From Airy to Abbe: quantifying the effects of wide-angle focusing for scalar spherical waves

Yitzi M Calm, Juan M Merlo, Michael J Burns and Michael J Naughton

Department of Physics, Boston College, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, United States of America

E-mail: naughton@bc.edu

Received 9 June 2017, revised 22 August 2017
Accepted for publication 31 August 2017
Published 22 September 2017

Abstract

Recent advances in optical microscopy have enabled imaging with spatial resolution beyond the diffraction limit. This limit is sometimes taken as one of several different criteria according to different conventions, including Rayleigh’s $0.61\lambda/NA$, Abbe’s $0.5\lambda/NA$, and Sparrow’s $0.47\lambda/NA$. In this paper, we perform a parametric study, numerically integrating the scalar Kirchhoff diffraction integrals, and we propose new functional forms for the resolution limits derived from scalar focusing. The new expressions remain accurate under wide angle focusing, up to $90^\circ$. Our results could materially impact the design of high intensity focused ultrasound systems, and can be used as a qualitative guideline for the design of a particular type of planar optical element: the flat lens metasurface.

Keywords: diffraction limit, scalar focusing, spatial resolution

(Some figures may appear in colour only in the online journal)

1. Introduction

1.1. Motivation

The role of diffraction in imaging systems has been studied since the days of Airy [1], Abbe [2, 3], and Rayleigh [4]. Textbook theory [5–9] says that conventional optical microscopes cannot resolve spatial features finer than approximately half the wavelength of light, $\lambda/2$. Pushing beyond the diffraction limit, near-field techniques based on scanning probe and plasmonic technologies collect information present only in evanescent waves [10, 11]. Far-field techniques based on fluorescence microscopy [12, 13] beat the diffraction limit with a priori information about the positions of fluorescent molecules (or about their concentration). Far-field super-resolution has even been demonstrated for non-fluorescent specimens, for example by using a super-oscillatory lens [14]. With this context, the motivation for the present work is to reexamine the textbook resolution limits of a conventional optical microscope:

$$d_x = \frac{\lambda}{NA}$$

(1)

$$d_z = \frac{\zeta \lambda n}{NA^2}$$

(2)

$d_x$ and $d_z$ are, respectively, the minimum transverse and longitudinal separations between two ‘barely resolved’ point sources, $\alpha$ is the half angle subtended by the objective lens, $n$ is the refractive index of the immersion medium, and $NA = n \sin \alpha$ is the numerical aperture. The numerical factors $\chi$ and $\zeta$ take on different values with different definitions of ‘barely resolved’. For example, $\chi \approx 0.61$ according Rayleigh, $\chi = 1/2$ according to Abbe, and $\chi \approx 0.47$ according to Sparrow (see page 474 of [5]); some sources [15] use $\zeta = 2$, while others [16] use $\zeta = 1.4$. Although the different numerical values stem from important conceptual differences, we present them here succinctly in order to highlight the
is the Fresnel constant, as shown in Figure 1. Figure 1 shows a sketch of the focusing problem.

There is an expansive body of work on the theory, computation, and experiments within the focal region, including textbooks [17] as well as volumes of journal articles that deal exclusively with the electromagnetic [18] and scalar [19] aspects of focusing. In broad terms, scalar and electromagnetic theory give the same result to within 5% for angles up to \( \alpha \sim 50^\circ \). For example, figure 8 of [20] shows less than a 1% difference for \( \alpha < 30^\circ \). For strong focusing (i.e. ‘wide angle’, \( \alpha > 50^\circ \)), polarization effects become increasingly important, see in particular figures 7 and 8 of [21].

We intentionally restrict the scope of the present work to scalar theory. As such, our results have limited quantitative applicability to high-NA optical microscopy, where objective lens angles can approach \( \alpha \sim 70^\circ \). For a detailed discussion of polarization effects, we refer the interested reader to the works cited in [18], chapters 3 and 4 of [9], or chapters 15 and 16 of [17]. However, scalar theory is often used as a qualitative guideline in the design of new optical elements [22, 23], and furthermore our scalar results can be applied quantitatively to the focusing of waves other than light, for example in high intensity focused ultrasound [24]. With our numerical results, we quantify the effects of wide angle focusing in a way that has not been previously reported, and our novel analytic formulae retain an accurate \( \alpha \)-dependence (within scalar theory) beyond the paraxial limit.

### 1.3. Analytic solutions to the focusing problem

For axial points, given by \( \rho = 0 \), the diffraction integrals can be evaluated exactly [26–30], and from the exact solution there are two important observations: first, the \( z \)-position of peak intensity is shifted toward the lens, \( z_{\text{peak}} < f \); and second, the intensity distribution generally does not have inversion symmetry about any \( z = \text{const} \) plane. The focal shift [31–35] is a robust feature for any coherent, converging wave [36], and inversion symmetry emerges for lenses with sufficiently large \( N \) [37].

For off-axis points, given by \( \rho > 0 \), the diffraction integrals cannot be evaluated exactly. Analytically approximate solutions [38–40] typically employ the Debye approximation [8, 17], which is equivalent to the Kirchhoff approximation for sufficiently large \( N \) [41–44].

Debye’s approximation in the paraxial limit (i.e. Fraunhofer diffraction), yields the seminal result of Airy [1]. Airy’s formulae, dropping all factors of \( \lambda \) and \( N A \) and retaining only the dependence on position \((x, z)\), give the intensity in the focal plane as a \( \text{jinc}^2 x \) distribution. The intensity distribution, \( I = |I|^2 \), is called the point spread function (PSF), and it is from Airy’s PSF that the Rayleigh criterion [4] is derived, \( \chi \approx 0.61 \) in equation (1). Similarly, \( \zeta = 2 \) in equation (2) also comes from Airy’s PSF, for which axial points produce a \( \text{sinc}^2 z = (\sin (z)/z)^2 \) dependence.

### 1.4. Numeric solutions to the focusing problem

For numerically approximate solutions, the wave at the input plane, \( u_{\text{sphere}}(\rho) \), is sampled discretely. To compute the wave function in the focal region, one commonly employs the discrete Fourier transform (DFT), wherein the computation is
done by either the summation of a discrete summand or by a
direction. A noncomprehensive selection of DFT
implementations can be found in
[45–48]. In the present work
we employ direct numeric integration
(DNI), in which the
computation is done by numerically integrating a continuous
integrand
(which is sampled recursively).
Although a DFT is
generally orders of magnitude faster than a DNI
[45], one
must take extra care when employing DFT to avoid aliasing
[49] and to mitigate errors
[50] and artifacts
[51], which result
from the discrete sampling and tiling.

2. Numerical methods

All numerical computation is performed using the built-in
functionality of commercially available software (Mathematica
and MATLAB) running on personal computers. To save time
on lengthy calculations, we network several computers toge-
ther for parallel computation on a grid environment. In total,
we have access to 64 processing cores at nominally 3 GHz and
400 GB of local RAM.

We use the angular spectrum method, as described
elsewhere [7, 17], and the underlying uniaxial symmetry
allows one to expand \( u(\mathbf{r}) \) in a basis of cylindrical waves
using the Fourier–Bessel transform (FBT):

\[
U(k) = \int_0^\infty d\rho \rho J_0(k\rho) u(\rho),
\]

where \( k \) gives the transverse (i.e. radial) component of the
wave’s momentum, \( U(k) \) is the reciprocal space spectrum,
and \( J_0(x) \) is a Bessel function of the first kind, order 0. With
equations (3)–(5), (7) and (8), one can compute the wave
function for \( z > 0 \) (denoting the inverse FBT by \( \text{iFBT} \)):

\[
\begin{align*}
\text{FBT} \left[ \exp \left( i z \sqrt{n k_0^2 - k^2} \right) \times \text{FBT} \left[ u_{\text{sphere}}(\rho_0) \right] \right].
\end{align*}
\]

3. Characteristics of a nonparaxial PSF

Consider a lens with \( f = a = 75 \lambda \) and \( n = 1 \). The output
wave, plotted in figure 2(a), has been normalized against its
peak amplitude, \( u_{\text{peak}} \), and has been computed on a mesh of
points within the dashed green boundary. Plotting the PSF on
a logarithmic scale in figure 2(b) clearly displays the structure
outside of the geometrical optics cone, and isophotes (con-
tours of constant intensity) are drawn at each decade in order
to reveal the asymmetry in Kirchhoff’s result.

From the nonparaxial PSF, we define the transverse and
the longitudinal resolution criteria by the positions of the
first zeros in the respective directions (\( d_x = \rho_{\text{zero}} \) and
\( d_z = z_{\text{peak}} - z_{\text{zero}} \)). Note that there are many different con-
tentions for how to define the width of the PSF [5]. We

Figure 2. Nonparaxial point spread function. (a) A plot of Re \( |u| \) shows the wavefronts. (b) A plot of \( |u|^2 \) shows the PSF. (c) Longitudinal and
(d) transverse cuts of the PSF are plotted on a 100x scale, and with coordinates normalized using equations (1) or (2). The insets show magnified views of the peak and of the first transverse zero. The numeric result (red circles) are plotted at a reduced density of points for clarity.
follow the choice of Rayleigh (using the position of the first zero), which is addressed further in the discussion section. For the lens shown in figure 2, where \( \alpha = 45^\circ \), we find \( \chi \approx 0.58 \) and \( \zeta \approx 1.65 \) (as opposed to the paraxial values \( \chi \approx 0.61 \) and \( \zeta = 2 \)). We also compute the encircled energy (integrated intensity) in the central spot of the PSF at the focal plane \( (z = z_{\text{peak}}) \), as a fraction of the total encircled energy at the aperture, \( E_{\text{spot}}/E_{\text{tot}} \). For Airy’s distribution, where \( \alpha \to 0 \), the encircled energy is \( E_{\text{spot}} = 83.8\% \), but at \( \alpha = 45^\circ \) the energy is reduced to \( E_{\text{spot}} = 77.8\% \). In the inset of figure 2(d), one sees that the side lobes are proportionately brighter at \( \alpha = 45^\circ \) than in the paraxial case.

Thus we recognize two general characteristics of a nonparaxial PSF: first, the central spot is generally ‘tighter’ than in the paraxial case (smaller \( \chi \) and \( \zeta \)); second, energy is redistributed outward from the central spot and into the side lobes. One observes that these nonparaxial characteristics (as well as focal shift from Kirchhoff’s approximation) are present in both scalar and electromagnetic diffraction, and therefore we consider the two diffraction theories qualitatively similar in this regard.

4. Results

The general characteristics of nonparaxial PSFs mentioned above have been identified previously, for example see figures 12.7 and 12.11 of [17]. The novel aspect of our work is that we identify a trend, tracking these features as we vary the three parameters which uniquely specify a lens \((f, a, \text{and } n)\). The criteria \( d = d(\alpha, n) \) in equations (1) and (2) are valid only in the paraxial limit, and for nonparaxial focusing they are inaccurate even in their functional form (i.e. independent of the choice in \( \chi \) and \( \zeta \)). In our numeric study, we analyze the PSFs for \( \sim 2500 \) sets of unique lens parameters.

The parameters \((f, a, n)\) form a 3D space, and figure 3(a) shows that in the first part of our two-part parametric study, we keep \( n = 1 \) constant and choose points in the 2D \((f, a)\)-plane. In this plane, \( \alpha \) is the clockwise polar angle (just like \( \theta \)) in a Cartesian \((x, y)\)-plane. For each lens, we compute a longitudinal and transverse cut in the PSF, and a relative error is calculated by comparing the longitudinal cut from the numeric result against the exact solution. Among all of the results, the median relative error is on the order of parts per thousand \((10^{-3})\), and we find that this error, which aggregates from the multiple numeric steps of our computational process, generally can be controlled with a tradeoff in computational time. In figure 3(b), one can see that as \( \alpha \to 90^\circ \) and energy gets redistributed outward to the side lobes, the fraction of energy remaining in the central spot falls precipitously to 0.

Two general properties emerge from the results of our parametric study: first, in the transverse cut of the PSF (at the focal plane), we find good agreement between the DNI of Kirchhoff’s and Debye’s theories, as expected [41–44] since we have intentionally restricted our study to lenses with \( N > 50 \); second, if we plot some characteristic of the nonparaxial PSF versus \( \alpha \) and take the limit \( \alpha \to 0 \), then we always recover the result of Airy’s PSF, no matter what the characteristic.

Further results from our parametric study are shown in figure 4, where we plot \( d_1 \) and \( d_2 \) versus \( \alpha \). Although the vertical axis in figure 4(a) has been normalized against \( \lambda/NA \), the data still have a clear, monotonic trend in \( \alpha \). It is therefore evident that equation (1) for \( d_1 \) has an inaccurate functional form, specifically that \( NA^{-1} \) does not give an accurate \( \alpha \)-dependence. We observe a similar trend in \( d_2 \) which is shown in figure 4(b). The apparent spread in the numerical result in figure 4(b) is not due to the error of our computational process, but rather due to the fact that our 2D \((f, a)\)-plane covers a range of Fresnel numbers. Consistent with the findings of [37–40], we observe that Kirchhoff’s theory and Debye’s approximation agree for sufficiently large \( N \). Quite interestingly, the trend in the ‘constant’ \( \chi \), shown in figure 4(a), is from Airy’s limit \((\chi \approx 0.61)\) down to Abbe’s criterion \((\chi = 1/2)\) as \( \alpha \) varies from \( 0^\circ \) to \( 90^\circ \), and our
and 2 of our parametric study indicate unambiguously that the resolution criteria are accurately expressed as:

\[ d_j(\alpha, n) = \frac{\lambda}{n} F_j(\alpha), \] (10)

where \( F_j(\alpha) \) is some function of \( \alpha \), and \( j \) is either \( x \) or \( z \).

5. Discussion

It makes intuitive sense that that the \( d_j \) in equation (10) carry an \( n^{-1} \) dependence; after all, \( \lambda/n \) is the wavelength and therefore the only possible choice of length scale for oscillations in \( u(r) \). To find the correct \( \alpha \)-dependence, we employ Debye’s approximation (see chapter 12 of [17]).

5.1. Transverse resolution

Starting with the transverse resolution, \( d_x \), and following [52, 53], the natural choice for \( F_1(\alpha) \) to take the inverse of the spectral width, \((\Delta k_x)^{-1}\). One can write down an expression for the spectrum at the plane \( z = f \):

\[ U_{\text{Debye}}(k) = \frac{-i \sqrt{\text{peak}}}{1 - \cos \alpha} \frac{\sqrt{k}}{\alpha k_0 \sqrt{(n k_0)^2 - k^2}}. \] (11)

This expression can be contrasted against Airy’s spectrum:

\[ U_{\text{Airy}}(k) = \frac{2 \sqrt{\text{peak}}}{(N k_0)^2} \frac{k}{\sqrt{(N k_0)^2 - k^2}}. \] (12)

To find \( \Delta k_x \), one must normalize the spectrum \( U_{\text{Debye}}(k) \) against the total power. From Plancherel’s theorem, this is given by \( P = \pi a^2 I_{\text{plane}} \), and can be used to relate \( I_{\text{peak}} \) to \( I_{\text{plane}} \) (i.e. the intensity enhancement). The normalized power spectra \( |U|^2 \) are probability distributions, and by computing the moments of those distributions one gets \( \Delta k_x \). For Debye’s spectrum, one must evaluate integrals of the form:

\[ \langle k^m \rangle = (nk_0)^m \frac{1}{\ln(\sec(\omega))} \int_0^{\sin \omega} \frac{\sin \beta^{m+1}}{1 - \beta^2} \, d\beta, \] (13)

where \( m \) is an integer and \( \beta = k/nk_0 \). Noting that \( k \) is the radial component of the momentum and using \( k = \sqrt{k_x^2 + k_y^2} \), where \( k_x \) and \( k_y \) are the Cartesian components, one can finally obtain the width of the spectrum \( \Delta k_x \) using \( \Delta k_x = k_{\text{rms}} / \sqrt{2} \):

\[ \Delta k_x = \frac{nk_0}{2} \sqrt{2 + \frac{\sin^2 \alpha}{\ln(\cos \alpha)}}. \] (14)

To the best of our knowledge, equation (14) has not been previously published. A qualitatively similar expression for \( \Delta k_x \) was obtained using a ray optics approach [52, 53], and both expressions are compared against our numerical result in figure 5. It is interesting to note the limits of the radical on the right-hand side of equation (14): at \( \alpha \to 90^\circ \) one gets \( \sqrt{2} \); and at \( \alpha \to 0 \) one recovers the result of Airy, \( N k_0/2 \). The latter can be seen by expanding in powers of \( \sin \alpha \) about

![Figure 4. Nonparaxial resolution limits. (a) Plotting the normalized transverse resolution \( d_x \) versus \( \alpha \) reveals a monotonic trend. The numeric data (red dots, at reduced density for clarity) are compared against Debye’s approximation (blue solid line). (b) There is a similar trend in the longitudinal resolution \( d_z \). We show our analytic results (yellow, dashed lines) equations (16) and (17), and the black curves (circle, square, and diamond) are from numerical analysis of the exact solution at different Fresnel numbers.](image-url)
\( \alpha = 0 \), and keeping only the leading order term:

\[
\Delta k_x = \frac{n k_0}{2} \left( \sin \alpha + \frac{1}{12} \sin^3 \alpha + \ldots \right).
\]  

(15)

The radical on the right-hand side of equation (14) can be used to get a phenomenological \( \alpha \)-dependence. However, in order to maintain accuracy when comparing to our numerical result, we include two empirical parameters \( F_1(\alpha) \), namely \( \chi_0 \) and \( b \):

\[
F_1(\alpha) = \chi_0 b \left( 2 + \frac{\sin^2 (b \alpha)}{\ln (\cos (b \alpha))} \right)^{-1/2}.
\]  

(16)

Fixing \( \chi_0 = 0.61 \), we find a good fit to our numerical results for \( \alpha < 60^\circ \) with \( b = 1.11 \), which is shown as a dark yellow, dashed curve in figure 4(a).

With regard to the parameters \( \chi_0 \) and \( b \), we make two comments. The first comment, as pointed out in the methods section, is there are several alternative definitions [5] for the transverse width \( \Delta x \) of the PSF. For example, we have used the rms \( k \)-value to compute \( \Delta k_x \), but in general the rms \( \rho \)-value of the PSF cannot be computed; even for Airy’s \( \text{jinc}^2 \) the variance \( \langle x^2 \rangle \) diverges. For nonparaxial focusing, this divergence only increases with increasing \( \alpha \). So we have followed the convention of Rayleigh and chosen the positions of the first zeros: \( \rho_{\text{zero}} \) and \( z_{\text{zero}} \). But we could have chosen the half-widths at half maximum, followed Sparrow’s criterion, or a number of different positions instead; the different choices are just different conventions, and we find that the general trends of nonparaxial focusing are insensitive to the convention choice. With a different choice, the trends could equally well be described by a different set of values for \( \chi_0 \) and \( b \). In all cases, equation (14) can be used to obtain a phenomenological \( \alpha \)-dependence.

The second comment: independent of which convention is used to define \( \Delta x \), the space-bandwidth (SW) product \( \Delta x \Delta k_x \) is generally not independent of \( \alpha \). This peculiar feature of \( u_{\text{sphere}} \) is not present in all types of focused waves. For example, in the focusing of Gaussian beams, as long as the beam waist \( w \) at the input plane is much smaller than the lens radius \( a \) (i.e. an underfilled back aperture), the SW-product is always equal to 1/2, as shown in [54]. In that work, the authors’ parametric study showed that the focusing of \( u_{\text{sphere}} \) and of a Gaussian beam can be thought of as two limiting cases of the ratio \( w/a \) (the degree of filling). The main point here is that an \( \alpha \)-dependent SW-product means that taking \( (\Delta k_x)^{-1} \) does not account for the full \( \alpha \)-dependence in \( \Delta x \), and thus we introduce the parameter \( b \) to compensate for this.

5.2. Longitudinal resolution

Next, for the longitudinal resolution \( d_z \), we examine equations (12.21a)–(12.21d) of [17] and note that within Debye’s approximation, for points along the optical axis, the first zero in the PSF always satisfies:

\[
d_z = \frac{\lambda}{n} \frac{1}{1 - \cos \alpha}.
\]  

(17)

This can again be expanded in powers of \( \alpha \) to recover the paraxial limit:

\[
d_z = \frac{\lambda}{n} \left( \frac{1}{2} \sin^2 \alpha + \frac{1}{8} \sin^4 \alpha + \ldots \right).
\]  

(18)

Equation (17) is plotted alongside the numerical results in figure 4(b), and a nearly identical formula has been reported [52, 53] using a ray-optics approach. In figure 4(b), we also plot a family of curves which result from numerical analysis of the exact solution at different Fresnel numbers \((N = 50, 100, 200)\). In practice, the resolution limits predicted by our nonparaxial formulae, equations (16) and (17) differ by nominally 10%–30% from the limits predicted by the paraxial formulae, equations (1) and (2). Consider an \( NA = 1.4 \) oil-immersion objective with \( n = 1.52 \) at \( \lambda = 532 \text{ nm} \): equations (1) and (2) give \( d_z = 232 \text{ nm} \) and \( d_z = 825 \text{ nm} \), respectively; meanwhile, equations (16) and (17) give \( d_z = 208 \text{ nm} \) and \( d_z = 573 \text{ nm} \).

The wide-angle features of a PSF have also been explained in the context of scaling the coefficients \( \chi \) and \( \zeta \) in equations (1) and (2) [55], or by introducing appropriate optical coordinates [56]. Alternatively, one can compose a nonparaxial PSF from a superposition of prolate spheroidal wave functions [57]; these are eigenfunctions of a finite-bandwidth Fourier transform, and therefore the natural choice for the focusing problem [58]. These explanations are rich with mathematical rigor and quite accurate, however our novel, closed-form expressions offer accessible alternative means of understanding a nonparaxial PSF.

6. Conclusions

In conclusion, from our numeric and analytic calculations, we propose a new set of resolution formulae, equations (16) and (17), which remain accurate under nonparaxial focusing, and note that equations (1) and (2) are their paraxial-limiting forms. Although our scalar results cannot be applied quantitatively to...
wide-angle ($\alpha > 50^\circ$) electromagnetic focusing, we conjecture that the nonparaxial trends we have observed illuminate some qualitative features, which may also be present even when polarization effects are considered. It would be interesting to repeat this parametric study using a flat lens which obeys a different transformation rule, such as:

$$u_{\text{perfect}} = u_{\text{sphere}} \left(1 + i k_0 R \right)$$

(19)

where $u_{\text{perfect}}$ is Stamnes’ so-called ‘perfect’ wave [17], and to apply the focusing of perfect waves to high-intensity focused ultrasound [24]. Such an application could materially improve the imaging contrast (specifically, by preserving Airy’s 83.8% energy concentration). For electromagnetic focusing, Stamnes’ perfect wave could be used as a design goal for a new class of metasurface, the perfect lens [59, 60].

Acknowledgments

This work was supported in part by the W M Keck Foundation (MJN).

ORCID iDs

Yitzi M Calm https://orcid.org/0000-0002-6216-396X
Michael J Burns https://orcid.org/0000-0001-9804-405X
Michael J Naughton https://orcid.org/0000-0002-6733-2398

References

[17] Stamnes J J 1986 Waves in Focal Regions (Bristol: Adam-Hilger)
[40] Li Y and Wolf E 1984 Three-dimensional intensity distribution near the focus in systems of different Fresnel numbers J. Opt. Soc. Am. A 1 801–8
[55] Lin B J 2002 The k3 coefficient in nonparaxial λ/NA scaling equations for resolution, depth of focus, and immersion lithography J. Micro/Nanolithography, MEMS, MOEMS 1 7–12
[56] Sheppard C J R and Török P 1999 Effects of Fresnel number in focusing and imaging Proc. SPIE 3729 458–72