

Higher-order boundary-diffraction-wave formulation

ERIC L. SHIRLEY*

Optical Technology Division, NIST 100 Bureau Dr., MS 8441 Gaithersburg,
MD 20899-8441, USA

(Received 18 January 2006; in final form 28 June 2006)

We consider propagation of radiation through N apertures in series modelled using the Huygens–Kirchhoff Green’s function method. We present a method to evaluate the final radiation field using at most N -fold integrals over aperture perimeters instead of the $2N$ -fold integration over the aperture areas that one might anticipate. This generalizes the boundary-diffraction-wave formulation to higher orders. After deriving a formula for the boundary-diffraction wave at all orders of diffraction, we provide sufficient detail to realize its implementation at first, second and third order. We provide a sample calculation of first-, second- and third-order boundary diffraction waves, discuss ramifications of geometrical blocking effects that complicate boundary-diffraction-wave formulations in general, and indicate three key potential applications of the present work in radiometry.

1. Introduction

The Huygens–Kirchhoff Green’s function method [1] to describe Fraunhofer and Fresnel diffraction of light by lenses and apertures continues to foster insight into this important and interesting phenomenon. Despite its formal inconsistencies, the method can predict irradiance in a detector plane quantitatively. Most applications of the method consider diffraction by one optical element (first-order diffraction), but diffraction by multiple elements in series (higher-order diffraction) is also of interest. This work presents an extension of the boundary-diffraction-wave (BDW) formulation [2] to higher-order diffraction effects for a *series* of apertures, so that one can determine these effects more easily because of simplified integration. This work also furnishes insight into the asymptotic properties of higher-order diffraction effects at small λ . Arrangements of apertures in series arise, for instance, in radiometric calibrations of reference blackbodies [3] and in prototypical set-ups designed to analyse diffraction effects [4]. We do not consider focusing optics in this work, and generalization of this work to set-ups containing such optics is a topic of interest.

The philosophy motivating this work can be stated as follows. If optical radiation passes through N apertures in series, naïve iterative application of the

*Email: eric.shirley@nist.gov

Huygens–Kirchhoff method might suggest that $2N$ -fold integration is required to calculate the end-to-end propagation of radiation. Of course, the propagation can be done stage-wise for a given set of values of the radiation field and its derivative at one end of the optics. Furthermore, such propagation can often be accelerated by exploiting separation of variables in the case of cylindrical symmetry or use of fast-Fourier transform techniques. However, this must be done repeatedly for each set of values. In many instances, though, the analytical properties of end-to-end propagation are of interest, such as in the limit of small wavelengths, where numerical calculations can become prohibitive, but asymptotic behaviour of end-to-end propagation of radiation becomes simpler to describe. Then the ability to express end-to-end propagation using a lower-dimensional integral is obviously advantageous. A reduction of the above $2N$ -fold integration to an N -fold integration is derived in what follows, in what can be considered the generalization of the boundary-diffraction-wave formulation to higher-order diffraction effects.

2. Background

Consider a scalar wave, with wavelength λ and angular wave number $q = 2\pi/\lambda$, originating at $\mathbf{P} = (x_0, y_0, z_0)$ as $u(\mathbf{r}) = U_0 \exp(iq|\mathbf{r} - \mathbf{P}|)/|\mathbf{r} - \mathbf{P}|$. It passes through N apertures to reach $\mathbf{Q} = (x_{N+1}, y_{N+1}, z_{N+1})$ (see figure 1). We adopt the Fresnel, paraxial (or gaussian optics) approximation [5] for distances in exponents and denominators. We assume that apertures can be treated as flat and normal to z . The distance between elements μ and $\mu + 1$ is $d_\mu = z_{\mu+1} - z_\mu$, with \mathbf{P} corresponding to element $\mu = 0$, and \mathbf{Q} to element $\mu = N + 1$. With $\alpha = [(x_0 - x_{N+1})^2 + (y_0 - y_{N+1})^2]/[2(d_0 + \dots + d_N)]$ and $M = [(i\lambda)^N d_0, \dots, d_N]^{-1} \exp[iq(d_0 + \dots + d_N)]$, this gives

$$u(\mathbf{Q}) = U_0 M \int_{A_1} dx_1 dy_1 \dots \int_{A_N} dx_N dy_N \exp \left\{ iq \left[\alpha + \sum_{\mu, v=1}^N B_{\mu v} (x'_\mu x'_v + y'_\mu y'_v) \right] \right\}. \quad (1)$$

Here A_μ is the area of aperture μ , which the line segment connecting \mathbf{P} and \mathbf{Q} crosses at $(x_{i\mu}, y_{i\mu}, z_\mu)$, and one has $x'_\mu = x_\mu - x_{i\mu}$, $y'_\mu = y_\mu - y_{i\mu}$, and $B_{\mu v} = [\delta_{\mu v}(1/d_{\mu-1} + 1/d_\mu) - \delta_{\mu+1, v}/d_\mu - \delta_{\mu, v+1}/d_v]/2$. Equation (1) maintains the self-consistency of

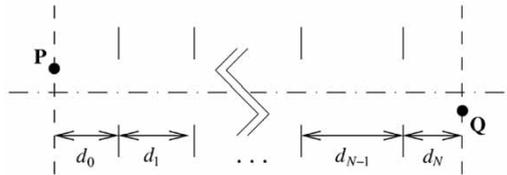


Figure 1. Optical set-up with N apertures between points \mathbf{P} and \mathbf{Q} at indicated positions along the z axis. Actual distances and aperture sizes are not indicated.

gaussian optics [6]. For $N=0$, one has $u(\mathbf{Q}) = \exp[iq(d_0 + \alpha)]/d_0$. Having $N=1$ gives Kirchhoff's result, and equation (1) generalizes that result to $N > 1$.

The BDW formulation for $N=1$ obviates double integration over x_1 and y_1 , which can streamline determination of $u(\mathbf{Q})$. One instead has $u(\mathbf{Q}) = u_G(\mathbf{Q}) + u_B(\mathbf{Q})$: a 'geometrical' (spherical) wave $u_G(\mathbf{Q})$ plus a boundary wave $u_B(\mathbf{Q})$ given by a single line integral around S_1 , the perimeter of A_1 . This is related to the geometrical theory of diffraction (GTD) that is helpful at small λ [7]. Here, extending the BDW formulation to $N > 1$ correspondingly simplifies calculation of higher-order diffraction effects.

3. Derivation of formula for higher-order BDW

Let $L(\{x_\mu\}, \{y_\mu\})$ denote the 'reduced' path length from \mathbf{P} to \mathbf{Q} , which is the length of the $(N+1)$ -segment path from \mathbf{P} to \mathbf{Q} through points $\{(x_\mu, y_\mu, z_\mu)\}$, minus $d_0 + \dots + d_N$. A given value of $L(\{x_\mu\}, \{y_\mu\})$, l , occurs with a frequency described by the distribution function,

$$\begin{aligned} f(l) &= \int_{A_1} dx_1 dy_1 \dots \int_{A_N} dx_N dy_N \delta(l - L(\{x_\mu\}, \{y_\mu\})) \\ &= \int d\Theta_1 R_1^2(\Theta_1) \dots \int d\Theta_N R_N^2(\Theta_N) \int_0^1 ds_1 s_1 \dots \int_0^1 ds_N s_N \delta(l - \alpha - \mathbf{s} \cdot \mathbf{M} \cdot \mathbf{s}). \end{aligned} \quad (2)$$

This expresses $f(l)$ using Cartesian coordinates and polar transverse coordinates. The latter are angles $\{\Theta_\mu\}$ and radii $\{s_\mu R_\mu(\Theta_\mu)\}$, with $s_\mu = 0$ at $(x_{i\mu}, y_{i\mu}, z_\mu)$ and $s_\mu = 1$ on S_μ . A function $R_\mu(\Theta_\mu)$ may or may not be single-valued, and range(s) of integration for Θ_μ can vary and be sampled forwards and/or backwards, as illustrated in figure 2. The metric \mathbf{M} depends on $\{\Theta_\mu\}$, with $\mathbf{s} = (s_1, \dots, s_N)$. One can obtain $u(\mathbf{Q})$ from $f(l)$:

$$u(\mathbf{Q}) = U_0 M \lim_{\eta \rightarrow 0^+} \int_{-\infty}^{\infty} dl \exp(iql - \eta|l|) f(l). \quad (3)$$

Damping ensures sensible results when partitioning $f(l)$ and $u(\mathbf{Q})$ into a geometrical part and a part resulting from diffraction.

Like the traditional BDW formulation, the present BDW formulation must be reconsidered if the distance travelled by diffracted light can equal the distance travelled by undiffracted light. More generally, we assume that path lengths represented in contributions to $u(\mathbf{Q})$ from each order of diffraction cannot equal path lengths represented in contributions from adjacent orders. This means that the angles of bends introduced into paths by diffraction at edges never approach zero. This still permits treatment of many optical systems, and we qualitatively discuss generalization to other systems later. The above condition ensures that $\mathbf{s} \cdot \mathbf{M} \cdot \mathbf{s}$ is of one sign. This sign is positive here. If it were negative, one could adapt the present analysis accordingly.

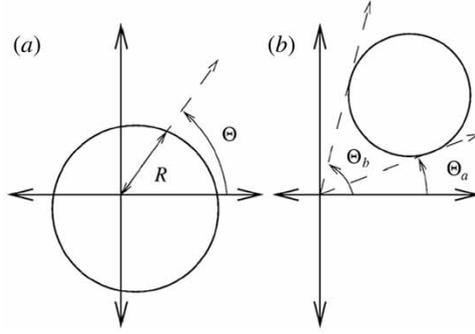


Figure 2. Two scenarios for $R(\Theta)$ discussed in the text: (a) Θ should run from 0 to 2π , and $R(\Theta)$ is a single-valued function; (b) Θ should run from Θ_a to Θ_b and from Θ_b to Θ_a , and $R(\Theta)$ has different values over these two intervals.

For any function $F(\mathbf{s})$, one can show

$$\int_0^1 ds_1 \dots \int_0^1 ds_N F(\mathbf{s}) = \sum_{k=1}^N \int_0^1 d\lambda \lambda^{N-1} \int d\{s_{\mu \neq k}\} F(\lambda \mathbf{s}), \quad (4)$$

with $s_k = 1$ and integration from 0 to 1 for all $\{s_{\mu \neq k}\}$ in each term. With $F(\mathbf{s}) = s_1, \dots, s_N \delta(l - \alpha - \mathbf{s} \cdot \mathbf{M} \cdot \mathbf{s})$, we have $F(\lambda \mathbf{s}) = \lambda^N s_1, \dots, s_N \delta(l - \alpha - \lambda^2 \mathbf{s} \cdot \mathbf{M} \cdot \mathbf{s})$ and

$$\begin{aligned} f(l) &= \int d\{\Theta_\mu\} \sum_{k=1}^N \int d\{s_{\mu \neq k}\} \left(\prod_{v=1}^N s_v R_v^2(\Theta_v) \right) \int_0^1 d\lambda \lambda^{2N-1} \delta(l - \alpha - \lambda^2 \mathbf{s} \cdot \mathbf{M} \cdot \mathbf{s}) \\ &= \frac{(l - \alpha)^{N-1} \theta(l - \alpha)}{2} \int d\{\Theta_\mu\} \sum_{k=1}^N \int d\{s_{\mu \neq k}\} \left(\prod_{v=1}^N s_v R_v^2(\Theta_v) \right) \frac{\theta(\mathbf{s} \cdot \mathbf{M} \cdot \mathbf{s} + \alpha - l)}{(\mathbf{s} \cdot \mathbf{M} \cdot \mathbf{s})^N}. \end{aligned} \quad (5)$$

Rearrangement gives

$$\begin{aligned} f(l) &= \frac{(l - \alpha)^{N-1} \theta(l - \alpha)}{2} \sum_{k=1}^N \int d\Theta_k \\ &\quad \times \left[\int d\{\Theta_{\mu \neq k}\} \int d\{s_{\mu \neq k}\} \left(\prod_{v=1}^N s_v R_v^2(\Theta_v) \right) \frac{\theta(\mathbf{s} \cdot \mathbf{M} \cdot \mathbf{s} + \alpha - l)}{(\mathbf{s} \cdot \mathbf{M} \cdot \mathbf{s})^N} \right]. \end{aligned} \quad (6)$$

One may write this differently as

$$f(l) = \frac{(l - \alpha)^{N-1} \theta(l - \alpha)}{2} \sum_{k=1}^N \int d\Theta_k R_k^2(\Theta_k) \left[\int d\{x_{\mu \neq k}, y_{\mu \neq k}\} \frac{\theta(l_u + l_d - l)}{(l_u + l_d - \alpha)^N} \right]. \quad (7)$$

Here l_u or l_d respectively denotes the reduced path length for the ‘subpath’ from \mathbf{P} to \mathbf{S}_k (a point on S_k) or from \mathbf{S}_k to \mathbf{Q} . In terms of distribution functions $f_{uk}(l_u)$ and $f_{dk}(l_d)$, which are counterparts of $f(l)$, the above result can be written

$$f(l) = \frac{(l-\alpha)^{N-1}\theta(l-\alpha)}{2} \sum_{k=1}^N \int d\Theta_k R_k^2(\Theta_k) \int_{-\infty}^{\infty} dl_u \int_{-\infty}^{\infty} dl_d f_{uk}(l_u) f_{dk}(l_d) \frac{\theta(l_u + l_d - l)}{(l_u + l_d - \alpha)^N}. \quad (8)$$

Noting

$$\theta(l_u + l_d - l) = 1 - \theta(l - l_u - l_d), \quad (9)$$

we arrive at

$$f(l) = f_G(l) - \frac{(l-\alpha)^{N-1}\theta(l-\alpha)}{2} \times \sum_{k=1}^N \int d\Theta_k R_k^2(\Theta_k) \int_{-\infty}^{\infty} dl_u \int_{-\infty}^{\infty} dl_d f_{uk}(l_u) f_{dk}(l_d) \frac{\theta(l - l_u - l_d)}{(l_u + l_d - \alpha)^N}, \quad (10)$$

with (assuming $N > 0$)

$$f_G(l) = D(l-\alpha)^{N-1}\theta(l-\alpha). \quad (11)$$

From phase-space arguments, use of a Feynman integral [8], and a general formula for $\det B$ obtained using mathematical induction, we have

$$D = \frac{\pi^N}{\Gamma(N) \det B} = \frac{(2\pi)^N d_0 d_1, \dots, d_N}{\Gamma(N)(d_0 + d_1 + \dots + d_N)} \quad (12)$$

if the line segment connecting \mathbf{P} and \mathbf{Q} is unobstructed and $D=0$ otherwise. Introducing $f_B(l)$, defined implicitly by equation (10) making the partition, $f(l) = f_G(l) + f_B(l)$, the geometrical part $f_G(l)$ determines $u_G(\mathbf{Q})$, whereas $f_B(l)$ determines $u_B(\mathbf{Q})$, the diffraction effects on $u(\mathbf{Q})$. As a check, integrating equation (10) with respect to l yields the expected gaussian-optics BDW result for $N=1$.

Maintaining the correspondence between $f(l)$, $f_{uk}(l_u)$ and $f_{dk}(l_d)$, one can iterate the analysis described in equations (2)–(12). This leads to the counterpart to equation (10) for $f_{uk}(l_u)$, $f_{dk}(l_d)$, and all subsequent generations of their subpaths. For a given Θ_k , just as $f_{uk}(l_u)$ (or $f_{dk}(l_d)$) is a reduced path-length distribution for subpaths from \mathbf{P} to \mathbf{S}_k (or from \mathbf{S}_k to \mathbf{Q}); α_{uk} (or α_{dk}) denotes the shortest reduced subpath length; giving $f_{uk}(l_u) = \delta(l_u - \alpha_{uk})$ for $k=1$, $f_{uk}(l_u) = D_{uk}(l_u - \alpha_{uk})^{k-2}\theta(l_u - \alpha_{uk}) + \dots$ for $k > 1$, and corresponding formulas for f_{dk} . The pattern established in equation (12) sets the values of D_{uk} and D_{dk} . The iterative process must terminate when every remaining subpath has its $N=0$.

Each iteration of equation (10) partitions the geometrical and diffraction parts of a given $f(l)$. The overall $f(l)$ and $u(\mathbf{Q})$ are sums of terms at each order m of diffraction from $m=0$ to $m=N$, with $f(l) = \sum_{m=0}^N f_m(l)$, where $f_m(l)$ truncates the iterative process by including only the first term on the right-hand side of equation (10) on

iteration $m + 1$. For instance, such a truncation on the first ($m = 0$) iteration gives $f_0(l) = f_G(l)$, which includes no diffraction effects. Finally, as a cautionary remark, we note that the radius function used on each iteration depends on all preceding members of $\{\Theta_\mu\}$ in the iterative process. This is because the counterparts of $\{(x_{i\mu}, y_{i\mu}, z_\mu)\}$ are on the line segment connecting the counterparts of \mathbf{P} and \mathbf{Q} on each iteration.

Suppose that one truncates the calculation by including only the geometrical part on the second iteration of equation (10), and that the system has no geometrical blocking effects. That is, line segments between any two members of the set of points, \mathbf{P} , \mathbf{Q} and those on $\{S_\mu\}$, pass through the interiors of all intervening apertures. In this case, the approximate result obtained for $u_B(\mathbf{Q})$ is the sum of the BDW results that would be obtained for each aperture if it were the only one. This is often a useful approximation for treating first-order diffraction effects in multi-staged optical systems, and this analysis motivates it explicitly [3, 4, 9].

A prescription for determining diffraction effects at all orders follows from equation (10). (We still assume that there are no geometrical blocking effects.) Consider the figurative paths depicted in figure 3, chosen at random for the sake of illustration. Three iterations of the type described above are indicated diagrammatically by three schematic bends introduced in turn at the perimeters of various optical elements. The first bend introduced at element $k(1)$ replaces a solid segment with two dashed segments k_1 . Successive bends are introduced at elements $k(2)$ and $k(3)$. Excluding the geometrical part on each iteration yields one of five contributions by these three elements to $f_3(l)$ and related quantities. Avoiding erroneous multiple counting, there can be up to $N(N - 1) \cdots (N - m + 1)$ contributions to $f_m(l) = \sum_p f_{m,p}(l)$, where p is an index that specifies a contribution, for which bends are introduced in sequence at element $k(1)$, element $k(2)$, and so forth up to element $k(m)$. (The dependence of $\{k(n)\}$ on p is suppressed.) A subset of m apertures can lead to up to $m!$ contributions. (Fewer contributions result in cases where one has $k(b) < k(a)$ and $k(c) > k(a)$ for any a , b and c with $b > a$ and $c > a$.)

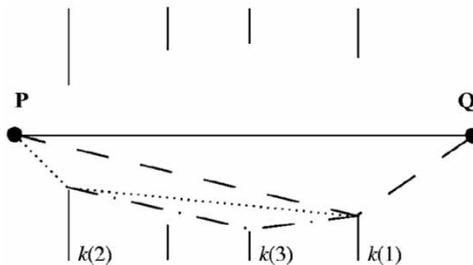


Figure 3. Schematic drawing of creation of a four-segment path by introduction of three bends in three steps: (1) bend at element $k(1)$ converts a solid segment into two dashed segments, (2) bend at element $k(2)$ converts a dashed segment into two dotted segments, and (3) bend at element $k(3)$ converts a dotted segment into two dot-dashed segments.

A contribution $f_{m,p}(l)$ involves integration over $\{\Theta_{k(n)}\}$ for $n=1$ through $n=m$ and accompanying radius functions $\{R_n^{(m)}(\mathbf{P}, \mathbf{Q}; \{k(t)\}, \{\Theta_{k(t)}\})\}$. For a given point on $S_{k(n)}$, a function $R_n^{(m)}$, whose dependences as well as the dependences of many other quantities are henceforth suppressed, depends on the optical set-up, $\mathbf{P}, \mathbf{Q}, p$ through $\{k(t)\}$, and $\{\Theta_{k(t)}\}$ for $t=1$ through $t=n$.

After iterating equation (10), evaluating a given $f_{m,p}(l)$ requires integration with respect to m pairs of length variables as equation (10) dictates. One way to do this exploits properties of the family of functions $\{g_s(x)\}$, given by $g_0(x) = \delta(x)$ and $g_s(x) = x^{s-1}\theta(x)/\Gamma(s)$ for integer $s > 0$, which are scaled versions of shifted geometrical path-length distributions in the case of s intervening apertures. If one has $\beta + \gamma > \alpha$, one can use

$$(L - \alpha)^{\beta+c}\theta(L - \alpha) \int_{-\infty}^{+\infty} dl \int_{-\infty}^{+\infty} dl' \frac{g_b(l - \beta)g_c(l' - \gamma)\theta(L - l - l')}{(l + l' - \alpha)^{\beta+c+1}} = \frac{g_{b+c+1}(L - \beta - \gamma)}{\beta + \gamma - \alpha} \quad (13)$$

to perform integrations over length variables analytically beginning with the innermost integrals and proceeding outward. With all prefactors in path-length distributions included, one obtains

$$f_{m,p}(l) = \frac{(2\pi)^N (-1)^m d_0, \dots, d_N}{(4\pi)^m Z_0, \dots, Z_m} \int d\{\Theta_{k(n)}\} \left[\prod_{n=1}^m \frac{(R_n^{(m)})^2}{l_n^{(m)} - s_n^{(m)}} \right] g_N(l - l_1^{(m)}). \quad (14)$$

By defining $P_m = U_0(-1)^m \exp[iq(d_0 + \dots + d_N)]/(4\pi)^m$ and integrating with respect to l , we have this result for a contribution to a general-order BDW:

$$u_{m,p}(\mathbf{Q}) = \frac{P_m}{Z_0, \dots, Z_m} \int d\{\Theta_{k(n)}\} \left[\prod_{n=1}^m \frac{(R_n^{(m)})^2}{l_n^{(m)} - s_n^{(m)}} \right] \exp(iql_1^{(m)}). \quad (15)$$

The factor $(-1)^m$ results from the θ -function on the right-hand side of equation (9) being subtracted on each iteration. Each denominator $l_n^{(m)} - s_n^{(m)}$ arises in turn from the denominator on the right-hand side of equation (13), so that $l_n^{(m)}$ is the reduced length of a subpath between two elements in the optical systems (where the list of elements includes \mathbf{P}, \mathbf{Q} and the N apertures), and $s_n^{(m)}$ is the reduced length of the line segment connecting the endpoints of that subpath. The difference $l - l_n^{(m)}$ appears in order of decreasing n in the argument in the numerator on the right-hand side of equation (13) when one iterates that equation. The $\{Z_i\}$, which depend on m and p , are intervening distances along the z axis between \mathbf{P} , the m elements and \mathbf{Q} . As a consequence, $u_{m,p}(\mathbf{Q})$ can be evaluated with at most an N -fold integral in equation (15) instead of the $2N$ -fold integral in equation (1).

3.1 Implementation of formula

To enhance clarity, the following illustrates how the geometrical-optics and first- and higher-order BDW results follow from equation (15). For $m=0$, evaluating

equation (15) involves $Z_0 = d_0 + \dots + d_N$ and $l_1^{(0)} = \alpha$, to yield the familiar geometric optics result,

$$u_0(\mathbf{Q}) = \frac{U_0 \exp [iq(d_0 + \dots + d_N + \alpha)]}{d_0 + \dots + d_N}. \tag{16}$$

For $m = 1$, and for each optical element, a contribution which reflects the familiar first-order BDW result is obtained, with $Z_0 Z_1 = (d_0 + \dots + d_{k(1)-1})(d_{k(1)} + \dots + d_N)$ and

$$\left. \begin{aligned} l_1^{(1)} &= d_{0,k(1)} + d_{k(1),N+1}, \\ s_1^{(1)} &= d_{0,N+1}. \end{aligned} \right\} \tag{17}$$

Here $d_{k,k'}$ denotes the reduced distance between points \mathbf{S}_k and $\mathbf{S}_{k'}$, with $\mathbf{S}_0 = \mathbf{P}$ and $\mathbf{S}_{N+1} = \mathbf{Q}$. In particular, we have $d_{0,N+1} = \alpha$. This gives

$$u_{1,p}(\mathbf{Q}) = -\frac{U_0 \exp [iq(d_0 + \dots + d_N)]}{4\pi Z_0 Z_1} \int d\Theta_{k(1)} \left(\frac{(R_1^{(1)})^2 \exp (iq l_1^{(1)})}{l_1^{(1)} - s_1^{(1)}} \right), \tag{18}$$

which can be related to the traditional BDW result for $m = 1$.

For $m = 2$, and for each pair of optical elements, there are two different contributions $\{u_{2,p}(\mathbf{Q})\}$, according to the ordering of $k(1)$ and $k(2)$. For $m > 1$, we do not restate the entire result that follows from equation (15). Instead, we note that, for $k(1) < k(2)$, one has $Z_0 Z_1 Z_2 = (d_0 + \dots + d_{k(1)-1})(d_{k(1)} + \dots + d_{k(2)-1})(d_{k(2)} + \dots + d_N)$ and

$$\left. \begin{aligned} l_1^{(2)} &= d_{0,k(1)} + d_{k(1),k(2)} + d_{k(2),N+1}, \\ s_1^{(2)} &= d_{0,N+1}, \\ l_2^{(2)} &= d_{k(1),k(2)} + d_{k(2),N+1}, \\ s_2^{(2)} &= d_{k(1),N+1}, \end{aligned} \right\} \tag{19}$$

whereas, for $k(1) > k(2)$, one has an analogous result.

For $m = 3$, and for each set of three optical elements, there are five different contributions $\{u_{3,p}(\mathbf{Q})\}$, even though $k(1)$, $k(2)$ and $k(3)$ can have six orderings. (In all cases, the values of $\{Z_i\}$ should now be clear and are not discussed further.) For $k(1) < k(2) < k(3)$, $\{l_n^{(m)}\}$ and $\{s_n^{(m)}\}$ are given by

$$\left. \begin{aligned} l_1^{(3)} &= d_{0,k(1)} + d_{k(1),k(2)} + d_{k(2),k(3)} + d_{k(3),N+1}, \\ s_1^{(3)} &= d_{0,N+1}, \\ l_2^{(3)} &= d_{k(1),k(2)} + d_{k(2),k(3)} + d_{k(3),N+1}, \\ s_2^{(3)} &= d_{k(1),N+1}, \\ l_3^{(3)} &= d_{k(2),k(3)} + d_{k(3),N+1}, \\ s_3^{(3)} &= d_{k(2),N+1}. \end{aligned} \right\} \tag{20}$$

An analogous result applies for $k(1) > k(2) > k(3)$. For $k(1) < k(3) < k(2)$, we have

$$\left. \begin{aligned} l_1^{(3)} &= d_{0,k(1)} + d_{k(1),k(3)} + d_{k(3),k(2)} + d_{k(2),N+1}, \\ s_1^{(3)} &= d_{0,N+1}, \\ l_2^{(3)} &= d_{k(1),k(3)} + d_{k(3),k(2)} + d_{k(2),N+1}, \\ s_2^{(3)} &= d_{k(1),N+1}, \\ l_3^{(3)} &= d_{k(1),k(3)} + d_{k(3),k(2)}, \\ s_3^{(3)} &= d_{k(1),k(2)} \end{aligned} \right\} \quad (21)$$

and the analogous result applies for $k(1) > k(3) > k(2)$. Of the orderings, $k(2) < k(1) < k(3)$ and $k(3) < k(1) < k(2)$, only one of the two should be considered, because they correspond to the same $u_{3,p}(\mathbf{Q})$. Choosing $k(2) < k(1) < k(3)$, one has

$$\left. \begin{aligned} l_1^{(3)} &= d_{0,k(2)} + d_{k(2),k(1)} + d_{k(1),k(3)} + d_{k(3),N+1}, \\ s_1^{(3)} &= d_{0,N+1}, \\ l_2^{(3)} &= d_{0,k(2)} + d_{k(2),k(1)}, \\ s_2^{(3)} &= d_{0,k(1)}, \\ l_3^{(3)} &= d_{k(1),k(3)} + d_{k(3),N+1}, \\ s_3^{(3)} &= d_{k(1),N+1}. \end{aligned} \right\} \quad (22)$$

Similar reasoning can be used to deduce $u_m(\mathbf{Q})$ for all higher values of m .

4. Numerical demonstration

As a concrete example, we consider the set-up specified by $N=3$, $d_0 = d_1 = d_2 = d_3 = 500$ mm, and three circular apertures centred and normal to the z axis with radii $r_1 = 3$ mm, $r_2 = 4$ mm and $r_3 = 3$ mm. For visible radiation with $\lambda = 580$ nm and $x_0 = -(1/2)$ mm, $y_0 = 0$ mm, the ratios $|u_m(\mathbf{Q})/u_0(\mathbf{Q})|^2$ for $m=1$ to $m=3$ are indicated in figure 4 for $y_{N+1} = 0$ mm but a range of x_{N+1} . At such a small wavelength, the typical ratios fall off rapidly with increasing m , so the ratios are shown using different scales. (In practice, the sum $|\sum_{m=0}^N u_m(\mathbf{Q})|^2$ is also of interest.)

The local maxima in $|u_1(\mathbf{Q})/u_0(\mathbf{Q})|^2$, near $x_{N+1} = +(3/2)$ mm, $x_{N+1} = +(1/2)$ mm and $x_{N+1} = +(1/6)$ mm, result from the Poisson bright spot associated with diffraction on the respective edges, S_1 , S_2 and S_3 . The features in $|u_2(\mathbf{Q})/u_0(\mathbf{Q})|^2$ can be understood qualitatively by considering the points in $(\Theta_{k(1)}, \Theta_{k(2)})$ -space where at least one $l_1^{(2)}$ is stationary with respect to both angles, and evaluation of equation (15) using the method of stationary phase. It appears that higher-order counterparts of Poisson bright spots can arise, some of which are conjectured to appear on the line sampled in figure 4. Germane stationary-phase analysis shall be considered in

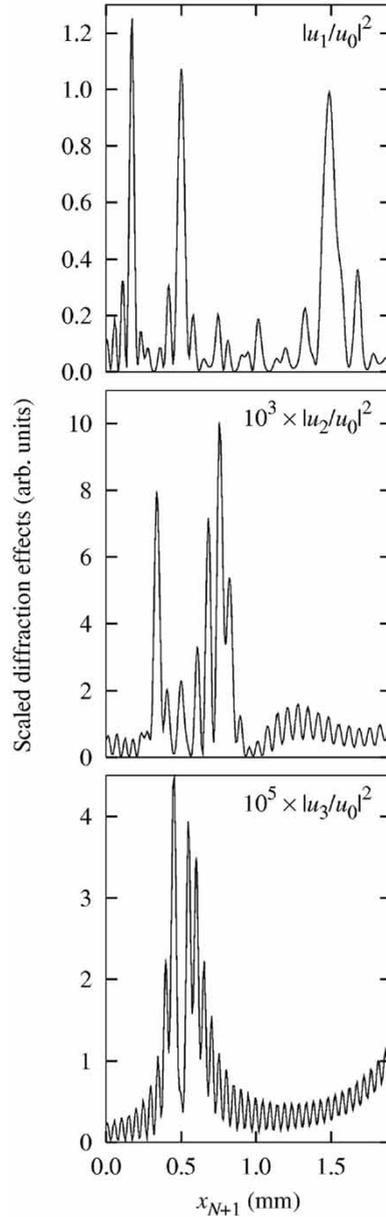


Figure 4. Relative magnitudes of boundary diffraction waves at first, second and third order for a set-up discussed in the text.

future work. The presence of a non-zero baseline plus oscillations is understandable from $|u_2(\mathbf{Q})|^2 = |\sum_p u_{2,p}(\mathbf{Q})|^2$, associating the baseline with $\sum_p |u_{2,p}(\mathbf{Q})|^2$, and the oscillations with the cross-terms. Presumably one can motivate $|u_3(\mathbf{Q})/u_0(\mathbf{Q})|^2$ along similar lines.

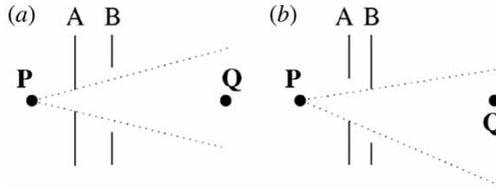


Figure 5. Two scenarios of geometrical blocking for $N=2$: (a) perimeter of B is totally blocked from lines of sight to \mathbf{P} and cannot participate in first-order diffraction; (b) perimeter B is partially blocked from lines of sight to \mathbf{P} .

5. Ramifications of blocking effects

One can identify geometrical blocking effects as belonging to one of two categories. In the first, simpler category, complete geometrical blocking prevents an element or elements from being involved in a contribution at a given order altogether. This is because at least one pertinent D -factor is always zero, making the contribution zero. This is illustrated in figure 5(a). In the second, more complicated case, lines of sight from a perimeter or perimeters to \mathbf{P} , \mathbf{Q} and/or other perimeters are partially blocked, as shown in figure 5(b). The second category arises when aperture perimeters or \mathbf{Q} intersect geometrical shadow boundaries or the $N > 1$ generalizations thereof. In contrast to our present analysis, the first- and higher-order BDW formulations then involve arbitrarily slight bends of paths. This situation requires extreme care, because various expressions in the above analysis can exhibit singularities and/or discontinuities that can cancel in the total $u(\mathbf{Q})$.

The mutually cancelling singularities and/or discontinuities can even arise within terms that formally occur at different orders of diffraction. For instance, at $m=1$, both $u_G(\mathbf{Q}) = u_0(\mathbf{Q})$ and $u_B(\mathbf{Q}) = u_1(\mathbf{Q})$ are discontinuous at a geometrical shadow boundary, whereas $u_G(\mathbf{Q}) + u_B(\mathbf{Q}) = u_0(\mathbf{Q}) + u_1(\mathbf{Q})$ is continuous there. Better understanding of the second category and extension of the current method to treating focusing elements are both topics of interest.

6. Potential applications

There are many potential applications for a higher-order BDW formulation, and three situations from practical radiometry are illustrated in figure 6. We have studied such situations for $m=2$ by performing angular integrations numerically and using the method of stationary phase. For $m=1$, the method of stationary phase was already applied in [3]. There, it was found that the method breaks down when \mathbf{P} , \mathbf{Q} and the centre of a circular aperture are sufficiently close to being collinear. For $m=2$, the method of stationary phase can be applied when $L(\{x_\mu\}, \{y_\mu\})$ is simultaneously stationary with respect to $\Theta_{k(1)}$ and $\Theta_{k(2)}$. By finding the points in the two-dimensional angle space where this condition holds, the stationary-phase method can be used according to standard techniques [10]. Determining this

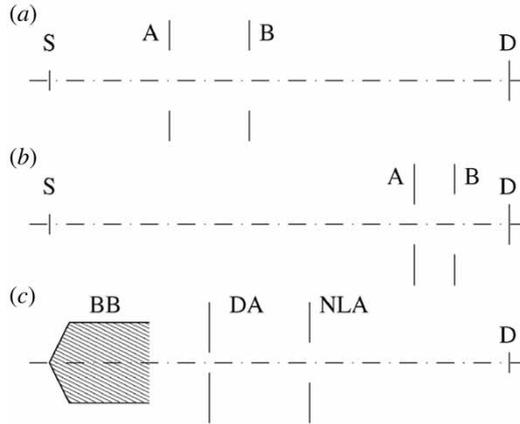


Figure 6. Three specific optical set-ups discussed in the text.

method's range of validity for $m > 1$ is beyond the scope of this work, but it is a topic of interest.

Figure 6(a) depicts a set-up used by Boivin [4] to study diffraction effects on light from the source (S) reaching the detector (D). The non-limiting apertures (A and B) can lead to first- and second-order diffraction effects, all of which can be treated as described here. In figure 6(b), a source (S) illuminates a detector (D) through two non-limiting apertures (A and B). The one labelled B is large, unilluminated, and very close to the one labelled A and the detector. Computing second-order diffraction effects of B on the radiation falling on D by equation (1) can be numerically intensive. Despite degradation of paraxial analysis in such a situation, the second-order effects are presumed to be small, so it can be sufficient to estimate them using the present higher-order BDW.

In figure 6(c), a blackbody (BB) is calibrated by measuring total power at the detector (D) that passes through the defining aperture (DA). A non-limiting aperture (NLA) is placed between the defining aperture and detector to reduce the stray radiation that reaches the detector. One may first consider a simple combination of first- and second-order diffraction by the DA and NLA, treating the DA as an extended Planck source in the latter case. Such $N=1$ calculations involving three elements (BB + DA + D or DA + NLA + D) have been studied extensively [11]. One can determine remaining diffraction effects on measured power by letting **P** sample the portion of the plane of the blackbody cavity opening that is not on the opening and **Q** sample the detector, and subtracting the second-order part of this last result.

7. Summary

In summary, we have presented a way to extend the boundary-diffraction-wave formulation to higher-order diffraction effects for radiation propagating through N apertures in series. This permits the $2N$ -fold integration that Kirchhoff's theory

suggests to be replaced by integrations that are at most N -fold, which can be numerically important. Furthermore, this work helps organize diffraction effects arising at different orders and helps facilitate analysis of their asymptotic properties. At present, we are seeking to derive simple expressions for the small-wavelength properties of diffraction effects mentioned in section 7, or, correspondingly, for the high-temperature properties of the average effects for broadband, thermal sources.

References

- [1] M. Born and E. Wolf, *Principles of Optics*, 3rd ed. (Pergamon, Oxford, 1965), pp. 375–386.
- [2] K. Miyamoto and E. Wolf, *J. Opt. Soc. Am.* **52** 615 (1962), and references therein.
- [3] E.L. Shirley, *Appl. Opt.* **43** 735 (2004).
- [4] L.P. Boivin, *Appl. Opt.* **17** 3323 (1978).
- [5] See, for example, J.W. Goodman, *Introduction to Fourier Optics* (The McGraw-Hill Companies, Inc., New York, 1996), pp. 66–67.
- [6] A. Walther, *The Ray and Wave Theory of Lenses* (Cambridge University Press, Cambridge, 1995), p. 152.
- [7] J.B. Keller, *J. Opt. Soc. Am.* **52** 116 (1962).
- [8] B.J. Bjorken and S.D. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill Book Company, New York, 1964), p. 170.
- [9] E.L. Shirley and M.L. Terraciano, *Appl. Opt.* **40** 4463 (2001).
- [10] See, for instance, the appendices of [1].
- [11] As an example, see P. Edwards and M. McCall, *Appl. Opt.* **42** 5024 (2003), and references therein.