Journal of Quantitative Spectroscopy & Radiative Transfer 210 (2018) 225–239

Contents lists available at ScienceDirect

Journal of Quantitative Spectroscopy & Radiative Transfer

journal homepage: www.elsevier.com/locate/jqsrt

A review and reassessment of diffraction, scattering, and shadows in electrodynamics

Matthew J. Berg*, Christopher M. Sorensen

Kansas State University, Department of Physics, 1228 North 17th Street, Manhattan, KS 66502-2601, USA

ARTICLE INFO

Article history:
Received 30 October 2017
Revised 26 January 2018
Accepted 2 February 2018
Available online 14 February 2018

ABSTRACT

The concepts of diffraction and scattering are well known and considered fundamental in optics and other wave phenomena. For any type of wave, one way to define diffraction is the spreading of waves, i.e., no change in the average propagation direction, while scattering is the deflection of waves with a clear change of propagation direction. However, the terms “diffraction” and “scattering” are often used interchangeably, and hence, a clear distinction between the two is difficult to find. This review considers electromagnetic waves and retains the simple definition that diffraction is the spreading of waves but demonstrates that all diffraction patterns are the result of scattering. It is shown that for electromagnetic waves, the “diffracted” wave from an object is the Ewald–Oseen extinction wave in the far-field zone. The intensity distribution of this wave yields what is commonly called the diffraction pattern. Moreover, this is the same Ewald–Oseen wave that cancels the incident wave inside the object and thereafter continues to do so immediately behind the object to create a shadow. If the object is much wider than the beam but has a hole, e.g., a screen with an aperture, the Ewald–Oseen extinction wave creates the shadow behind the screen and the incident light that passes through the aperture creates the diffraction pattern. This point of view also illustrates Babinet’s principle. Thus, it is the Ewald–Oseen extinction theorem that binds together diffraction, scattering, and shadows.

1. Introduction

Diffraction can be thought of as the spreading of a wave into the geometrical shadow behind an impervious obstacle [1–3]. The *mechanism* of diffraction depends upon the type of wave. Generally, waves divide into two types; those that require a material medium in which to propagate and those that do not. For those propagating in a material medium, e.g., water and sound waves, a wave is blocked by an obstacle and the portion of the wave passing near the edge of the obstacle spreads into the geometrically shaded region due to the elastic nature of the medium. In this context, “blocking” refers to a discontinuity in the medium that supports the wave propagation wherein propagation is not allowed. Waves that require no material medium, such as electromagnetic (EM) waves, fundamentally cannot be blocked because a discontinuity in a medium does not change the fact that these waves require no medium to propagate. Said less formally, there is no medium to be blocked. What then is the mechanism that creates an optical diffraction pattern? Here, it is shown that secondary radiation from an obstacle in the path of incident light, which is induced by that light, produces a scattering pattern identical to the diffraction pattern predicted by Huygens’ description. Thus, for EM waves, *secondary radiation is the mechanism of diffraction*.

A *clear* definition of what optical diffraction *is* and, in particular, how it may be different, or not, from scattering is rare in the literature. One could propose that diffraction relates to waves at sharp edges of two-dimensional (2D) objects, while scattering relates to three-dimensional (3D) objects. Such delineation, however, leads to ambiguity. For example, it would be difficult to understand the striking, albeit qualitative, similarity of the angular spread of light in the far-field from an opaque circular *disk* and a transparent *sphere* of the same diameter. Indeed, some references state that there is no logical separation between the two concepts [1,4]. An aim of this review is to clearly illustrate that the general concepts of diffraction and scattering relate to the *same* physical phenomenon.

The focus here is on EM waves due to the enduring interest in the topic and because these waves require no medium to propagate. As a consequence, optical shadows can form from destructive interference only, and definitely not due to obstacles in the medium “blocking” the wave in a mechanical-like sense. A novel insight revealed by this description is that the interference process creating shadows is always active, whether an object is absorbing

* Corresponding author.
E-mail address: matt.berg@phys.ksu.edu (M.J. Berg).

https://doi.org/10.1016/j.jqsrt.2018.02.002
0022-4073/

or not, is larger than the wavelength or not, and it is fundamentally linked to the observed phenomena associated with diffraction or scattering. It is also shown that these phenomena are produced from secondary radiation emanating throughout the *entire* volume of an obstacle. Thus, statements often encountered in the literature like "light diffracts around the obstacle," are misleading as they imply only a surface effect. Given the extensive amount of study on these concepts, this review cannot summarize all previous work. Rather, the focus will be on the mathematical treatments involved and their physical interpretations.

2. Huygens–Fresnel and Babinet principles: conceptual basis

It is helpful to first review the common description of optical diffraction. Begin with the familiar example of light of wavelength λ incident upon a rectangular aperture Σ_a, or slit, of width $2w$ and length 2ℓ with $\ell \gg w$ in an otherwise opaque screen Σ_s infinite in extent. The term "opaque" will refer to a perfectly conducting screen. Alternatively, a perfectly absorbing screen could be considered, but due to complications with the concept of a perfect absorber in electrodynamics, this case is not considered, cf. [5,6]. Suppose that the incident light is a well-collimated laser beam propagating along the positive z-axis. A good approximation for this wave is a Gaussian beam with a waist-width of $2w_0$ [7]. At the beam waist, the wave fronts are planar, but the beam profile is finite in size. In Section 6, this will allow the intrinsic angular spreading of the beam with distance from the waist to be incorporated in the analysis. At the waist, the beam encounters the aperture, which is much smaller than the waist, $w_0 \gg w$ and $w_0 \gg \ell$. Consequently, the aperture may – despite use of a beam – be regarded as uniformly illuminated by a plane wave following the customary treatment. The first objective is to examine the distribution of light beyond the aperture across an observation plane σ that is parallel to Σ_s located a distance $z=d$ from it as shown in Fig. 1. Also, σ will be assumed to be in the far-field zone of the aperture, which is defined by $d \gg kw^2/2$, where $k=2\pi/\lambda$ [8]. This condition is commonly known as the Fraunhofer approximation [1].

The incident beam at the aperture appears blocked by Σ_s and across σ one observes a spread of light intensity modulated by a series of band-like interference maxima and minima, i.e., fringes, commonly called the single-slit diffraction pattern. It is thus customary to say that the light "diffracts into the shadow." The diffraction pattern can be approximately calculated in the far-field zone from the Huygens-Fresnel principle [1,9,10]. One imagines fictitious point sources of light that span the aperture Σ_a where each radiates a spherical wave of wavelength λ into the $z > 0$ region. A point source, located at $\mathbf{r}'$ in Fig. 1, is driven in-phase and in magnitude with the incident wave across the aperture; this is the Kirchhoff approximation [4]. Adding the contributions from these sources at $\mathbf{r}$ on σ approximates the observed diffraction pattern. With reference to Fig. 1 and following the treatment of [1], the normalized pattern in the far-field zone is given by:

$$\frac{I(\mathbf{r})}{I_0} = \left| \iint_{\Sigma_a} e^{i\mathbf{k}\hat{\mathbf{r}}\cdot\mathbf{r}'} \mathrm{d}a' \right|^2, \tag{1}$$

where I is the diffracted light-intensity (irradiance) and I_0 is the intensity along the z-axis (beam direction). The fringe structure of the pattern is then explained from the phase difference introduced by each source-point's differing location within Σ_a. In other words, the Huygens-Fresnel principle explains the diffraction pattern as interference from radiation emitted across a *free-space region*, i.e., the aperture. Eq. (1) also shows that the diffraction pattern is the absolute square of the Fourier transform of the aperture; another characteristic of the Huygens–Fresnel treatment.

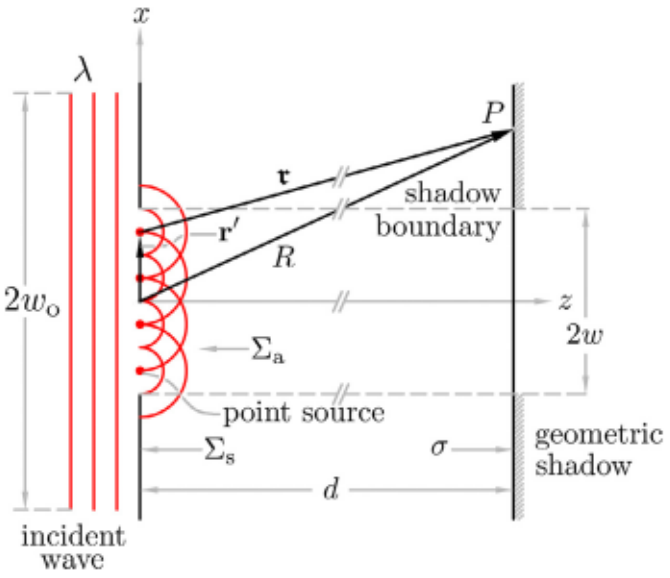

Fig. 1. Diffraction from a slit aperture. An aperture Σ_a in an opaque screen Σ_s is illuminated by a plane wave traveling along the positive z-axis. The aperture length is much larger than its width, i.e., $\ell \gg w$, whereas both dimensions are smaller than the beam waist w_0. Spanning the aperture are fictitious "Huygens" point-sources (red dots) that each emit a spherical wave into the region beyond the screen. Adding these waves across σ in the far-field zone gives an approximate description for the linear fringes of diffracted light observed on σ. In particular, the outcome predicts the observation of light in the geometric shadow of the aperture, shown in dash, and is a classic phenomenon associated with diffraction [1]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Diffraction is said to occur not only through apertures but also around barriers [1]. In 2D, any barrier may be envisioned as a screen that is the inverse of, or compliment to, an aperture. The complimentary screen to the rectangular aperture is a thin, opaque rectangular strip with dimensions $2w \times 2\ell$. If a screen Σ_s is combined with its complimentary screen Σ_s', the result is a complete opaque screen Σ, i.e., $\Sigma_s + \Sigma_s' = \Sigma$. Diffraction of a beam from an aperture and its complimentary screen is related by Babinet's principle [1]. As stated by [11],"the diffraction patterns which are produced by two complementary screens are identical excepting the central spot, which is diffraction angle zero."

A demonstration of Babinet's principle is instructive wherein two distinct scenarios, labeled 1 and 2, are compared. In scenario 1, an infinite screen $\Sigma_s^{(1)}$ containing aperture $\Sigma_a^{(1)}$ is illuminated by the beam and the resulting pattern is observed on σ. In scenario 2 however, only the complimentary screen is present, $\Sigma_s^{(2)}$, which is illuminated by the same beam and the resulting pattern is again observed on σ. If E_1 and E_2 are the *scalar* light-fields at the same point $\mathbf{r}$ on σ in scenario 1 or 2 respectively, then Babinet's principle states that [1,3]

$$E_1(\mathbf{r}) + E_2(\mathbf{r}) = E_0(\mathbf{r}). \tag{2}$$

Here, $E_0(\mathbf{r})$ is the (complex-valued) scalar light-field amplitude of the beam at $\mathbf{r}$ on σ when *neither* screen is present, i.e., when the beam is freely illuminating σ. Conceptually, one can understand Eq. (2) from the Huygens-Fresnel principle. As stated, the diffracted light (E_1) for scenario 1 from screen $\Sigma_s^{(1)}$ is given in Eq. (1) by an integral over the aperture opening, $\Sigma_a^{(1)}$. In scenario 2, Eq. (1) provides the diffracted light (E_2) as an integral over the planar region of free space *not* occupied by the complementary screen $\Sigma_s^{(2)}$, which could be regarded as a large aperture $\Sigma_a^{(2)}$, see Fig. 2. Adding these two surface integrals in Eq. (2) amounts to an integral of fictitious Huygens point sources over a complete plane Σ in empty space, and thus, reproduces the incident beam. Note

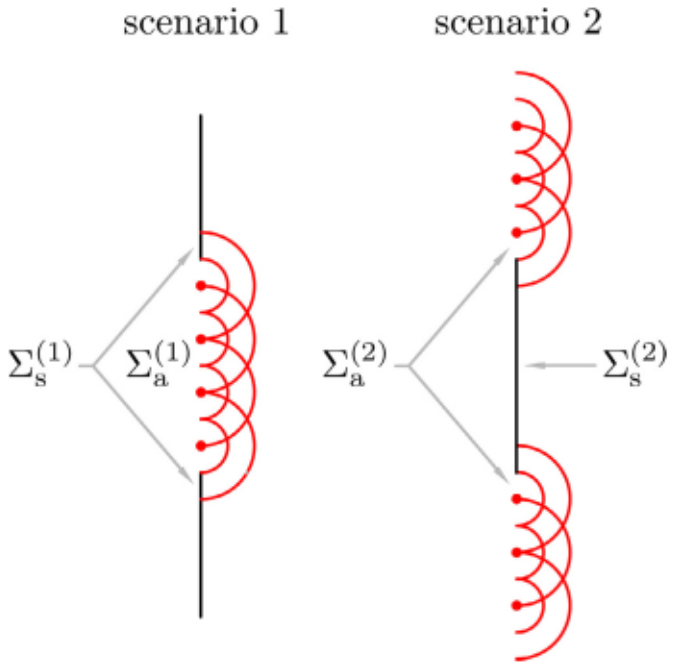

Fig. 2. Babinet's principle. Scenarios 1 and 2 contain complimentary screens. The first shows a screen with an aperture as in Fig. 1, while the second shows the compliment. If combined, the two screens form a complete opaque sheet Σ, while the combined apertures form a plane of empty space filled with Huygens point sources.

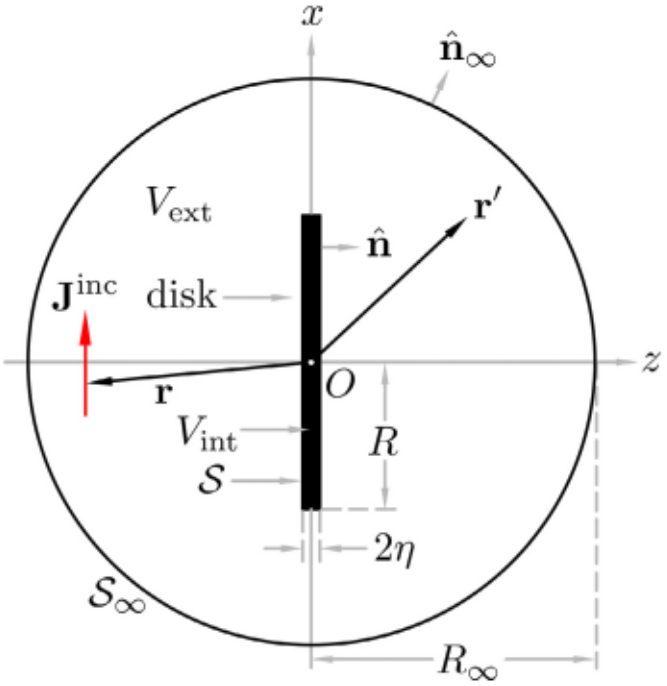

Fig. 3. Division of space in the Green function description of diffraction from an aperture in a screen. The disk shown here, which is later deformed to become a thin screen, spans a portion of the x-y plane and is normal to the incident wave. The external region V_{ext} is defined by the volume outside of this disk and the remainder of space. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

that Babinet's principle is an exact relationship despite its illustration here through the approximate Huygens–Fresnel principle.

Now, consider application of Eq. (2) to the rectangular aperture where $\Sigma_s^{(1)} = \Sigma_s$. The complementary screen would be a rectangular strip of the same dimensions, $\Sigma_s^{(2)}$. When the beam is unobstructed, i.e., no screens are present, and if $\mathbf{r}$ is *not* along the beam direction (z-axis), but is rather at an angle $\theta \gg \lambda/w_0$, then $E_0 = 0$ at that point. Eq. (2) then shows that $E_1 = -E_2$ where the negative sign implies that the two fields are 180° out of phase. Their magnitudes, and hence intensities, are equal so the diffraction patterns will be the *same* in both scenarios. The rectangular aperture and its complimentary screen (strip) show the same diffraction pattern. Note that Eq. (2) makes no stipulation on the type of wave, e.g. EM, elastic, etc., and so Babinet's principle should apply to all such waves, e.g., see [12,13]. Babinet's principle is only qualitative when the screen and its compliment reside in spaces of differing dimensions, i.e., 2D vs. 3D. For example, although the fringe pattern for a slit and the scattering pattern for an opaque fiber look qualitatively similar, the intensity minima for a slit's pattern go to zero whereas the minima in scattered intensity from the fiber never reach zero at any angle [14].

3. Electromagnetic formulation of diffraction

Eq. (1) is an approximate description of the field intensity across σ, even in the far-field zone. Nevertheless, the approximation is often excellent, agreeing well with the observed patterns for arbitrarily shaped apertures larger than λ, e.g., see [1,15]. Eq. (1) is also heuristically pleasing in that it gives one a way to understand the spreading of light into the geometric shadow; it links diffraction to radiation from the *open space* in a screen, whether the space is an aperture Σ_a or the free space surrounding a complimentary screen Σ_s'. These spaces, however, are precisely where there are no *physical* sources to radiate. This apparent inconsistency is recognized by some authors but often goes unaddressed [1]. When it is addressed, the treatment is generally qualitative. Thus, it is worthwhile to see how the Huygens–Fresnel description originates from a rigorous treatment wherein the salient effects are linked only to physical sources of radiation within the framework of the macroscopic Maxwell equations. Similar analyses can be investigated analytically such as diffraction from ribbons [16], wide slits [17], circular disks [18,19], subwavelength apertures and arrays thereof [20,21], knife edges [22], and arbitrary apertures [22–24].

3.1. Green function description

While the objective is to ultimately consider the infinitesimally thin, infinitely wide, screen Σ_s in Section 2, the treatment here begins with a finite-sized, thick, disk-shaped object for mathematical expediency. Later, this disk will be deformed to become Σ_s through a limiting process. Using this disk will also enable a direct connection to Section 4 where 3D objects are considered. Fig. 3 shows the disk with radius R_s that is oriented perpendicular to the z-axis and centered on the origin. The thickness of the disk is 2η along the axis and the surface and volume are S and V_{int}, respectively. The disk material is assumed uniform and characterized by permittivity ε and permeability μ and is surrounded by free space (ε_0, μ_0). Far from the disk is current density $\mathbf{J}^{\text{inc}}$ that produces the incident beam, and as such, is considered a known quantity independent of the presence of the disk. Enclosing both the source and disk is a large spherical surface S_∞ of radius R_∞ defining an exterior region V_{ext}. In the limit $R_\infty \to \infty$, all space is enclosed. Lastly, all field quantities in the following are assumed to contain the same time-harmonic dependence $e^{-i\omega t}$ where $\omega = kc$; these are suppressed for brevity.

The time-harmonic (or frequency domain [25]) Maxwell equations for $\mathbf{r}$ in regions V_{int} or V_{ext} are [26]

$$\left.\begin{aligned} \nabla \times \mathbf{E}(\mathbf{r}) - i\omega \mathbf{B}(\mathbf{r}) &= 0 \\ \nabla \times \mathbf{B}(\mathbf{r}) + i\omega\epsilon\mu \mathbf{E}(\mathbf{r}) &= 0 \end{aligned}\right\} \quad \mathbf{r} \in V_{\text{int}}, \tag{3}$$

$$\left.\begin{aligned}\nabla\times\mathbf{E}(\mathbf{r})-i\omega\mathbf{B}(\mathbf{r})&=0\\ \nabla\times\mathbf{B}(\mathbf{r})+i\omega\epsilon_0\mu_0\mathbf{E}(\mathbf{r})&=\mu_0\mathbf{J}^{\text{inc}}(\mathbf{r})\end{aligned}\right\}\quad \mathbf{r}\in V_{\text{ext}}. \tag{4}$$

Given that no free charge is present ($\nabla\cdot\mathbf{E}=0$), applying the curl operation to Eqs. (3) and (4) yields the following wave equations in each region:

$$\left.\begin{aligned}\nabla\times\nabla\times\mathbf{E}(\mathbf{r})-m^2k^2\mathbf{E}(\mathbf{r})&=0\\ \nabla\times\nabla\times\mathbf{B}(\mathbf{r})-m^2k^2\mathbf{B}(\mathbf{r})&=0\end{aligned}\right\}\quad \mathbf{r}\in V_{\text{int}}, \tag{5}$$

where $m=\sqrt{\epsilon\mu/\epsilon_0\mu_0}$ is the refractive index of the disk and

$$\left.\begin{aligned}\nabla\times\nabla\times\mathbf{E}(\mathbf{r})-k^2\mathbf{E}(\mathbf{r})&=i\omega\mu_0\mathbf{J}^{\text{inc}}(\mathbf{r})\\ \nabla\times\nabla\times\mathbf{B}(\mathbf{r})-k^2\mathbf{B}(\mathbf{r})&=\mu_0\nabla\times\mathbf{J}^{\text{inc}}(\mathbf{r})\end{aligned}\right\}\quad \mathbf{r}\in V_{\text{ext}}. \tag{6}$$

The fields $\mathbf{E}$ and $\mathbf{B}$ satisfy the usual boundary conditions at the disk surface S [26],

$$\hat{\mathbf{n}}\times\left[\mathbf{E}^{\text{ext}}(\mathbf{r})-\mathbf{E}^{\text{int}}(\mathbf{r})\right]=0,\qquad \mathbf{r}\in S \tag{7}$$

$$\hat{\mathbf{n}}\times\left[\mathbf{B}^{\text{ext}}(\mathbf{r})-\mathbf{B}^{\text{int}}(\mathbf{r})\right]=\mu_0\mathbf{K}(\mathbf{r}),\qquad \mathbf{r}\in S \tag{8}$$

where $\mathbf{E}^{\text{int}}$ and $\mathbf{E}^{\text{ext}}$ are, respectively, the fields in V_{int} and V_{ext} with the unit-vector normal $\hat{\mathbf{n}}$ directed outward from V_{int} as shown in Fig. 3. Here, $\mathbf{K}$ is any induced surface current that may reside on S.

Because the disk will eventually be turned into a screen, which has no volume, the focus will be on Eq. (6) rather than Eq. (5). Here, the wave equations are inhomogeneous and a solution can be formulated in terms of Green's functions. These functions satisfy the point-source analogs of Eq. (6) i.e., [26]

$$\nabla\times\nabla\times\ddot{\mathbf{G}}_{\text{e}}(\mathbf{r},\mathbf{r}')-k^2\ddot{\mathbf{G}}_{\text{e}}(\mathbf{r},\mathbf{r}')=\ddot{\mathbf{I}}\delta(\mathbf{r}-\mathbf{r}'), \tag{9}$$

$$\nabla\times\nabla\times\ddot{\mathbf{G}}_{\text{m}}(\mathbf{r},\mathbf{r}')-k^2\ddot{\mathbf{G}}_{\text{m}}(\mathbf{r},\mathbf{r}')=\nabla\times\left[\ddot{\mathbf{I}}\delta(\mathbf{r}-\mathbf{r}')\right]. \tag{10}$$

Here, $\ddot{\mathbf{G}}_{\text{e}}$ and $\ddot{\mathbf{G}}_{\text{m}}$ are the dyadic Green functions of the electric and magnetic type, respectively, $\ddot{\mathbf{I}}$ is the 3 × 3 identity dyad, and δ is the 3D Dirac delta function. The Green functions are found by solving Eqs. (9) and (10) subject to specified boundary conditions. Once known, the solution to Eq. (6) is given as a convolution integral of the Green function with the source $\mathbf{J}^{\text{inc}}$ and a surface integral over S involving the Green function and field boundary-values. This will be seen below. Note that the intent is usually to impose the same boundary conditions on Eqs. (9) and (10) that apply to the fields themselves such that the surface integral either vanishes or is known. Another option, pursued here, is to employ the free-space dyadic Green functions, which are the solutions to Eqs. (9) and (10) in the absence of any boundaries, i.e., no disk S in Fig. 3, and therefore take on a simplified mathematical form. Importantly, one will see that the functions facilitate a clear physical interpretation. The price paid, however, is that if S (the disk) is to be reintroduced into the analysis, the resulting surface integral must then be evaluated by other means.

As shown by [26,27], the solutions to Eqs. (9) and (10) for a *boundary-free system* are the free-space dyadic Green functions:

$$\ddot{\mathbf{G}}_{\text{e}}(\mathbf{r},\mathbf{r}')=\left(\ddot{\mathbf{I}}+\frac{1}{k^2}\nabla\otimes\nabla\right)G_0(\mathbf{r},\mathbf{r}'), \tag{11}$$

and

$$\ddot{\mathbf{G}}_{\text{m}}(\mathbf{r},\mathbf{r}')=\nabla\times\left[\ddot{\mathbf{I}}G_0(\mathbf{r},\mathbf{r}')\right], \tag{12}$$

where

$$G_0(\mathbf{r},\mathbf{r}')=\frac{1}{4\pi}\frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{|\mathbf{r}-\mathbf{r}'|}, \tag{13}$$

is the (scalar) free-space Green function and in Eqs. (11) and (12), $\otimes$ represents the direct product. Physically, $\ddot{\mathbf{G}}_{\text{e}}$ and $\ddot{\mathbf{G}}_{\text{m}}$ are 3 × 3 tensors where each column represents the electric or magnetic field, respectively, at $\mathbf{r}$ produced by an infinitesimal current element located at $\mathbf{r}'$ oriented along one of the Cartesian directions [26]. As such, they describe outward traveling vector spherical waves from three orthogonal point sources that may be superimposed to represent any solution to the Maxwell equations. Equivalently, the Green functions can be thought of as propagators for the electric and magnetic field from $\mathbf{r}'$ to $\mathbf{r}$. Consequently, $\ddot{\mathbf{G}}_{\text{e}}$ satisfies the Sommerfeld radiation condition on S_∞:

$$\lim_{r\to\infty}\left[\nabla\times\ddot{\mathbf{G}}_{\text{e}}(\mathbf{r},\mathbf{r}')-ik\hat{\mathbf{r}}\times\ddot{\mathbf{G}}_{\text{e}}(\mathbf{r},\mathbf{r}')\right]=0. \tag{14}$$

The same condition holds for $\ddot{\mathbf{G}}_{\text{m}}$ [26]. The scattered fields also must also satisfy a similar relation at infinity known as the Silver–Müller radiation condition, see [26]. Note that both $\ddot{\mathbf{G}}_{\text{e}}$ and $\ddot{\mathbf{G}}_{\text{m}}$ are singular at $\mathbf{r}=\mathbf{r}'$; this will be an important consideration in Section 4.

Now consider Green's second vector identity

$$\begin{aligned}&\int_V\{\mathbf{A}\cdot[\nabla\times\nabla\times\mathbf{C}]-\mathbf{C}\cdot[\nabla\times\nabla\times\mathbf{A}]\}\,dv\\ &=\oint_S\{\mathbf{C}\times[\nabla\times\mathbf{A}]+[\nabla\times\mathbf{C}]\times\mathbf{A}\}\cdot\hat{\mathbf{n}}\,da,\end{aligned} \tag{15}$$

which, in essence, is a generalization of the divergence theorem [28]. To connect this to the discussion above, focus on the electric-field part of Eq. (6) and its Green function companion, Eq. (9). Then, let $\mathbf{A}=\mathbf{E}$ and $\mathbf{C}=\ddot{\mathbf{G}}_{\text{e}}\cdot\mathbf{a}$ in Eq. (15) where $\mathbf{a}$ is a constant "pilot" vector. Applying Eq. (15) to V_{ext}, which is bounded by the *two* surfaces S and S_∞ gives:

$$\begin{aligned}&\int_{V_{\text{ext}}}\left\{\mathbf{E}(\mathbf{r})\cdot\left[\nabla\times\nabla\times\ddot{\mathbf{G}}_{\text{e}}(\mathbf{r},\mathbf{r}')\cdot\mathbf{a}\right]-\left[\ddot{\mathbf{G}}_{\text{e}}(\mathbf{r},\mathbf{r}')\cdot\mathbf{a}\right]\cdot[\nabla\times\nabla\times\mathbf{E}(\mathbf{r})]\right\}dv\\ &=\oint_{S_\infty}\left\{\left[\ddot{\mathbf{G}}_{\text{e}}(\mathbf{r},\mathbf{r}')\cdot\mathbf{a}\right]\times[\nabla\times\mathbf{E}(\mathbf{r})]+\left[\nabla\times\ddot{\mathbf{G}}_{\text{e}}(\mathbf{r},\mathbf{r}')\cdot\mathbf{a}\right]\times\mathbf{E}(\mathbf{r})\right\}\cdot\hat{\mathbf{n}}_\infty\,da\\ &-\oint_S\left\{\left[\ddot{\mathbf{G}}_{\text{e}}(\mathbf{r},\mathbf{r}')\cdot\mathbf{a}\right]\times[\nabla\times\mathbf{E}(\mathbf{r})]+\left[\nabla\times\ddot{\mathbf{G}}_{\text{e}}(\mathbf{r},\mathbf{r}')\cdot\mathbf{a}\right]\times\mathbf{E}(\mathbf{r})\right\}\cdot\hat{\mathbf{n}}\,da.\end{aligned} \tag{16}$$

The last surface integral has a negative sign because the surface normal $\hat{\mathbf{n}}$ is directed into V_{ext} for that integral. The first integral over S_∞ is zero via the radiation condition, Eq. (14) [26,27]. With Eqs. (4), (6), (9), (11) and (13), Eq. (16) simplifies to

$$\begin{aligned}&\int_{V_{\text{ext}}}\left\{\mathbf{E}(\mathbf{r})\cdot\mathbf{a}\,\delta(\mathbf{r},\mathbf{r}')-\left[\ddot{\mathbf{G}}_{\text{e}}(\mathbf{r},\mathbf{r}')\cdot\mathbf{a}\right]\cdot i\omega\mu_0\mathbf{J}^{\text{inc}}(\mathbf{r})\right\}dv\\ &=-\oint_S\left\{i\omega\left[\ddot{\mathbf{G}}_{\text{e}}(\mathbf{r},\mathbf{r}')\cdot\mathbf{a}\right]\times\mathbf{B}(\mathbf{r})+\left[\ddot{\mathbf{G}}_{\text{m}}(\mathbf{r},\mathbf{r}')\cdot\mathbf{a}\right]\times\mathbf{E}(\mathbf{r})\right\}\cdot\hat{\mathbf{n}}\,da.\end{aligned} \tag{17}$$

Using $[(\ddot{\mathbf{G}}_{\text{e}}\cdot\mathbf{a})\times\mathbf{B}]\cdot\hat{\mathbf{n}}=-(\ddot{\mathbf{G}}_{\text{e}}\cdot\mathbf{a})\cdot(\hat{\mathbf{n}}\times\mathbf{B})$, $[(\ddot{\mathbf{G}}_{\text{m}}\cdot\mathbf{a})\times\mathbf{E}]\cdot\hat{\mathbf{n}}=-(\ddot{\mathbf{G}}_{\text{m}}\cdot\mathbf{a})\cdot(\hat{\mathbf{n}}\times\mathbf{E})$, evaluating the delta-function term, and canceling the (arbitrary) pilot vector $\mathbf{a}$ from each side of Eq. (17), one finds:

$$\begin{aligned}\left.\begin{matrix}\mathbf{E}(\mathbf{r}') & \text{if} & \mathbf{r}'\in V_{\text{ext}}\\ 0 & \text{if} & \mathbf{r}'\in V_{\text{int}}\end{matrix}\right\}\quad&=i\omega\mu_0\int_{V_{\text{ext}}}\ddot{\mathbf{G}}_{\text{e}}(\mathbf{r},\mathbf{r}')\cdot\mathbf{J}^{\text{inc}}(\mathbf{r})\,dv\\ &+\oint_S\left\{i\omega\ddot{\mathbf{G}}_{\text{e}}(\mathbf{r},\mathbf{r}')\cdot\left[\hat{\mathbf{n}}\times\mathbf{B}(\mathbf{r})\right]+\ddot{\mathbf{G}}_{\text{m}}(\mathbf{r},\mathbf{r}')\cdot\left[\hat{\mathbf{n}}\times\mathbf{E}(\mathbf{r})\right]\right\}da.\end{aligned} \tag{18}$$

Eq. (18) has several implications. To see this, first return to Eq. (16) and remove the boundary S by removing the disk from

Fig. 3. With Eq. (14) in mind, the result is

$$\int_{\mathbb{R}^3} \left\{ \mathbf{E}(\mathbf{r}) \cdot \left[\nabla \times \nabla \times \ddot{\mathbf{G}}_e(\mathbf{r}, \mathbf{r}') \cdot \mathbf{a} \right] - \left[\ddot{\mathbf{G}}_e(\mathbf{r}, \mathbf{r}') \cdot \mathbf{a} \right] \cdot [\nabla \times \nabla \times \mathbf{E}(\mathbf{r})] \right\} \, dv = 0 \tag{19}$$

where $V_{ext} \to \mathbb{R}^3$ (all space). Use of Eqs. (6) and (9), and again removing the pilot vector shows that Eq. (19) becomes,

$$\int_{\mathbb{R}^3} \left\{ \mathbf{E}(\mathbf{r}) \, \delta(\mathbf{r}, \mathbf{r}') - i\omega\mu_0 \ddot{\mathbf{G}}_e(\mathbf{r}, \mathbf{r}') \cdot \mathbf{J}^{inc}(\mathbf{r}) \right\} \, dv = 0,$$

or, evaluating the delta-function integral, this becomes

$$\mathbf{E}^{inc}(\mathbf{r}') = i\omega\mu_0 \int_{\mathbb{R}^3} \ddot{\mathbf{G}}_e(\mathbf{r}, \mathbf{r}') \cdot \mathbf{J}^{inc}(\mathbf{r}) \, dv. \tag{20}$$

Eq. (20) is the Green function solution to Eq. (6) when no disk is present and all that exists is the source $\mathbf{J}^{inc}$, i.e., this is the *incident field*. Next, assume that $\mathbf{J}^{inc}$ is nonzero only in a well-defined volume within V_{ext} that excludes regions where the field is desired such that the singularity in Eq. (20) at $\mathbf{r} = \mathbf{r}'$ is avoided. Physically, this would correspond to the source of the incident wave being localized in a region far removed from where the fields are wanted. Then, the volume integral in Eq. (18) is identical to that in Eq. (20) and the former equation becomes,

$$\left. \begin{matrix} \mathbf{E}(\mathbf{r}') & \text{if} \quad \mathbf{r}' \in V_{ext} \\ 0 & \text{if} \quad \mathbf{r}' \in V_{int} \end{matrix} \right\} = \mathbf{E}^{inc}(\mathbf{r}') + \oint_S \left\{ i\omega\ddot{\mathbf{G}}_e(\mathbf{r}, \mathbf{r}') \cdot [\hat{\mathbf{n}} \times \mathbf{B}(\mathbf{r})] + \ddot{\mathbf{G}}_m(\mathbf{r}, \mathbf{r}') \cdot [\hat{\mathbf{n}} \times \mathbf{E}(\mathbf{r})] \right\} \, da. \tag{21}$$

Notice that Eq. (21) now applies when the disk *is* present. However, the expression utilizes a somewhat unconventional notation where the field point $\mathbf{r}'$ is expressed in the primed coordinate system while the source integrals, Eq. (20) and the surface integral in Eq. (21), are expressed in unprimed coordinates $\mathbf{r}$. The standard notation is retrieved by interchanging the coordinates, $\mathbf{r} \to \mathbf{r}'$ and $\mathbf{r}' \to \mathbf{r}$ and using the symmetry relations of the Green functions, $\ddot{\mathbf{G}}_e(\mathbf{r}', \mathbf{r}) = \ddot{\mathbf{G}}_e(\mathbf{r}, \mathbf{r}')$ and $\ddot{\mathbf{G}}_m(\mathbf{r}', \mathbf{r}) = -\ddot{\mathbf{G}}_m(\mathbf{r}, \mathbf{r}')$, giving [26]

$$\left. \begin{matrix} \mathbf{E}(\mathbf{r}) & \text{if} \quad \mathbf{r} \in V_{ext} \\ 0 & \text{if} \quad \mathbf{r} \in V_{int} \end{matrix} \right\} = \mathbf{E}^{inc}(\mathbf{r}) + \oint_S \left\{ i\omega\ddot{\mathbf{G}}_e(\mathbf{r}, \mathbf{r}') \cdot [\hat{\mathbf{n}}' \times \mathbf{B}(\mathbf{r}')] - \ddot{\mathbf{G}}_m(\mathbf{r}, \mathbf{r}') \cdot [\hat{\mathbf{n}}' \times \mathbf{E}(\mathbf{r}')] \right\} \, da'. \tag{22}$$

This interchange of coordinates is not simply a choice of notation. That is, taking $\mathbf{r} \to \mathbf{r}'$ and $\mathbf{r}' \to \mathbf{r}$ in Eqs. (9) and (10) would correspond to placing the (Green function) point current-source at $\mathbf{r}$ rather than at $\mathbf{r}'$, and thus, the derivatives would switch from ∇ to ∇'. Ultimately, it is a consequence of the electromagnetic reciprocity theorem that yields the symmetry properties leading to Eq. (22) [26].

Eq. (22), and its magnetic-field companion [via Eqs. (12) and (13) but not shown] are the central result of this section. The Maxwell equations for the fields in V_{ext} are now given in an exact way by the incident field $\mathbf{E}^{inc}$ and a surface integral of the tangential components of the fields across the disk S. As is stands, Eq. (22) does not directly provide the total fields $\mathbf{E}$ and $\mathbf{B}$ since their surface components are generally not known. Thus, Eq. (22) is an integral equation for the unknown fields, which can be solved through various methods including the extended boundary conditions method [29-31]. This is the price paid for using the free-space Green functions mentioned earlier. However, Eq. (22) permits use of approximations for the surface-field components, one of which is the vector analog of the Kirchhoff approximation introduced in Section 2. More importantly, there is a clear distinction between the portion of the total field $\mathbf{E}$ due to the incident wave and that attributed solely to the disk, with the latter being encapsulated by the surface integral in Eq. (22). This will have implications for physical interpretations below.

3.2. Example: a mirror

As the next step on the way to describing diffraction through Eq. (22), consider a limiting operation that transforms the 3D disk into an opaque, flat 2D sheet. This is done by taking $R \to \infty$, $\eta \to 0$ in Fig. 3, $\mu = \mu_0$, and $\varepsilon \to \infty$. The result is the same infinitesimally thin 2D sheet Σ of perfectly conducting material as in Section 2 that has no enclosed volume, i.e., $V_{int} \to 0$. Physically, this could correspond to a non-magnetic metal film of negligible thickness that can be regarded as a perfect electric conductor, i.e., simply an uncoated first-surface mirror [32,33]. An aperture Σ_a is still not present, but will be incorporated later. For now, focus on the implications of Eq. (22) for $\mathbf{r}$ in V_{ext}.

The sheet splits space into two regions, $V^{(-)}$ for locations with $z < 0$ and $V^{(+)}$ for $z > 0$. Given that the incident wave approaches Σ from $V^{(-)}$, these two regions will be called the illuminated and shaded sides, respectively. Yet, before Eq. (22) can be applied, a subtle point must be addressed. The disk surface S is a *closed* surface. Upon drawing the disk thickness η to zero, the *two* surfaces constituting the illuminated and shaded side of the (thick) disk are brought into coincidence [9]. Let these be called $S^{(-)}$ and $S^{(+)}$ with surface normals $\hat{\mathbf{n}}^{(-)} = -\hat{\mathbf{z}}$ and $\hat{\mathbf{n}}^{(+)} = \hat{\mathbf{z}}$, respectively. When the limiting process is complete, $S^{(-)}$ and $S^{(+)}$ become the $z = 0$ plane and the surface integral in Eq. (22) splits:

$$\begin{aligned} \mathbf{E}(\mathbf{r}) = \mathbf{E}^{inc}(\mathbf{r}) &+ \iint_{S^{(+)}} \left\{ i\omega\ddot{\mathbf{G}}_e(\mathbf{r}, \mathbf{r}') \cdot \left[\hat{\mathbf{n}}^{(+)} \times \mathbf{B}(\mathbf{r}') \right] \right. \\ &\left. - \ddot{\mathbf{G}}_m(\mathbf{r}, \mathbf{r}') \cdot \left[\hat{\mathbf{n}}^{(+)} \times \mathbf{E}(\mathbf{r}') \right] \right\} \, da' \\ &+ \iint_{S^{(-)}} \left\{ i\omega\ddot{\mathbf{G}}_e(\mathbf{r}, \mathbf{r}') \cdot \left[\hat{\mathbf{n}}^{(-)} \times \mathbf{B}(\mathbf{r}') \right] \right. \\ &\left. - \ddot{\mathbf{G}}_m(\mathbf{r}, \mathbf{r}') \cdot \left[\hat{\mathbf{n}}^{(-)} \times \mathbf{E}(\mathbf{r}') \right] \right\} \, da' \end{aligned} \tag{23}$$

Meanwhile, the boundary conditions of Eqs. (7) and (8) are

$$\hat{\mathbf{n}}^{(+)} \times \left[\mathbf{E}^{(+)}(\mathbf{r}') - \mathbf{E}^{(-)}(\mathbf{r}') \right] = 0, \qquad \mathbf{r}' \in \Sigma, \tag{24}$$

$$\hat{\mathbf{n}}^{(+)} \times \left[\mathbf{B}^{(+)}(\mathbf{r}') - \mathbf{B}^{(-)}(\mathbf{r}') \right] = \mu_0 \mathbf{K}(\mathbf{r}'), \qquad \mathbf{r}' \in \Sigma, \tag{25}$$

where $\mathbf{E}^{(+)}$ and $\mathbf{E}^{(-)}$ are, respectively, the *total* (and physical) fields on the $V^{(+)}$ and $V^{(-)}$ sides of Σ. Noting that $\hat{\mathbf{n}}^{(-)} = -\hat{\mathbf{n}}^{(+)}$ and $S^{(-)} = S^{(+)} = \Sigma$, the integrals in Eq. (23) combine via Eqs. (24) and (25) to give

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}^{inc}(\mathbf{r}) + i\omega\mu_0 \iint_{\Sigma} \ddot{\mathbf{G}}_e(\mathbf{r}, \mathbf{r}') \cdot \mathbf{K}(\mathbf{r}') \, da', \quad \mathbf{r} \in V^{(-)}, \, V^{(+)} \tag{26}$$

Eq. (26) now has a clear interpretation. With the Green function envisioned as the field propagator for a point-current source, the surface integral describes the field *radiated* by the *physical* surface current on the sheet. Thus, the influence of the sheet on the total field is distinguished from the incident field as

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}^{inc}(\mathbf{r}) + \mathbf{E}^{rad}(\mathbf{r}), \tag{27}$$

where $\mathbf{E}^{rad}$ is the *integral* in Eq. (26).

Next, take the incident wave source $\mathbf{J}^{inc}$ to be such that it produces a plane wave at the sheet. Then, the fields are

$$\mathbf{E}^{inc}(\mathbf{r}) = E_0 e^{ik\hat{\mathbf{z}}\cdot\mathbf{r}} \hat{\mathbf{x}}, \quad \mathbf{B}^{inc}(\mathbf{r}) = \frac{k}{\omega} \hat{\mathbf{z}} \times \mathbf{E}^{inc}(\mathbf{r}), \quad \mathbf{r} \in V^{(-)}, \, V^{(+)}. \tag{28}$$

Indeed, this case is simple enough that the *total* field [Eq. (27)] is known *a priori*; it includes Eq. (28) and a reflected plane wave in $V^{(-)}$ [1],

$$\mathbf{E}^{ref}(\mathbf{r}) = -E_0 e^{-ik\hat{\mathbf{z}}\cdot\mathbf{r}} \hat{\mathbf{x}}, \quad \mathbf{B}^{ref}(\mathbf{r}) = -\frac{k}{\omega} \hat{\mathbf{z}} \times \mathbf{E}^{ref}(\mathbf{r}), \quad \mathbf{r} \in V^{(-)}. \tag{29}$$

Meanwhile, the total fields on the shaded side of the sheet are zero,

$$\mathbf{E}^{(+)}(\mathbf{r}) = \mathbf{B}^{(+)}(\mathbf{r}) = 0, \ \mathbf{r} \in V^{(+)}. \tag{30}$$

Thus, on the illuminated side of the sheet, the field is the superposition of Eqs. (28) and (29), which forms a standing wave [11],

$$\mathbf{E}^{(-)}(\mathbf{r}) = 2iE_0 \sin\left(k\hat{\mathbf{z}}\cdot\mathbf{r}\right)\hat{\mathbf{x}}, \ \mathbf{B}^{(-)}(\mathbf{r}) = \frac{2k}{\omega}E_0 \cos\left(k\hat{\mathbf{z}}\cdot\mathbf{r}\right)\hat{\mathbf{y}}, \ \mathbf{r} \in V^{(-)}. \tag{31}$$

Using Eqs. (28) and (29) in Eq. (25) reveals the surface current that must be present:

$$\mathbf{K}(\mathbf{r}) = \frac{2}{\mu_0}\frac{k}{\omega}E_0\hat{\mathbf{x}}, \ \mathbf{r} \in \Sigma_s. \tag{32}$$

This is consistent with the expectation that the incident electric field drives current via Ohm's law, i.e., uniformly along the polarization direction $\hat{\mathbf{x}}$.

Eq. (32) can now be used in Eq. (26) to describe the fields, or lack thereof, on either side of the sheet in terms of *radiation* from $\mathbf{K}$. To do so, let $\boldsymbol{\tau} = \mathbf{r} - \mathbf{r}'$ and evaluate the derivatives in Eq. (11) while retaining only the terms with τ^{-1} dependence, which describe the radiation contribution. With $\hat{\boldsymbol{\tau}} = \boldsymbol{\tau}/|\boldsymbol{\tau}|$, his gives [34]:

$$\ddot{\mathbf{G}}_e(\mathbf{r},\mathbf{r}') = \frac{1}{4\pi}\left(\ddot{\mathbf{I}} - \hat{\boldsymbol{\tau}}\otimes\hat{\boldsymbol{\tau}}\right)\frac{e^{ik\tau}}{\tau}. \tag{33}$$

Then, the field radiated by the induced current follows from the integral in Eq. (26):

$$\mathbf{E}^{\text{rad}}(\mathbf{r}) = \frac{ik}{2\pi}E_0\iint_{-\infty}^{\infty}\frac{e^{ik\tau}}{\tau}\left(\ddot{\mathbf{I}} - \hat{\boldsymbol{\tau}}\otimes\hat{\boldsymbol{\tau}}\right)\cdot\hat{\mathbf{x}}\,\mathrm{d}x'\mathrm{d}y', \ \mathbf{r} \in V^{(-)}, V^{(+)} \tag{34}$$

This can be simplified by realizing that $(\hat{\boldsymbol{\tau}}\otimes\hat{\boldsymbol{\tau}})\cdot\hat{\mathbf{x}} = \hat{\boldsymbol{\tau}}\cos\beta$ where β is the angle formed by $\hat{\boldsymbol{\tau}}$ and the positive x-axis. Because $\mathbf{K}$ is a constant over the x'-y' plane, i.e., Σ, $\mathbf{r}$ may be placed arbitrarily along the z-axis in $V^{(-)}$ or $V^{(+)}$. Thus, β ranges from $\beta = 0$ when $x = -\infty$ to $\beta = \pi$ when $x = \infty$. Then, $\cos\beta$ is an odd function in x' and y' so its contribution to Eq. (34) cancels, leaving:

$$\mathbf{E}^{\text{rad}}(\mathbf{r}) = \frac{ik}{2\pi}E_0\hat{\mathbf{x}}\iint_{-\infty}^{\infty}\frac{e^{ik\tau}}{\tau}\mathrm{d}x'\mathrm{d}y', \ \mathbf{r} \in V^{(-)}, V^{(+)}. \tag{35}$$

This integral can be evaluated by transforming to cylindrical coordinates (ρ', ϕ', z') where $\rho'^2 = x'^2 + y'^2$ giving

$$\mathbf{E}^{\text{rad}}(\mathbf{r}) = ikE_0\hat{\mathbf{x}}\int_0^{\infty}\frac{e^{ik\sqrt{\rho'^2+z^2}}}{\sqrt{\rho'^2+z^2}}\rho'\mathrm{d}\rho' \ \mathbf{r} \in V^{(-)}, V^{(+)}, \tag{36}$$

where the ϕ' integral has been evaluated. Next, use the substitution $\xi(\rho) = \sqrt{\rho'^2 + z^2}$, which converts Eq. (36) into

$$\mathbf{E}^{\text{rad}}(\mathbf{r}) = ikE_0\hat{\mathbf{x}}\int_{\xi(0)}^{\infty} e^{ik\xi}\,\mathrm{d}\xi. \tag{37}$$

As it stands, Eq. (37) will yield an indeterminate result due to the integral's upper limit. One way to resolve this is to add a small imaginary part α to the wave number such that $k \to k + i\alpha$, evaluate the integral, and then take the limit $\alpha \to 0$, i.e. [32],

$$\mathbf{E}^{\text{rad}}(\mathbf{r}) = E_0\lim_{\alpha\to 0}\left[e^{ik\infty}e^{-\alpha\infty} - e^{ik\xi(0)}e^{-\alpha\xi(0)}\right]\hat{\mathbf{x}}. \tag{38}$$

Lastly, one must recognize that the value for $\xi(0)$ appearing in the lower limit in Eq. (37) tacitly depends on the sign of z. If $z > 0$, then $\xi(0) = z$, while if $z < 0$, then $\xi(0) = -z$. With these limits and Eqs. (28) and (29), Eq. (38) yields

$$\mathbf{E}^{\text{rad}}(\mathbf{r}) = \begin{cases} -E_0 e^{-ikz}\,\hat{\mathbf{x}} = \mathbf{E}^{\text{ref}}(\mathbf{r}), & z < 0, \\ -E_0 e^{ikz}\,\hat{\mathbf{x}} = -\mathbf{E}^{\text{inc}}(\mathbf{r}), & z > 0. \end{cases} \tag{39}$$

The total field present on either side of the sheet is then, via Eq. (27),

$$\mathbf{E}(\mathbf{r}) = \begin{cases} \mathbf{E}^{\text{inc}}(\mathbf{r}) + \mathbf{E}^{\text{ref}}(\mathbf{r}), & z < 0, \\ \mathbf{E}^{\text{inc}}(\mathbf{r}) - \mathbf{E}^{\text{inc}}(\mathbf{r}) = 0, & z > 0. \end{cases} \tag{40}$$

Eq. (40) reveals the remarkable fact that the *shadow* behind the sheet ($z > 0$) can be exactly described by *radiation from the induced current on the sheet*. That is, the current radiates a copy of the incident field travelling into $z > 0$ but 180° out of phase, which extinguishes the incident field, and hence, is often called the *extinction wave*. On the illuminated side, $z < 0$, the sheet again radiates a copy of the incident field 180° out of phase, but that wave travels in the opposite direction. Thus, the radiated field superposes with the incident field to create a standing wave, as expected from Eq. (31).

This interference explanation of the sheet's *shadow* may seem unnecessary from the Huygens-Fresnel perspective. Because the sheet is a perfect conductor, it is natural to think that the sheet simply "blocks" the incident light in the same way that a water wave is blocked, then reflected, by a rigid barrier. However, remember that EM waves require no medium to support their propagation. This is obvious from the ability of light to propagate in vacuum and is consistent with the absence of a luminiferous aether. Moreover, it is the non-existence of the aether that (partly) underlies the special theory of relativity and accounts for the different descriptions of the Doppler shift for light and sound waves [1]. Consequently, *there is no mechanical-like mechanism of blocking a light wave*. As such, the only way that light can be *removed* from any region of space is through destructive interference, and this requires the appearance of a secondary, or induced, source of radiation. In the example above, this source is the physical surface current on the sheet.

3.3. Electromagnetic origin of the Huygens–Fresnel principle

Given Section 3.2, a direct way to incorporate an aperture in the above treatment is to null the surface current where the aperture Σ_a is to be. Of course, this is equivalent to removing the portion of the conductor occupying this region. Then, an alternative to the Huygens-Fresnel explanation for light spreading into an aperture's shadow becomes possible, see Fig. 4. The secondary radiation from the current induced on the *remainder* of the sheet, which is now a screen Σ_s, cannot completely cancel the incident wave in $V^{(+)}$. Moreover, the missing current in Σ_a will alter the current that does develop across Σ_s, meaning that the cancellation of the incident wave remains incomplete even within the geometric shadow. As one moves *along* Σ_s far from Σ_a, the effect of this missing current on the current that is induced decays and one regains Eq. (32). Thus, points "deep" in the geometric shadow but near the screen show negligible field because the current is nearly able to produce the extinction wave.

One now has two seemingly different ways to understand diffraction from an aperture: The Huygens-Fresnel principle describes the light on σ as radiation from fictitious sources across the empty space of the aperture. Meanwhile, Eq. (26) describes this same distribution on σ as the superposition of the original (incident) wave with that radiated by (real) induced surface current on the screen. These two views are related, and in fact, the familiar Huygens-Fresnel principle is contained in Eq. (26) via Babinet's principle.

To see how, recall that Babinet's principle compares two separate scenarios, 1 and 2, as described in Section 2. The former involves a screen $\Sigma_s^{(1)}$ with aperture $\Sigma_a^{(1)}$, while the latter contains only the aperture's complimentary screen $\Sigma_s^{(2)}$, recall Fig. 2. Implicit in scenario 2 is the free space region that would otherwise

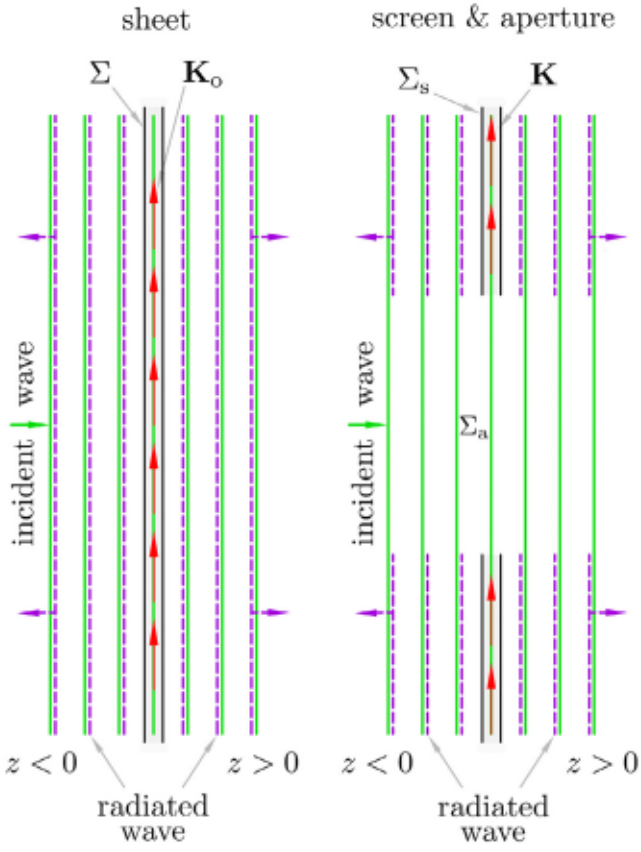

Fig. 4. Radiation description of shadows and diffraction. In (a) is the sheet Σ with induced surface current $\mathbf{K}_o$. This current (red), induced by the incident wave (green), radiates a copy (dashed purple) of the incident wave that interferes destructively on the shaded side ($z > 0$) of the sheet and a counter-propagating copy to produce a standing wave on the illuminated side ($z < 0$). In (b) is the screen Σ_s with an aperture Σ_a. Here the induced surface current $\mathbf{K}_o$ is modified by the aperture to become $\mathbf{K}$ and the radiation is incomplete with regard to its ability to cancel the incident wave in ($z > 0$). The result is an uncancelled, i.e., diffracted wave surviving in $z > 0$. The sheet and screen are shown with finite width for clarity, but are infinitesimally thick in reality. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

be occupied by $\Sigma_s^{(1)}$ to form a complete sheet; this can be envisioned as the aperture $\Sigma_a^{(2)}$ in scenario 2. Each scenario has associated fields:

$$\mathbf{E}_1 = \mathbf{E}_1^{\text{inc}} + \mathbf{E}_1^{\text{rad}}, \quad \mathbf{B}_1 = \mathbf{B}_1^{\text{inc}} + \mathbf{B}_1^{\text{rad}}, \tag{41}$$

$$\mathbf{E}_2 = \mathbf{E}_2^{\text{inc}} + \mathbf{E}_2^{\text{rad}}, \quad \mathbf{B}_2 = \mathbf{B}_2^{\text{inc}} + \mathbf{B}_2^{\text{rad}}. \tag{42}$$

Take $\mathbf{r} \in V^{(+)}$ in both scenarios but assume that $\mathbf{r} \notin \Sigma_s^{(1)}$ in scenario 1 and $\mathbf{r} \notin \Sigma_s^{(2)}$ in scenario 2. Then, Eq. (4) will not involve the current source-term and Eqs. (3) and (4) exhibit duality. That means, e.g., that Eq. (4) is obtained from Eq. (3) upon making the substitutions $\mathbf{E} \to (\omega/k)\mathbf{B}$ and $\mathbf{B} \to -(k/\omega)\mathbf{E}$ and vice versa. Note that transformation of the volumes of Section 3.2 to the surfaces $V_{\text{int}} \to \Sigma_s^{(1,2)} \cup \Sigma_a^{(1,2)}$ and $V_{\text{ext}} \to V^{(-)} \cup V^{(+)}$ is assumed here, where the superscript (1,2) represents either scenario 1 or 2. This duality can be exploited if the polarization of the incident beam is rotated by 90° from scenario 1 to 2, i.e., $\mathbf{E}_2^{\text{inc}} = -(\omega/k)\mathbf{B}_1^{\text{inc}}$ and $\mathbf{B}_2^{\text{inc}} = (k/\omega)\mathbf{E}_1^{\text{inc}}$ as shown in Fig. 5.

Using the boundary conditions on the screens, [9] shows that the electrodynamic version of Babinet's principle is realized by the following substitutions:

$$\mathbf{E}_2^{\text{rad}} \to \frac{\omega}{k}\mathbf{B}_1 \text{ for } \Sigma_a^{(2)} \to \Sigma_a^{(1)}, \tag{43}$$

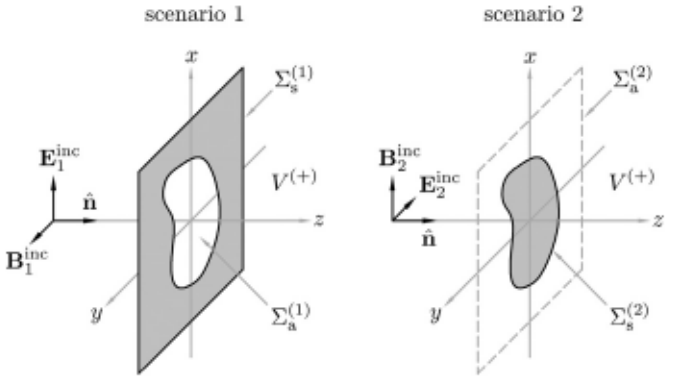

Fig. 5. Babinet's principle for EM waves. The concept here is identical to that in Fig. 2 except that the vector nature of the waves is taken into account. To do so, the polarization of the incident wave is rotated 90° as shown between scenario 1 and 2. Note that this rotation means that the vector-based treatment of Babinet's principle is different from that for the scalar case in Sec. 2 in that the illumination source is different between the two scenarios with regard to its polarization. As such, the induced surface current is also rotated, e.g., Eq. (15) below [9].

$$\mathbf{B}_2^{\text{rad}} \to -\frac{k}{\omega}\mathbf{E}_1 \text{ for } \Sigma_s^{(2)} \to \Sigma_s^{(1)}. \tag{44}$$

For example, if one knows the radiated field for scenario 2, then the *total* field for scenario 1 is given by Eqs. (43) and (44). Laboratory results demonstrating this relationship are given in [35] .One should note, however, that this specific relation only holds for points in $V^{(+)}$ although it is possible to find analogs in $V^{(-)}$ [9].

Now, let the surface current induced on $\Sigma_s^{(1)}$ be $\mathbf{K}_1$ and that on the complimentary screen $\Sigma_s^{(2)}$ be $\mathbf{K}_2$. Then, the radiated fields in each scenario are given by the integral term in Eq. (26). For scenario 2, that is

$$\mathbf{E}_2^{\text{rad}}(\mathbf{r}) = i\omega\mu_o \iint_{\Sigma_s^{(2)}} \ddot{\mathbf{G}}_e(\mathbf{r}, \mathbf{r}') \cdot \mathbf{K}_2(\mathbf{r}')\, d\sigma'. \tag{45}$$

Babinet's principle connects this radiation to scenario 1 via Eq. (43) as

$$\mathbf{B}_1(\mathbf{r}) = ik\mu_o \iint_{\Sigma_s^{(2)}} \ddot{\mathbf{G}}_e(\mathbf{r}, \mathbf{r}') \cdot \mathbf{K}_2(\mathbf{r}')\, d\sigma'. \tag{46}$$

Using Eq. (4) for points *not* on the screen such that no source term is present, Eq. (46) can be re-expressed to yield the total electric field for scenario 1 in terms of $\mathbf{K}_2$ on the complimentary screen as

$$\mathbf{E}_1(\mathbf{r}) = -\mu_o c \iint_{\Sigma_s^{(2)}} \ddot{\mathbf{G}}_m(\mathbf{r}, \mathbf{r}') \cdot \mathbf{K}_2(\mathbf{r}')\, d\sigma', \tag{47}$$

where Eqs. (12) and (13) have also been used. Simplification is achieved via Eq. (11) to show that

$$\ddot{\mathbf{G}}_m(\mathbf{r}, \mathbf{r}) = \frac{1}{4\pi}\left(\nabla \times \ddot{\mathbf{I}}\right)\frac{e^{ik\mathfrak{r}}}{\mathfrak{r}}. \tag{48}$$

With the relations

$$\nabla \times \ddot{\mathbf{I}} f(\mathbf{r}, \mathbf{r}') = \nabla f(\mathbf{r}, \mathbf{r}') \times \ddot{\mathbf{I}} \text{ and } \nabla \frac{e^{ikr}}{r} = \left(ik - \frac{1}{r}\right)\frac{e^{ikr}}{r}\hat{\mathbf{r}},$$

where f is any well-behaved scalar function, and using the far-field limit [4],

$$\frac{1}{\mathfrak{r}} \to \frac{1}{r} \text{ and } e^{ik\mathfrak{r}} \to e^{ik(r - \hat{\mathbf{r}}\cdot\mathbf{r}')},$$

Eq. (48) becomes

$$\ddot{\mathbf{G}}_m(\mathbf{r}, \mathbf{r}') \simeq \frac{ik}{4\pi}\frac{e^{ikr}}{r}\left(\hat{\mathbf{r}} \times \ddot{\mathbf{I}}\right)e^{-ik\hat{\mathbf{r}}\cdot\mathbf{r}'}. \tag{49}$$

Combining Eqs. (47) and (49) yields,

$$\mathbf{E}_1(\mathbf{r}) = -\frac{i\omega\mu_0}{4\pi}\frac{e^{ikr}}{r}\iint_{\Sigma_s^{(2)}}\left(\hat{\mathbf{r}}\times\ddot{\mathbf{I}}\right)\cdot\mathbf{K}_2(\mathbf{r}')\,e^{-ik\hat{\mathbf{r}}\cdot\mathbf{r}'}da', \quad \mathbf{r}\in V^{(+)}\ \text{in far-field}. \tag{50}$$

Lastly, employ the electrodynamic equivalent of the Kirchhoff approximation. This means that the surface current on the complimentary screen $\Sigma_s^{(2)}$is approximated by Eq. (32) with the recognition that the direction of the current is now along the negative y-axis via Fig. 5, i.e.,

$$\mathbf{K}_2(\mathbf{r}) = -\frac{2}{\mu_0}\frac{k}{\omega}E_0\hat{\mathbf{y}},\ \ \mathbf{r}\in\Sigma_s^{(2)}. \tag{51}$$

Noting that $(\hat{\mathbf{r}}\times\ddot{\mathbf{I}})\cdot\hat{\mathbf{y}} = \hat{\mathbf{r}}\times\hat{\mathbf{y}}$, Eq. (50) via Eq. (51) becomes:

$$\mathbf{E}_1(\mathbf{r}) = \frac{ik}{2\pi}E_0\frac{e^{ikr}}{r}\left(\hat{\mathbf{r}}\times\hat{\mathbf{y}}\right)\iint_{\Sigma_s^{(2)}}e^{-ik\hat{\mathbf{r}}\cdot\mathbf{r}'}da', \quad \mathbf{r}\in V^{(+)}. \tag{52}$$

Returning to Eq. (46), using Eqs. (33) and (51), and implementing the far-field limit in $\ddot{\mathbf{G}}_e$ yields the expression for the magnetic field

$$\mathbf{B}_1(\mathbf{r}) = -\frac{ik^2}{2\pi\omega}E_0\frac{e^{ikr}}{r}\left[\hat{\mathbf{y}}-(\hat{\mathbf{r}}\cdot\hat{\mathbf{y}})\hat{\mathbf{r}}\right]\iint_{\Sigma_s^{(2)}}e^{-ik\hat{\mathbf{r}}\cdot\mathbf{r}'}da', \quad \mathbf{r}\in V^{(+)}\ \text{in far-field}, \tag{53}$$

where the identity $(\ddot{\mathbf{I}}-\hat{\mathbf{r}}\otimes\hat{\mathbf{r}})\cdot\hat{\mathbf{y}} = \hat{\mathbf{y}}-(\hat{\mathbf{r}}\cdot\hat{\mathbf{y}})\hat{\mathbf{r}}$ is used. Before proceeding, notice that the integrals in Eqs. (52) and (53) are nearly Fourier transforms of the complimentary screen $\Sigma_s^{(2)}$. They can be cast into the familiar form by defining an aperture function

$$A(\mathbf{r}) = \begin{cases}1, & \mathbf{r}\in\Sigma_s^{(2)},\\ 0, & \mathbf{r}\notin\Sigma_s^{(2)}.\end{cases} \tag{54}$$

With the recognition that $\Sigma_s^{(2)}$ resides in the x-y plane, the integrand can be multiplied by $1 = e^{ik\hat{\mathbf{z}}\cdot\mathbf{r}'}$, since $\hat{\mathbf{z}}\cdot\mathbf{r}' = 0$ on $\Sigma_s^{(2)}$, and introducing the scattering wave vector $\mathbf{q} = k(\hat{\mathbf{z}}-\hat{\mathbf{r}})$, one finds that:

$$\iint_{\Sigma_s^{(2)}}e^{-ik\hat{\mathbf{r}}\cdot\mathbf{r}'}da' = \iint_{\Sigma_s^{(2)}}e^{ik(\hat{\mathbf{z}}-\hat{\mathbf{r}})\cdot\mathbf{r}'}da' = \iint_{-\infty}^{\infty}A(\mathbf{r}')e^{ik\mathbf{q}\cdot\mathbf{r}'}da'.$$

With this, the total fields can be used to calculate the electromagnetic energy flow from the time-averaged Poynting vector as $\langle\mathbf{S}\rangle_t = (1/2\mu_0)\mathrm{Re}\{\mathbf{E}_1\times\mathbf{B}_1^*\}$, where the asterisk denotes complex conjugation. The radiated intensity then follows from $I = \langle\mathbf{S}_t\rangle_t$, which is most readily expressed in spherical coordinates (r, θ, ϖ). Finally, if I is evaluated on a large hemispherical surface of radius $r = d$ (for $z > 0$) where r is large enough that the small angle approximations for $\sin\theta$ and $\cos\theta$ may be used, the normalized radiated intensity becomes

$$\frac{I(\theta,\phi)}{I_0} = \left|\iint_{-\infty}^{\infty}A(\mathbf{r}')e^{ik\mathbf{q}\cdot\mathbf{r}'}da'\right|^2. \tag{55}$$

where $(\hat{\mathbf{r}}\times\hat{\mathbf{y}})\times[(\hat{\mathbf{r}}\cdot\hat{\mathbf{y}})\hat{\mathbf{r}}-\hat{\mathbf{y}}] = [1-(\hat{\mathbf{r}}\cdot\hat{\mathbf{y}})^2]\hat{\mathbf{r}}$ is used and the small angle approximations are equivalent to assuming that the hemispherical surface approximately coincides with the observation screen σ in Section 2.

Eq. (55) is the classic result that the diffraction pattern for $\Sigma_a^{(1)}$ in the far-field zone is the absolute square of the Fourier transform of the (open) *aperture*, i.e., Eq. (1). Yet, notice the significance of this result. Eq. (55) derives from a formal solution to the Maxwell equations. The duality of the Maxwell equations allows formulation of the fields radiated by the induced current on a *screen* in terms of the current induced on its complimentary screen via Babinet's principle. The Kirchhoff and far-field approximations then lead to the Fourier transform property of the end result without the need to introduce fictitious Huygens point sources. In other words, *when employing Huygens point sources in the Huygens-Fresnel principle, one is tacitly using both Babinet's principle and the Kirchhoff approximation*. Lastly, note that the treatment above is not the only way to formulate an EM analog of the Huygens-Fresnel principle, e.g., see [4,27,28]. The general conclusions, however, remain the same.

Before proceeding a subtle point in this treatment should be addressed. When the aperture is cut in the sheet Σ, the induced current must, obviously, vanish in the free-space region of the resulting aperture. This means that the current on the screen is discontinuous upon crossing the aperture rim, which is defined by a contour, C. Yet Eqs. (45) and (46) derive from Green's second vector identity, Eq. (15), which is a generalization of the divergence theorem. Thus, Eqs. (45) and (46) appear to be invalid due to the discontinuity of $\mathbf{K}$ across C [36]. Moreover, this concern extends to the fields as well via Eqs. (24) and (25). Stratton and Chu recognize this issue and propose a resolution by adding line integrals of the fields along C to variants of Eq. (22) [36]. Physically, the idea is that in order for the current to discontinuously vanish at the aperture rim, there must be an accumulation of charge density there. For example, in the case of the slit aperture in Section 2, this charge density looks like the slit is a strip of electric dipole moments oscillating in antiphase with the incident electric field. Eventually, one finds that the behavior of the fields across C is such that the line-integral corrections are not needed *if* the exact (correct) values of the fields are used [37]. That is, using Eq. (32), which is tantamount to the Kirchhoff approximation, means that Eqs. (45) and (46) are only approximate descriptions for the radiated fields and the line-integral corrections should apply. However, as will be mentioned in Section 4, there now exist methods to find the (numerically) exact fields in the aperture, and thus Eqs. (45) and (46) are valid in practice.

4. Ewald–Oseen theorem

The preceding section highlights two important concepts. First, diffraction can be viewed as secondary radiation from sources of current induced on screens. Second, the special case of a complete sheet shows that its shadow arises from this same radiation, where it both perfectly cancels the incident wave in the shaded region and produces the reflected wave in the illuminated region. Thus, *the shadow is a consequence of interference rather than a mechanical-like blocking of the incident light* as may be implied by the Huygens-Fresnel principle. In fact, this concept is general and applies to transparent or opaque 3D objects in addition to the 2D screens above. The aim now is to show how and examine several implications.

To begin, replace the disk in Section 3.1 by a single 3D particle of arbitrary size, shape, and uniform refractive index m, which may correspond to *either* absorbing or non-absorbing materials. For simplicity, assume that the particle is nonmagnetic with $\mu = \mu_0$. Again, a source $\mathbf{J}^{\mathrm{inc}}$ is present far from the particle to produce an incident wave. The exact same treatment in Section 3.1 may be applied to this situation, culminating in Eq. (22). First consider locations $\mathbf{r}$ that are *inside* the particle volume V_{int},

$$\oint_S\left\{i\omega\ddot{\mathbf{G}}_e(\mathbf{r},\mathbf{r}')\cdot\left[\hat{\mathbf{n}}'\times\mathbf{B}(\mathbf{r}')\right]-\ddot{\mathbf{G}}_m(\mathbf{r},\mathbf{r}')\cdot\left[\hat{\mathbf{n}}'\times\mathbf{E}(\mathbf{r}')\right]\right\}da' = -\mathbf{E}^{\mathrm{inc}}(\mathbf{r}),\quad \mathbf{r}\in V_{\mathrm{int}}. \tag{56}$$

In view of Eq. (56), the integral over S can be interpreted as radiation emitted from the particle surface in analogy to that from the sheet in Eq. (26). Yet, one must not take this literally as the actual source of radiation is less obvious than $\mathbf{K}$ above; more will

be said on this point below. Eq. (56) also reveals that this radiation is an exact copy of the incident field but 180° out of phase, i.e., such that the incident field is extinguished everywhere inside the particle. This result is known as the Ewald-Oseen (EO) extinction theorem [27]. More precisely, it is a particle-surface formulation of the EO theorem. Notice that Eq. (56) does not rely on an assumed value for m, *thus the cancellation of the incident wave occurs for absorbing and non-absorbing particles alike*. It follows that the theorem should also apply to any wavelength, including X-rays, and evidence for this is given in [38].

Outside the particle this same integral yields the particle's scattered wave,

$$\mathbf{E}^{\mathrm{sca}}(\mathbf{r}) = \oint_S \left\{ i\omega \ddot{\mathbf{G}}_{\mathrm{e}}(\mathbf{r},\mathbf{r}') \cdot \left[\hat{\mathbf{n}}' \times \mathbf{B}(\mathbf{r}')\right] - \ddot{\mathbf{G}}_{\mathrm{m}}(\mathbf{r},\mathbf{r}') \cdot \left[\hat{\mathbf{n}}' \times \mathbf{E}(\mathbf{r}')\right] \right\} \mathrm{d}a', \quad \mathbf{r} \in V_{\mathrm{ext}}. \tag{57}$$

Both Eqs. (56) and (57) are exact relationships with much physical meaning. Recall from Section 3.1 that the appearance of the surface integral here is essentially a consequence of applying the divergence theorem to the fields in V_{ext}. Thus, in this case where S bounds V_{int}, the surface integral is a "short-hand" representation of a volume effect in the same way that Gauss's law in electrostatics equates a surface expression involving the normal components of the field to a volume quantity, the enclosed charge density. While Eq. (56) shows that the incident field is canceled inside the particle, it is important to realize that does not mean the total field in the particle is zero. Because Eq. (56) results from integrating throughout V_{ext}, any source of radiation in V_{int} is not explicitly captured in the treatment, and one will see below that the particle's (material) polarization fills this conceptual void.

Pursuing the analogy with Gauss's law, the source responsible for the radiation expressed in Eqs. (56) and (57) should appear in the volume integral representation of these equations. One way to see this is to recast Eq. (5) to build-in an *ad hoc* current density $\mathbf{J}^{\mathrm{int}}$ that is nonzero only in V_{int} and will be seen later to relate to the polarization of the particle's material [39]. Define

$$\mathbf{J}^{\mathrm{int}}(\mathbf{r}) = -i\omega\epsilon_0\left(m^2-1\right)\mathbf{E}(\mathbf{r}) \quad \mathbf{r} \in V_{\mathrm{int}}, \tag{58}$$

where $\mathbf{E}$ is the total (macroscopic) electric field inside the particle. Then the Maxwell equations for the electric field in either region can be written as

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r}) - k^2\mathbf{E}(\mathbf{r}) = \begin{cases} i\omega\mu_0\mathbf{J}^{\mathrm{int}}(\mathbf{r}) & \mathbf{r} \in V_{\mathrm{int}} \\ i\omega\mu_0\mathbf{J}^{\mathrm{inc}}(\mathbf{r}) & \mathbf{r} \in V_{\mathrm{ext}}. \end{cases} \tag{59}$$

Note that k here is the vacuum wave number and while $\mathbf{J}^{\mathrm{inc}}$ is an external specified current, $\mathbf{J}^{\mathrm{int}}$ relates to the internal field, and thus, ultimately depends on $\mathbf{J}^{\mathrm{inc}}$. Now Eq. (59) has the correct form to be solved via the method of dyadic Green functions as defined by Eq. (9). Apply Greens second identity, Eq. (15)–(59) for the volume V_{int}. If the vector $\mathbf{r}'$ is restricted to V_{ext}, then the delta function makes no contribution and after implementing the symmetry properties following Eq. (17), the result is

$$\int_{V_{\mathrm{int}}} \ddot{\mathbf{G}}_{\mathrm{e}}(\mathbf{r},\mathbf{r}') \cdot \mathbf{J}^{\mathrm{int}}(\mathbf{r})\, \mathrm{d}v = \oint_S \left\{ i\omega \ddot{\mathbf{G}}_{\mathrm{e}}(\mathbf{r},\mathbf{r}') \cdot \left[\hat{\mathbf{n}} \times \mathbf{B}(\mathbf{r})\right] + \ddot{\mathbf{G}}_{\mathrm{m}}(\mathbf{r},\mathbf{r}') \cdot \left[\hat{\mathbf{n}} \times \mathbf{E}(\mathbf{r})\right] \right\} \mathrm{d}a, \quad \mathbf{r}' \in V_{\mathrm{ext}} \tag{60}$$

where the pilot vector $\mathbf{a}$ has been canceled as in Eq. (18). Next, with Eq. (58), interchanging the coordinates, $\mathbf{r} \to \mathbf{r}'$ and $\mathbf{r}' \to \mathbf{r}$, and again using the symmetry relations of the Green functions where $\ddot{\mathbf{G}}_{\mathrm{e}}(\mathbf{r}',\mathbf{r}) = \ddot{\mathbf{G}}_{\mathrm{e}}(\mathbf{r},\mathbf{r}')$ and $\ddot{\mathbf{G}}_{\mathrm{m}}(\mathbf{r}',\mathbf{r}) = -\ddot{\mathbf{G}}_{\mathrm{m}}(\mathbf{r},\mathbf{r}')$, Eq. (60) yields

$$k^2\left(m^2-1\right)\int_{V_{\mathrm{int}}} \ddot{\mathbf{G}}_{\mathrm{e}}(\mathbf{r},\mathbf{r}') \cdot \mathbf{E}(\mathbf{r}')\, \mathrm{d}v' = \oint_S \left\{ i\omega \ddot{\mathbf{G}}_{\mathrm{e}}(\mathbf{r},\mathbf{r}') \cdot \left[\hat{\mathbf{n}}' \times \mathbf{B}(\mathbf{r}')\right] - \ddot{\mathbf{G}}_{\mathrm{m}}(\mathbf{r},\mathbf{r}') \cdot \left[\hat{\mathbf{n}}' \times \mathbf{E}(\mathbf{r}')\right] \right\} \mathrm{d}a' \quad \mathbf{r} \in V_{\mathrm{ext}}. \tag{61}$$

The key step now is to notice that the surface integral in Eq. (61) is the same as in Eq. (57), showing that

$$\mathbf{E}^{\mathrm{sca}}(\mathbf{r}) = k^2\left(m^2-1\right)\int_{V_{\mathrm{int}}} \ddot{\mathbf{G}}_{\mathrm{e}}(\mathbf{r},\mathbf{r}') \cdot \mathbf{E}(\mathbf{r}')\, \mathrm{d}v', \quad \mathbf{r} \in V_{\mathrm{ext}}. \tag{62}$$

Eq. (62) is equivalent to the surface integral expression of Eq. (57). Aside from a constant prefactor, the internal field $\mathbf{E}$ is viewed as a radiating source in analogy to $\mathbf{K}$ in Section 3.2. In fact, this interpretation can be made explicit by relating the field to the particle's polarization $\mathbf{P}$ as $\mathbf{P} = \varepsilon_0(m^2-1)\mathbf{E}$. Dividing the volume integral into a collection of differential elements $\mathrm{d}v'$ of polarization, the scattered field can then be viewed as collective radiation from electric dipole moments $\mathbf{p}$ each related to the internal field within $\mathrm{d}v'$.

If Green's second identity is next applied to V_{ext}, the source is now $\mathbf{J}^{\mathrm{inc}}$ and the result is exactly the same as Eq. (22). Exchanging the resulting surface integral by Eq. (62), the complete solution to Eq. (59) for $\mathbf{r} \in V_{\mathrm{ext}}$ is obtained as all space is now integrated, i.e., V_{int} and V_{ext}, giving

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}^{\mathrm{inc}}(\mathbf{r}) + \frac{k^2}{\epsilon_0}\int_{V_{\mathrm{int}}} \ddot{\mathbf{G}}_{\mathrm{e}}(\mathbf{r},\mathbf{r}') \cdot \mathbf{P}(\mathbf{r}')\, \mathrm{d}v', \quad \mathbf{r} \in V_{\mathrm{ext}}, \tag{63}$$

This is one form of the volume integral equation (VIE) and is valid for locations *outside* of the particle. It shows that the *total* field is the superposition of the incident and scattered fields, the latter of which is due to radiation from the polarization of the particle material.

The question now is what is the volume-integral equivalent of Eq. (57), which would be the volume analog to the EO theorem. One may think that all that is required is to allow $\mathbf{r}'$ [the field point in the derivation leading to Eq. (60)] to reside in V_{int}. Remember that $\mathbf{r}'$ and $\mathbf{r}$ are not exchanged until after Eq. (60). However, because the volume integral in Eq. (60) runs over V_{int} in the coordinate $\mathbf{r}$, it will include the point $\mathbf{r}' = \mathbf{r}$ where the Green function becomes singular, recall Eq. (11). In that case, Green's second identity, Eq. (15), cannot be applied. This is resolved by excluding the singular point by removing a small volume V_δ, called the principle volume, at $\mathbf{r}' = \mathbf{r}$ and then taking the limit that this volume vanishes. After exchanging $\mathbf{r}'$ and $\mathbf{r}$, Eq. (63) for points ($\mathbf{r}$) inside the particle becomes [39–41]

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}^{\mathrm{inc}}(\mathbf{r}) + \frac{k^2}{\epsilon_0}\lim_{V_\delta \to 0}\int_{V_{\mathrm{int}}-V_\delta} \ddot{\mathbf{G}}_{\mathrm{e}}(\mathbf{r},\mathbf{r}') \cdot \mathbf{P}(\mathbf{r}')\, \mathrm{d}v' + \frac{\ddot{\mathbf{L}}(\mathbf{r}) \cdot \mathbf{J}^{\mathrm{int}}(\mathbf{r})}{i\omega\epsilon_0}, \quad \mathbf{r} \in V_{\mathrm{int}}. \tag{64}$$

That is, the field inside the particle is *still* given by the VIE except with the principle volume removed and the addition of a correction term to account for this missing contribution. The correction term involves the so-called self-interaction dyadic [40]

$$\ddot{\mathbf{L}}(\mathbf{r}) = \frac{1}{4\pi}\oint_{S_\delta} \frac{\hat{\mathbf{n}}'\hat{\boldsymbol{\tau}}}{\tau^2}\, \mathrm{d}a', \tag{65}$$

where $\boldsymbol{\tau} = \mathbf{r} - \mathbf{r}'$ [recall Eq. (33)] and S_δ is the surface bounding V_δ with outward normal $\hat{\mathbf{n}}'$. Note that Eq. (64) requires yet another dyadic term if the limit in Eq. (64) is not taken and V_δ remians finite; this happens when one employs Eq. (64) to solve EM scattering problems in practice [39-41].

If S_δ is taken to be a sphere of radius R_δ, the self-interaction dyadic is $\ddot{\mathbf{L}} = \ddot{\mathbf{I}}/3$ [40]. Moreover, the current appearing in the cor-

rection term can be expressed as a time-derivative of the polarization within V_δ, $\mathbf{J}^{\text{int}} = -i\omega\mathbf{P}$. Then, Eq. (64) becomes

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}^{\text{inc}}(\mathbf{r}) + \frac{k^2}{\epsilon_0}\lim_{R_\delta\to 0}\int_{V_{\text{int}}-V_\delta}\ddot{\mathbf{G}}_e(\mathbf{r},\mathbf{r}')\cdot\mathbf{P}(\mathbf{r}')\,d\nu' - \frac{1}{3\epsilon_0}\mathbf{P}(\mathbf{r}), \quad \mathbf{r}\in V_{\text{int}} \tag{66}$$

With these parameters specified, the terms in Eq. (66) can be interpreted. Since the limit $R_\delta \to 0$ means that the principle volume is much smaller than the refracted wavelength $\lambda/\text{Re}\{m\}$, the polarization within this volume may be assumed uniform. Thus, the last term describes the field at the center of a uniformly polarized sphere, representing the self-field in the microscopic description of dielectrics [40]. The sum of the first two terms on the right-hand side of Eq. (66) is then the Lorentz, or local field, i.e., the field at $\mathbf{r}$ due to the incident field and that radiated from all the volume elements throughout the particle other than the principle volume.

A utility of Eq. (66) is that one can use it to solve for the unknown internal field $\mathbf{E}$. In short, the particle volume is discretized into volume elements much smaller than $\lambda/\text{Re}\{m\}$, where the polarization within each element can reasonably be replaced with a single, point electric dipole moment $\mathbf{p}$ [41]. By assuming a linear relationship between $\mathbf{p}$ and the so-called exciting field $\mathbf{E}^{\text{exc}} = \mathbf{E} + \mathbf{P}/3\varepsilon_0$ through a polarizability factor α, Eq. (66) is recast as a system of *linear* algebraic equations for the unknown dipole moments. Solving the system yields the dipole moments, and thus, the field $\mathbf{E}$. This is known as the discrete dipole approximation (DDA) [39,41].

In the context of the DDA, Eq. (66) can be understood in an intuitive way by imagining the interaction between the dipoles in a *contrived* temporal context. First consider the dipoles, which are distributed throughout V_{int} on a lattice, in the case when the incident plane wave is not present, i.e., $\mathbf{J}^{\text{inc}} = 0$. In this case, the moments have zero magnitude because there is no electric field to polarize them. Now suppose that $\mathbf{J}^{\text{inc}}$ is "turned on." Each dipole becomes polarized and oscillates with the incident wave. A given dipole then radiates a secondary wave that propagates through the vacuum filling the lattice and influences the polarization of the other dipoles. The secondary radiation from these other dipoles then acts back on the first dipole, and so on, *ad infinitum*. This interaction, enabled by the secondary waves, couples the moments. Eventually, a "steady state" is reached, where the dipoles' magnitude, direction, and phase stops evolving under this coupling action and the dipoles collectively attain a state of common time-harmonic oscillation. The end result is that each dipole behaves (aside from a constant prefactor) as if it is driven by the internal electric field $\mathbf{E}$ that would otherwise be found by solving the Maxwell equations as a boundary value problem (BVP), e.g., Mie theory for a spherical particle. From this point of view, it is the coupling that causes the phenomenon one associates with refraction.

Notice that this picture is a different description of EM scattering than what one may see from the BVP point of view. The concept of a refractive index is in a sense absent here, replaced by the requirement that the coupling between the dipoles results in a self-consistent solution for the dipole-dipole interactions. In the BVP approach, one simply scales the wavenumber in the particle by the refractive index, i.e., $k_{\text{int}} = mk$, and the fields in V_{int} and V_{ext} are expanded over a complete set of vector spherical wave functions [8]. At that point, it is only a matter of imposing the boundary conditions at S to relate the expansion coefficients and have a complete solution. It is remarkable then, that the complicated coupling between the dipoles results in the same fields and associated phenomena of reflection, refraction, and scattering, as are also provided in the BVP approach [42,43].

The strength of the coupling between the dipoles depends, in part on the polarizability α, and on the angular structure of the dipole fields. The polarizability does depend on m, and overall, the dipoles are more strongly coupled as the real part of m increases. If the real part of m is close to one, α is small and the coupling between the dipoles is weak. This means that the incident wave will dominate the polarization of the particle, resulting in a situation where $\mathbf{E}$ deviates little from $\mathbf{E}^{\text{inc}}$. This weak-refraction limit is often called the Rayleigh-Debye-Gans (RDG) or first Born-approximation [2]. If the real part of m is increased, the coupling also increases and the dipole moments become further deviated from the RDG expectation.

Elementary texts often describe refraction as the "slowing down" of light as it enters a medium from vacuum [1,32]. The speed of light is reduced to $c/\text{Re}\{m\}$ and the wavelength is reduced to $\lambda/\text{Re}\{m\}$. However, in the context of Eq. (66) where the effects of refraction are caused by dipole-dipole coupling, one could propose that light actually travels inside the particle medium at the same speed as it does through vacuum, c [44]. This is supported by the presence of the vacuum wave number k in the Green function in Eq. (66), which acts as the *free space* propagator for the field. Indeed, this is expected if one views the medium as Democritus does; "...there are only atoms and the void," (and no medium) [45].

The apparent reduction of the wavelength of light upon entering a medium from vacuum can be observed, e.g., from thin-film interference [1]. So, if the waves exchanged between the particle's dipoles travel at c, how may one to account for the observed reduced internal wavelength? The key is to realize that it is the dipole moments, or equivalently the polarization $\mathbf{P}$, that constitute the radiating *source* of the scattered wave, which is the wave that is observed. The distribution of $\mathbf{P}$ has a wavelength less than λ due to the shift in phase, relative to the incident wave, of the dipole oscillations brought about by coupling, which itself depends on m [1]. Yet, the coupling interactions that establish $\mathbf{P}$ are realized by the exchange of dipole waves of wavelength λ and speed c [46]. Note, however, that there is some debate on this interpretation, see [47].

Also, consider the specific region of space V_{int} when the particle is absent and when it is present. When it is absent, only $\mathbf{E}^{\text{inc}}$ occupies V_{int}. When it is present, $\mathbf{E}$ occupies V_{int}. Thus, $\mathbf{E}^{\text{inc}}$ is somehow *replaced* by the internal field $\mathbf{E}$, the latter of which is equivalent to polarization $\mathbf{P} = \varepsilon_0(m^2-1)\mathbf{E}$. Because EM waves require no medium to support their propagation, there is no mechanism for mechanical-like blocking of a wave and the only way this replacement can occur is through interference [44,48–50]. That is, the secondary radiation emitted by $\mathbf{P}$ as described by Eq. (66) must cancel $\mathbf{E}^{\text{inc}}$ throughout V_{int} via destructive interference. While this is not directly revealed in Eq. (66), it *is* seen in Eq. (56) – the EO theorem – where the same $\mathbf{E}$ (and $\mathbf{B}$) leading to $\mathbf{P}$ act as the source of this radiation. Notice though, that it is not just the fields at the particle surface in Eq. (56) that are responsible for the extinction of $\mathbf{E}^{\text{inc}}$ since those surface fields are established from secondary radiation emitted *throughout* V_{int} [44]. The EO extinction of the incident field should be physically regarded as originating from the complete volume of the particle, although the practically useful form of the EO theorem is usually the surface integral, Eq. (56).

5. The role of Ewald–Oseen in scattering and diffraction

Section 4 hints at a connection between several apparently independent effects. The EO-theorem's cancellation of $\mathbf{E}^{\text{inc}}$ within the particle, the appearance of $\mathbf{E}$ in its place displaying the characteristics of refraction, and $\mathbf{E}^{\text{sca}}$ outside of the particle all originate from secondary radiation from an induced current source $\mathbf{J}^{\text{int}}$. Consequently, whether it is light (seemingly) "diffracting" through an

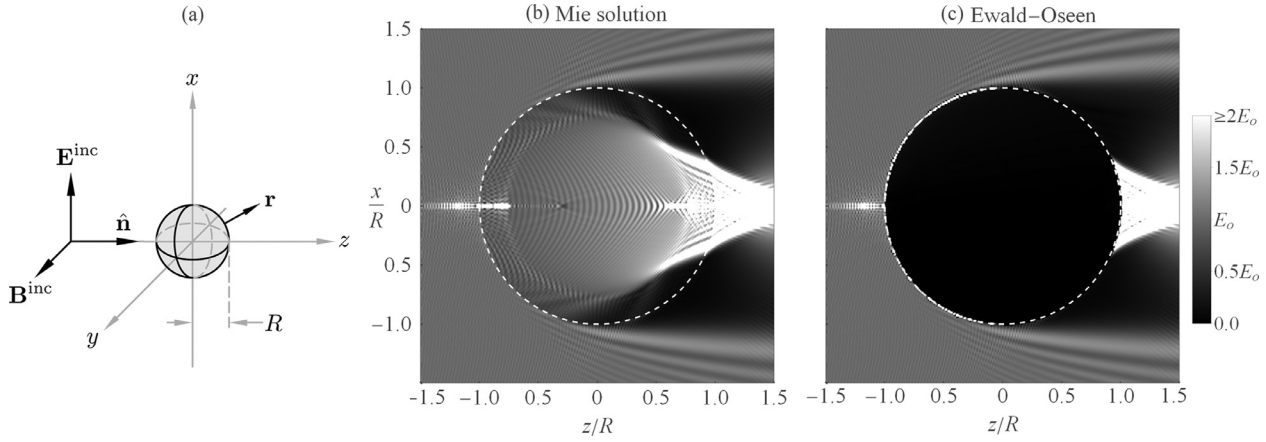

Fig. 6. Demonstration of EO extinction and its connection to scattering for a spherical particle illuminated by a linearly polarized wave as shown in (a). The size parameter of the particle is $kR = 151.5$ and the refractive index is $m = 1.33 + 0i$. Plot (b) shows the total electric-field magnitude $|\mathbf{E}^{inc} + \mathbf{E}^{sca}|$ in the x-z plane through the origin outside the particle and the internal field $|\mathbf{E}|$ inside the particle as calculated from Mie theory. In (c) however, is shown the superposition of the incident wave and the EO extinction wave of Eq. (56), i.e., $|\mathbf{E}^{inc} + \text{Eq. (56)}|$. *The internal field is not shown here* although Eq. (56) is evaluated from it. The dark appearance of the interior region demonstrates that the extinction wave cancels the incident wave in the particle. Outside of the particle, Eq. (56) becomes Eq. (57), and the total-field magnitude is again shown except calculated from Eq. (57) i.e., $|\mathbf{E}^{inc} + \text{Eq. (57)}|$. One can see agreement with this external field and that from Mie theory in (b).

aperture in an opaque screen or "scattering" from a particle, the observed distribution of light is due to, and completely explained by, radiation from one $\mathbf{J}^{int}$ or the other in either case. And, accompanying this radiation is the inherent process of extinguishing the incident wave for all points within the particle.

One can see refraction, diffraction, scattering, and EO extinction all in Fig. 6. Here, Mie theory is used to find the internal fields, $\mathbf{E}$ and $\mathbf{B}$, in a spherical particle with $m = 1.33 + 0i$ and size parameter $kR = 151.5$ where R is the particle radius. The particle is illuminated by a linearly polarized plane wave as shown in Fig. 6(a). Fig. 6(b) shows the magnitude of the *total* electric field $|\mathbf{E}^{inc} + \mathbf{E}^{sca}|$ *outside* the particle in the near-field zone as calculated from Mie theory. Inside the particle, the physical field present is the internal field $\mathbf{E}$ and $|\mathbf{E}|$ is shown. From the internal field, it is possible to demonstrate that radiation from the associated polarization $\mathbf{P}$ extinguishes the incident field inside the particle. This is done in Fig. 6(c) by plotting the superposition of $\mathbf{E}^{inc}$ and the *integral* in Eq. (56) using the Mie-calculated internal fields $\mathbf{E}$ and $\mathbf{B}$ in the integral. The entire interior of the particle appears dark, showing that the incident field is extinguished by the integral in Eq. (56). Remember that this (surface) integral physically represents radiation from the induced polarization throughout the entire particle volume, i.e., the surface dependence in Eq. (56) merely results from use of the divergence theorem. Also realize that the dark particle-interior does not mean that there is no field present in reality; the internal field is present and (with $\mathbf{B}$) leads to this cancellation of $\mathbf{E}^{inc}$ via Eq. (56). Note that this EO extinction is also fundamentally connected to the extinction cross section of the particle but that is beyond the scope of this discussion, see [51].

Outside of the particle in Fig. 6(c), the total electric-field magnitude $|\mathbf{E}^{inc} + \mathbf{E}^{sca}|$ is shown, except now $\mathbf{E}^{sca}$ is calculated by Eq. (57), which is what Eq. (56) becomes for locations outside of the particle. There is exact agreement between the external field magnitude here and the Mie result in Fig. 6(b). Looking more closely at Fig. 6(c), one can see that the particle's dark interior "leaks out" of the particle parallel to the z-axis on either side of the high magnitude focused feature. What this is communicating is that the scattered field in this external region has much the same character of the extinction wave inside the particle and is what may be called the near-field shadow for this *transparent* particle.

Comparing Eqs. (56) and (57) shows that exactly the same integral appears in each, demonstrating that it is radiation from the *same* source, the internal field (or more appropriately its polarization), that establishes the scattered field *and* extinguishes the incident field inside the particle. Of course, this does not mean there is no field inside the particle, just that the incident field has been canceled. Indeed, the refractive index in this example corresponds to a strongly refractive transparent particle for visible light and one can clearly see the expected characteristics of refraction in Fig. 6(b).

It is now possible to illustrate the fundamental equivalence of diffraction and scattering. Notice from Fig. 6(c) that the end effect of the EO theorem is to extinguish a portion of the incident wavefront across the area of the geometric cross section $C^{geo} = \pi R^2$ of the particle as seen along $\hat{\mathbf{n}}$. With respect to this extinction area, one has exactly the same situation that would occur for an opaque 2D screen with the same size and shape as C^{geo}. In this analogy, the screen would be a disk of radius R. As demonstrated by Eqs. (26) and (57), exactly the same source is responsible for extinguishing the incident wave and generating the scattered wave for the screen [Eq. (26)] or the particle [Eq. (57)]. From the recognition in Sections 3.1 and 3.2 that the finite size of the source producing the extinction wave means that the incident wave's cancellation is not complete over *all* space, the spread of light across the observation screen σ for small θ will be the same for a spherical particle or its corresponding screen. This is because the induced source for each of these objects extends over the same cross-sectional area C^{geo}. Moreover, via Babinet's principle, this spread will also be the same as a circular aperture in an infinite screen. There will, however, be differences between the angular distribution of scattered light for larger θ for a sphere and the disk-shaped screen (or its complimentary aperture) due to the fact that the sphere has finite extent along the z-axis while the screen does not [15]. This can be understood from the z-dependence present in Eq. (57) that is absent in Eq. (26).

6. How diffraction works in electrodynamics

In the title of this section the term "diffraction" follows the common usage to mean a spreading wave scattered from an object to create a diffraction pattern. Now recall that earlier, diffraction was defined as simply the spreading of a propagating wave. However, the common usage has value in its brevity for the complex scattering mechanism, i.e., re-radiation, that forms a diffraction pattern. Yet, the goal here is to clearly describe the relationship between diffraction, both as defined and as used, with scat-

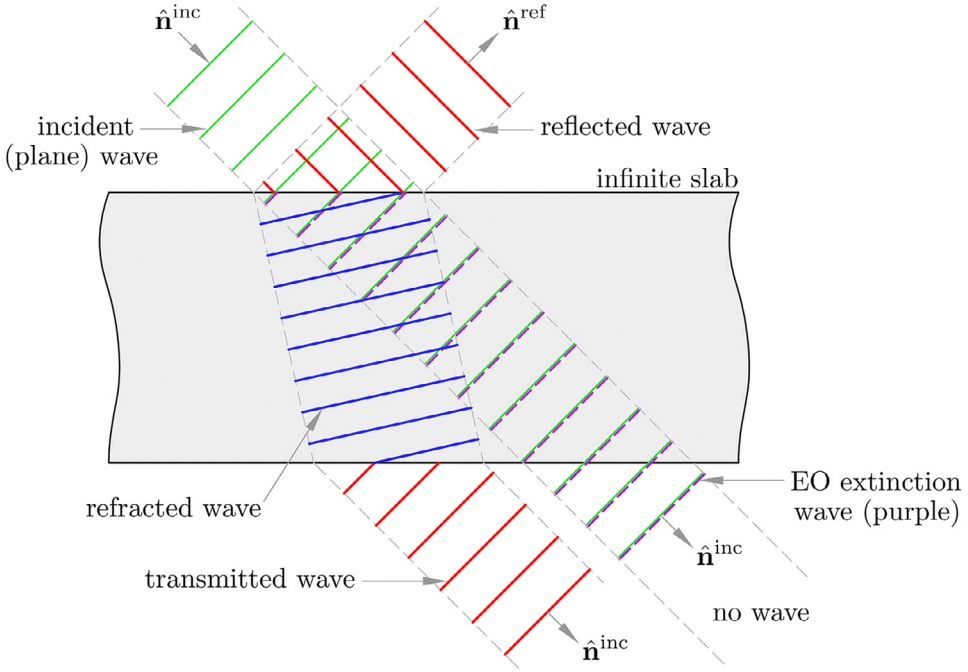

Fig. 7. A wave (green) incident upon a semi-infinite slab of glass. To aid discussion, the incident wave is imagined to have a finite width much larger than the wavelength such that it can be thought of qualitatively as a plane wave. This wave causes polarization in the glass (blue) which radiates an EO extinction wave (dashed purple) 180° out of phase with the incident wave both inside the slab and after where both propagate out the exit side. The interior polarization wave (blue) also radiates the exterior reflected and transmitted waves, which are scattered waves (red). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

tering. The exposition above demonstrates that diffraction patterns are generated by scattering and they are the intensity distribution of the diffracting (spreading) EO extinction wave. With this sound mathematical foundation, a heuristic explanation of "diffraction" is now possible.

This explanation begins by examining the EO extinction theorem in a familiar example. Fig. 7 shows a semi-infinite slab of glass illuminated by a wave with finite width much greater than the wavelength such that it can be approximated by a plane wave. From the EO point-of-view this wave passes unperturbed into the glass, through it, and back out the exit side. However, when it is in the slab, it induces the EO extinction wave (through polarization of the glass) that propagates along with the incident wave but is 180° out of phase. Thus, the well-known EO extinction occurs within the slab via destructive interference between these two waves. Note, however, that these waves continue to propagate together out of the slab on the exit side as drawn such that the destructive interference continues as well. One might say that on the exit side of the slab, the EO wave destructively interferes with the incident wave to create a shadow. Furthermore, because these two waves have the same width, they diffract, i.e., spread, the exact same way and the complete destructive interference continues into the far field.

The polarization of the glass induced by the incident wave propagates as to constitute the interior refracted wave as drawn. This polarization radiates a reflected wave at the entrance side of the slab and an exterior transmitted wave on the exit side. One might say that these two waves outside the slab are the waves that are scattered by it.

Now consider the same plane wave incident on a glass sphere with diameter $2R$ much less than the width of the wave $2w_o$ as drawn in Fig. 8. From Section 4, there are several components to the *total* wave present either inside or outside the particle, which are tagged according to the source producing them. Specifically, outside the particle, there is the incident wave shown in green and the scattered wave shown in red. Inside the particle, there is again the incident wave (green), and two additional waves: the EO extinction wave (dashed purple), and the internal, i.e., polarization, wave (blue). Notice that all of these have complements in Fig. 7 for the glass slab.

From the EO point of view the incident wave passes into the sphere and continues forward out of the sphere unperturbed. Yet, when inside the sphere it induces the EO extinction wave that propagates along with it, but 180° out of phase. Thus, EO extinction occurs within the sphere via destructive interference between these two waves. As with the glass slab, these waves continue to propagate together out of the sphere along the forward direction as drawn in the figure, and the destructive interference continues as well. Thus, immediately behind the sphere, this interference creates a shadow equal in extent to the circular geometric cross section of the sphere. Now the situation differs from that of the slab because the sphere diameter is smaller than the width of the incident beam. Given this, the geometric shadow is surrounded by a bright halo of the incident beam. Remarkably, as the two waves continue to propagate into the far field, the narrower EO wave will diffract (spread) more than the incident wave to the point where wave overlap, and hence destructive interference, will be minimal; recall that this is not the situation for the slab. The end result is the diffraction pattern of a circular obstacle of diameter $2R$ on the distant screen σ with a bright central spot, i.e., "Drude's spot" [5]. The diffraction pattern is the intensity of the EO extinction wave in the far field; the Drude spot is the lesser diffracted (lesser spread) remainder of the incident wave.

To use more common terminology, the wider incident beam diffracts less than the narrower EO extinction wave so the complete destructive interference between the two (occurring inside the particle and immediately after it) cannot continue as one proceeds along the propagation axis. The narrower incident beam will form a diffraction pattern consistent with the (perhaps Gaussian) beam profile. Because the diffraction is "narrow" for this

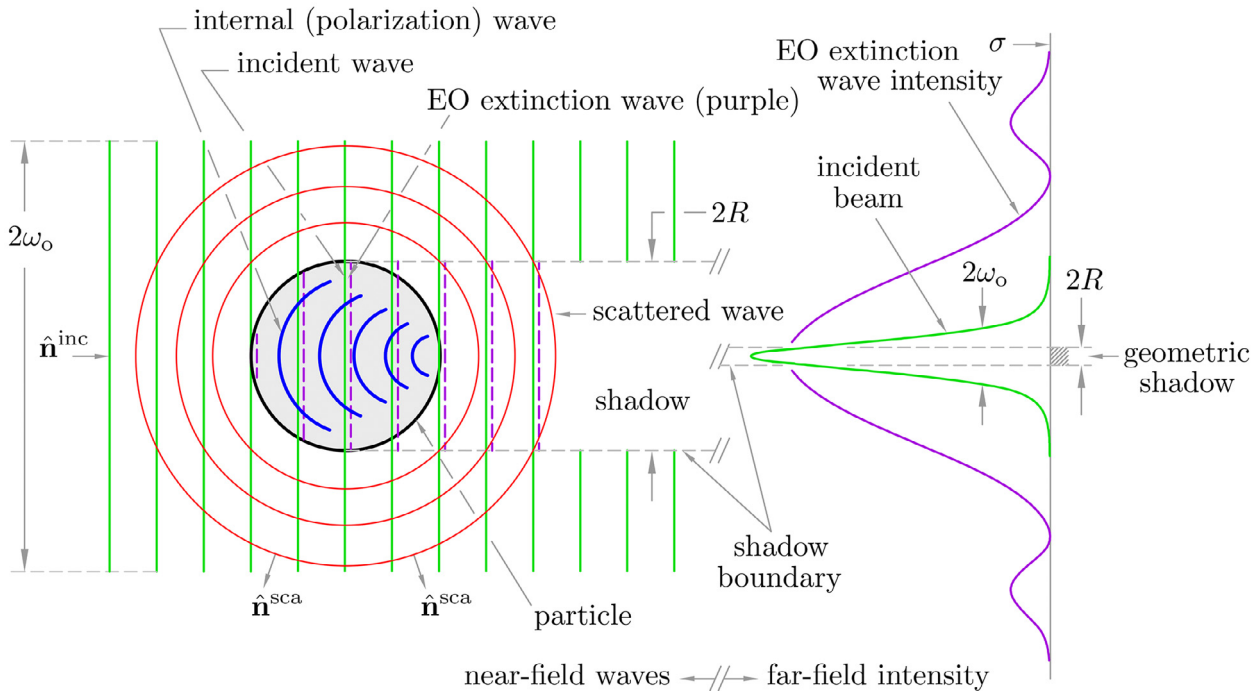

Fig. 8. Diffraction revisited. Shown is a diagram of a beam of light with width $2w_o$ incident upon a spherical particle of radius R. The diagram is split with the left portion showing the particle interior and near-field zone with component *waves*, and the right portion shows the far-field zone *intensities*. The component waves inside and outside of the particle are indicated by the color coding as follows: incident wave green, internal (polarization) wave blue, EO extinction wave dashed purple, and scattered wave (outside) red. Note that the shadow region shown may, or may not, constitute a complete field-free region. If the particle is large (compared to λ) and strongly absorbing, it will, whereas if the particle is smaller and/or partly absorbing, it would not. The intent here is to emphasize the extinction, via destructive interference, of the incident wave by the EO extinction wave inside and immediately behind the particle and to contrast that to the lack of extinction – due to the significantly different amounts of diffraction – between these same two waves in the far field. The far-field intensities on σ show the classic diffraction pattern from the particle, while the remainder of the incident beam is seen as the brighter central peak, i.e., the Drude spot. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

wide beam, the intensity will remain relatively large; this is, again, Drude's spot. The narrower EO extinction wave, which is also the scattered wave when outside the particle, will spread more than the incident wave. Given the circular profile of the particle from which it is radiated, the diffraction pattern will resemble that of a circular aperture. The key point with this description of diffraction by a particle is that light is never blocked. Furthermore, light does not "diffraction *around* the particle" ; it is in fact radiated light from the particle itself. Finally, Drude's spot results from what remains of the incident beam after the EO extinction wave removes a portion of it across C^{geo}.

It is now useful to consider Babinet's principle from this perspective. Fig. 9 shows a screen with a circular aperture. This screen is the complement to a circular disk, hence, via Babinet's principle, it should have approximately the same diffraction pattern in the far field. In Fig. 9 we see the incident wave passing unperturbed through the screen. At the screen, the EO extinction wave is created. Immediately behind the screen, the destructive interference between the incident and EO waves creates an annular shadow with outer diameter equal to the diameter of the incident wave and inner diameter equal to the aperture diameter (see dashed purple in Fig. 9). Immediately behind the aperture, there is no EO wave, and hence, no destructive interference for the incident wave. As the two waves in the annulus continue to propagate into the far field, they diffract (spread) by the same amount, and thus, continue to completely, destructively interfere. On the other hand, the narrower incident wave part that passed through the aperture will diffract (spread) to create a circular aperture diffraction pattern in the far field as drawn. This is the same diffraction pattern as for the complimentary circular-disk and approximately the same as for a sphere, and hence, is consistent with Babinet's principle [52,53]. Note, however, that wave that is responsible for the diffraction pattern of the aperture is the remainder of the incident wave, whereas the particle's diffraction pattern was due to the EO extinction wave. Also, note that for the screen there is no Drude spot.

7. Summary

Diffraction is the spreading of waves. However, common usage often applies the word "diffraction" to a scattering phenomenon in which the scattered waves spread, especially around the forward direction, to create a diffraction pattern. In this review, Maxwell's equations are applied to understand how electromagnetic radiation interacts with an object to create the resulting diffraction patterns. In particular, it is shown that the diffracted wave from an object, such as a particle, smaller that the width of the incident beam is the Ewald–Oseen extinction wave radiated from the object into the far field zone. This same Ewald–Oseen extinction wave extinguishes the incident wave inside the object and immediately behind the object to create a shadow.

We have also shown that if the object is wider than the incident beam and contains an aperture, the Ewald–Oseen extinction wave radiated from the screen destructively interferes with the incident wave immediately behind the screen to create a shadow. Meanwhile, the portion of the incident wave that passes through the aperture propagates into the far field to create the diffraction pattern. If the aperture and the obstacle have the same cross-sectional shape, i.e., are complimentary, the diffraction patterns are the same, thus giving another explanation of Babinet's principle. However, this review adds that while the diffraction patterns of

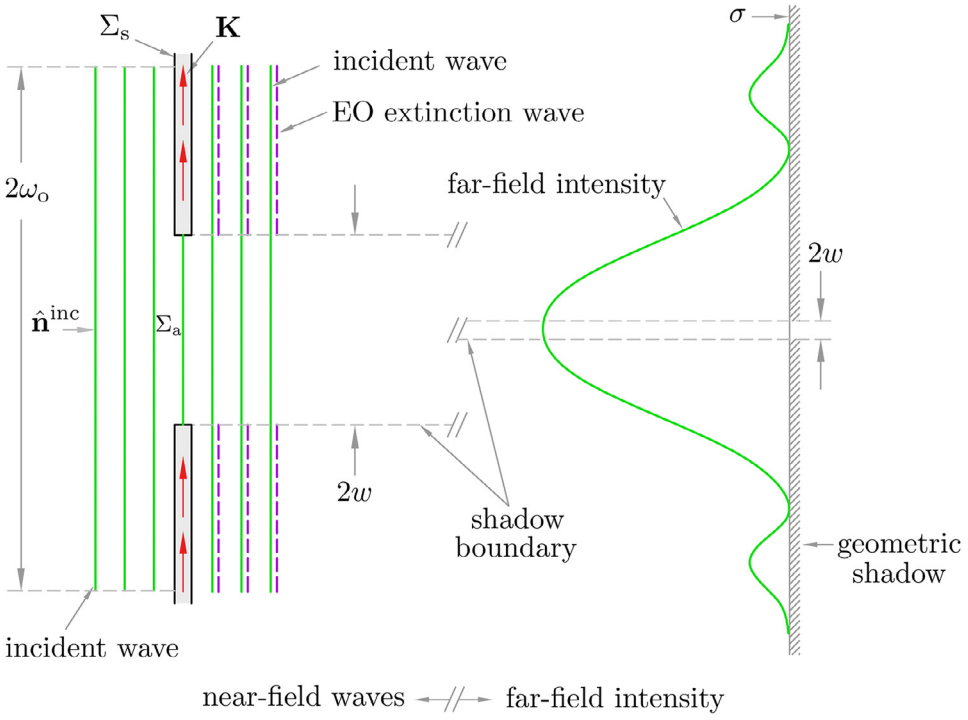

Fig. 9. Diffraction of a beam of light with width $2w_o$ incident upon a wider screen with a circular aperture of radius R smaller than the beam width. The sketch is split with the left portion showing the screen and near-field zone with the component *waves*, and the right portion showing the far-field zone *intensities*. The component waves are indicated by the color coding as: incident wave is green and the EO extinction wave is dashed purple. The intent here is to emphasize the extinction, via destructive interference, of the incident wave by the EO extinction wave in an annulus immediately behind the screen and to show the remaining circular incident wave. The screen does not block the incident wave, instead the screen radiates the EO wave that eliminates the incident wave except for that portion passing through the aperture. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

complimentary objects are the same, the waves that lead to the observed diffraction patterns are different.

These explanations rely on the radiation of the Ewald–Oseen extinction wave induced by the incident wave's interaction with the entire object. Thus, this is essentially a scattering process and the incident wave is not "blocked" by the object. Rather, the object radiates a wave, the Ewald–Oseen extinction wave, which can cancel the incident wave and create near field shadows and far-field diffraction patterns. Furthermore, with this explanation, the incident beam does not diffract around the edge of objects. Instead, the differing degrees of spreading, i.e., diffraction, of the incident and Ewald–Oseen waves lead to finite wave-intensities within the far-field projections of the near field shadows. In summary, diffraction patterns are created by the manipulation of the incident wave through the Ewald–Oseen extinction wave.

Acknowledgments

The authors are grateful for helpful discussions with O. Kemppinen, M. I. Mishchenko, A. J. Yuffa, G. Videen, and suggestions form two anonymous reviewers. This work was supported by the U.S. Army Research Office(W911NF-15-1-0549) and the National Science Foundation (1665456, 1623808).

References

[1] Hecht E. Optics. 4th ed. Reading, Mass: Addison-Wesley; 2002.
[2] Bohren CF, Huffman DR. Absorption and scattering of light by small particles. Wiley; 1983.
[3] Born M, Wolf E. Principles of optics; electromagnetic theory of propagation, interference, and diffraction of light. Cambridge University Press; 1999.
[4] Jackson JD. Classical electrodynamic. 3rd ed. Wiley; 1999.
[5] Nye JF, Liang W. Near-field diffraction by two slits in a black screen. Proc R Soc Lon. A 1998;454:1635–58.
[6] Umul YZ. Diffraction by a black half plane: Modified theory of physical optics approach. Opt Ex 2005;13:1–12.
[7] Siegman AE. Lasers. University Science Books; 1986.
[8] Mishchenko MI, Travis LD, Lacis AA. Scattering, absorption, and emission of light by small particles. Cambridge University Press; 2002.
[9] Smith GS. An introduction to classical electromagnetic radiation. Cambridge University Press; 1997.
[10] Baker BB, Copson ET. The mathematical theory of Huygens' principle. AMS Chelsea Publishing; 1987.
[11] Drude P. The theory of optics. New York: Dover; 1959.
[12] Gangi AF, Mohanty BB. Babinet's principle for elastic waves. J. Acoust. Soc. Am. 1973;53:525–34.
[13] Kälbermann G. Single- and double-slit scattering of wavepackets. J Phys A Math Gen 2002;35:4599–616.
[14] Gilliar W, Bickel WS, Videen G, Hoar D. Light scattering from fibers: an extension of a single-slit diffraction experiment. Am J Phys 1987;55:555–9.
[15] Kathuria YP. Fresnel and far-field diffraction due to an elliptical aperture. J Opt Soc Am A 1985;2:852–7.
[16] Morse PM, Rubenstein PJ. The diffraction of waves by ribbons and slits. Phys Rev 1938;54:895–8.
[17] Karp SN, Russek A. Diffraction by a wide slit. J Appl Phys 1956;27:886–94.
[18] Flammer C. The vector wave function solution of the diffraction of electromagnetic waves by circular disks and apertures. II. The diffraction problems. J Appl Phys 1953;24:1224–31.
[19] Flammer C. The vector wave function solution of the diffraction of electromagnetic waves by circular disks and apertures. I. Oblate spheroidal vector wave functions. J Appl Phys 1953;24:1218–23.
[20] Carcía de Abajo FJ, Gómez-Medina R, Sáenz JJ. Full transmission through perfect-conductor subwavelength hole arrays. Phys Rev E 2005;72:016608.
[21] Carcía de Abajo FJ. Light scattering by particle and hole arrays. Rev Mod Phys 2007;79:1267–90.
[22] Sommerfeld A. Mathematical theory of diffraction. Birkhäuser; 2004.
[23] Levine H, Schwinger J. On the theory of diffraction by an aperture in an infinite plane screen. I. Phys Rev 1948;74:958–74.
[24] Levine H, Schwinger J. On the theory of diffraction by an aperture in an infinite plane screen. II. Phys Rev 1949;75:1423–32.
[25] Rothwell EJ, Cloud MJ. Electromagnetics. 2nd ed. CRC Press; 2009.
[26] Tai C-T. Dyadic green functions in electromagnetic theory. IEEE press; 1994.
[27] Tsang L, Kong JA, Ding K-H, Ao CO. Scattering of electromagnetic waves: theories and applications. John Wiley & Sons; 2001.
[28] Zangwill A. Modern electrodynamics. Cambridge University Press; 2012.
[29] Bouwkamp CJ. Diffraction theory. Rep Prog Phys 1954;17:302.
[30] Agarwal GS. Relation between Waterman's extended boundary condition and the generalized extinction theorem. Phys Rev D 1976;14:1168–71.

[31] Mishchenko MI, Martin PA. Peter Waterman and T-matrix methods. J Quant Spectrosc Radiat Transfer 2013;123:2–7.
[32] Feynman RP, Leighton RB, Sands M. The Feynman lectures on physics, 1. Addison-Wesley Publishing; 1964.
[33] Arnoldus HF. Reflection off a mirror. J Mod Opt 2007;54:45–66.
[34] van Bladel J. Singular electromagnetic fields and sources. Oxford University Press; 1991.
[35] Ehrlich MJ, Silver S, Held G. Studies of the diffraction of electromagnetic waves by circular apertures and complimentary obstacles: the near-zone field. J Appl Phys 1955;26:336–45.
[36] Stratton JA, Chu LJ. Diffraction theory of electromagnetic waves. Phys Rev 1939;56:99–107.
[37] Meixner J. The behavior of electromagnetic fields at edges. IEEE Trans Antennas Prop 1972;20:442–6.
[38] Oxtoby DW, Novak F, Rice SA. The Ewald-Oseen theorem in the x-ray frequency region: a microscopic analysis. J Chem Phys 1982;76(9):5278–82.
[39] Kahnert FM. Numerical methods in electromagnetic scattering theory. J Quant Spectrosc Radiat Transfer 2003;79-80:775–824.
[40] Yaghjian AD. Electric dyadic Green's functions in the source region. Proc IEEE 1980;68(2):248–63.
[41] Yurkin MA, Hoekstra AG. The discrete dipole approximation: An overview and recent developments. J Quant Spectrosc Radiat Transfer 2007;106:558–89.
[42] James MB, Griffiths DJ. Why the speed of light is reduced in a transparent medium. Am J Phys 1992;60(4):309–13.
[43] Arnoldus HF. Boundary conditions in an integral approach to scattering. J Opt Soc Am A 2006;23:3063–71.
[44] Wolf E. Progress in optics, x. North-Holland; 1977.
[45] Durant W. The story of civilization: Vol. 2: the life of Greece. Simon & Schuster; 1939. Ch. XVI.
[46] Berman DH. An extinction theorem for electromagnetic waves in a point dipole model. Am J Phys 2003;71:917–24.
[47] Enders P. The Ewald-Oseen extinction theorem in the light of Huygens' principle. Ele J Theoretical Phys 2011;8:127–36.
[48] Oseen CW. Über die Wechselwirkung zwischen zwei elektrischen Dipolen und üer die Drehung der Polarisationsebene in Kristallen und Flüssigkeiten. Ann Phys 1915;48:1–56.
[49] Faxén H. Der zusammenhang zwischen den Maxwellschen gleichungen für dielektrika und den atomistischen ansätzen von H. A. Lorentz u. a. Z Physik 1920;2:218–29.
[50] Ewald PP. Über den brechungsindex für röntgnstrahlen und die abweichungen vom Braggschen reflexionsgesetz. Z Physik 1924;30:1–13.
[51] Berg MJ, Sorensen CM, Chakrabarti A. A new explanation of the extinction paradox. J Quant Spectrosc Radiat Transfer 2011;112:1170–81.
[52] Sorensen CM, Maughan JB, Chakrabarti A. The partial light scattering cross section of spherical particles. J Opt Soc Am A 2017;34:681–4.
[53] Heinson WR, Chakrabarti A, Sorensen CM. Crossover from spherical particle Mie scattering to circular aperture diffraction. J Opt Soc Am A 2014;31:2362–4.