a^2(1+\rho)^2} + \sqrt{z^2 + a^2(1-\rho)^2}}, \quad (28)$$

which is consistent with the previous results (as in Equation (16)). For $z \gg a$, if we also have $\rho \gg 1$, then

$$v = \frac{ka^2\rho}{\sqrt{z^2 + a^2\rho^2}}, \quad (29)$$

which agrees with FrA2. If, in addition, $\rho \ll z/a$, then this also agrees with FrA1.

The amplitude and phase of the diffraction pattern can be calculated from Equation (4) using the values of u, v, w and I_L to provide gFrA-HF. These same values can alternatively be used together with RSI, RSII, or Kp diffraction integrals. The integral should be pre-multiplied by a factor $(1 - i/kr_0) \cos \theta$ for gFrA-RSI, a factor $(1 + i/kr_0)$ for gFrA-RSII, and a factor $\cos^2(\theta/2) + (i/kr_0) \sin^2(\theta/2)$ for gFrA-Kp, where θ is the angle subtended at the axis by the line from the centre of the aperture to the point of observation. For a distant observation point, $kr_0 \ll 1$ and can be neglected; however, there will still be distinct angular differences between RSI and Kp relative to the Huygens–Fresnel result. In fact, as is well known in antenna theory, RSI acts as an axial dipole rather than as a simple source. Closer to the aperture, gFrA-RSI reproduces the exact RSI results better than the gFrA-HF predictions [64].

4. The Focused Case

4.1. Focal Length and Fresnel Number

We now consider the case of diffraction of a convergent wave by a circular aperture, considering a spherical wave convergent on an opaque screen with a circular aperture of radius a (Figure 5). We stress that this is different from the case of a microscope objective used to focus light to a small spot, where the aperture is situated in the front focal plane of the lens and Fresnel number effects are not present. We define the focal length as the radius of curvature f of the wave front at the aperture. This is different from the usual convention in paraxial optics, which is the distance d of the focus from the plane of the aperture. If the focal length of the lens is long, then these tend to the same value; however, in the limiting case of a hemisphere of focused wavefront, $d = f \cos \alpha \rightarrow 0$, where α is the semi-angular aperture of the focusing system. We choose to specify either α or the NA $NA = \sin \alpha$ rather than the f -number $f/\# = d/2a$, as the f -number becomes zero for the hemispherical case in which $NA = 1$.

We define the Fresnel number N_f as the number of half-wavelengths in the difference between the distances of the geometrical focus from the edge and centre of the aperture; thus, $\pi N_f = k(f - d) = kf(1 - \cos \alpha) = 2kf \sin^2(\alpha/2) = ka \tan(\alpha/2)$. This definition is different from the definitions used by Li and by Hsu and Barakat (which are also different from each other) [53,67]. Note that N_f is related to the surface area of the cap of the spherical wavefront $A_{\text{cap}} = 4\pi f^2 \sin^2(\alpha/2)$ by $\pi N_f = A_{\text{cap}}/f\lambda$, which determines the amplitude at the focal spot. The significance of the Fresnel number of half-period zones is that the field for an even number of zones tends to cancel by destructive interference to give a zero, while

for an odd number of zones there tends to be close to a secondary maximum in intensity. This definition of Fresnel number is consistent with Welford's proposal for generating diffractive lenses on spherical surfaces to eliminate spherical aberrations [40]. For a low NA, α is small, meaning that $2 \sin(\alpha/2) \approx \sin \alpha$ and $N_f \approx a^2 / \lambda f = N$; however, this latter expression does not accurately provide the number of Fresnel zones for higher NA. A consequence is that the radii a_i of the regions in a zone plate, conventionally $a_i = \sqrt{\lambda f}$, need to be modified at higher NAs [40]. Indeed, for the hemispherical case, $N_f = 2f / \lambda$, while the paraxial form $N = a^2 / \lambda d$ becomes infinite. Even for the complete spherical case, N_f is still defined, providing $N_f = 4f / \lambda$; however, N is only defined for $\alpha < 90^\circ$. Moreover, as kf must be non-zero, $N_f = 0$ implies that $\alpha = 0$.

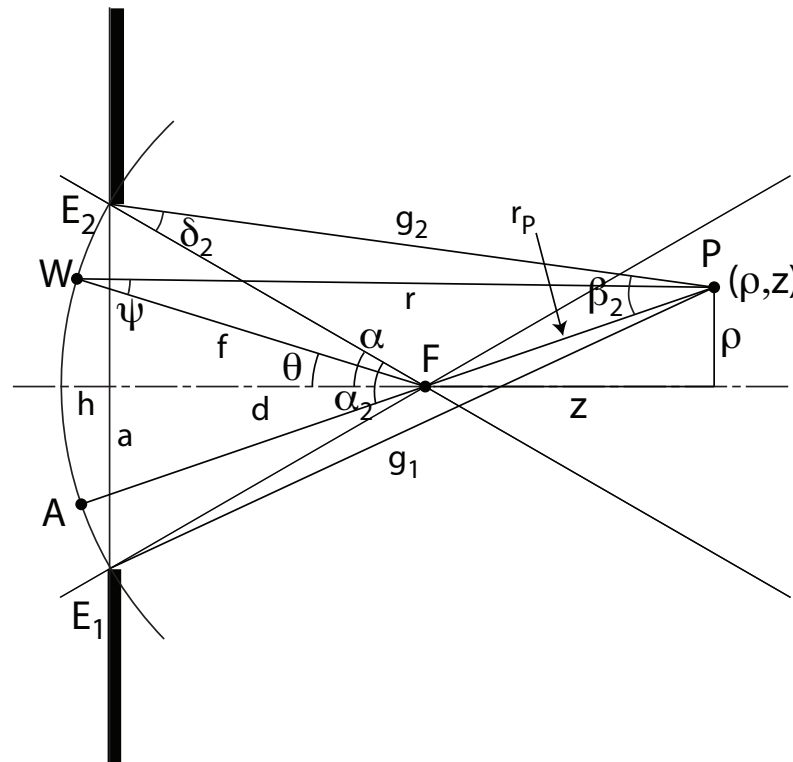


Figure 5. Schematic diagram of a spherical wave diffracted by a circular aperture of radius a in an opaque screen at distance d from the geometrical focal point F , with sagitta $h = f - d$; the angle α is the semi-angular aperture, and θ is the angle subtended at the optical axis by a ray from a general point W on the wavefront to F . The points E_1, E_2 on the aperture edge correspond to the maximum and minimum distances g_1, g_2 from the observation point P , with cylindrical coordinates ρ, ϕ, z relative to F and distant r_P from F and r from W . The triangle ΔFPE_2 has an exterior angle $\angle E_2FA = \alpha_2$ and interior angles $\angle FPE_2 = \beta_2$ and $\angle FE_2P = \delta_2$. The angle rays between rays from point W to F and P are $\angle FWP = \psi$.

Li argued that RSI is not valid for the case of high NA focusing by a system of finite Fresnel number [4], as the calculated focal shift can be positive, as calculated by Osterberg and Smith [3], whereas in practice negative values are always observed. In addition, negative focal shift would be intuitively expected by analogy with Gaussian beams, where the waist of a focused wavefront occurs before its centre of curvature. The problem with RSI is that, although it is exact, the assumed amplitude within the aperture (equal to the amplitude if the screen with aperture were absent) is not correct. For this reason we agree with Li, and choose not to use RSI.

The amplitude in the focal region of a lens can be alternatively calculated using the Kirchhoff diffraction integral, which is known to be exact when performed over an arbitrary closed surface. We choose to integrate over the spherical wave front (Figure 1), an approach

we call K_s (for Kirchhoff, spherical). Note that while this has similarities with the approach of Kraus, it is different in that we apply integration over the wave front for a converging wave rather than for a diverging one [62]. We assume that the field on this surface is equal to the incident spherically-convergent wave (Kirchhoff boundary condition), and the field on the remainder of the closed surface is taken as zero. We argue that this assumption is more accurate than either of the RSI and RSII Rayleigh–Sommerfeld approaches, where the integral is performed over a planar surface (or, in the Kirchhoff approximation, performed by integrating over a plane, which we call K_p) because the presented method incorporates effects of diffraction during the propagation of the wave before reaching the aperture plane. The obliquity factor for this geometry is close to unity, and as such can be neglected for observation points not too distant from the geometrical focus, even though the convergence angle is appreciable. This is very different from the geometry for a planar integration surface or diverging wave as studied by Kraus [62,63]. We continue by neglecting a term of order $1/kf$. Then, the amplitude at a point P with cylindrical coordinates ρ, z relative to the geometrical focus is

$$U_{Ks}(\rho, z) = -\frac{ikf^2}{2\pi} \exp(-ikf) \int_0^\alpha \int_0^{2\pi} \frac{\exp(ikr)}{r} \sin \theta \, d\phi d\theta, \quad (30)$$

where

$$\begin{aligned} r &= [f^2 + z^2 + \rho^2 - 2f\rho \sin \theta \cos \phi + 2fz \cos \theta]^{1/2} \\ &= [(f+z)^2 + \rho^2 - 2f\rho \sin \theta \cos \phi - 4fz \sin^2(\theta/2)]^{1/2}. \end{aligned} \quad (31)$$

4.2. Behaviour along the Optical Axis

We first consider the amplitude U at a point on the axis at a distance z from the geometrical focus when

$$U_{Ks}(z) = -ikf^2 \exp(-ikf) \int_0^\alpha \frac{\exp(ikr)}{r} \sin \theta \, d\theta \quad (32)$$

and

$$r = [(f+z)^2 - 4fz \sin^2(\theta/2)]^{1/2}. \quad (33)$$

Li and Wolf presented an evaluation based on series expansion of the square root, resulting in a form of the Fresnel approximation [50,72]. However, rather than approximating the square root, which can introduce inaccuracies, we choose to evaluate the integral in Equation (32) directly for an observation point on the optical axis. This approach was described previously in [74]. Thus, changing the variable of integration from θ to r using $r \, dr = -fz \sin \theta \, d\theta$ from Equation (33), we have

$$\begin{aligned} U_{Ks}(z) &= -\frac{ikf}{z} \exp(-ikf) \int_g^{f+z} \exp(ikr) \, dr, \\ &= -\frac{f}{z} \{\exp(ikz) - \exp[ik(g-f)]\}, \end{aligned} \quad (34)$$

where the distance g of P from the centre and edge of the aperture is

$$g = [(f+z)^2 - 4fz \sin^2(\alpha/2)]^{1/2}. \quad (35)$$

Geometrically, Equation (34) can be recognized as comprising two components, representing the sum of two waves from the centre and edge of the aperture, respectively (as in Young's BDW approach) added in antiphase. The axial amplitude is then provided by

$$U_{Ks}(z) = -\frac{2if}{z} \exp\left\{i\left[\frac{k(g+z-f)}{2}\right]\right\} \sin\left[\frac{k(f+z-g)}{2}\right]. \quad (36)$$

Therefore, to describe the axial amplitude, we introduce the axial (defocus) optical coordinate u , defined in terms of the argument of the sine function in Equation (36) as follows:

$$u = 2k(f+z-g). \quad (37)$$

This definition for the axial optical coordinate has previously been discussed elsewhere [64,71,74].

Then, for the axial amplitude we have

$$\begin{aligned} U_{Ks}(z) &= -\frac{ifu}{2z} e^{ikz} e^{-iu/4} \left[\frac{\sin(u/4)}{u/4} \right] \\ &= -\frac{ifu}{2z} e^{i\psi} \left[\frac{\sin(u/4)}{u/4} \right], \end{aligned} \quad (38)$$

where the phase variation (unwrapped) along the axis relative to the geometrical focal point is

$$\psi = kz - \frac{u}{4} = \frac{k}{2}(g+z-f). \quad (39)$$

This expression for the axial amplitude is a generalization of the paraxial Debye expression ([14], p. 441) but is valid for both large angles of convergence and finite values of the Fresnel number. Thus, the axial intensity is

$$I_{Ks}(z) = \left(\frac{fu}{2z}\right)^2 \left[\frac{\sin(u/4)}{(u/4)}\right]^2. \quad (40)$$

From Equation (37), it is difficult to see the behaviour of u when z is small; we can rewrite the equation as

$$\begin{aligned} u &= \frac{4kz(f-d)}{f+z+g} = \frac{8kfz \sin^2(\alpha/2)}{f+z+g}, \\ &= \frac{8kfz \sin^2(\alpha/2)}{(f+z) + \sqrt{(f+z)^2 - 4fz \sin^2(\alpha/2)}}. \end{aligned} \quad (41)$$

Now, $u = 4kZ \sin^2(\alpha/2)$, with a mapping

$$z \rightarrow Z; \quad Z = \frac{2fz}{f+z+g}. \quad (42)$$

In a similar way, ψ is provided by

$$\begin{aligned} \psi &= \frac{kz[(g+z-f) + 2f \cos^2(\alpha/2)]}{f+z+g} \\ &= \frac{kz[(z-f \cos \alpha) + \sqrt{(f+z)^2 - 4fz \sin^2(\alpha/2)}]}{(f+z) + \sqrt{(f+z)^2 - 4fz \sin^2(\alpha/2)}}. \end{aligned} \quad (43)$$

We can obtain an equation between u and z by eliminating g from Equations (35) and (37):

$$u^2 - 4ku(f+z) + 8k^2fz \sin^2(\alpha/2) = 0, \quad (44)$$

which can be solved for kz as a function of u :

$$kz = \frac{u \left(1 - \frac{u}{4kf}\right)}{\left[1 - \frac{u}{4kf \sin^2(\alpha/2)}\right]} \frac{1}{4 \sin^2(\alpha/2)}. \quad (45)$$

We now introduce the Fresnel number N_f , where $\pi N_f = 2kf \sin^2(\alpha/2)$. For $u/2\pi N_f$ in terms of z/f , we have

$$\frac{u}{2\pi N_f} = \frac{2z}{f + z + g} = \frac{Z}{f} = \frac{z}{f} D = \frac{2(z/f)}{(1 + z/f) + \sqrt{(1 + z/f)^2 - 4(z/f) \sin^2(\alpha/2)}}, \quad (46)$$

which is plotted in Figure 6a for different values of α . In Equation (46), we have introduced a dimensionless parameter $D = Z/z = 2f/(z + f + g) = u/[4kz \sin^2(\alpha/2)]$. All the curves exhibit the same slope for small z/f . For small z/f (or for large kf , and consequently large N_f), we have $u = 4kz \sin^2(\alpha/2)$, agreeing with the result of the pseudo-paraxial treatment [57], while for $\alpha \rightarrow 0$ u reduces to the form of Li and Wolf for the finite Fresnel number case in the paraxial approximation [50]. For large values of kz , $u/2\pi N_f \rightarrow 1$, meaning that there is a finite number $[N_f/2]$ of zeros in intensity beyond the geometrical focus. In the centre of the aperture $z = -f \cos \alpha$, unlike in the paraxial case, $u = -2\pi N_f [\cot(\alpha/2) - 1]$ is also finite. Interestingly, for $\alpha = \pi$, corresponding to extreme case of a complete spherically focused wave, $u/4 = kz$ and $\psi = 0$, agreeing with the correct values for a spherical standing wave [57]. This gives us confidence that the diffraction integral is providing meaningful results.

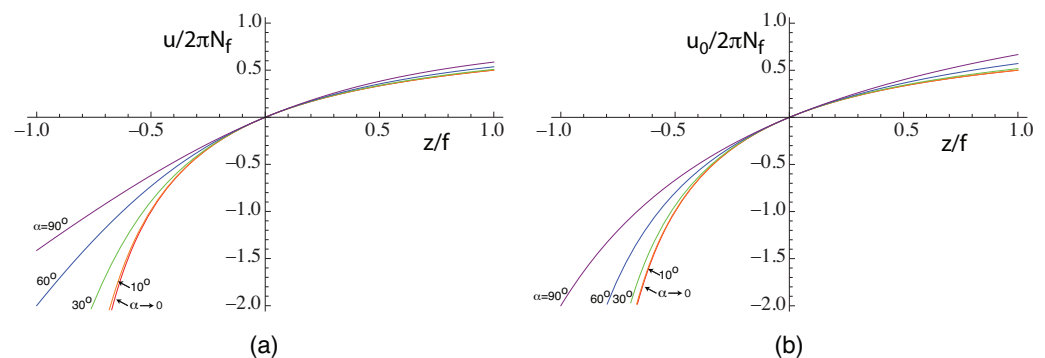


Figure 6. The behaviour of (a) $u/2\pi N_f$ (from Ks) and (b) $u_0/2\pi N_f$ (from aKs), with z/f for different semi-angular apertures. Lines are coloured red for small α , orange for $\alpha = 10^\circ$, green for $\alpha = 30^\circ$, blue for $\alpha = 60^\circ$, and purple for $\alpha = 90^\circ$. For small α , the behaviour of $u/2\pi N$ and $u_0/2\pi N_f$ agrees with that of Li and Wolf [50]. Along the optical axis, $u/2\pi N_f = Z/f = D(z/f)$, with Z as in Equation (42) and D as in Equation (46).

We now have the following for z/f as a function of $u/2\pi N_f$ and α :

$$\frac{z}{f} = \frac{\frac{u}{2\pi N_f} \left(1 - \frac{u}{2\pi N_f} \sin^2 \frac{\alpha}{2}\right)}{\left(1 - \frac{u}{2\pi N_f}\right)}, \quad (47)$$

which was plotted in [74].

We have expressed u in terms of f , α , N , and z , with the first three parameters characterizing the optical system. There are many alternative ways to specify the geometry. The triangle ΔFPE_2 has sides (r_P, g_2, f) in general (and similarly for triangle ΔFPE_1); for the special case when P is on the axis, it has sides (z, g, f) . The angles of the triangle are $(\beta_2, \delta_2, \alpha_2)$,

where $\alpha_2 = (\beta_2 + \delta_2)$ is the exterior angle. For P on the axis, $\alpha_2 = \alpha, \beta_2 = \beta, \delta_2 = \delta$. On the optical axis [74],

$$u = 2\pi N \left(1 - \frac{\tan(\beta/2)}{\tan(\alpha/2)} \right). \quad (48)$$

Using the sine rule, we have

$$u = 4kz \sin(\alpha/2) \sin(\beta/2) \sec(\delta/2), \quad (49)$$

which, if α, β, δ are small, gives us

$$u = kz \sin \alpha \sin \beta, \quad (50)$$

which is the same as derived for the paraxial case [55]. Further, if we defines a Fresnel number N_g for the observation point, then $\pi N_g = k(g - d - z)$, meaning that $\pi N_g = 2kg \sin^2(\beta/2)$. Then,

$$N_g \cot(\beta/2) = N_f \cot(\alpha/2) \quad (51)$$

and

$$u = 2\pi(N_f - N_g). \quad (52)$$

The intensity at the geometrical focal point is

$$I_0 = \left(2kf \sin^2 \frac{\alpha}{2} \right)^2 = (\pi N_f)^2. \quad (53)$$

Then, the axial intensity is

$$I = I_0 \left(\frac{2f}{f + z + g} \right)^2 \left[\frac{\sin(u/4)}{u/4} \right]^2 = I_0 D^2 \left[\frac{\sin(u/4)}{u/4} \right]^2, \quad (54)$$

or in terms of u and N_f ,

$$I = I_0 \left\{ \frac{1 - \frac{u}{2\pi N_f}}{1 - \frac{u}{2\pi N_f} \sin^2(\frac{\alpha}{2})} \right\}^2 \left[\frac{\sin(u/4)}{u/4} \right]^2. \quad (55)$$

For large N , Equation (55) reduces to the standard form for the paraxial Debye approximation as provided by Born and Wolf [14], except that the optical coordinate u is provided by the more correct high NA form $u = 4kz \sin^2(\alpha/2)$ [57]. The focal shift effect is caused by the factor in the braces in Equation (55). It can be seen that this factor becomes weaker as NA increases, becoming a constant for a complete spherical wave (i.e., corresponding to no focal shift).

4.3. Focal Shift

The optical coordinate u provided by Equations (37) or (46) is of a more complicated form than that of Li and Wolf [50]. However, if we apply the condition that $z/f \ll 1$, which we call aKs (approximate Kirchhoff spherical), we have $g \approx f + z \cos \alpha$ and can obtain approximate expressions u_0 for u from Equation (41):

$$u_0 = \frac{4kz \sin^2(\alpha/2)}{1 + (z/f) \cos^2(\alpha/2)} = \frac{2\pi N_f z}{f + z \cos^2(\alpha/2)}. \quad (56)$$

Note that this equation cannot be derived accurately from Equation (37) in the same way, as $(g - f)$ is the small difference between two large quantities; thus, we must be careful when using the approximation $g \approx f + z \cos \alpha$. A better approximation is $g \approx f + z \cos \alpha + (z^2/2f) \sin^2 \alpha$. The first equality of Equation (56) reduces to provide the

usual high NA expression if z/f is negligible [57], while the second form reduces to the finite Fresnel number expression of Li and Wolf when α is small [50]. Thus, Equation (56) can be regarded as providing more accurate expressions for the optical coordinate in the regime with finite Fresnel number and high NA that are valid for small distances from the geometrical focus. We can further introduce Z_0, D_0 , approximate values of Z, D , valid near the geometrical focal point. Then, we have $Z_0 = u_0/4k \sin^2(\alpha/2) = fz/[f + z \cos^2(\alpha/2)]$, $D_0 = u_0/4kz \sin^2(\alpha/2) = f/[f + z \cos^2(\alpha/2)]$. It should be noted that Hsu and Barakat [67] assumed the paraxial form for the optical coordinates in their high-aperture vectorial treatment based on the Stratton–Chu formula, which is not justifiable for high NAs. The behaviour of $u_0/2\pi N_f = Z_0/f = D_0(z/f)$ with z/f is shown in Figure 6b. It can be seen that the range of values of z/f for which u_0 is a good approximation for u increases as NA decreases.

If z/f is small, then for the analogous approximate expression ψ_0 for ψ we have

$$\psi_0 = \frac{kz(f+z) \cos^2(\alpha/2)}{f + z \cos^2(\alpha/2)}, \quad (57)$$

which reduces to the high NA axial phase variation $\psi \approx kz \cos^2(\alpha/2)$ for $z \ll f$ [57].

The approximate form u_0 for u is valid for small focal shifts, meaning that the principal maximum in intensity is located at

$$u \approx \frac{24}{\pi N_f} \cos^2 \frac{\alpha}{2}. \quad (58)$$

Then, the fractional focal shift is provided by

$$\frac{\Delta f}{f} \approx -\frac{\cos^2(\alpha/2)}{\cos^4(\alpha/2) + \pi^2 N_f^2/12}, \quad (59)$$

and the increase in intensity relative to that at the geometrical focus by

$$\frac{I - I_0}{I_0} \approx \frac{12 \cos^4(\alpha/2)}{\pi^2 N_f^2}. \quad (60)$$

Equations (58)–(60) are generalizations of those of Li and Wolf for low NAs [47,50]. Further, Equation (59) for the fractional focal shift is much simpler than the empirical formula provided by Li for the nonparaxial case [53]. As we have seen that $N_f = 0$ implies that $\alpha = 0$, in this case Equation (59) provides $\Delta f/f = -1$, meaning that the axial intensity decays monotonically from the aperture plane [79]. For non-zero α , as the value of $\Delta f/f$ cannot be less than $-\cos \alpha$ from geometrical arguments, there is a minimum Fresnel number predicted by Equation (59) for any given value of α . For example, the minimum Fresnel number is $N_f = 1$ for $\alpha = 57.03^\circ$, corresponding to a value of $kf = 6.92$ and $ka = 5.78$.

Figure 7 compares the fractional focal shift predicted from the Ks expression for intensity in Equation (40) using Equation (46) and by the analytic expression in Equation (59) provided by aKs. The behaviour calculated from Equation (40) and aKs show good agreement. The fractional focal shift $\Delta f/f$ is negative, showing that the diffraction focus is shifted towards the aperture for all NAs, in contrast to Osterberg and Smith's results using RSI for high NA ($NA > 0.943$, $\alpha = 70.56^\circ$) [3] (and confirmed by [4,74]). The absolute value of the fractional focal shift increases as the Fresnel number decreases. For Ks, the absolute value of the fractional focal shift decreases with NA for Fresnel numbers greater than about 0.62 and increases for smaller values of Fresnel number. The behaviour predicted by the aKs analytic expression in Equation (40) is qualitatively similar, but the effect of NA is smaller, with the cross-over for no variation in focal shift with aperture occurring at a Fresnel number of $2\sqrt{3}/\pi = 1.10$ for low NAs. For high NA, as N_f decreases the size of

the aperture also decreases, and eventually the diffraction calculations become invalid. We indicate where in Figure 7 the value of $a/\lambda = 5$ ($ka = 10\pi$) by coloured circles.

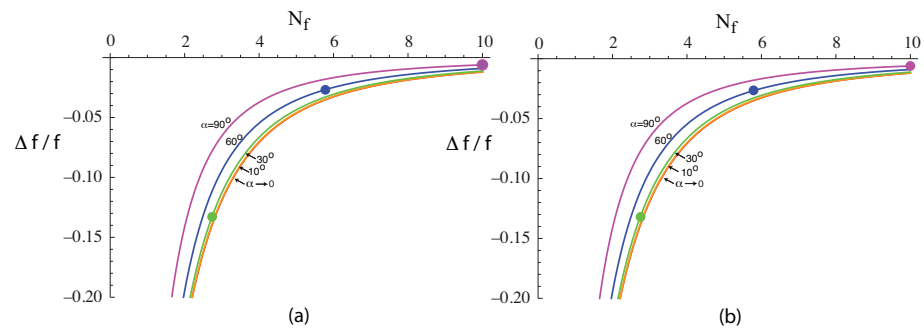


Figure 7. The fractional focal shift as a function of the Fresnel number N for different NAs: (a) for the Ks model (Equations (46) and (55)) and (b) from the analytic aKs expression in Equation (59). Lines are coloured red for small α , orange for $\alpha = 10^\circ$, green for $\alpha = 30^\circ$, blue for $\alpha = 60^\circ$, and purple for $\alpha = 90^\circ$. For small α , the behaviour agrees with that of Li and Wolf [50]. The curves for small α and $\alpha = 10^\circ$ are very close to each other. The small circles indicate where $a/\lambda = 5$: for small a/λ , the Kirchhoff boundary conditions tend to break down.

4.4. Off-Axis Behaviour

We now wish to generalize to off-axis observation points. Although Equation (30) can be evaluated exactly for points on the axis, unfortunately this cannot be done in general; thus, we wish to approximate it. In [57,64], we approximated it to the Fresnel diffraction integral, which has analytic solutions in terms of Lommel functions. We called this the pseudo-paraxial approximation. However, as the NA increases, the field calculated by the Fresnel diffraction integral departs from that predicted by the Debye integral (angular spectrum) method. Therefore, here we choose to approximate to the scaled Debye–Wolf (sDW) approximation, as the (unscaled) Debye–Wolf (DW) form provides better agreement for higher NAs and large Fresnel numbers than the pseudo-paraxial approximation [57,71]:

$$U_{sDW} = -\frac{ik^2 f^2}{2(w + kf)} \exp(iw) \int_0^\alpha 2J_0\left(v \frac{\sin \theta}{\sin \alpha}\right) \exp\left[-\frac{iu \sin^2(\theta/2)}{2 \sin^2(\alpha/2)}\right] \sin \theta d\theta, \quad (61)$$

where the optical coordinates u, v, w are scaled as follows:

$$u = 4kZ \sin^2(\alpha/2), \quad v = kR \sin \alpha, \quad w = kW, \quad (62)$$

by the mappings $z \rightarrow Z, \rho \rightarrow R, z \rightarrow W$ to approximate r in Equation (31) by

$$r - f \approx W - R \sin \theta \cos \phi - 2Z \sin^2 \frac{\theta}{2} = (W - Z) - R \sin \theta \cos \phi + Z \cos \theta. \quad (63)$$

Note the opposite sign of u in Equation (61) as compared with the collimated case in Equation (11). This is consistent with the fact that for the collimated case u was defined as positive and increasing towards the aperture, as in Figure 3. The diffraction integral no longer has variables in the Bessel and complex exponential functions of the simple forms ρ, ρ^2 , as they are in the paraxial Debye or the Fresnel approximations. Along the axis, $v = 0$, it reduces to the Ks of Equation (38).

For the (unscaled) DW equation, then, $Z = z, R = \rho$, and $W = z$, and it is valid for high NAs in the scalar approximation, though only for large values of the Fresnel number. Thus, the normalized axial intensity is $[\sin(u/4)/(u/4)]^2$ regardless of the value of α , even up to $\alpha = 180^\circ$; however, the transverse intensity in the focal plane varies from an Airy disk $[2J_1(v)/v]^2$ for small α to $(\sin v/v)^2$ for both $\alpha = 90^\circ$ and $\alpha = 180^\circ$.

For smallish α such that $\sin \theta \approx 2 \sin(\theta/2)$, Equation (61) reduces to the pseudo-paraxial form, with analytic solutions provided in terms of Lommel functions. For large Fresnel

4.6. The Scaled Debye–Wolf (sDW) Case

In the sDW case, for the illuminated region we match r and its approximation at the three critical points in the stationary phase expansion of the integral; note, however, that we do not use an asymptotic evaluation of the integral. For observation points close to the geometrical focus, even though the stationary phase asymptotic method cannot be used, the expansion is still valid. The two critical points of the second kind give

$$\begin{aligned} k(g_1 - f) &= w - \frac{u}{2} + v = kW - 2kZ \sin^2 \frac{\alpha}{2} + kR \sin \alpha, \\ k(g_2 - f) &= w - \frac{u}{2} - v = kW - 2kZ \sin^2 \frac{\alpha}{2} - kR \sin \alpha, \end{aligned} \quad (66)$$

where g_1, g_2 are the maximum and minimum distances of the observation point from the aperture edge

$$g_{1,2} = \left[(f + z)^2 + \rho^2 - 4fz \sin^2 \frac{\alpha}{2} \pm 2f\rho \sin \alpha \right]^{1/2}. \quad (67)$$

Adding and subtracting, we have

$$\begin{aligned} k(g_1 + g_2 - 2f) &= 2\left(w - \frac{u}{2}\right) = 2k\left(W - 2Z \sin^2 \frac{\alpha}{2}\right) = 2kA, \\ k(g_1 - g_2) &= 2v = 2kR \sin \alpha, \end{aligned} \quad (68)$$

where $A = W - 2Z \sin^2(\alpha/2)$. The second of these equations defines R and v valid for the whole of the focal region, whether illuminated or shadow:

$$R = \frac{g_1 - g_2}{2 \sin \alpha} = \frac{2f\rho}{g_1 + g_2}; \quad v = k(g_1 - g_2) = \frac{2kf\rho \sin \alpha}{g_1 + g_2}. \quad (69)$$

Surfaces of constant R are hyperboloids of one sheet:

$$\frac{\rho^2}{R^2} - \frac{(z + f \cos \alpha)^2}{f^2 - R^2} = \sin^2 \alpha. \quad (70)$$

Close to the axis, such that $(g_1 + g_2) \approx 2g$, we have $R \approx f\rho/g$, $v = kf\rho \sin \alpha/g$. If α is small, then the surfaces of constant R tend to cones with the vertex at the centre of the aperture and $R \approx \pm \rho f / (f + z)$, agreeing with Li and Wolf [50].

The first line of Equation (68) contains both W and Z , and as such cannot be solved independently for these. We note that on the axis it reduces to Equation (39), meaning that $w_{\rho=0} = kz$ and $\psi = w - u/4$. However, for the illuminated region there is a third critical point corresponding to matching up the values at the maximum or minimum of the two sides of Equation (63), resulting in an additional equation. For the left-hand side, this corresponds to a direct undiffracted ray passing through the geometrical focal point along the line FP in Figure 5. For the right-hand side, we find the maximum or minimum of the right-hand side of Equation (63). Therefore, the third equation is

$$\pm \sqrt{z^2 + \rho^2} = (W - Z) \pm \sqrt{Z^2 + R^2}. \quad (71)$$

Eliminating W using the first line of Equation (68), we have

$$f \pm \sqrt{z^2 + \rho^2} - (g_1 + g_2)/2 = m = \pm \sqrt{Z^2 + R^2} - Z \cos \alpha, \quad (72)$$

where we have introduced the variable m , a function of the observation point, which is zero at the geometrical focal point. We take $+\sqrt{z^2 + \rho^2}$ for $Z \geq 0$ and $-\sqrt{z^2 + \rho^2}$ for $Z < 0$. Equation (72) can then be written as a quadratic equation in Z :

$$Z^2 \sin^2 \alpha - 2mZ \cos \alpha + (R^2 - m^2) = 0 \quad (73)$$

and solved for Z in terms of m using Equation (69):

$$Z = \frac{m \cos \alpha \pm \sqrt{m^2 - R^2 \sin^2 \alpha}}{\sin^2 \alpha} = \frac{1}{4 \sin^2(\alpha/2)} \left[m \left(1 - \tan^2 \frac{\alpha}{2} \right) \pm \left(1 + \tan^2 \frac{\alpha}{2} \right) \sqrt{m^2 - \frac{4f^2 \rho^2 \sin^2 \alpha}{(g_1 + g_2)^2}} \right], \quad (74)$$

where we take the positive square root for $Z > 0; \rho \leq z \tan \alpha$ (inside the shadow edge) or $Z < 0; \rho > z \tan \alpha$ (outside the shadow edge) and the negative root for $Z > 0; \rho > z \tan \alpha$ (outside) or $Z < 0; \rho \leq z \tan \alpha$ (inside). On the shadow edge, $m = \pm R \sin \alpha = Z \tan \alpha \sin \alpha$.

The corresponding quadratic equation in the axial optical coordinate u is

$$u^2 - 2km \left(1 - \tan^2 \frac{\alpha}{2} \right) u + \left(1 + \tan^2 \frac{\alpha}{2} \right)^2 v^2 = 4(km)^2 \tan^2 \frac{\alpha}{2}, \quad (75)$$

and its solution for u is

$$u = km \left(1 - \tan^2 \frac{\alpha}{2} \right) \pm k \left(1 + \tan^2 \frac{\alpha}{2} \right) \sqrt{m^2 - \frac{4f^2 \rho^2 \sin^2 \alpha}{(g_1 + g_2)^2}}. \quad (76)$$

Along the axis, $g_1 = g_2 = g$ and $v = 0$. Then, $w = kz$ and $u = 2k(f + z - g)$, as in Equation (37) for the Ks theory. Simply, $u = 2km$ or $Z = m/[2 \sin^2(\alpha/2)]$, where we take the positive/negative roots in Equations (74) and (76) for positive/negative z , respectively.

In addition, for W we have

$$W = \frac{g_1 + g_2}{2} - f + \frac{1}{2} \left[m \left(1 - \tan^2 \frac{\alpha}{2} \right) \pm \left(1 + \tan^2 \frac{\alpha}{2} \right) \sqrt{m^2 - \frac{4f^2 \rho^2 \sin^2 \alpha}{(g_1 + g_2)^2}} \right]. \quad (77)$$

Contour plots for R/f and Z/f are shown in Figure 9 for different values of α , while contours of W/f are shown in Figure 10. The plots cover the domain within a sphere of radius f . For R/f and Z/f , contours for constant Z/f and W/f are shown in green and blue for $Z < 0$ and $Z > 0$, respectively. In particular, we note that for higher NA the sDW version of Ks can predict the field strength before (i.e., to the left of) the aperture plane, whereas RSI, RSII, and Kp cannot.

Although the third critical point is only present for the illuminated region, we can consider extrapolation into the shadow region as a type of analytic continuation. The slope of the contours for Z/f is continuous at the shadow edge. For $Z/f = 0$ or $W/f = 0$, the two contours do not match up completely in each case, but agree well for $\rho/f < 0.2$. This value of ρ/f corresponds to a larger value of v as N_f increases or α decreases. A magnified contour plot of a smaller region is shown in Figure 11. On this scale, the two contours for $Z/f = 0$ are almost indistinguishable; only one, $Z/f = 0^{(-)}$, is shown in the plot. However, even for $\alpha = 10^\circ$ the contours of constant Z/f are noticeably curved, unlike in the paraxial treatment of Li and Wolf [50].

As we have mentioned, the Debye approximation is only valid for regions near the focus [73]. Even for small values of α , e.g., 10° , the Debye approximation breaks down for distances from the focal point that are a substantial fraction of f , although of course for large f the intensity is very small in this case. For small α , the surfaces for $Z/f = 0$ and $W/f = 0$ are approximately spherical, with a radius of 1, agreeing with [52]. The radii of curvature increase with α , becoming planar for $\alpha = 90^\circ$. In fact, the radius of curvature of Z or W for small ρ is approximately equal to $f/\cos \alpha$. For small α , the contours become approximately flat for $z/f \approx 0.3$, which reduces to the aperture plane $z/f = 0$ for $\alpha = 90^\circ$.

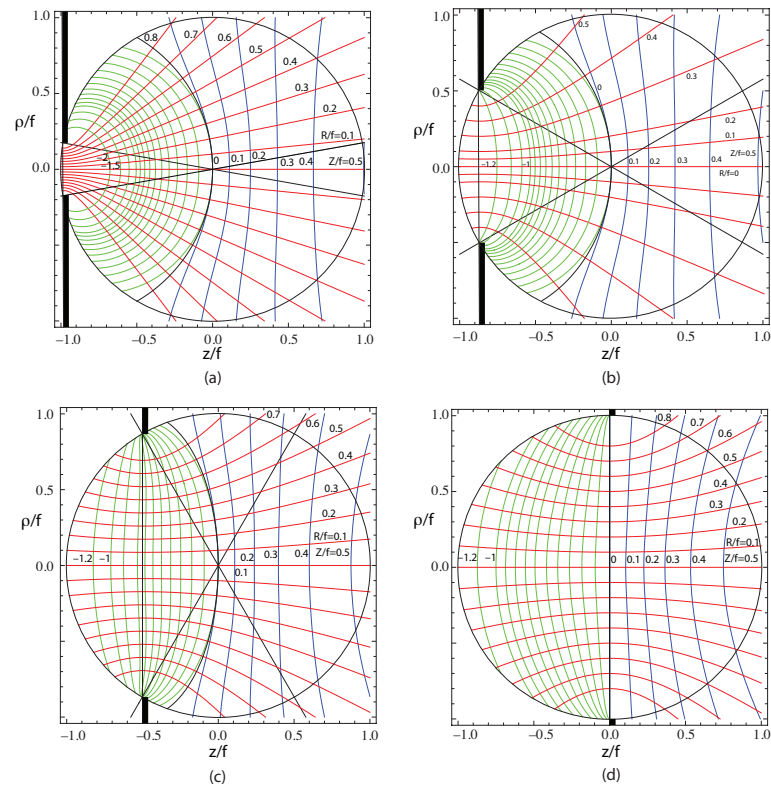


Figure 9. Contours of R/f (in red) and Z/f (in green for $Z/f < 0$ and blue for $Z/f > 0$) for the scaled Debye–Wolf theory for diffraction of a spherical wave by a circular aperture: (a) $\alpha = 10^\circ$, (b) $\alpha = 30^\circ$, (c) $\alpha = 60^\circ$, (d) $\alpha = 90^\circ$. The black contour is for $A/f = [W - 2Z \sin^2(\alpha/2)]/f = 0$. The shadow edge is also indicated in black.

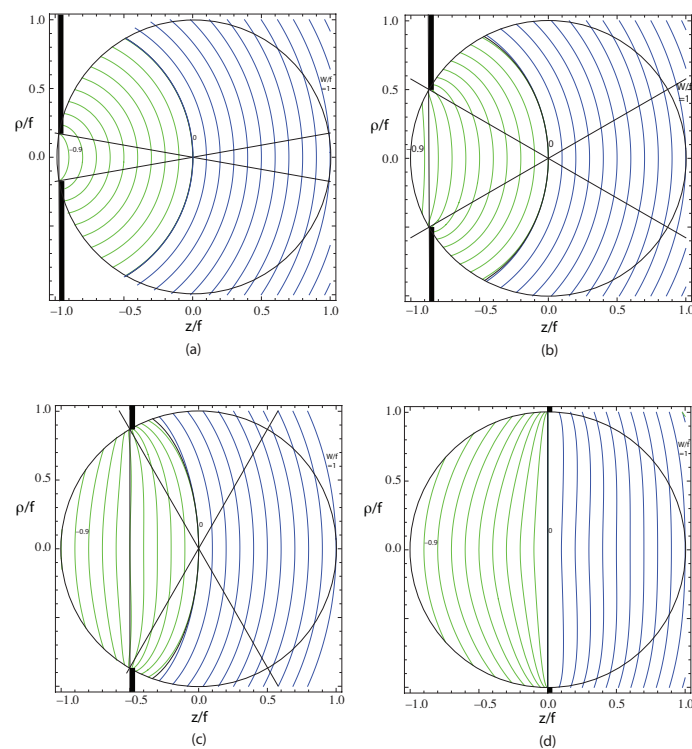


Figure 10. Contours of W/f (in green for $W/f < 0$ and blue for $W/f > 0$) for the scaled Debye–Wolf theory for diffraction of a spherical wave by a circular aperture: (a) $\alpha = 10^\circ$, (b) $\alpha = 30^\circ$, (c) $\alpha = 60^\circ$, (d) $\alpha = 90^\circ$. The black contour is for $A/f = 1$.

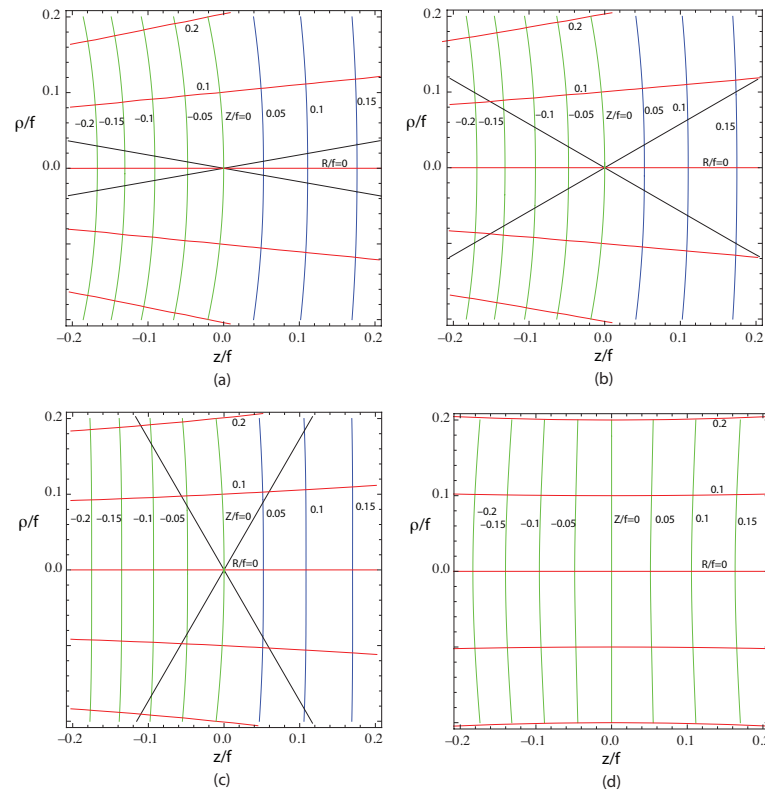


Figure 11. A magnified view of the contours of R/f (in red) and Z/f (in green for $Z/f \leq 0$ and blue for $Z/f > 0$) for the scaled Debye–Wolf theory for diffraction of a spherical wave by a circular aperture: (a) $\alpha = 10^\circ$, (b) $\alpha = 30^\circ$, (c) $\alpha = 60^\circ$, (d) $\alpha = 90^\circ$. The shadow edge is indicated in black. The contours for $Z/f = 0$ provided by the two alternative expressions in Equation (74) are almost identical at this scale.

The values of Z/f and W/f along the shadow edge are shown in Figure 12, where $r_p = \sqrt{z^2 + \rho^2}$. The behaviour of Z is qualitatively similar to the variation along the axis, as shown in Figure 6.

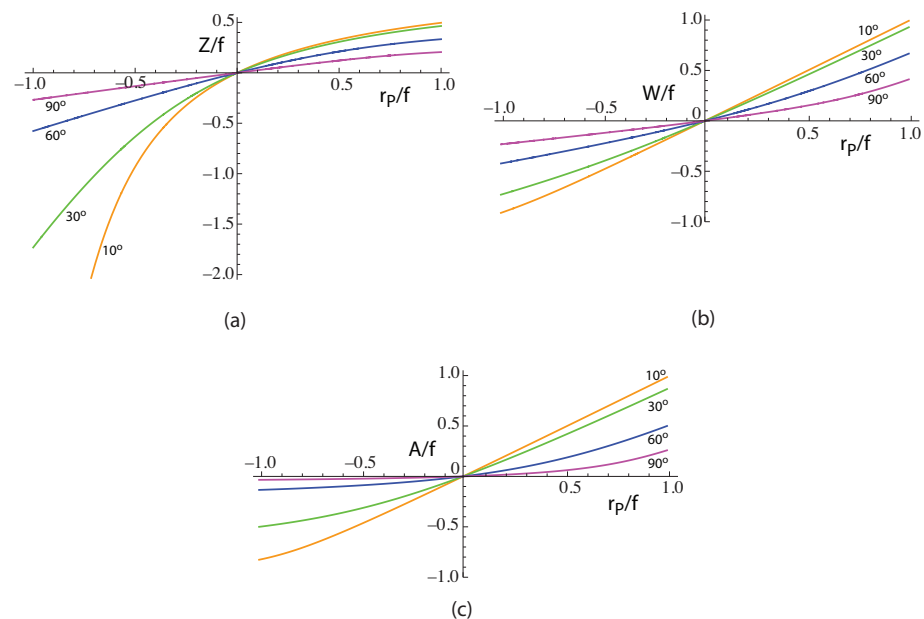


Figure 12. The variations in (a) Z/f , (b) W/f and (c) $A/f = [W - 2Z \sin^2(\alpha/2)]/f$ along the shadow edge. Here, $r_p = \sqrt{z^2 + \rho^2}$. Lines are coloured orange for $\alpha = 10^\circ$, green for $\alpha = 30^\circ$, blue for $\alpha = 60^\circ$, and purple for $\alpha = 90^\circ$.

In the far field, when $Z \gg f$ and $R \ll Z$, we have

$$m \approx 2Z \sin^2 \frac{\alpha}{2} + \frac{R^2}{2Z}, \quad (78)$$

which can be written as a quadratic equation in Z

$$4Z^2 \sin^2 \frac{\alpha}{2} - 2mZ + R^2 = 0 \quad (79)$$

and solved for Z in terms of m :

$$Z = \frac{m}{2 \sin^2(\alpha/2)} \left[1 - \frac{f^2 \rho^2 \sin^2(\alpha/2)}{g^2 m^2} \right], \quad (80)$$

where g is defined in Equation (35). In a similar way, the quadratic equation in the axial optical coordinate u is

$$u^2 - 2kmu + \left(1 + \tan^2 \frac{\alpha}{2} \right) v^2 = 0, \quad (81)$$

and its solution for u is

$$u = 2km \left\{ 1 - \frac{[1 + \tan^2(\alpha/2)] v^2}{4(km)^2} \right\}. \quad (82)$$

For m , in terms of the spatial coordinates, for $z \gg f$ and $\rho \ll z$ we have

$$\begin{aligned} m &= f + \sqrt{z^2 + \rho^2} - (g_1 + g_2)/2 \\ &\approx f + z + \frac{\rho^2}{2z} - g - \frac{(g^2 - f^2 \sin^2 \alpha) \rho^2}{2g^3} \\ &= (f + z - g) + \frac{\rho^2}{2g^3 z} [g^3 - z(f \cos \alpha + z)^2], \end{aligned} \quad (83)$$

meaning that for Z we have

$$Z = \frac{2fz}{f + z + g} \left\{ 1 + \frac{\rho^2(f + z + g)}{16fg^3z^2 \sin^2(\alpha/2)} [2g^3 - 2z(f \cos \alpha + z)^2 - fg(f + z + g)] \right\}. \quad (84)$$

This expression for Z predicts the behaviour of Z near the optical axis in the far field. We find that the curvature of the contours of constant Z reduce as z increases, then changes sign, reaches a maximum value, and decays as $1/z$. The position of the sign change agrees well with Figure 9 for $\alpha = 10^\circ$ or 30° . For $\alpha = 60^\circ$, it does not agree, probably because z/f is not sufficiently large for the approximations to be valid.

While the approach used thus far fares well for the illuminated region, the definitions of Z and W become more inaccurate as we move into the shadow region. This is because there are only two critical points in the shadow region. We have defined $A = W - 2Z \sin^2(\alpha/2)$, and can see from Equation (68) that because constant A corresponds to constant $(g_1 + g_2)$, the surfaces of constant A are oblate spheroids (surfaces of rotation of ellipses about their minor axes) with equations

$$\frac{\rho^2}{(A + f)^2} + \frac{(z + f \cos \alpha)^2}{A^2 + 2Af + f^2 \cos^2 \alpha} = 1. \quad (85)$$

The corresponding contours of A/f are shown in Figure 13. For small α , $A \approx W$; if $Z = 0$, then $A = W$. On the optical axis, from Equations (37), (68) and (62) we have

$$Z = \frac{f + z - g}{2 \sin^2(\alpha/2)}; \quad W = z; \quad A = g - f. \quad (86)$$

The value of A/f along the shadow edge is shown in Figure 12c.

If $A/f = 0$, then

$$z = \left(-f + \sqrt{f^2 - \rho^2} \right) \cos \alpha. \quad (87)$$

The contour for $A/f = 0$ is indicated in Figures 9 and 10. In all cases, this contour lies between the corresponding green and blue contours for both $Z/f = 0$ and $W/f = 0$. Z is small when $A = 0$; therefore, we propose taking as an approximation that $Z = 0, u = 0$ for points in the region between the two alternative contours for $Z = 0$, meaning that $W = A$. In this region, the intensity is very weak anyway, and small values of u result in only a small effect on the intensity. For the conventional Airy disk, the intensity for $u = 0$ falls off quickly, as $1/v^3$. Interestingly, we find that when $A = 0$, the two solutions for m are equal and opposite in sign; thus, their mean is zero.

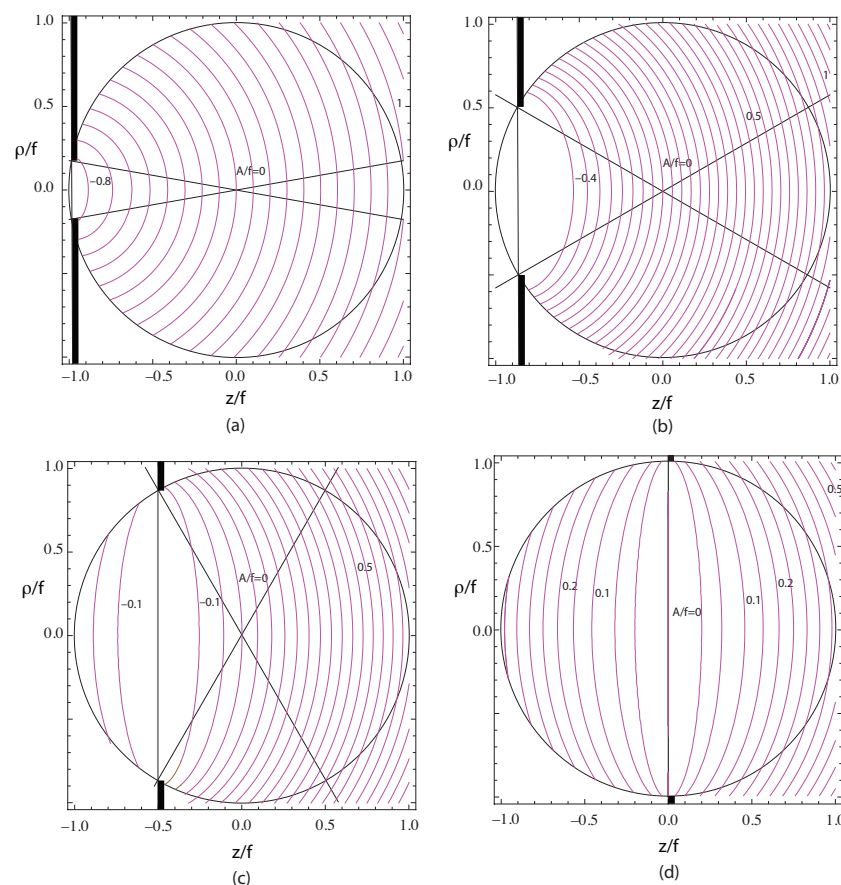


Figure 13. Contours of $A/f = [W - 2Z \sin^2(\alpha/2)]/f$ for the scaled Debye–Wolf theory for diffraction of a spherical wave by a circular aperture: (a) $\alpha = 10^\circ$, (b) $\alpha = 30^\circ$, (c) $\alpha = 60^\circ$, (d) $\alpha = 90^\circ$.

For this choice of values, Figure 14 shows the variation in R/f , Z/f , and W/f along three lines of constant z ($z = -0.2f, 0f, 0.2f$) and for four values of α equal to $10^\circ, 30^\circ, 60^\circ, 90^\circ$, as indicated by colour. For $z = 0$, there is a nonlinear variation in R as a function of ρ (except for $\alpha = 90^\circ$, when it is linear) and $Z = 0, W = 0$. The nonlinearity decreases as α increases. For $z = 0.2f$, Z/f and W/f exhibit break points at the shadow edge, marked by small circles; for $z = -0.2f$, the behaviour is more complicated for Z/f and W/f . There are up to three break points, located at the shadow edge and the two points where $Z = 0$. Between these, we assign $Z = 0$ and $W = A$.

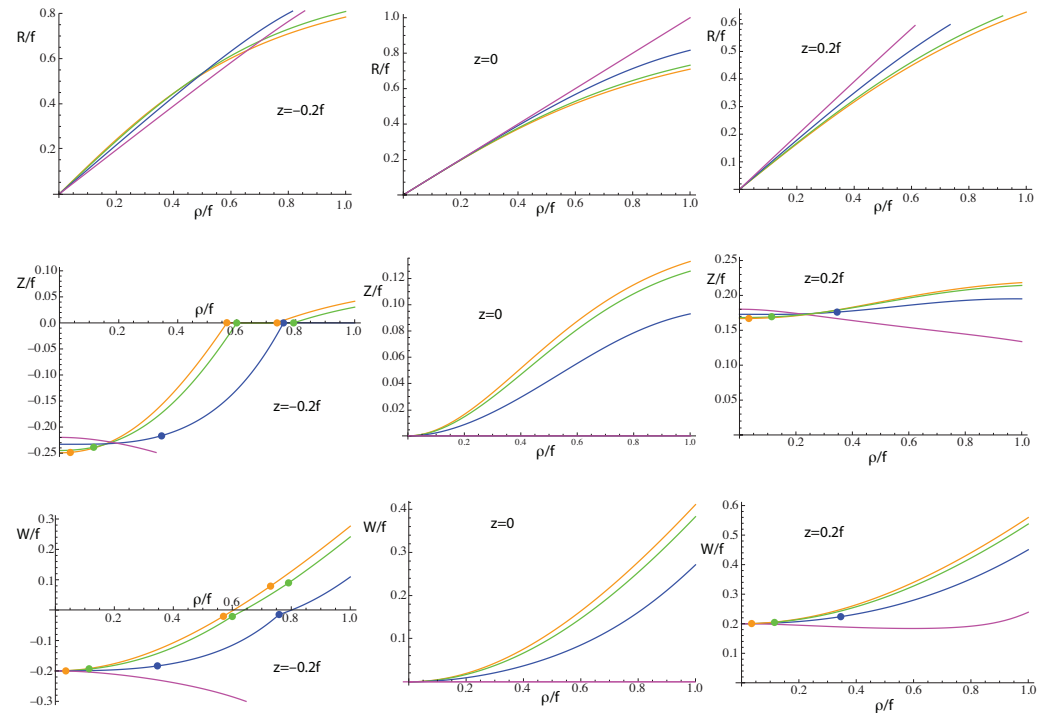


Figure 14. Plots of R/f (top row), Z/f (middle row), and W/f (bottom row) along lines of constant z/f ((left column) $z = -0.2f$, (middle column) $z = 0$, and (right column) $z = 0.2f$), shown for different values of α (indicated by colour, orange $\alpha = 10^\circ$, green $\alpha = 30^\circ$, blue $\alpha = 60^\circ$, and purple $\alpha = 90^\circ$). Break points are indicated by small circles.

This choice of Z/f and W/f in the shadow region is only one possible strategy. Another alternative is to take contours of both Z/f and W/f in the shadow region to be the same shape as contours for A/f , with the magnitude chosen such that the contours are continuous with those in the illuminated region at the shadow edge. We call this the corrected scaled Debye–Wolf theory (csDW). Figure 15 shows contours of constant Z/f and W/f for $\alpha = 30^\circ$ and 60° . Note that there are now no gaps in the coverage, unlike the previous strategy (sDW), which had a problem with the region around $Z/f = 0, W/f = 0$. csDW works well for negative z/f , as the contours are smooth through the shadow edge.

However, for positive z/f the contours of constant Z/f have a different shape, meaning that there is a discontinuity of the slope at the shadow edge. Preliminary numerical calculations have shown that the proposed choice of sDW results in better predictions than csDW near the shadow edge for positive z .

In fact, the shape of the contours in the illuminated region for $z/f < 0$ is very similar to that of the contours of A/f , and could be approximated by contours in A/f scaled to match on the optical axis. We call this the approximate sDW theory (asDW). Setting $g_s = (g_1 + g_2)/2$, the z coordinate of the contour on the optical axis is $z_a = -f \cos \alpha \pm \sqrt{g_s^2 - f^2 \sin^2 \alpha}$, where the negative sign is taken if $z_a < -f \cos \alpha$. Then, we have

$$2Z \sin^2(\alpha/2) = z_a + f - g_s; \quad W = z_a. \quad (88)$$

Contours of constant Z/f and W/f for $\alpha = 30^\circ$ and 60° are shown in Figure 16. For $z/f < 0$, the contours agree well with those for sDW and csDW. For $z/f > 0$, however, the contours for Z/f in particular, although agreeing on the optical axis, exhibit curvatures of opposite sign.

The case for $\alpha = 90^\circ$ is interesting. As there is no shadow region, the three critical points are valid everywhere, and Bertilone’s analytical solution for the field in the DW case is available. The equations for the optical coordinates also simplify, meaning that $Z = \pm \sqrt{m^2 - R^2}$ and $g_{1,2} = \sqrt{(f \pm \rho)^2 + z^2}$.

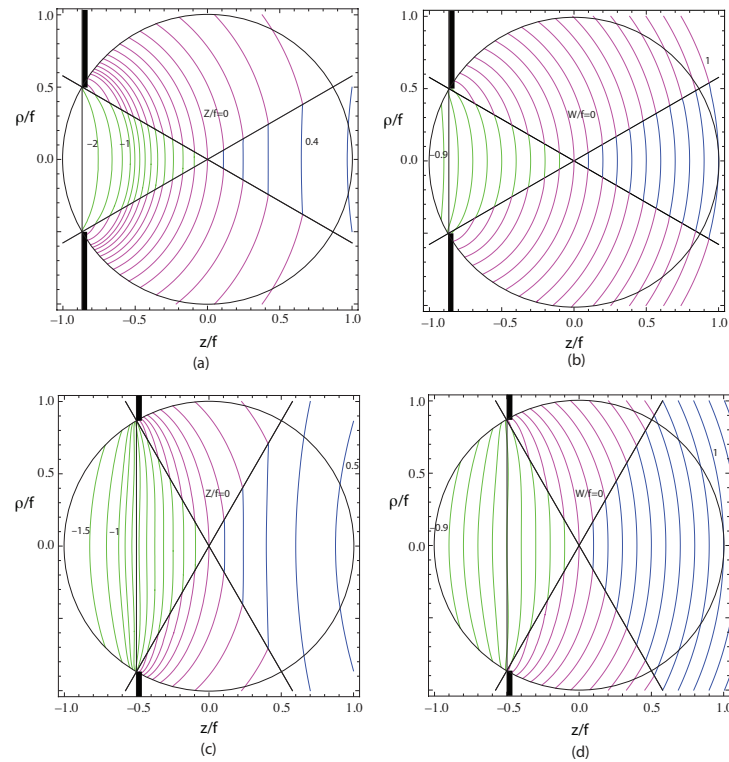


Figure 15. Plots showing contours of constant Z/f (left column: **a** and **c**) and W/f (right column: **b** and **d**) for $\alpha = 30^\circ$ (top row: **a** and **b**) and $\alpha = 60^\circ$ (bottom row: **c** and **d**) for the csDW theory. Colours: illuminated region, blue $z > 0$, green $z < 0$; shadow region, purple.

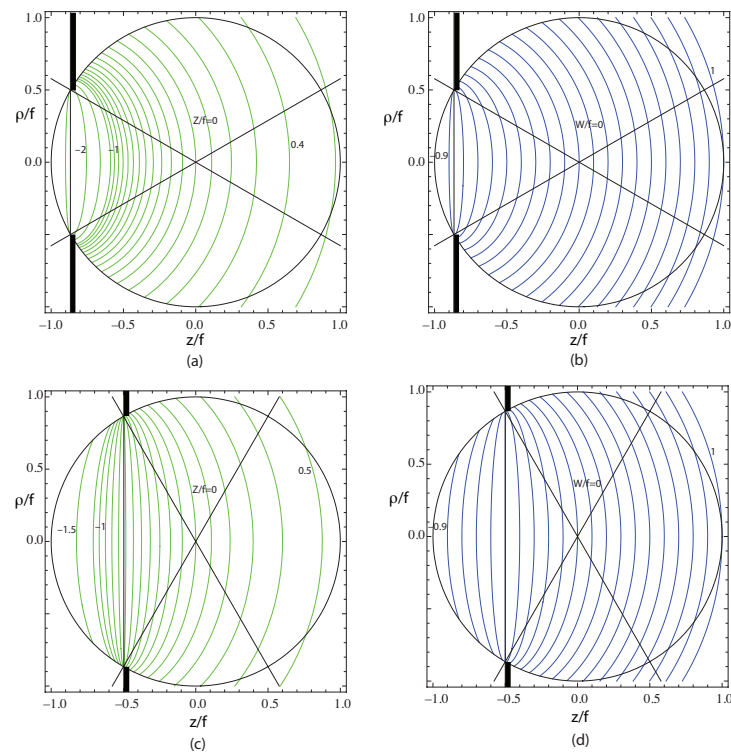


Figure 16. Plots showing contours of constant Z/f (in green) (left column: **a** and **c**) and W/f (in blue) (right column: **b** and **d**) for $\alpha = 30^\circ$ (top row: **a** and **b**) and $\alpha = 60^\circ$ (bottom row: **c** and **d**) for the asDW theory.

4.7. The Pseudo-Paraxial Approximation

In this paper, we have considered the scaled Debye–Wolf approach for approximating the focused field in the finite Fresnel number case as a scaled high-NA field. In a previous paper [64], we studied the pseudo-paraxial theory for a finite Fresnel number, which is not as accurate at high NAs but has the advantage that the analytical Lommel solution is available. This pseudo-paraxial theory is based on the generalized Fresnel approximation, i.e., a quadratic fit of the DW theory is made, matching up at the critical points. It is interesting to compare these two solutions, namely, the sDW and the pseudo-paraxial approximation to Ks. In the pseudo-paraxial approximation, the $k(r - f)$ in Equation (63) is matched to the parabola $-us^2/2 + vs \cos \phi + w$, where s is a radial variable normalized to unity at the aperture edge $\theta = \alpha$. The critical points corresponding to the aperture edge are the same as for sDW, as in Equation (66); thus, the values of R, v are the same as those in Equation (68). The third critical point satisfies the condition

$$\pm k\sqrt{z^2 + \rho^2} = w + \frac{v^2}{2u} = k\left(W + \frac{R^2 \cos^2(\alpha/2)}{2Z}\right); \quad (89)$$

thus, eliminating w ,

$$km = \frac{u^2 + v^2}{2u}. \quad (90)$$

In this way, we obtain quadratic equations for u or Z :

$$\begin{aligned} u^2 - 2kmu + v^2 &= 0, \\ 4Z^2 \sin^2 \frac{\alpha}{2} - 2mZ + R^2 \cos^2 \frac{\alpha}{2} &= 0, \end{aligned} \quad (91)$$

and solve for u or Z in terms of m :

$$\begin{aligned} u &= km \pm \sqrt{(km)^2 - v^2}, \\ Z &= \frac{m \pm \sqrt{m^2 - R^2 \sin^2 \alpha}}{4 \sin^2(\alpha/2)}, \end{aligned} \quad (92)$$

which reduce to the corresponding Equations (74) and (76) for either small α or on the axis. The intensity along the axis is identical to that in the sDW case, and the intensity along the shadow edge is similar for $\alpha = 10^\circ$ or $\alpha = 30^\circ$. However, plots of Z/f and W/f (Figure 17) show that the behaviour is quite different from the sDW case, even for $\alpha = 10^\circ$. Contours of constant small value of Z/f have an opposite sign of the curvature than the sDW case, which means that, although they are the same as sDW close to the axis where the contours are flat, there is actually only a single point, at the geometrical focus, where $Z/f = 0$ and $u = 0$. Unlike sDW, there is a region between the two contours for $W/f = 0$ corresponding to different signs of the square roots where the two solutions provide different values for W/f . These factors suggest that the pseudo-paraxial treatment is useful only in the illuminated region or close to it. The contours are similar, though more detailed, when compared to those shown in our previous paper (which were only shown for off-axis illumination) [64]. In our earlier paper, the expressions for Z, W were provided in terms of p, q , as in Equation (20) of this paper, whereas in the notation used in the present paper this is $(q^2 - p^2) = (g_1 - g_2)$ and $(p^2 + q^2) = 2m$.

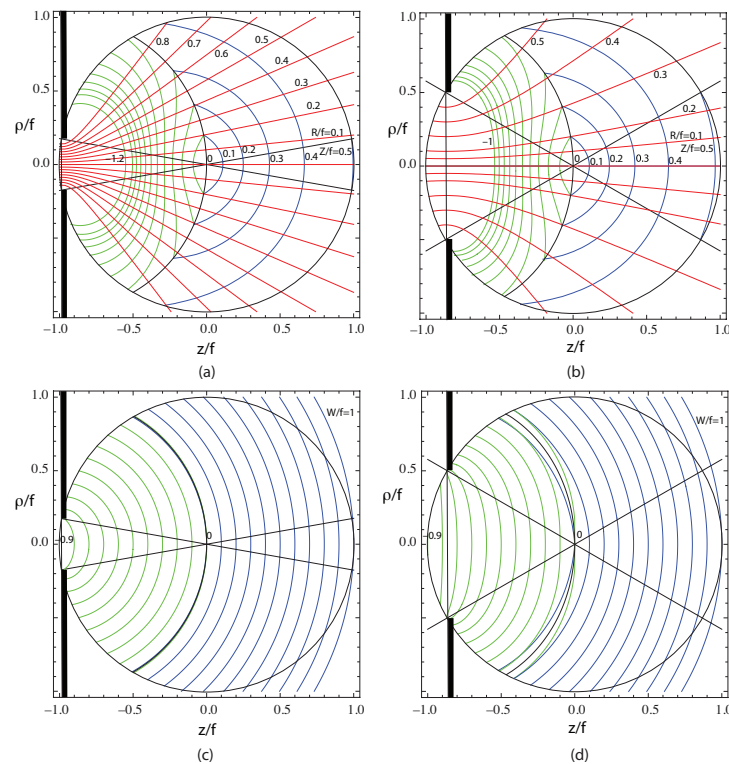


Figure 17. Contours of R/f (in red), Z/f and W/f (in green for $Z/f, W/f < 0$ and blue for $Z/f, W/f > 0$) for the pseudo-paraxial theory for diffraction of a spherical wave by a circular aperture: (a) R/f and Z/f , $\alpha = 10^\circ$, (b) R/f and Z/f , $\alpha = 30^\circ$, (c) W/f , $\alpha = 10^\circ$, (d) W/f , $\alpha = 30^\circ$. The black contours in (c,d) are for $A/f = 1$.

In the far field, $z \gg f$, near to the axis $\rho \ll z$,

$$Z \approx \frac{m}{2 \sin^2(\alpha/2)} \left[1 - \frac{f^2 \rho^2 \sin^2 \alpha}{4g^2 m^2} \right], \quad (93)$$

and

$$u \approx 2km \left\{ 1 - \frac{v^2}{4(km)^2} \right\}. \quad (94)$$

The parameter m is provided by Equation (83), meaning that Z becomes

$$Z = \frac{2fz}{f + z + g} \times \left\{ 1 + \frac{\rho^2(f + z + g)}{16fg^3z^2 \sin^2(\alpha/2)} \left[2g^3 - 2z(f \cos \alpha + z)^2 - fg(f + z + g) \cos^2(\alpha/2) \right] \right\}. \quad (95)$$

Although the curvature of the contours of Z for positive z in Figure 17 always has the same sign, we actually find that the curvature changes sign at about $z/f = 3.3$ (reducing as α increases).

4.8. The Kirchhoff Diffraction Integral Performed over the Plane of the Aperture (Kp)

For Ks in Equation (34), we found that the amplitude along the optical axis is provided by the destructive summation of equal contributions from a direct wave and a BDW, agreeing with the concepts of Young's BDW theory. Exact solutions for the amplitude along the optical axis predicted by RSI, RSII, and Kp have been published as well [3,4,39,74]. It is

found that all these solutions are of the same form, with only the relative strength of the BDW differing; thus, we now write a general form:

$$\begin{aligned} U(z) &= -\frac{f}{z} \{ \exp(ikz) - C(z/f) \exp[ik(g-f)] \} \\ &= -\frac{f}{z} e^{ikz} \left[1 - C(z/f) e^{-iu/2} \right], \end{aligned} \quad (96)$$

where $C(z/f)$ denotes the strength of the BDW. Here, C can alternatively be expressed in terms of u using Equation (47). As well as generalizing to these other diffraction theories, this approach allows for investigation of different apodization functions. Then, the intensity along the axis is

$$\begin{aligned} I &= \left(\frac{uf}{2z} \right)^2 \left\{ C^2 \left[\frac{\sin(u/4)}{u/4} \right]^2 + \left(\frac{2}{u} \right)^2 (1-C)^2 \right\} \\ &= I_0 D^2 \left\{ C^2 \left[\frac{\sin(u/4)}{u/4} \right]^2 + \left(\frac{2}{u} \right)^2 (1-C)^2 \right\}. \end{aligned} \quad (97)$$

For the different diffraction theories, $C(0) = 1$ such that $[1 - C(0)] = 0$, and the intensity at the geometrical focal point I_0 is unchanged from that described for Ks in Equation (53). Thus, the expressions for Ks in Equations (40) and (55) are easily modified.

RSI predicts positive focal shift at high NAs as a result of an incorrect assumed boundary condition [3]. This is contrary to observations; nevertheless, we consider RSI because it is a step towards calculating Kp. The value of C can be expressed in several different forms: in terms of the sides f, g, z or angles α, β, δ of the triangle $EF P$ with P on the optical axis. The value of C for RSI or RSII depends on the obliquity factor $\cos \alpha, \cos \beta$ for RSI and RSII, respectively, as well as on the change in the distance g . Wenow have

$$\begin{aligned} C_{RSI} &= \frac{z(d+z)}{g(g-f)} = \frac{z(d+z)(f+g)}{g(g^2-f^2)} = \frac{(f+g)(f \cos \alpha + z)}{g(2f \cos \alpha + z)} = \frac{2 \cos \beta}{\cos \alpha + \cos \beta} \cos^2 \frac{\delta}{2} \\ C_{RSII} &= \frac{zd}{f(g-f)} = \frac{zd(f+g)}{f(g^2-f^2)} = \frac{(f+g) \cos \alpha}{(2f \cos \alpha + z)} = \frac{2 \cos \alpha}{\cos \alpha + \cos \beta} \cos^2 \frac{\delta}{2} \\ C_{Kp} &= \frac{z[(f+g)d + fz]}{2fg(g-f)} = \frac{(f+g)[(f+g) \cos \alpha + z]}{2g(2f \cos \alpha + z)} = \cos^2 \frac{\delta}{2} \end{aligned} \quad (98)$$

The final expression for C_{Kp} is particularly simple, and is equal to the obliquity factor for the Kirchhoff diffraction integral performed over the spherical wavefront, as disussed by Kraus for the case of illumination by a diverging wave [63]. Then $(1 - C_{Kp})^2$ is $\sin^4(\delta/2)$. The expressions for C_{RSI} and C_{RSII} , however, do not seem to be simply related to the amplitudes of the components of the integrals RSI and RSII corresponding to the centre and edge of the aperture.

Plots of the variation in C^2, D^2 , and $(1 - C)^2$ with z/f are shown in Figure 18. D^2 is responsible for shifting the maximum axial intensity towards the screen for Ks. Its variation decreases as the NA increases. D^2 is multiplied by C^2 for RSI, RSII, and Kp. For RSI, C_{RSI}^2 increases for positive z (more strongly for larger α), which has the effect of producing a positive focal shift for high NA. On the other hand, C_{RSII}^2 increases for negative z , increasing the negative focal shift. For Kp, C_{Kp}^2 maintains a value of about one over a large variation in z/f ; thus, the focal shift is negative and close to (slightly greater than) that predicted by Ks. The value of $(1 - C)^2$ at the geometrical focal point is zero in all cases. Closer inspection shows that the relative strength of the complete second term of Equation (97) tends to zero at the geometrical focal point unless $N_f = 0$. Furthermore, it can be seen that $(1 - C_{Kp})^2$ remains smaller for a longer distance along the optical axis than $(1 - C_{RSI})^2$ or $(1 - C_{RSII})^2$.

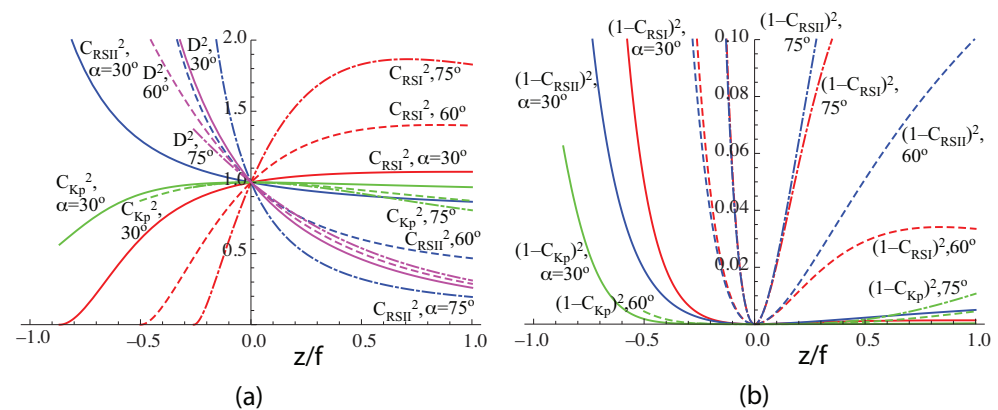


Figure 18. (a) The variation in C^2 with z/f for RSI (red), RSII (blue), and Kp (green). Curves are shown for $\alpha = 30^\circ$ (solid line), $\alpha = 60^\circ$ (dashed line), and $\alpha = 75^\circ$ (chained line). The variation in $D^2 = [2f/(f+z+g)]^2$ is shown in purple. (b) The variation in $(1-C)^2$ with z/f for RSI (red), RSII (blue), and Kp (green). Curves are shown for $\alpha = 30^\circ$ (solid line), $\alpha = 60^\circ$ (dashed line), and $\alpha = 75^\circ$ (chained line).

5. Discussion

It is commonly regarded that the RSI diffraction integral is superior to that of Kirchhoff, mainly because it is self-consistent; for example, Osterberg and Smith state that “Consequently, the authors have a strong preference for Rayleigh’s diffraction integral” [3]. While RSI does provide a rigorously correct result for a given input field, in the present case, where the input field is not known exactly, it provides results that are not in agreement with experiments. Thus, the present study is one where RSI (or RSII) is not the most appropriate solution. Many previous works have come to similar conclusions in cases where the input field is not known exactly, and consequently prefer the Kirchhoff solution [2]. In this paper, we have used an approximation $1/kf \ll 1$ to the Kirchhoff diffraction integral evaluated over a spherical wavefront, which we call Ks, to calculate the field along the axis for a convergent spherical wave diffracted by a circular aperture. The work is an extension of our previous studies [64,71,74]. We have found that Kp provides on-axis results that are quite similar to Ks. If we had not made the approximation in Ks, then integration over spherical or planar surface should provide exactly the same result, as discussed by Kraus for the case of diffraction of a divergent spherical wave [63]. This is because the amplitude provided by the sum of contributions from the undiffracted wave and an edge scattered wave are identical to the obliquity factors. In fact, Ks, Kp, RSI, and RSII all predict the same positions for the axial zeros, unlike the two versions of the Fresnel approximation, FrA1 and FrA2. This fact then leads to a generalized axial coordinate u that is valid for different Fresnel numbers or NAs (Equations (37), (41) and (46)). A further approximation to Ks, aKs, is valid for small axial distances from the geometrical focal point and has a simple algebraic form (Equation (56)), allowing an analytic expression for the fractional focal shift to be obtained (Equation (59)). The focal shift measured experimentally for low NA acoustic, microwaves, and visible light systems agrees well with theoretical predictions [21–24,32,54]. Li and Platzer provided results for Fresnel numbers less than 1 [54]. To the best of our knowledge, detailed measurements for high NA have not been reported. For high NA and low Fresnel number, the radius a of the aperture is necessarily quite small, and the theoretical treatment could be validated using finite-difference time domain (FDTD) calculations.

We then went on to develop the optical coordinates for off-axis observation points for nonparaxial focusing at high NA and finite Fresnel number. The optical coordinates v, u for the diffraction of a converging spherical wave in the paraxial Debye regime are orthogonal cylindrical coordinates proportional to ρ, z . The importance of the optical coordinates u and v is that they allow for prediction of geometrical distortion of the focal field distribution and axial scaling (for example, of the axial zeros in intensity) as system parameters are altered. For high NA and/or finite Fresnel number, the optical coordinates describe a geometrical

transformation that allows the position of different features of the diffraction pattern to be determined. Unlike a previous paper [64], we do not assume that the focused field is a geometrically distorted version of the Lommel (paraxial Debye) solution but rather of the non-paraxial Debye case [71]. The appropriate geometrical transformation for the illuminated region is derived for this sDW case. A notable observation is that surfaces of constant defocus optical coordinate u are curved and approximately spherical, rather than being planar, as predicted by Li and Wolf for the paraxial case [50]. Previously, we have found this transformation to provide good predictions of the focused field in the shadow region close to the shadow edge ($v/u < 1.2$), although it eventually fails further into the shadow region. Nonetheless, the transverse optical coordinate v remains valid, and for much of the shadow region where the intensity is not negligible the axial optical coordinate is small and can be taken as zero, as it does not appreciably affect the amplitude.

Forms for the transverse optical coordinate v valid for systems of both high NA and finite Fresnel number are provided in Equation (69). Forms for the optical coordinate u along the axis valid for systems of both high NA and finite Fresnel number are provided in Equations (37), (41) and (47). An approximate expression for u for points near the geometrical focus is provided in Equation (56). All of the theories, namely, Ks, Kp, RSI, and RSII, result in the same expression for the transverse optical coordinate v and for the on-axis variation in the optical coordinate u , as these follow directly from the difference in optical path between the axial and edge diffracted waves. Therefore, the optical coordinates are applicable for the vectorial case as well. These coordinates reduce to the well-known forms for both paraxial systems of finite Fresnel number and for nonparaxial systems in which the Fresnel number can be assumed to be infinite.

We additionally discussed a different approach, csDW, that avoids the problem of sDW near the contour for $A = 0$. This approach provides reasonable agreement for $z/f < 0$, but results in discontinuities of slope in the contours at the shadow edge for $z/f > 0$. Another approach, asDW, again agrees well for $z/f < 0$ but provides a curvature of the opposite sign compared with sDW for contours of Z/f for $z/f > 0$.

The optical coordinates for the sDW theory were then compared with those for the pseudo-paraxial approach based on the generalized Fresnel approximation. Although these agree with each other in the limiting case of $\alpha \rightarrow 0$, even for $\alpha = 10^\circ$, the pseudo-paraxial theory behaves well only within or near to the shadow edge. In particular, the value $Z/f = 0$ occurs only at the geometrical focal point.

The Kirchhoff diffraction integral performed over a spherical surface, Ks, was then compared with Kp integrated over a planar surface and the two Rayleigh–Sommerfeld integrals, RSI and RSII. On the axis, the field predicted by all of these is provided by the sum of a central undiffracted component and a wave scattered by the aperture edge. The intensity minima all agree with each other, but differ from the predictions of the conventional Fresnel approximation. Kp, RSI, and RSII all predict non-zero minima, unlike Ks, which includes the approximation $1/kf \ll 1$. As we have mentioned, applying the Kirchhoff diffraction formula over the wave front but without the approximation provides the same result as Kp on-axis. Off-axis, we are of the opinion that the Kirchhoff diffraction formula performed over the wave front without approximation should not provide exactly the same results for the field as Kp, as although the obliquity functions are the same, different Kirchhoff approximations to the incident field are used. Neither RSI or RSII provide predictions that agree with experimental observations as a result of the inaccuracy of the assumed field in the aperture plane. The off-axis behaviour of the predictions of the different approaches for high NA and low Fresnel number could be compared using FDTD.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

aKs	approximate Kirchhoff diffraction formula, integrated over a spherical wavefront
asDW	Approximate scaled Debye–Wolf
BDW	Boundary diffracted wave
csDW	Corrected scaled Debye–Wolf
DW	Debye–Wolf
FrA1	Fresnel approximation 1, dividing by z
FrA2	Fresnel approximation 2, dividing by r
gFrA	Generalized Fresnel approximation
HF	Huygens–Fresnel integral
Kp	Kirchhoff diffraction formula, integrated over a planar surface
Ks	Kirchhoff diffraction formula, integrated over a spherical wavefront
RSI	First Rayleigh–Sommerfeld diffraction formula
RSII	Second Rayleigh–Sommerfeld diffraction formula
sDW	Scaled Debye–Wolf

References

1. Daly, C.J.; Rao, N.A.H.K. *Scalar Diffraction from a Circular Aperture*; Kluwer Academic: Amsterdam, The Netherlands, 2000.
2. Stamnes, J.J. *Waves in Focal Regions*; Adam Hilger: Bristol, UK, 1986.
3. Osterberg, H.; Smith, L.W. Closed solutions of Rayleigh’s diffraction integral for axial points. *J. Opt. Soc. Am.* **1961**, *51*, 1050–1054. [[CrossRef](#)]
4. Li, Y. Predictions of Rayleigh’s diffraction theory for the effect of focal shift in high-aperture systems. *J. Opt. Soc. Am. A* **2008**, *25*, 1835–1842. [[CrossRef](#)] [[PubMed](#)]
5. Aime, C. Fresnel diffraction of multiple disks on axis: Application to coronagraphy. *Astron. Astrophys.* **2020**, *637*, A16. [[CrossRef](#)]
6. Basistiy, I.; Bazhenov, V.; Soskin, M.; Vasnetsov, M. Optics of light beams with screw dislocations. *Opt. Commun.* **1993**, *103*, 422–428. [[CrossRef](#)]
7. Khonina, S.N.; Ustinov, A.V.; Kovalyov, A.A.; Volotovskiy, S.G. Near field propagation of vortex beams: Models and computation algorithms. *Opt. Mem. Neural Netw.* **2014**, *23*, 50–73. [[CrossRef](#)]
8. Engelberg, J.; Levy, L. The advantages of metalenses over diffractive lenses. *Nat. Commun.* **2020**, *11*, 1991. [[CrossRef](#)]
9. Gonçalves, M.R.; Rozenman, G.G.; Zimmerman, M.; Efremov, M.A.; Case, W.B.; Arie, A.; Shemer, L.; Schleich, W.P. Bright and dark diffractive focusing. *Appl. Phys. B* **2022**, *128*, 51. [[CrossRef](#)]
10. Weisman, D.; Carmesin, C.M.; Rozenman, G.G.; Efremov, M.A.; Shemer, L.; Schleich, W.; Arie, A. Diffractive guiding of waves by a periodic array of slits. *Phys. Rev. Lett.* **2021**, *127*, 014303. [[CrossRef](#)]
11. Young, T. The Bakerian Lecture. Experiments and calculations relative to physical optics. *Philos. Trans. R. Soc. Lond.* **1804**, *94*, 1–16.
12. Airy, G.B. On the diffraction of an object-glass with circular aperture. *Trans. Camb. Philos. Soc.* **1835**, *5*, 283–291.
13. Lommel, E. Die Beugungserscheinungen einer kreisrunden Oeffnung und eines kreisrunden Schirmschens theoretisch und experimentell Bearbeitet. *Abh. Bayer. Akad.* **1885**, *15*, 233–328.
14. Born, M.; Wolf, E. *Principles of Optics*, 1st ed.; Pergamon: Oxford, UK, 1959.
15. Nijboer, B. The diffraction theory of optical aberrations. Part I: General discussion of the geometrical aberrations. *Physica* **1943**, *10*, 679–692. [[CrossRef](#)]
16. Boersma, J. On the computation of Lommel’s functions of two variables. *Math. Comput.* **1962**, *16*, 232–238.
17. Rayleigh, J.W.S. On pin-hole photography. *Philos. Mag.* **1891**, *31*, 87–91. [[CrossRef](#)]
18. Rayleigh, J.W.S. On the passage of waves through apertures in plane screens, and allied problems. *Philos. Mag.* **1897**, *43*, 259–272. [[CrossRef](#)]
19. Sommerfeld, A. *Optics*; Lectures on theoretical physics; Academic Press: New York, NY, USA, 1964.
20. Debye, P. Das Verhalten von Lichtwellen in der Nähe eines Brennpunktes oder einer Brennnlinie. *Ann. Der Phys.* **1909**, *30*, 755–776. [[CrossRef](#)]
21. Williams, A.O., Jr. Acoustic intensity distribution from a ‘piston’ source. II The concave piston. *J. Acoust. Soc. Am.* **1946**, *17*, 219–227. [[CrossRef](#)]
22. Fein, L. Ultrasonic radiation from curved quartz crystals. *J. Acoust. Soc. Am.* **1949**, *21*, 511–516. [[CrossRef](#)]
23. O’Neil, H.T. Theory of focused radiators. *J. Acoust. Soc. Am.* **1949**, *21*, 516–526. [[CrossRef](#)]
24. Lucas, B.G.; Muir, T.G. The field of a focusing source. *J. Acoust. Soc. Am.* **1982**, *72*, 1289. [[CrossRef](#)]
25. Andrews, C.L. Diffraction pattern of a circular aperture at short distances. *Phys. Rev.* **1947**, *71*, 777–786. [[CrossRef](#)]

26. Bouwkamp, C.J. On the freely vibrating circular disk and the diffraction by circular disks and apertures. *Physica* **1950**, *16*, 1–16. [[CrossRef](#)]
27. Bouwkamp, C.J. On the diffraction of electromagnetic waves by small circular disks and holes. *Philips Res. Rep.* **1950**, *5*, 401–422.
28. Bouwkamp, C.J. Theoretical and numerical treatment of diffraction through a circular aperture. *IEEE Trans. Antennas Propag.* **1970**, *AP-18*, 152–176. [[CrossRef](#)]
29. Bouwkamp, C.J. Diffraction theory. *Rep. Prog. Phys.* **1954**, *17*, 35–100. [[CrossRef](#)]
30. Linfoot, E.H.; Wolf, E. Diffraction images in systems with an annular aperture. *Proc. Phys. Soc. B* **1953**, *66*, 145–149. [[CrossRef](#)]
31. Farnell, G.W. Calculated intensity and phase distribution in the image space of a microwave lens. *Can. J. Phys.* **1957**, *35*, 777–783. [[CrossRef](#)]
32. Farnell, G.W. Measured phase distribution in the image space of a microwave lens. *Can. J. Phys.* **1958**, *36*, 935–943. [[CrossRef](#)]
33. Richards, B.; Wolf, E. Electromagnetic diffraction in optical systems. II Structure of the image field in an aplanatic system. *Proc. R. Soc. Lond. A* **1959**, *253*, 358–379.
34. McCutchen, C.W. Generalized aperture and the three-dimensional diffraction image. *J. Opt. Soc. Am.* **1964**, *54*, 240–244. [[CrossRef](#)]
35. Kogelnik, H.; Li, T. Laser beams and resonators. *Appl. Opt.* **1966**, *5*, 1550–1567. [[CrossRef](#)] [[PubMed](#)]
36. Sherman, G.C. Application of the convolution theorem to Rayleigh's integral formulas. *J. Opt. Soc. Am.* **1967**, *57*, 546–547. [[CrossRef](#)] [[PubMed](#)]
37. Dainty, J.C. The image of a point for an aberration free lens with a circular pupil. *Opt. Commun.* **1969**, *1*, 176–178. [[CrossRef](#)]
38. Zemanek, J. Beam behavior within the near-field of a vibrating piston. *J. Acoust. Soc. Am.* **1971**, *49*, 181–191. [[CrossRef](#)]
39. Heurtley, J.C. Scalar Rayleigh-Sommerfeld and Kirchhoff diffraction integrals: A comparison of exact evaluations for axial points. *J. Opt. Soc. Am.* **1973**, *73*, 1003–1008. [[CrossRef](#)]
40. Welford, W.T. Aplanatic hologram lenses on spherical surfaces. *Opt. Commun.* **1973**, *9*, 268–269. [[CrossRef](#)]
41. Papoulis, A. Ambiguity function in Fourier optics. *J. Opt. Soc. Am.* **1974**, *64*, 779–788. [[CrossRef](#)]
42. Carter, W. Band-limited angular spectrum approximating to a spherical wave field. *J. Opt. Soc. Am.* **1975**, *65*, 1054–1058. [[CrossRef](#)]
43. Arimoto, A. Intensity distribution of aberration-free diffraction patterns due to circular apertures in large f-number optical systems. *Opt. Acta* **1976**, *23*, 245–250. [[CrossRef](#)]
44. Harvey, J.E. Fourier treatment of near-field scalar diffraction theory. *Am. J. Phys.* **1979**, *47*, 974–980. [[CrossRef](#)]
45. Southwell, W.H. Validity of the Fresnel approximation in the near field. *J. Opt. Soc. Am.* **1981**, *71*, 7–14. [[CrossRef](#)]
46. Erkkila, J.H.; Rogers, M.E. Diffracted fields in the focal region of a convergent wave. *J. Opt. Soc. Am.* **1981**, *71*, 904–905. [[CrossRef](#)]
47. Li, Y.; Wolf, E. Focal shifts in diffracted converging spherical waves. *Opt. Commun.* **1981**, *39*, 211–215. [[CrossRef](#)]
48. Stamnes, J.J.; Spjelkavik, S. Focusing at small angular apertures in the Debye and Kirchhoff approximations. *Opt. Commun.* **1981**, *40*, 81–85. [[CrossRef](#)]
49. Sheppard, C.J.R.; Wilson, T. Effects of high angles of convergence on $V(z)$ in the scanning acoustic microscope. *Appl. Phys. Lett.* **1981**, *38*, 858–859. [[CrossRef](#)]
50. Li, Y.; Wolf, E. Three-dimensional intensity distribution near the focus in systems of different Fresnel numbers. *J. Opt. Soc. Am. A* **1984**, *1*, 801–808. [[CrossRef](#)]
51. Born, M.; Wolf, E. *Principles of Optics*, 7th (expanded) ed.; Pergamon: Oxford, UK, 1999.
52. Wilson, T.; Sheppard, C.J.R. Imaging with finite values of Fresnel number. *J. Opt. Soc. Am.* **1982**, *72*, 15–18. [[CrossRef](#)]
53. Li, Y. Dependence of the focal shift on Fresnel number and f number. *J. Opt. Soc. Am.* **1982**, *72*, 770–774. [[CrossRef](#)]
54. Li, Y.; Platzer, H. An experimental investigation of diffraction patterns in low-Fresnel-number focusing systems. *Opt. Acta* **1983**, *31*, 1621–1643. [[CrossRef](#)]
55. Sheppard, C.J.R. Imaging in optical systems of finite Fresnel number. *J. Opt. Soc. Am.* **1986**, *A3*, 1428–1432. [[CrossRef](#)]
56. Li, Y. Three-dimensional intensity distribution in low Fresnel number focusing systems. *J. Opt. Soc. Am. A* **1987**, *4*, 1349–1353. [[CrossRef](#)]
57. Sheppard, C.J.R.; Matthews, H.J. Imaging in high aperture optical systems. *J. Opt. Soc. Am. A* **1987**, *4*, 1354–1360. [[CrossRef](#)]
58. English, R., Jr.; George, N. Diffraction from a circular aperture: On axis field strength. *Appl. Opt.* **1987**, *26*, 2760–2763. [[CrossRef](#)] [[PubMed](#)]
59. Steane, A.M.; Rutt, H.N. Diffraction calculations in the near field and the validity of the Fresnel approximation. *J. Opt. Soc. Am. A* **1989**, *6*, 1809–1814. [[CrossRef](#)]
60. Bertilone, D.C. The contributions of homogeneous and evanescent plane waves to the scalar optical field: Exact diffraction formulae. *J. Mod. Opt.* **1991**, *38*, 865–875. [[CrossRef](#)]
61. Bertilone, D.C. Wave theory for a converging spherical incident wave in an infinite-aperture system. *J. Mod. Opt.* **1991**, *38*, 1531–1536. [[CrossRef](#)]
62. Kraus, H.G. Huygens-Fresnel-Kirchhoff wave-front diffraction formulation: Spherical waves. *J. Opt. Soc. Am. A* **1989**, *6*, 1196–1205. [[CrossRef](#)]
63. Kraus, H.G. Huygens-Fresnel-Kirchhoff wave-front diffraction formulations for spherical waves and Gaussian laser beams: Discussion and errata. *J. Opt. Soc. Am. A* **1992**, *9*, 1132–1134. [[CrossRef](#)]
64. Sheppard, C.J.R.; Hrynevych, M. Diffraction by a circular aperture: A generalization of Fresnel diffraction theory. *J. Opt. Soc. Am. A* **1992**, *A9*, 274–281. [[CrossRef](#)]

65. Hrynevych, M. Diffraction Effects in Michelson Stellar Interferometry. Ph.D. Thesis, University of Sydney, Sydney, NSW, Australia, 1992.
66. Andrés, P.; Martínez-Corral, M.; Ojeda-Castañeda, J. Off-axis focal shift for rotationally nonsymmetric screens. *Opt. Lett.* **1993**, *18*, 1290–1293. [[CrossRef](#)]
67. Hsu, W.; Barakat, R. Stratton–Chu vectorial diffraction of electromagnetic fields by apertures with application to small-Fresnel-number systems. *J. Opt. Soc. Am.* **1994**, *A11*, 623–629. [[CrossRef](#)]
68. Wang, W.; Friberg, A.T.; Wolf, E. Structure of focused fields in systems with large Fresnel numbers. *J. Opt. Soc. Am. A* **1995**, *12*, 1947–1953. [[CrossRef](#)]
69. Wang, W.; Wolf, E. Far-zone behavior of focused fields in systems with different Fresnel numbers. *Opt. Commun.* **1995**, *119*, 453–459. [[CrossRef](#)]
70. Forbes, G.W. Validity of the Fresnel approximation in the diffraction of collimated beams. *J. Opt. Soc. Am. A* **1996**, *13*, 1816–1826. [[CrossRef](#)]
71. Sheppard, C.J.R.; Török, P. Dependence of focal shift on Fresnel number and angular aperture. *Opt. Lett.* **1998**, *23*, 1803–1804. [[CrossRef](#)]
72. Sheppard, C.J.R.; Török, P. Effects of Fresnel number in focusing and imaging. *Proc. SPIE* **1999**, *3729*, 458–472.
73. Sheppard, C.J.R. Validity of the Debye approximation. *Opt. Lett.* **2000**, *25*, 1660–1662. [[CrossRef](#)] [[PubMed](#)]
74. Sheppard, C.J.R.; Török, P. Focal shift and the axial optical coordinate for high-aperture systems of finite Fresnel number. *J. Opt. Soc. Am. A* **2003**, *20*, 2156–2162. [[CrossRef](#)]
75. Teng, S.; Liu, L.; Liu, D. Analytic expression of the diffraction of a circular aperture. *Optik* **2005**, *116*, 568–572. [[CrossRef](#)]
76. Lin, J.; Yuan, X.C.; Kou, S.S.; Sheppard, C.J.R.; Rodríguez-Herrera, O.G.; Dainty, J.C. Direct calculation of a three-dimensional diffraction field. *Opt. Lett.* **2011**, *36*, 1341–1343. [[CrossRef](#)]
77. Kou, S.S.; Sheppard, C.J.R.; Lin, J. Evaluation of the Rayleigh–Sommerfeld diffraction formula with 3D convolution: The 3D angular spectrum (3D-AS) method. *Opt. Lett.* **2013**, *38*, 5296–5299. [[CrossRef](#)] [[PubMed](#)]
78. Glückstad, J.; Madsen, A.E.G. New analytical diffraction expressions for the Fresnel–Fraunhofer transition regime. *Optik* **2023**, *285*, 170950. [[CrossRef](#)]
79. Li, Y. Three-dimensional intensity distribution in the far zone of focused fields in systems with different Fresnel numbers. *J. Opt. Soc. Am. A* **2023**, *40*, 2197–2204. [[CrossRef](#)]
80. Murty, M.V.R.K. On the theoretical limit of resolution. *J. Opt. Soc. Am.* **1957**, *47*, 667–668. [[CrossRef](#)]
81. Zverev, V.A. Illumination distribution in the diffraction image of an off-axis point. *Sov. J. Opt. Technol.* **1986**, *53*, 451–454.
82. Gibson, S.F.; Lanni, F. Diffraction by a circular aperture as a model for three-dimensional optical microscopy. *J. Opt. Soc. Am. A* **1989**, *6*, 1357–1367. [[CrossRef](#)] [[PubMed](#)]
83. Sheppard, C.J.R.; Roberts, P.P.; Gu, M. Fresnel approximation for off-axis illumination of a circular aperture. *J. Opt. Soc. Am.* **1993**, *A10*, 984–986. [[CrossRef](#)]
84. Sheppard, C.J.R.; Hegedus, Z. Resolution for off-axis illumination. *J. Opt. Soc. Am. A* **1998**, *15*, 622–624. [[CrossRef](#)]
85. Sheppard, C.J.R.; Cooper, I.J. Fresnel diffraction by a circular aperture with off-axis illumination and its use in deconvolution of microscope images. *J. Opt. Soc. Am. A* **2004**, *21*, 540–545. [[CrossRef](#)]
86. Sheppard, C.J.R.; Kou, S.S.; Lin, J. The Hankel transform in n -dimensions and its applications in optical propagation and imaging. In *Advances in Imaging and Electron Physics*; Hawkes, P.W., Ed.; Academic Press: Burlington, NJ, USA, 2015; Volume 188, pp. 135–184.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.