CHAPTER ONE

The Microscope in a Computer: Image Synthesis from Three-Dimensional Full-Vector Solutions of Maxwell's Equations at the Nanometer Scale

İlker R. Çapoğlu[a], Jeremy D. Rogers[a], Allen Taflove[b] and Vadim Backman[a]

[a]Biomedical Engineering Department, Northwestern University, Evanston, IL, USA
[b]Electrical Engineering and Computer Science Department, Northwestern University, Evanston, IL, USA

Contents

Progress in Optics, Volume 57
ISSN 0079-6638, http://dx.doi.org/10.1016/B978-0-44-459422-8.00001-1

1. INTRODUCTION

Optical imaging systems have traditionally been analyzed using well-established approximations such as ray-based geometrical optics (Born & Wolf, 1999) and scalar Fourier theory (Goodman, 1996). However, there has recently been increased interest in applying the rigorous framework of Maxwell's-equations-based electromagnetic theory and numerical modeling to the analysis of optical imaging systems. The availability of more powerful computer hardware and more efficient computational algorithms has obviously contributed to this interest. Although the basic principles of light scattering encoded in Maxwell's equations had been around for decades, the widespread application of these principles to the complete modeling of an optical imaging system had to wait until the 1990s, at which time the personal computers were getting powerful enough to process megabytes of data in their memory. This allowed the modeling of objects that are comparable in size to the wavelength of the illuminating light (400–800 nm). With the arrival of these computational capabilities, the possibility of bypassing most of the traditional simplifying approximations and numerically calculating the optical image of an arbitrary object was at hand; and the demand for this accuracy was already present. Some engineering applications require the control of all the aspects of the optical imaging system down to sub-wavelength precision. Examples of such applications can be found in many subfields of physics and engineering. Historically, the earliest work on the numerical simulation of optical imaging was for modeling integrated-circuit production via photolithography (Cole, Barouch, Conrad, & Yeung, 2001; Neureuther, 2008), integrated-circuit inspection (Neureuther, 1992), and mark alignment (Nikolaev & Erdmann, 2003). More recently, there has been increasing interest in modeling optical microscopy modalities (Capoglu et al., 2011; Hollmann, Dunn, & DiMarzio, 2004; Sierra, DiMarzio, & Brooks, 2008; Simon & DiMarzio, 2007; Tanev, Pond, Paddon, & Tuchin, 2008; Tanev, Sun, Pond, Tuchin, & Zharov, 2009). If realized to its full potential, this technique could have immediate benefit on the optical detection of early stage nanoscale alterations in precancerous cells (Subramanian et al., 2008, 2009). This review/tutorial paper is primarily aimed as a reference for the numerical algorithms and techniques necessary for implementing a purely virtual imaging system, which we will refer to as a "microscope in a computer." Since the basic principles are also applicable to any other optical imaging system, this paper could also be consulted for modeling photolithography and metrology systems.

Although Maxwell's-equations-based electromagnetic principles have been successfully applied to the characterization of optical systems, the literature on the subject is fragmented across several independent lines of research, resulting in considerable overlap and inefficiency. This is a consequence of the fact that different forms of optical imaging systems are employed in many independent branches of engineering, sometimes based on similar principles but for diverse purposes. This fragmented literature has not yet been compiled and categorically documented for the benefit of the general engineering community. In this paper, we present a coherent and self-contained account of the numerical electromagnetic simulation of optical imaging systems, and review the body of work amassed in this rapidly growing field. We place special emphasis on numerical modeling issues such as discretization, sampling, and signal processing. Although the majority of the paper is tailored for optics, most of the concepts and formulas given in Section 2 and Sections 3.1–3.3 are applicable to a broader range of electromagnetics problems involving antennas, antenna arrays, metamaterials, RF, and microwave circuits and radars. The refocusing concept in Section 3.4, however, is a defining characteristic of an optical imaging system, with few exceptions such as focused antenna arrays in RF electromagnetics (Hansen, 1985).

The remainder of the paper is organized as follows. In Section 2, the basic principles of electromagnetics and optical coherence are reviewed. In Section 3, the optical imaging system is divided into fundamental components, and the numerical simulation of each component is described in detail. In Section 4, an optical imaging simulation system based on the finite-difference time-domain method is introduced, and several microscopy simulation examples are presented. A summary of our review and some concluding remarks are given in Section 5.

2. BASIC PRINCIPLES OF ELECTROMAGNETICS AND OPTICAL COHERENCE

An integral part of the numerical electromagnetic analysis of optical imaging systems is based on a set of vectorial relationships called *Maxwell's equations* that explain the propagation of light and its behavior in material media. These equations describe the nature and interrelationship of two vectorial quantities, the *electric* and *magnetic* field vectors $\boldsymbol{\mathcal{E}}(\boldsymbol{r}, t)$ and $\boldsymbol{\mathcal{H}}(\boldsymbol{r}, t)$, in free space and matter. The interaction of these vectors with matter is specified by two scalar material properties, the relative permittivity $\epsilon_r(\boldsymbol{r})$ and

permeability $\mu_{\mathrm{r}}(\boldsymbol{r})$. In crude terms, these two material properties quantify the response of matter to the electric and magnetic fields, respectively. In free space, these parameters are both equal to unity ($\epsilon_{\mathrm{r}} = \mu_{\mathrm{r}} = 1$). In differential form, Maxwell's equations are written as

$$\nabla \times \boldsymbol{\mathcal{E}} = -\mu_{\mathrm{r}}\mu_0 \frac{\mathrm{d}\boldsymbol{\mathcal{H}}}{\mathrm{d}t}, \tag{1}$$

$$\nabla \times \boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{J}} + \epsilon_{\mathrm{r}}\epsilon_0 \frac{\mathrm{d}\boldsymbol{\mathcal{E}}}{\mathrm{d}t}, \tag{2}$$

$$\nabla \cdot \boldsymbol{\mathcal{E}} = \rho, \tag{3}$$

$$\nabla \cdot \boldsymbol{\mathcal{H}} = 0, \tag{4}$$

where the symbol "$\nabla\times$" denotes the curl operator, which locally quantifies the amount and orientation of the "vorticity" in the vector field, and "$\nabla\cdot$" denotes the "div" operator, which quantifies the local magnitude of the "source" or "sink" associated with the vector field. Both definitions are in analogy to a velocity field in a fluid-dynamics context. In these equations, the electric current density $\boldsymbol{\mathcal{J}}(\boldsymbol{r}, t)$ acts as the excitation for the electromagnetic field. If the response of a system at a particular frequency of operation ω is of interest, Maxwell's equations simplify to their time-harmonic versions in which the time dependence is factored out in the form $\exp(j\omega t)$:

$$\nabla \times \boldsymbol{E} = -j\omega\mu_{\mathrm{r}}\mu_0 \boldsymbol{H}, \tag{5}$$

$$\nabla \times \boldsymbol{H} = \boldsymbol{J} + j\omega\epsilon_{\mathrm{r}}\epsilon_0 \boldsymbol{E}, \tag{6}$$

$$\nabla \cdot \boldsymbol{E} = \rho, \tag{7}$$

$$\nabla \cdot \boldsymbol{H} = 0. \tag{8}$$

Here and in what follows, calligraphic fonts $\mathcal{A}$, $\mathcal{B}$ will be used to denote general time dependence, while Roman fonts A, B will be used to denote time-harmonic quantities for which the time dependence $\exp(j\omega t)$ is implicit. In the engineering literature, it is customary to refer to Equations (1)–(4) as being in the *time domain*, and the time-harmonic versions (5)–(8) as being in the *frequency domain*.

In optics, the parameter $n = (\epsilon_{\mathrm{r}}\mu_{\mathrm{r}})^{1/2}$ is called the *refractive index* of the medium. It relates the light velocity v in the medium to the velocity c in the vacuum as $v = c/n$. In electromagnetics, the expression $W_{\mathrm{E}} = \epsilon_{\mathrm{r}}\epsilon_0|\boldsymbol{E}(\boldsymbol{r})|^2/2$ is the average electrical energy density at a point in space (in SI units). In the geometrical-optics (small-wavelength) approximation, the radiated

power per unit area in the local direction of propagation is equal to $I = 2(c/n)W_{\mathrm{E}}$ (Born & Wolf, 1999). Assuming non-magnetic media ($\mu_{\mathrm{r}} = 1$), this becomes

$$I = n|\boldsymbol{E}(\boldsymbol{r})|^2/\eta_0, \tag{9}$$

in which $\eta_0 = (\mu_0/\epsilon_0)^{1/2}$ is the *wave impedance* of free space. Although alternative terminologies do exist, we will use the term *light intensity* or simply *intensity* for the radiated power per unit area. The light intensity is a direct measure of the signal collected by recording media that convert light energy to other forms of energy. Examples of these recording media include photoresists, CCD cameras, and the retina. We will assume non-magnetic media throughout the paper and define the light intensity as in (9).

In most practical situations, the excitation in the optical system (whether it be a filament or a laser source) has a certain random character. This creates randomness in the resulting optical electromagnetic field in both space and time. If this is the case, the electromagnetic field may only be representable as a *random field* that possesses certain statistical properties. Fortunately, we are almost always concerned with time averages of optical parameters such as intensity or polarization, because these are the only parameters that most optical instruments can measure. If an adequate statistical model is constructed for the random electromagnetic field, the average quantities measured at the output of the system can be inferred mathematically. The categorization and rigorous mathematical description of these matters is the subject of *optical coherence* (Born & Wolf, 1999; Goodman, 2000). Although optical illumination systems almost always have a random character, the numerical electromagnetic simulation methods considered in this paper operate on deterministic field values that are known precisely in space and time. Numerical solutions of differential equations that operate directly on statistically averaged values [such as the radiative transfer equation (Ishimaru, 1999)] are outside the scope of this paper; see (Arridge & Hebden, 1997) for a review of these methods. The question arises, therefore, as to whether it is possible to compute statistical averages belonging to infinite random processes using completely deterministic numerical electromagnetic simulation methods. It turns out that this is possible, provided that the physical system satisfies certain conditions. One of the simplest of such situations is when the excitation is *statistically stationary* in time. Stationarity, in its strictest form, means that the statistical properties of the waveforms anywhere in the system *do not* change in time. This is a reasonable assumption for many forms of optical sources and will be made throughout this paper. The

study of non-stationary, spectrally partially coherent sources are outside the scope of this review. Interested readers may consult references (Christov, 1986; Lajunen, Vahimaa, & Tervo, 2005; Wang, Lin, Chen, & Zhu, 2003). The importance of stationarity is manifested when the response of a linear system to a stationary time waveform is sought. This is the case in our analysis, because both Maxwell's equations (5)–(8) and the scattering materials are assumed to be linear. Let us consider an input waveform $x_\mathrm{i}(t)$ exciting the system in some way and an output waveform $x_\mathrm{o}(t)$ measured somewhere else. If $x_\mathrm{i}(t)$ is the only excitation, the relation between these is a convolution with the impulse response $h(\tau)$ of the system:

$$x_\mathrm{o}(t) = \int_{-\infty}^{\infty} h(\tau) x_\mathrm{i}(t-\tau)\mathrm{d}\tau. \tag{10}$$

The *transfer function* $H(\omega)$ is defined as the Fourier transform of the impulse response $h(\tau)$,

$$H(\omega) = \int_{\tau=-\infty}^{\infty} h(\tau)\mathrm{e}^{-j\omega\tau}\mathrm{d}\tau. \tag{11}$$

It can be shown that the *power-spectral densities* $S_\mathrm{i}(\omega)$ and $S_\mathrm{o}(\omega)$ of the input and output waveforms are related linearly by the absolute square of the transfer function (Born & Wolf, 1999; Goodman, 2000; Haykin, 2001; Papoulis, 1991):

$$S_\mathrm{o}(\omega) = |H(\omega)|^2 S_\mathrm{i}(\omega). \tag{12}$$

The power-spectral density is an optically relevant and directly measurable quantity, defined as the power at the output of a narrowband filter centered at ω. The Wiener–Khintchine theorem (Born & Wolf, 1999) states that it is also the Fourier transform of the correlation function associated with the stationary waveform. The relation (12) is the central result that connects random waveforms in optics with the deterministic numerical methods of electromagnetics. In a given problem, the power-spectral density of the source $S_\mathrm{i}(\omega)$ is usually known, and the power-spectral density of the output $S_\mathrm{o}(\omega)$ is desired. The necessary link is provided by the absolute square of the transfer function $H(\omega)$. A numerical electromagnetic method can be used to find $H(\omega)$ by sending deterministic signals through the optical system, and calculating the response. Although the majority of the formulas in this review will be given for a fixed frequency ω, the response to a broadband stationary waveform can easily be obtained by repeating the analysis for different ω and using the power-spectral density relation (12).

This repetition becomes unnecessary if a time-domain method is used to obtain the scattering response. In such a case, $H(\omega)$ can be directly obtained at a range of frequencies via temporal Fourier transform of the time-domain response.

3. STRUCTURE OF THE OPTICAL IMAGING SYSTEM

An optical imaging system can be decomposed into several subsystems, each performing a self-contained task that is simple enough to model theoretically. Once the theoretical underpinnings of each subsystem are laid out, the numerical computation of actual physical parameters concerning the subsystem (transmission coefficients, far-field intensities, aberrations, etc.) becomes a matter of approximating the analytical equations in a suitable manner. We represent the optical imaging system as a combination of four subsystems: illumination, scattering, collection, and refocusing. These subsystems are drawn schematically in Figure 1.

3.1 Illumination

The light source and the lens system (usually called the condenser) that focuses the light created by the source onto the object are included in this subsystem. The last lens in the condenser system is shown on the left-hand side of Figure 1, along with the wavefront W_{i} incident on the object. We will base our review of illumination systems on whether they are *spatially coherent* or *incoherent*. Temporal coherence is a secondary concern since the sources considered in this review are always stationary (see Section 2). Once the responses to all the frequencies in the temporal spectrum of the source

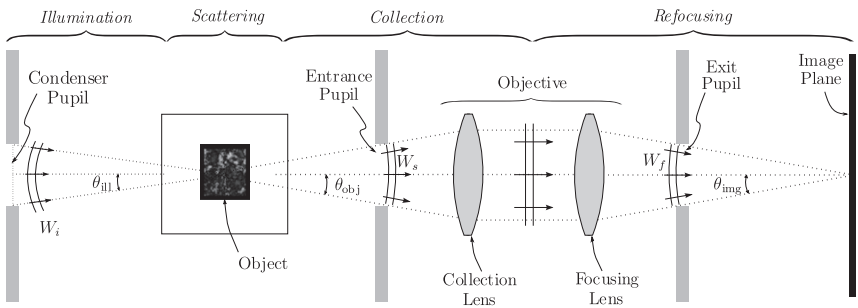

Figure 1 The four subcomponents of an optical imaging system: illumination, scattering, collection, and refocusing.

are found, then the synthesis of the output intensity is simply a matter of adding the intensities of the responses at each frequency.

3.1.1 Coherent Illumination

Spatially coherent illumination means that different points on the illumination beam are fully coherent. This kind of illumination can be created by an infinitesimally small light source, or by an atomic process called stimulated emission, as with lasers. Numerical models with varying degrees of complication are used to represent coherent beams. The simplest coherent illumination method used in numerical modeling is the *plane-wave illumination*. Being invariant in all but one dimension, the plane wave is one of the most basic solutions to Maxwell's equations, wherein the planes of constant phase are all perpendicular to the direction of propagation $\hat{\boldsymbol{k}}_i$. The electric and magnetic field vectors of the plane wave are perpendicular to each other and $\hat{\boldsymbol{k}}_i$. Individually, the plane wave can approximate a more complicated coherent illumination scheme over a very small illumination angle θ_{ill} (Salski & Gwarek, 2009b; Tanev, Tuchin, & Paddon, 2006). Full treatments of some of these illumination schemes in large-θ_{ill} cases have also been considered in the literature, albeit with less popularity. This is primarily because non-planar coherent beams are often difficult to compute and/or implement numerically. One of the more popular coherent illumination beams is the Gaussian beam (Smith, 1997). Although it has an approximate closed-form analytical expression that can be used in limited cases (Salski, Celuch, & Gwarek, 2010; Salski & Gwarek, 2008, 2009a), it is often decomposed into its plane-wave components; resulting in a more accurate description than the more limited closed-form expression (Yeh, Colak, & Barber, 1982). This method has the additional advantage of permitting the use of efficient and readily available plane-wave algorithms, such as the total-field/scattered-field (TF/SF) algorithm in FDTD. Since the Gaussian beam is defined at a single frequency, it is readily adapted to frequency-domain methods (Huttunen & Turunen, 1995; Wei, Wachters, & Urbach, 2007; Wojcik et al., 1991b). However, it can also be used in conjunction with the FDTD method in time-harmonic operation (Choi, Chon, Gu, & Lee, 2007; Judkins, Haggans, & Ziolkowski, 1996; Judkins & Ziolkowski, 1995; Simon & DiMarzio, 2007). The plane-wave spectrum (or the angular spectrum) method can also be used to synthesize arbitrary coherent illumination beams of non-Gaussian shape (Aguilar & Mendez, 1994; Aguilar, Mendez, & Maradudin, 2002). A practical example of a coherent beam is the electromagnetic field distribution around the focal

region of an aplanatic lens excited by a plane wave, derived by Richards and Wolf (Richards & Wolf, 1959; Wolf, 1959) using the angular-spectrum method. This beam has been used to simulate the coherent illumination in scanning-type confocal or differential-interference contrast (DIC) microscopes (Munro & Török, 2005; Török, Munro, & Kriezis, 2008). An extension of this technique to time-domain focused pulses was described in (Capoglu, Taflove, & Backman, 2008), which can be used to simulate either ultrafast optical pulses (Davidson & Ziolkowski, 1994; Gu & Sheppard, 1995; Ibragimov, 1995; Kempe, Stamm, Wilhelmi, & Rudolph, 1992; Veetil, Schimmel, Wyrowski, & Vijayan, 2006), or stationary broadband systems via temporal Fourier analysis. The latter type of systems have recently become feasible with the development of white-light laser sources (Booth, Juskaitis, & Wilson, 2008; Coen et al., 2002).

The plane-wave illumination is also sufficient when the scatterer under consideration is very thin compared to the wavelength and/or the range of illumination angles is sufficiently narrow. For example, under the thin-mask assumption (see Section 3.2) in photolithography, scattering from any plane wave from an arbitrary direction is completely determined by the scattering from a plane-wave incident normally on the thin mask. This is because the thin mask is assumed to simply impart a position-dependent phase shift on the plane wave upon transmission. If the scattered wave is decomposed into its angular spectrum (which is continuous if the mask is non-periodic, and discrete if it is periodic), it can easily be shown that this angular spectrum will *rotate* in the same direction that the incident plane wave is rotated. Therefore, it is only necessary in numerical computation to consider a single normally incident plane wave and calculate the "diffracted orders," as the Fourier components of the scattered wave are commonly called in photolithography. In passing, it is worthwhile to note that this "angular-shift invariance" property of the scattered field from a thin mask is a direct result of the Fourier relationship between the scattered field near the mask and the associated angular spectrum of the scattered field. This can easily be seen by comparison to a linear time-invariant (LTI) system, whose effect on its input is a multiplication by a transfer function in the Fourier (or frequency) domain. Similarly, the angular-shift invariance of the scattered field in the Fourier (or angular) domain is a result of the multiplicative action of the mask on the incident field in the spatial domain.

Illumination modeling generally becomes a harder task when the object space is multilayered. The total-field/scattered-field (TF/SF) algorithm in FDTD has been generalized to handle multilayered spaces (Capoglu &

Smith, 2008; Winton, Kosmas, & Rappaport, 2005; Zhang & Seideman, 2010). The plane-wave TF/SF techniques can be used as building blocks for injecting arbitrary coherent beams into a multilayered space, since any beam can in principle be decomposed into a plane-wave spectrum.

3.1.2 Incoherent Illumination

The term "incoherent illumination" is traditionally used to designate an illumination scheme that exhibits partial spatial coherence over the illumination area. Incoherent illumination occurs when the light source has finite spatial extent, with every point on the source radiating in an incoherent fashion. This is an adequate model for many natural and artificial light sources such as the sun, a xenon arc lamp, or a tungsten filament. Incoherence also requires that the excitation source have a finite bandwidth, however small it may be. In fact, the converse of this requirement (strict monochromaticity) is very hard to achieve, for even the most coherent laser sources have a finite bandwidth.

Perhaps the most prominent incoherent illumination scheme in use today is called *Köhler illumination* (Born & Wolf, 1999; Nolte, Pawley, & Höring, 2006, chap. 6), named after August Köhler who designed it in the late 1800s. One of the key advantages of this scheme is that it provides spatially uniform illumination throughout the sample, regardless of the inhomogeneities of the light source. This is accomplished by sending a collimated beam on the sample for every infinitesimal point on the light source. The details of Köhler illumination are shown schematically in Figure 2. The light source on the left of Figure 2 is imaged on the aperture stop by an auxiliary lens. The image of the light source on the aperture stop acts as a secondary source for the succeeding portion of the system. Unlike the original light source, the spatial coherence length on this secondary

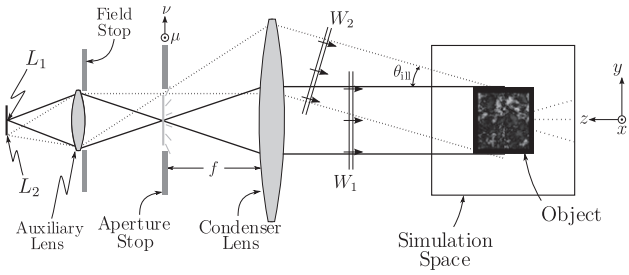

Figure 2 Schematical illustration of Köhler illumination.

source is not zero; in other words, the secondary source is technically a partially coherent source. Fortunately, if the aperture stop is much larger than the size of the diffraction spot (also called the Airy disc) associated with the auxiliary lens, there is little accuracy lost if every point on this secondary source is also assumed incoherent (Born & Wolf, 1999). The remaining analysis of Köhler illumination is always based on this assumption. Two rays emanating from each of two mutually incoherent infinitesimal point sources L_1 and L_2 on the light source are shown in Figure 2 by solid and dotted lines, respectively. Since the aperture stop is situated at the front focal plane of the condenser lens, every point source on the aperture stop creates a collimated beam illuminating the object from a different direction. Since the secondary point sources on the aperture stop are assumed incoherent, these beams are also incoherent. The flat wavefronts created by L_1 and L_2 are denoted by W_1 and W_2 in Figure 2. The aperture stop limits the angles from which the incoherent beams hit the object within an illumination cone, defined by θ_{ill}. In general, the image of the source on the aperture stop may be inhomogeneous, therefore the beams hitting the object may have different amplitudes. If the source is of uniform intensity, these amplitudes are also uniform. The spatial extent of the illumination, on the other hand, is controlled by the field stop in Figure 2. Let the field stop be imaged by the condenser lens at the plane S at distance d from the back focal plane F, as shown in Figure 3. For illustration purposes, the distance d in Figure 3 is drawn much larger than usual. The field stop is usually at several focal lengths in front of the aperture stop, so S is usually pretty close to F. It is clearly seen from Figure 3 that the optimum position for the sample is at S, since any forward or backward movement of the sample will cause the

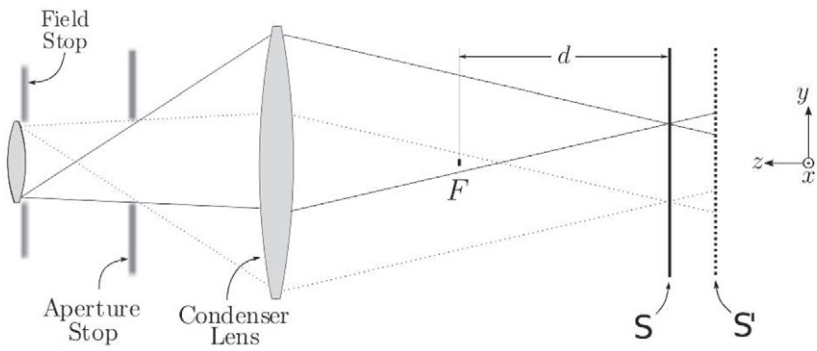

Figure 3 Optimal placement of the sample for spatially uniform illumination.

elimination of some rays incident from certain directions. As the sample is moved away from focus (say to S'), the illumination area gets larger and starts blurring at the edges. This undesirable effect is avoided by focusing the field stop sharply over the sample at all times.

The annotations "field stop" and "aperture stop" in Figure 2 are based on the implicit understanding that the field behind the auxiliary lens (on the field stop) is taken as the "source" to be imaged onto the object by the condenser system. If the actual physical source on the left-hand side of Figure 2 (e.g., a xenon lamp) is considered to be the "source" of the illumination system, then the plane S in Figure 3 coincides with the *exit pupil* of the condenser. This is because the "field stop" in Figure 2 now acts as the aperture stop, and its image (the exit pupil) lies at S. One can therefore say equivalently that the optimal position S for the sample is the exit pupil of the condenser, if the source is understood to be the actual physical source. In photolithography, there might not be any intermediate stops between the physical source and the condenser lens, in which case the exit pupil lies at the plane of the condenser lens (Goodman & Rosenbluth, 1988; Mack, 2007; Thompson, Willson, & Bowden, 1994; Tirapu Azpiroz, 2004; Yeung, 1988).

If the sample is close enough to the center of the illumination area on S, the collimated beams can be very well approximated by *plane waves*. In numerical computation, the continuum of mutually incoherent plane waves over the illumination cone has to be approximated by a finite sum. This is, in effect, a two-dimensional numerical quadrature problem, for which unfortunately no universally optimum method exists (Press, Flannery, Teukolsky, & Vetterling, 1992). A heuristic and straightforward method that is applicable regardless of the actual shape of the source image on the aperture stop is an equally spaced arrangement of point sources, combined with the midpoint rule (Press et al., 1992). The corresponding placement of the plane waves incident on the sample in Figure 3 can be found from geometrical optics (Born & Wolf, 1999, Section 10.6.2). Within the accuracy of Gaussian optics (small off-axis distances, small angles around the axis), every position (μ, ν) on the aperture stop corresponds to a plane wave with direction cosines $(s_x, s_y) = (\mu/f, \nu/f)$ at the back focal plane of the condenser, where f is the focal length of the condenser. The direction cosines are defined as

$$\begin{aligned} s_x &= \sin\theta\cos\phi = \cos\chi, \\ s_y &= \sin\theta\sin\phi = \cos\eta, \end{aligned} \tag{13}$$

in which the angles χ, η, θ, and ϕ are as shown in Figure 4. The angles θ, ϕ are the usual longitudinal and azimuthal angles in the spherical coordinate system centered around the z axis. An equal spacing of point sources on the aperture stop results in the equal spacing of the direction cosines (s_x, s_y) at the back focal plane of the condenser. An example of equally spaced arrangement of the direction cosines is shown in Figure 5. The maximum value that either s_x or s_y can attain is $\sin\theta_{\text{ill}}$, where θ_{ill} is the illumination half-angle in Figure 2. The quantity $\text{NA}_{\text{ill}} = n\sin\theta_{\text{ill}}$, where n is the refractive index of the medium, is called the *illumination numerical aperture*.

As every plane wave in Figure 5 propagates to the sample plane S (at distance d from F), it acquires a phase shift that will also be preserved in the scattered field due to the linearity of the system. If the intensities of the scattered field are of interest (as is the case in Köhler illumination),

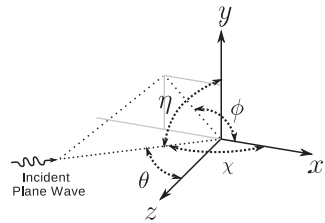

Figure 4 Definitions of certain angles associated with plane-wave incidence.

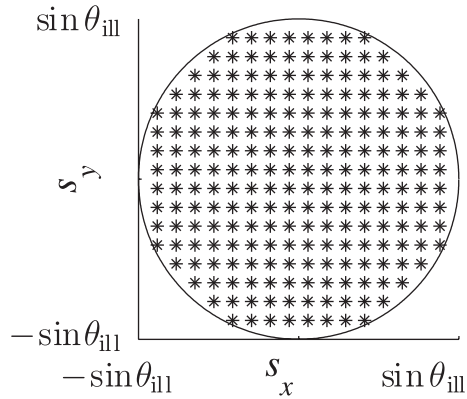

Figure 5 Equal spacing of plane waves in Köhler illumination. Two orthogonal polarizations (+) and (x) are shown for each direction of incidence.

this extra phase shift will not have any effect on the output because of the mutual incoherence of the incident plane waves.

One can quantify the quality of the approximation that results from the discrete arrangement of the plane waves in Figure 5. Let us consider quasi-monochromatic illumination with mean wavenumber $k = nk_0$, where k_0 is the mean wavenumber in free space. Let $J(x_1, y_1; x_2, y_2)$ denote the mutual coherence function at the sample plane (x, y), which quantifies the statistical correlation between two points with coordinates (x_1, y_1) and (x_2, y_2). As far as second-order quantities (intensity, two-point correlation, power-spectral density, etc.) are concerned, the mutual coherence function $J(x_1, y_1; x_2, y_2)$ completely specifies the excitation. Any illumination scheme that results in the same $J(x_1, y_1; x_2, y_2)$ will yield the same second-order quantities at the output. For the Köhler-illumination scheme considered here, $J(x_1, y_1; x_2, y_2)$ is given by (Born & Wolf, 1999)

$$J(x_1, y_1; x_2, y_2) = J(x_d; y_d) = \iint_{\Omega_{\text{ill}}} \mathrm{e}^{-jk(s_x x_d + s_y y_d)}\, \mathrm{d}\Omega, \tag{14}$$

in which $x_d = x_1 - x_2$, $y_d = y_1 - y_2$, Ω_{ill} is the illumination solid angle bounded by $s_x^2 + s_y^2 < \sin^2\theta_{\text{ill}}$, and the differential solid angle $\mathrm{d}\Omega$ is equal to $\mathrm{d}s_x \mathrm{d}s_y / \cos\theta$. Assuming moderate θ_{ill} values and neglecting the $\cos\theta$ term, this expression can also be written as

$$J(x_d; y_d) = \iint_{-\infty}^{\infty} P(s_x, s_y) \mathrm{e}^{-jk(s_x x_d + s_y y_d)}\, \mathrm{d}s_x \mathrm{d}s_y, \tag{15}$$

in which $P(s_x, s_y)$ is equal to unity within the circle $s_x^2 + s_y^2 < \sin^2\theta_{\text{ill}}$ and zero elsewhere. Let us label the discrete directions in Figure 5 with indices (m, n), with the direction cosines

$$s_{x_m} = m\Delta s_x, \quad s_{y_n} = n\Delta s_y. \tag{16}$$

The indices m and n can be assumed to run from $-\infty$ to ∞. The discrete plane waves should be weighed by $(\Delta s_x \Delta s_y)^{1/2}$ (the square root of the differential area in the direction-cosine space), so that the mutual coherence function is weighed by the differential area $\Delta s_x \Delta s_y$ in the direction-cosine space. With these weights, the arrangement in Figure 5 results in the following mutual coherence function:

$$J^*(x_d; y_d) = \Delta s_x \Delta s_y \sum_{m,n} P(s_{x_m}, s_{y_n}) \mathrm{e}^{-jk(s_{x_m} x_d + s_{y_n} y_d)}. \tag{17}$$

In Appendix A, it is shown that $J^*(x_d; y_d)$ is a sum of shifted copies of the original mutual coherence function $J(x_d; y_d)$:

$$J^*(x_d; y_d) = \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} J\left(x_d + r\frac{2\pi}{k\Delta s_x}; y_d + s\frac{2\pi}{k\Delta s_y}\right). \tag{18}$$

This is called *aliasing* in signal processing (Oppenheim, Schafer, & Buck, 1999). For $J^*(x_d; y_d)$ to represent $J(x_d; y_d)$ in a faithful manner, the shifted copies must not overlap, i.e.,

$$\Delta s_x < \frac{2\pi}{kW_c}, \quad \Delta s_y < \frac{2\pi}{kW_c}, \tag{19}$$

where W_c is defined as the distance $(x_d^2 + y_d^2)^{1/2}$ at which $J(x_d; y_d)$ falls below a negligible value. Using (15), a closed-form expression can be found for $J(x_d; y_d)$, with a W_c value of the order of $1/(k \sin\theta_{\text{ill}}) = 1/(k_0 \text{NA}_{\text{ill}})$. If the sample dimension D is larger than W_c, then D must be substituted for W_c in (19). Otherwise, the mutual coherence function $J^*(x_d; y_d)$ evaluated between two most distant points on the sample will be aliased and incorrect. A more general form of the non-aliasing condition (19) is therefore

$$\Delta s_x < \frac{2\pi}{k \max\{D, W_c\}}, \quad \Delta s_y < \frac{2\pi}{k \max\{D, W_c\}}. \tag{20}$$

For a stationary broadband excitation, the largest wavenumber k (the smallest wavelength λ) present in the illumination waveform determines the non-aliasing condition (20).

If the illumination numerical aperture NA_{ill} is not very large (or the scattering object is very thin), one can invoke Hopkins' approximation (Hopkins, 1953), in which case there is no need to consider all the plane waves in Figure 5. Under Hopkins' approximation, the response of the scatterer to any plane wave is completely determined by its response to the normally incident plane wave. Under even more stringent conditions, this response can also be obtained in a very simple manner. More on this will be said in Section 3.2. Here, we assume that the response of the object to each plane wave in Figure 5 is distinct and needs separate consideration. This procedure is commonly known as the *source-point* or *Abbe* integration (Erdmann & Evanschitzky, 2007; Kirchauer & Selberherr, 1997; Marx, 2007; Smith & Mack, 2003; van Haver et al., 2008; Wojcik, Mould, Monteverde, Prochazka, & Frank, 1991a; Yang, Milster, Zhang, & Chen, 2010; Zhang, Kim, Yang, & Milster, 2010). Since the plane waves in Figure 5 are all

mutually incoherent, a separate simulation should be run for each of them. The resulting image intensities (not field values) of each simulation are then added to yield the final image intensity (see Section 3.4).

The treatment so far has been for a scalar field. It turns out that two orthogonal, mutually incoherent polarizations for the electric field of the plane wave can always be chosen for every direction (s_x, s_y), as will be seen shortly. The two polarizations for each plane wave are denoted by $+$ and $\times$, and shown superposed at each (s_x, s_y) direction in Figure 5. If polarization information is critical, these two polarizations should also be simulated separately, as they are mutually incoherent. The overall number of simulations is therefore twice the number of direction cosines in Figure 5. This brute-force repetition of the entire simulation for incoherent illumination is a consequence of the deterministic nature of the numerical methods within the scope of our review. A clever, more efficient way of reducing this burden may be a topic for future research.

The determination of the two orthogonal, mutually incoherent polarization states for the plane waves in Figure 5 requires the knowledge of the polarization properties of the source on the aperture stop. We restrict ourselves to sources that exhibit uniform polarization properties throughout the aperture stop. Denoting the coordinates on the aperture stop as (μ, ν), we can express the uniform second-order polarization properties of the source using the *cross-spectral coherency matrix* $\boldsymbol{J}(\mu, \nu; \omega)$ (Born & Wolf, 1999):

$$\boldsymbol{J}(\mu, \nu; \omega) = \begin{bmatrix} \langle E_\mu^2 \rangle & \langle E_\mu E_\nu^* \rangle \\ \langle E_\mu^* E_\nu \rangle & \langle E_\nu^2 \rangle \end{bmatrix}, \tag{21}$$

where E_μ, E_ν are the tangential components of the electric field on the aperture stop, $\langle \cdot \rangle$ denotes temporal averages (or statistical expectation values), and $\exp(j\omega t)$ dependence is implicit. Since $\boldsymbol{J}(\mu, \nu; \omega)$ is Hermitian, it can be represented as a weighted sum of two orthogonal coherency matrices (Mandel, 1963; Tervo, Setälä, & Friberg, 2004):

$$\boldsymbol{J}(\mu, \nu; \omega) = \begin{bmatrix} A & B \\ B^* & C \end{bmatrix} + D \begin{bmatrix} C & -B \\ -B^* & A \end{bmatrix}, \tag{22}$$

subject to $A, C, D \geqslant 0$ and $AC = |B|^2$. This corresponds to decomposing the partially polarized field on the aperture stop into two orthogonal, mutually incoherent, fully polarized fields. The directions of these

polarization states coincide with the orthogonal eigenvectors of the coherency matrix $\boldsymbol{J}(\mu, \nu; \omega)$, and their relative weights are determined by the eigenvalues. Explicit formulas for the parameters A, B, C, D can be found in Tervo et al. (2004). A, B, C determine the angles and the ellipticity of the two polarization states at each (μ, ν), while D determines the relative powers of these components. Once the two orthogonal, mutually incoherent polarization states are determined, they should be treated individually in separate numerical simulations. The problem is thus reduced to fully polarized excitation, in which the electric field on the aperture stop is uniformly polarized in a certain direction. Since the general case of elliptical polarization can be handled as a complex superposition of two linearly polarized states, we only consider linear polarization. A good approximation for the polarization of the resulting plane waves in the back focal plane of the condenser can be obtained using the construction in Figure 6. An arbitrary point A on the plane of the aperture stop (μ, ν) is shown on the left-hand side of the figure. Let B denote the point on the lens such that $\overline{AB}$ is parallel to the optical axis OF. Let $\boldsymbol{\alpha}$ denote the ray emanating from A, hitting the lens at B, and intersecting the optical axis at the back focal point F. The plane including the ray $\boldsymbol{\alpha}$ and the optical axis OF is called the meridional plane, which, in our case, makes an angle $\boldsymbol{\phi}$ with the μ axis. The key observation is thus: the vector electric field on the rays in the neighborhood of $\boldsymbol{\alpha}$, which are parallel to $\boldsymbol{\alpha}$ around the focus F, will be almost the same as that of the ray $\boldsymbol{\alpha}$. Therefore, if the illuminated sample at the

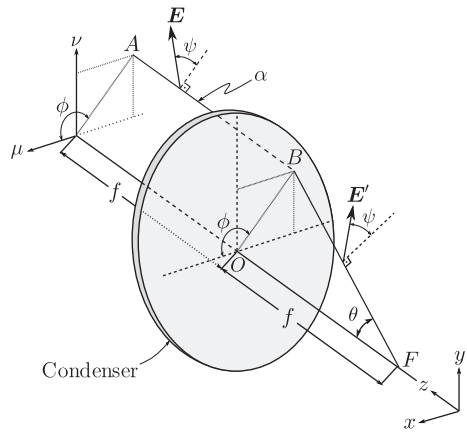

Figure 6 Polarization of the plane wave created by the condenser at focus *F* due to an infinitesimal source at point *A* on the aperture stop.

back focal plane at F is confined to a small area with dimensions $D \ll f$, the polarization and magnitude of the electric field on the sample at F is determined by the ray $\boldsymbol{\alpha}$. The magnitude of the electric field at F follows from the intensity law of geometrical optics. The infinitesimal source at A creates a spherical wavefront centered around A. The magnitude of the electric field at B due to this source is proportional to $|\boldsymbol{E}|/f$, where $\boldsymbol{E}$ is the *strength factor* of the ray $\boldsymbol{\alpha}$, depending only on the magnitude and polarization of the source at A but not on f. For a more detailed description of the strength factor of a ray, see (40) and the following discussion. Let us now consider an infinitesimal bundle of rays emanating from A and spread over an infinitesimal area around B. These rays are collimated by the condenser into a parallel tube of rays intersecting the optical axis around the back focal point F. The infinitesimal area subtended by this parallel tube of rays is $(\cos\theta)$ times the infinitesimal area subtended by the ray bundle on the other side of the condenser. From the intensity law of geometrical optics (Born & Wolf, 1999), it follows that the magnitude of the electric field $\boldsymbol{E}'$ at the back focal point F is given by

$$|\boldsymbol{E}'| = (\cos\theta)^{-1/2}|\boldsymbol{E}|/f. \tag{23}$$

The polarization of $\boldsymbol{E}'$ still remains to be found. Let ψ denote the angle that $\boldsymbol{E}$ makes with the meridional plane, as shown in Figure 6. If the angles of incidence at every surface of refraction through the lens are small, the angle ψ between the electric-field vector on the ray and the meridional plane stays constant (Born & Wolf, 1999; Richards & Wolf, 1959). This fact has been previously used in similar Köhler-illumination constructions (Totzeck, 2001; Yang, Milster, Park, & Zhang, 2010; Zhang et al., 2010), as well as the synthesis of a coherent converging light pulse in the FDTD method (Capoglu et al., 2008). Apart from the factor $\cos\theta^{-1/2}/f$, the electric field vector $\boldsymbol{E}'$ is then a *rotation* of $\boldsymbol{E}$ by an angle θ around an axis perpendicular to the meridional plane (Totzeck, 2001).

An important special case of incoherent Köhler-style illumination is when the two orthogonal components E_μ and E_ν of the electric field on the aperture stop are of equal power and are completely uncorrelated: $|E_\mu|^2 = |E_\nu|^2$ and $\langle E_\mu E_\nu^* \rangle = \langle E_\mu^* E_\nu \rangle = 0$. The source on the aperture stop is then said to be *natural*, or *completely unpolarized*. The cross-spectral coherency matrix $\boldsymbol{J}(\mu, \nu)$ in (21) is then proportional to the identity matrix, which amounts to $D = 1$ in the decomposition (22). This means that the A, B, C values for the decomposition in (22) are not constrained

by anything but the coherency condition $AC = |B|^2$. As a result, the choice of the two orthogonal polarization states + and × for each and every plane wave in Figure 5 becomes *completely arbitrary*.

It should be noted that there are other aperture shapes besides the circular shape in Figure 5 employed in practice. Depending on the geometry of the aperture, the discretization scheme for the incidence directions can be slightly modified. For example, the annular and quasar-shaped apertures commonly employed in photolithography can be accommodated using an equal spacing of the incidence angles, rather than the direction cosines (Pistor, 2001; Tirapu Azpiroz, 2004). The annular aperture is also a characteristic element of phase-contrast microscopy (Tanev et al., 2009; Tanev, Tuchin, & Pond, 2008).

3.2 Scattering

The difficulty of obtaining an exact expression for the light scattered from the illuminated object depends on the overall size and refractive-index distribution of the object. The simplest scattering geometry results when the object is so thin that it can be modeled as a phase object. A phase object is completely characterized by its *complex transmittance* $T(x, y)$ at the object plane (x, y), defined as the amplitude and phase imparted upon an incident waveform upon its exit on the other side of the object. This approximation is a crucial element of the early analytical models of partially coherent imaging (Barouch, Cole, Hollerbach, & Orszag, 1993; Cole, Barouch, Hollerbach, & Orszag, 1992a; Cole et al., 1992b; Hopkins, 1951, 1953; Kintner, 1978), which have been extensively applied to the numerical modeling of photolithography. Assuming $\exp(j\omega t)$ time dependence, a phase object modifies the incident amplitude $A_-(x, y)$ on the object plane (x, y) by the complex transmittance $T(x, y)$, yielding the transmitted amplitude $A_+(x, y)$:

$$A_+(x, y) = T(x, y)A_-(x, y). \tag{24}$$

As discussed in Section 3.1.1, this is equivalent to the thin-mask assumption in photolithography. For thicker masks, the validity of this approximation may become questionable. However, some rigor can be maintained in this case by calculating only the response $A_+^0(x, y)$ of the thick mask to a normally incident plane wave, and approximating the response at other incidence directions (s_x, s_y) by *angularly shifted* versions of $A_+^0(x, y)$ (Erdmann, Evanschitzky, Citarella, Fuhner, & De Bisschop, 2006; Guerrieri, Tadros,

Gamelin, & Neureuther, 1991; Lucas, Tanabe, & Strojwas, 1996; Pistor, 2001; Wong, 1994; Wong & Neureuther, 1994, 1995; Wong, Guerrieri, & Neureuther, 1995):

$$A_+(x, y) = A_+^0(x, y)\mathrm{e}^{-jk(s_x x + s_y y)}. \tag{25}$$

This approximation can be applied if the angular dimensions of the source as seen from the object are very small, and the rays incident on the object are nearly vertical. This is true for the optical systems in photolithography, where illumination NAs of ~ 0.1 ($\theta_{\mathrm{ill}} \sim 4 - 5°$) are quite common (Yeung, Lee, Lee, & Neureuther, 1993) (see Section 3.4, Figure 14(a)).

If the refractive-index contrast between the scatterer and the surrounding medium is very small, then the total electromagnetic field inside the scatterer can be approximated in the first order by the incident field, resulting in the *Born approximation*, or the *weak-scattering approximation*. This approximation was first used by the German physicist Max Born (1882–1970) in his analysis of atomic particle scattering. It has since found application in other areas where the scattering is also governed by wave equations, such as electromagnetics and optics (Born & Wolf, 1999; Chew, 1990; Ishimaru, 1999).

In many situations of practical interest, the phase-object assumption (24) or the weak-scattering approximation is not valid. Even if the angular-shift invariance property (25) is assumed to hold, the response $A_+^0(x, y)$ of the object to a normally incident plane wave needs to be calculated using more rigorous methods. If the spatial features of the scattering object are much larger than the wavelength, ray-based geometrical-optics methods (Lam, 2005; Lam & Neureuther, 2004) or the first-order physical-optics approximation (Yeung & Neureuther, 1995) may be used. For objects consisting of a small number of sharp edges that do not give rise to much resonance or multiple scattering, asymptotic high-frequency solutions based on the physical theory of diffraction (PTD) may also yield satisfactory results (Tirapu Azpiroz, 2004). However, if the scattering object has key structural details comparable in size to the wavelength of the incident light, ray-based or asymptotic methods completely fail to describe the scattering process. Some examples of the latter are biological cells, photonic crystals, and phase-shift masks in lithography. In this situation, one has no other choice but to seek an approximate numerical solution to the Maxwell's equations (1)–(4). In the following, we will present a brief overview of the numerical methods used in the solution of Maxwell's equations. This overview is by no means a complete account of all the numerical methods in

electromagnetics; but is rather a preliminary introduction to the methods that have been applied to the modeling of optical imaging systems.

An approximate numerical solution to Maxwell's equations starts with the expansion of the electromagnetic field into a set of basis functions. Maxwell's equations (either in differential or integral form) are then transformed into algebraic equations involving the coefficients of this expansion. Depending on the characteristics of the basis functions used in the expansion, the numerical solution method can take a wide variety of forms. A fundamental distinction can be made between numerical methods that use basis functions that have *global* or *finite* support. In the former case, the basis functions are nonzero over the entire problem geometry or a significant portion of it. The resulting numerical method is commonly called a *modal* method. When the basis functions have finite support, they are defined over a small enough region of the geometry that resolves the spatial (or temporal) features of the scattering object. This type of numerical method is referred to as a *finite* method.

3.2.1 Modal Methods

Modal methods usually result in an infinite series or an infinite matrix problem, which is solved approximately by truncation. Although they require less computational power than finite methods, their application is very much limited to simple (usually planar and periodic) materials. Modal methods are also the oldest and most established ones, with a large body of analytical literature since Maxwell's time. A comprehensive review of all modal methods in electromagnetics is far beyond the scope of this paper. Here, we will only present an introductory survey of the most common methods in existence.

Several representative examples of scattering structures that are amenable to modal numerical analysis are shown in Figure 7. In all subfigures, a 2D medium is assumed with period d in the x direction. The extension to 3D is straightforward, as it introduces no conceptual novelty. Interested readers will find relevant information in the references. The scattering medium is divided into three planar regions, indexed from 0 to 2 from top to bottom. Without much loss of generality, the uppermost region (with index 0) is assumed to be free space, and the lowermost region (with index 2) is assumed to have homogeneous relative permittivity ϵ_2. The inhomogeneous permittivity in region 1 is denoted by $\epsilon(x, z)$. In Figure 7(a), a planar grating is shown with equi-permittivity lines $\epsilon(x, z) = [\text{const.}]$ forming parallel straight lines at an angle ϕ with the x axis. If the slant angle $\phi = 90°$

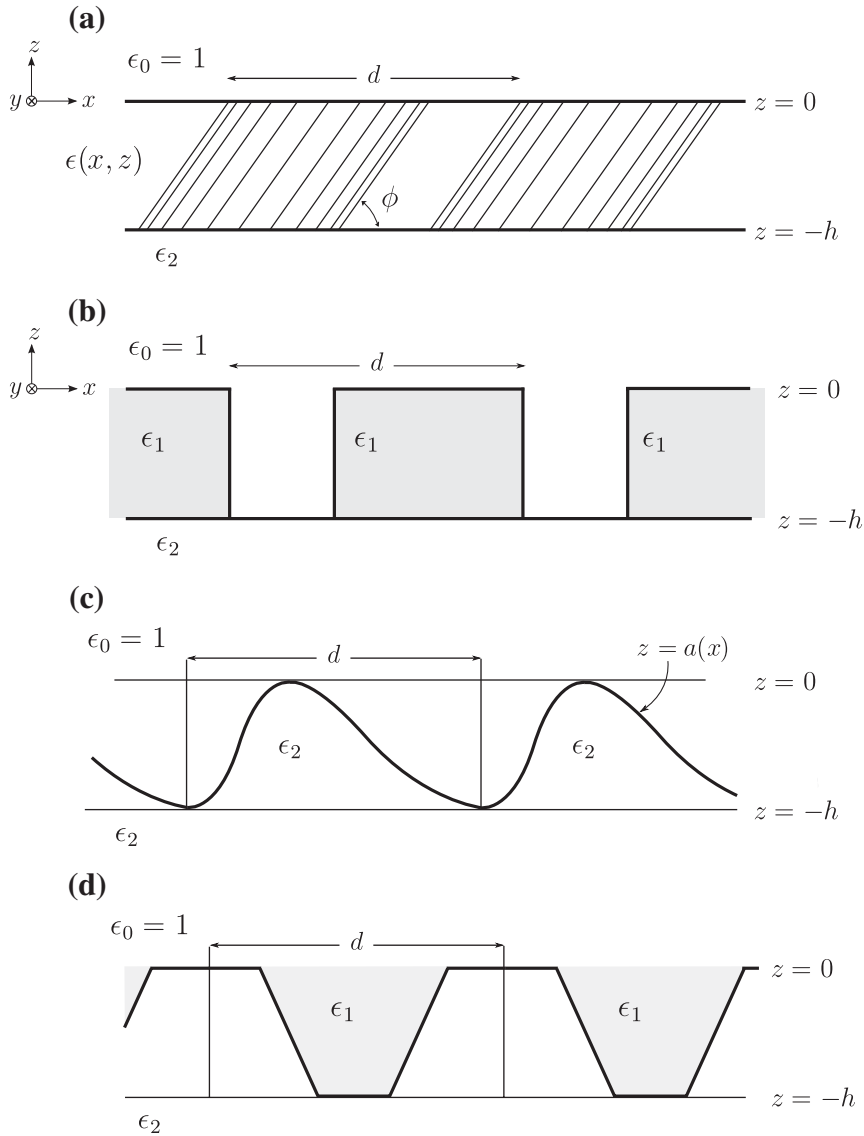

Figure 7 Examples of periodic structures that can be analyzed using modal methods. (a) Planar grating with slanted permittivity fringes. (b) Lamellar planar grating. (c) Two-layered surface-relief grating. (d) Inhomogeneous surface-relief grating.

and $\epsilon(x, z)$ is piecewise continuous in x, the lamellar (or laminar, binary) grating in Figure 7(b) is obtained. A more general distribution of $\epsilon(x, z)$ results in the surface-relief grating shown in Figure 7(c) with surface profile

$z = a(x)$. An even more general permittivity distribution is represented by the structure in Figure 7(d). Although the structures in Figure 7(b) and (c) are special cases of that in Figure 7(d), the numerical approaches for dealing with them, as well as their respective applications, may be quite different.

Let us assume monochromatic excitation with time dependence $\exp(j\omega t)$, and a TE polarized plane wave (electric vector perpendicular to the grating, parallel to the y axis) incident at an angle θ with respect to the z axis. Maxwell's equations in regions 0, 1, and 2 then reduce to the following second-order Helmholtz equations for the y component of the electric field E_y:

$$\nabla^2 E_{y_0} + k_0^2 E_{y_0} = 0, \tag{26}$$

$$\nabla^2 E_{y_1} + k_0^2 \epsilon(x, z) E_{y_1} = 0, \tag{27}$$

$$\nabla^2 E_{y_2} + k_2^2 E_{y_2} = 0, \tag{28}$$

where $\nabla^2 = (\partial^2/\partial x^2) + (\partial^2/\partial z^2)$, $k_{\{0,2\}} = (\omega/c)\epsilon_{\{0,2\}}^{1/2}$ are the wavenumbers in regions 0 and 2, respectively. From Floquet's theorem (Peterson, Ray, & Mittra, 1998), the periodicity of the scattering region 1 limits the electric field E_{y_0}, E_{y_2} in regions 0 and 2 to a discrete sum of plane waves

$$E_{y_0} = \mathrm{e}^{-jk_0(x\sin\theta - z\cos\theta)} + \sum_{p=-\infty}^{\infty} R_p \mathrm{e}^{-j(\beta_p x + \rho_p z)}, \tag{29}$$

$$E_{y_2} = \sum_{p=-\infty}^{\infty} T_p \mathrm{e}^{-j(\beta_p x + \tau_p z)}, \tag{30}$$

in which R_p and T_p are the reflection and transmission coefficients for the pth Floquet mode with lateral wavenumber β_p:

$$\beta_p = k_0 \sin\theta - p\frac{2\pi}{d} \tag{31}$$

and the perpendicular wavenumbers result from β_p as

$$\rho_p = \sqrt{k_0^2 - \beta_p^2}, \tag{32}$$

$$\tau_p = \sqrt{k_2^2 - \beta_p^2}. \tag{33}$$

In the above, the negative and positive signs are chosen for ρ_p and τ_p, respectively, if $\beta_p > k_0$ or $\beta_p > k_2$. The plane-wave expansions in (29) and (30)

are also called *Rayleigh expansions*. All modal numerical methods follow the same route up to this point. It is in the representation of the electric field in the inhomogeneous region 1 and the expression of the boundary conditions at $z = 0$ and $z = -h$ that various modal methods differ from each other. The *coupled-wave* method (Kogelnik, 1969; Moharam & Gaylord, 1981) is the most prominent and flexible method, capable of handling a multitude of periodic-structure types. In its most basic form, the coupled-wave method is applicable to gratings of the type shown in Figure 7(a) with straight, slanted, and parallel equi-permittivity lines $\epsilon(x, z) = [\text{const.}]$. The lateral period d is related to the slanted fringe period Λ by $d = \Lambda / \sin\phi$. The grating vector is defined as

$$\begin{aligned} \boldsymbol{K} &= K_x \hat{\boldsymbol{x}} + K_z \hat{\boldsymbol{z}} \\ &= K \sin\phi \hat{\boldsymbol{x}} + K \cos\phi \hat{\boldsymbol{z}}, \end{aligned} \tag{34}$$

where $K = 2\pi/\Lambda$. At the heart of the coupled-wave formulation is the expansion of the electric field in region 1 in the infinite series

$$E_{y_1} = \sum_{p=-\infty}^{\infty} S_p(z) \mathrm{e}^{-j\boldsymbol{\sigma}_p \cdot \boldsymbol{r}}, \tag{35}$$

where $\boldsymbol{r} = x\hat{\boldsymbol{x}} + z\hat{\boldsymbol{z}}$ and

$$\boldsymbol{\sigma}_p = \beta_p \hat{\boldsymbol{x}} - pK_z \hat{\boldsymbol{z}}. \tag{36}$$

The x component of $\boldsymbol{\sigma}_p$ is dictated by the phase-matching condition at $z = 0$ and $z = -h$, while the z component is chosen to yield simpler forms for the equations to follow (Moharam & Gaylord, 1983). Upon substitution of (35) into the Helmholtz equation (27), it is found that the z-dependent coefficients $S_p(z)$ with different p are coupled into each other through an infinite set of ordinary second-order differential equations in z. Thanks to the judicious choice of the z component in (36), the resulting differential equations have constant coefficients. In general every component $S_p(z)$ in the expansion is coupled with every other component $S_{p-i}(z)$, $i = \pm 1, \pm 2, \ldots$ Minimum coupling occurs when the permittivity fringes vary sinusoidally, in which case $S_p(z)$ is coupled only to $S_{p\pm1}(z)$ (Moharam & Gaylord, 1981). The constant-coefficient set of coupled differential equations involving the waves $S_p(z)$ can be put into a matrix form. First, the vector state variable $\boldsymbol{S}$ is defined as follows:

$$\boldsymbol{S}(z) = \begin{bmatrix} \boldsymbol{S_1}(z) \\ \boldsymbol{S_2}(z) \end{bmatrix}, \tag{37}$$

where $\boldsymbol{S_1}(z)$ and $\boldsymbol{S_2}(z)$ are the column vectors of $S_p(z)$ and $\mathrm{d}S_p(z)/\mathrm{d}z$. The coupled differential equations involving $S_p(z)$ then take the matrix form

$$\frac{\mathrm{d}}{\mathrm{d}z}\boldsymbol{S}(z) = \boldsymbol{B}\boldsymbol{S}(z), \tag{38}$$

in which the infinite matrix $\boldsymbol{B}$ is independent of z. Its elements are related simply to the parameters of the geometry. In practice, the expansion (35) is approximated by a finite sum of N modes, resulting in a $2N \times 2N$ matrix $\boldsymbol{B}$. The resulting finite-dimensional differential matrix equation (38) is then decoupled by projecting $\boldsymbol{S}(z)$ onto the $2N$ eigenvectors $\boldsymbol{w}_m$ of $\boldsymbol{B}$, yielding the solution

$$\boldsymbol{S}(z) = \sum_{m=0}^{2N-1} C_m \boldsymbol{w}_m \mathrm{e}^{q_m z}, \tag{39}$$

where q_m is the eigenvalue corresponding to the eigenvector w_m. The coefficients C_m of this expansion, as well as the reflection and transmission coefficients R_p and T_p, are determined from the boundary conditions at $z = 0$ and $z = -h$ ensuring the continuity of the tangential electric field E_y and magnetic field H_x. The N unknown reflection coefficients R_p, N unknown transmission coefficients T_p, and $2N$ unknown expansion coefficients C_m add up to $4N$ unknowns. On the other hand, two continuity equations at $z = 0$ and $z = -h$, each involving N modes, result in $4N$ equations in total. The number of equations thus equals the number of unknowns.

There is another modal formulation that is intimately related to the coupled-wave analysis, based on an *eigenmode* expansion (also called coordinate separation or separation of variables) of the fields inside region 1 (Burckhardt, 1966; Chu & Kong, 1977; Kaspar, 1973; Noponen & Turunen, 1994). Within the context of planar gratings of the form shown in Figure 7(a), this method is completely equivalent to the coupled-wave analysis (Gaylord & Moharam, 1982; Magnusson & Gaylord, 1978). The only difference here is that the field inside region 1 is represented in terms of the allowable modes of the periodic medium, which was not imposed a priori in (35). This makes the coupled-wave formulation slightly less complicated, and often the preferred method. Because of their equivalence, the coupled-wave and eigenmode approaches are sometimes both referred to as the *Fourier-modal method* in more recent literature (Ichikawa, Masuda, & Ueda, 2009; Li, 1997; Vallius, 2002; Vallius & Turunen, 2006).

If the permittivity profile has a rectangular cross-section as in Figure 7(b), the coupled-wave method is the simplest and fastest modal solution method (Knop, 1978; Lee & Degertekin, 2004; Moharam & Gaylord, 1986; Moharam, Grann, Pommet, & Gaylord, 1995). The piecewise-constant permittivity profile also allows an exact eigenmode analysis inside region 1 without recourse to an infinite Fourier expansion (Botten, Craig, McPhedran, Adams, & Andrewartha, 1981; Tayeb & Petit, 1984). The general surface-relief profile can also be approximated as a finite stack of lamellar gratings using a staircase approximation as shown in Figure 8 (Chateau & Hugonin, 1994; Moharam & Gaylord, 1982; Popov, Neviere, Gralak, & Tayeb, 2001).

There is a modal method that is especially suitable for the analysis of the surface-relief structure in Figure 7(c), called the Chandezon method, or the C-method (Chandezon, Dupuis, Cornet, & Maystre, 1982; Cotter, Preist, & Sambles, 1995; Li, 1999). The method is also applicable to multicoated gratings with layer interfaces positioned at $z = a(x) - L_i$, where $L_i,\ i = 1, 2, \ldots$ are a monotonically increasing sequence of positive constants. In this method, the interface curves $z = a(x) - L_i$ are mapped to parallel lines by a coordinate transformation, and Maxwell's equations are then solved in the planar multilayered medium in the new coordinate system. For gratings with homogeneous layers separated by smooth interfaces, the C-method outperforms the coupled-wave method paired with the staircase approximation of the surface-relief profile (Li, Chandezon, Granet, & Plumey, 1999). Even for discontinuous permittivity profiles, some algorithmic improvements can enable the C-method to yield comparable performance to the coupled-wave method (Vallius, 2002).

For the most general permittivity profile in Figure 7(d), the coupled-wave expansion in (35) is still valid. However, the resulting equations for $S_p(z)$ are much more complicated. They are still an infinite set of coupled ordinary differential equations for $S_p(z)$, but their coefficients are no longer independent of z. The most direct approach is to solve these equations

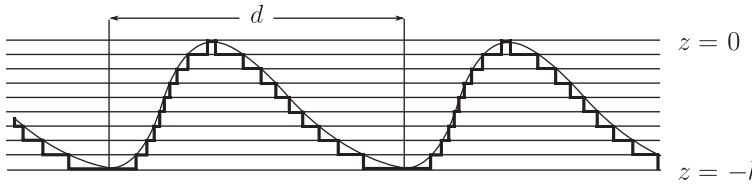

Figure 8 The surface-relief grating approximated by a finite stack of lamellar gratings.

directly using numerical quadrature. This approach, called the *differential* method, was first introduced by Nevière for studying the diffraction from surface-relief gratings (Neviere, Vincent, & Petit, 1974). Although its formulation is very general, the differential method has mostly been applied to homogeneous gratings with smooth profiles (Maystre & Neviere, 1978; Popov & Bonod, 2007; Popov & Neviere, 2000). The numerical difficulties associated with the direct quadrature solution of variable-coefficient coupled differential equations are rather formidable. On the other hand, the differential method is evidently superior to staircase-based approximate methods for smooth permittivity profiles (Popov et al., 2001). The latter methods suffer from serious convergence issues due to sharp local maxima at the ridges of the staircase profile.

One of the pioneering applications of modal methods to the simulation of optical imaging was reported by Nyyssonen (1982). This study is in fact one of the earliest applications of rigorous Maxwell's-equations-based analyses of the imaging of complex objects without the thin-film approximation. In this work, the eigenmode variation of the coupled-wave method was applied to optical edge detection. It has also found use in modeling integrated-circuit/grating/mask inspection (Nyyssonen & Kirk, 1988; Sheridan & Sheppard, 1993; Totzeck, 2001; Yang, Milster, Zhang, et al., 2010; Yuan & Strojwas, 1991b; Zhang et al., 2010), surface profilometry (Gale, Pether, & Dainty, 1996; Huttunen & Turunen, 1995), and defect detection (Rafler, Schuster, Frenner, Osten, & Seifert, 2008). The coupled-wave method was readily adopted and heavily used by the photolithography community, where it is more commonly referred to as the *waveguide* method. It has been used for mask modeling (Erdmann et al., 2006; Kundu, Mathur, & Gupta, 1986; Lucas et al., 1996; Lucas, Yuan, & Strojwas, 1992), line-width measurement (Davidson, 1999; Kundu et al., 1986; Yuan, 1992; Yuan & Strojwas, 1991a), and alignment (Nikolaev & Erdmann, 2003; Yuan, 1992; Yuan & Strojwas, 1991a).

The modal methods mentioned so far have been designed to handle periodic, infinite structures. There are also some modal methods that are applicable to spatially bounded, aperiodic scatterers. The most popular modal method for an isolated scatterer is the *Mie solution* for electromagnetic wave scattering from a sphere (Born & Wolf, 1999), later generalized to multiple spheres (Xu, 1995). Modal methods are only suitable for scatterers that have highly symmetric shapes such as spheres, ellipsoids, planar sheets, etc., that have a straightforward geometric representation in a reasonably simple coordinate system (Yang, Taflove, & Backman, 2011).

3.2.2 Finite Methods

Finite numerical methods can be broadly separated into two categories: differential-equation and integral-equation methods. Differential-equation methods are based on the direct discretization of the differential form of Maxwell's equations inside a volumetric solution space. Two prominent approaches in this category are the finite-difference (FD) and finite-element (FE) methods. In the FD approach, the field values are placed on regular (frequently Cartesian) grids in such a way that Maxwell's divergence equations (3) and (4) are automatically satisfied. Such "divergence-free" grids prevent the emergence of late-time instabilities in time-domain methods, and spurious modes in frequency-domain methods (Teixeira & Chew, 1999). The simplest and most famous divergence-free grid was developed by Yee (1966), and set the basis for the widely popular finite-difference time-domain (FDTD) method (Taflove, 1980; Taflove & Hagness, 2005). The most desirable property of the FDTD method is its ease of implementation. In its most basic form, the fields are updated in time using a simple leapfrog updating procedure, without any matrix inversions. In more complicated time-domain FD schemes, as well as frequency-domain FD schemes (called FDFD), matrix inversions may become necessary. In most cases, however, the leapfrog updating scheme of the FDTD method is preferred due to its simplicity and intuitiveness. The FDTD method also has the advantage of yielding direct time-domain data, allowing immediate broadband analysis. In spite of their simplicity, the regular FD grids can be overly restrictive when local mesh refinement or conformal gridding is required. Furthermore, staircase approximations have to be made for modeling curved surfaces in regular FD grids. When these constraints are too stringent, one might prefer to use highly irregular triangular meshes that allow a much finer discretization of round surfaces and much easier mesh refinement. In spite of the latitude they offer in representing different geometries, irregular grids require much more effort to ensure the consistency and stability of the numerical solution algorithm. Collectively, these algorithms are referred to as finite-element (FE) methods. The acronym "FEM" is commonly used in electromagnetics (Jin, 2002). Although they can be regarded as more general and less limiting than FD methods, they are considerably more difficult to implement. Finite-element formulations always involve the solution of a large matrix equation with the field values in the three-dimensional volume as the unknowns. Fortunately, the matrix is always sparse, banded, and in many cases, symmetric. A considerable part of the difficulty in implementing FE algorithms lies in solving this system

of equations efficiently, while avoiding spurious modes. As mentioned above, spurious modes arise when there are inconsistencies in the construction of the grid. Consistency in a FE scheme can be expressed as a combination of two conditions: a divergence-free topological construction of the grid, and preservation of the reciprocal nature of the continuum Maxwell's equations (Teixeira & Chew, 1999). Although frequency-domain problems constitute the majority of FE applications, time-domain FE variants also exist (Teixeira, 2008). There are even time-domain FE formulations featuring explicit time-updating schemes like the FD methods; however, this feature comes with its own disadvantages (Lee, Lee, & Cangellaris, 1997). FD and FE methods share some inherent drawbacks that are a result of their very construction. Since both methods operate on field values in a finite volume, they require auxiliary techniques for handling sources that radiate in an unbounded region. Many so-called "absorbing boundary conditions" (ABCs) have been developed for truncating the FD and FE solution spaces (Taflove & Hagness, 2005). These conditions emulate the extension of the simulation grid to infinity by absorbing most of the energy incident on the outer boundary of the grid. The most popular ABC in use today is Berenger's "perfectly matched layer" (Berenger, 1994), which constitutes a milestone in the development of differential-equation methods. Grid dispersion and grid anisotropy are two other major problems caused by the finite size of the grid voxels and their lack of rotational symmetry. These problems can never be completely eliminated, but can be partially alleviated by choosing finer spatial steps and/or employing more rotationally symmetric discretization schemes (Taflove & Hagness, 2005). Hybrid algorithms have been proposed for merging useful features of FD and FE methods. For example, the hybrid FDTD-FEM method described in Wu and Itoh (1997) makes use of both the simplicity of the FDTD method and the power of the finite-element method in modeling curved surfaces. This hybrid method was used in Yeung and Barouch (1999) to model curved mask surfaces in photolithography.

An integral-equation method operates on an equivalent integral equation that completely embodies the differential form of Maxwell's equations *and* the associated boundary conditions, including the ones at infinity (Chew, 1990; Peterson et al., 1998). Barring a few exceptions, integral-equation methods are used mainly to solve the frequency-domain Maxwell's equations (5)–(8). The integral equation is typically expressed on a two-dimensional boundary that separates two homogeneous material regions, with the unknowns representing the field values on the boundary.

Consequently integral-equation methods are sometimes referred to as "boundary integral" (or "boundary element") methods. The reduction in the dimensionality of the problem allows for a much finer spatial discretization. However, the applicability of integral-equation methods is limited to geometries consisting of piecewise-homogeneous material regions for which the Green's function is at least conceptually available in each homogeneous region. Consequently, integral-equation methods cannot be applied when nonlinear or highly inhomogeneous materials exist in the problem geometry. For the latter case, volume-integral equations can be formulated (Chew, 1990); however, these are less common. An alternative method is to adopt a finite-element or a finite-difference discretization scheme inside the inhomogeneous regions and an integral-equation method outside these regions, with boundary conditions to tie these regions together (Kotlyar & Nesterenko, 2001; Mirotznik, Prather, & Mait, 1996; Prather, Shi, & Sonstroem, 2002).

The primary difficulty associated with an integral-equation method is the necessity of solving a dense matrix system whose diagonal elements include singular integrals. Usually, some iterative method with an acceleration algorithm such as the fast multipole method (FMM) is used for solving this system (Coifman, Rokhlin, & Wandzura, 1993). On the other hand, integral-equation methods have the distinct advantage that the radiation boundary conditions are inherent in their formulation, rendering absorbing boundary conditions unnecessary. Furthermore, integral-equation methods do not suffer from some discretization artifacts that are inherent to differential-equation methods such as grid dispersion and grid anisotropy.

With the widespread availability of powerful computational resources, rigorous numerical approximations to Maxwell's equations reached and surpassed the analytical series solutions in popularity. Among these, the FDTD method seems to have gained wider acceptance than others primarily because it is conceptually simpler, physically more intuitive, and easier to implement. One of the earliest applications of FDTD to numerical imaging is UC Berkeley's TEMPEST software developed for photolithography simulation (Guerrieri et al., 1991). Originally designed to handle 2D mask patterns, it was later generalized to 3D (Wong, 1994; Wong & Neureuther, 1994; Wong et al., 1995) and further enhanced to handle extreme-ultraviolet (EUV) photolithography simulation (Brukman, Deng, & Neureuther, 2000; Pistor, 2001; Pistor, Adam, & Neureuther, 1998; Pistor & Neureuther, 1999a, 1999b). In addition to its initial purpose of simulating the scattering response of photomasks (Adam & Neureuther, 2002; Tirapu Azpiroz,

2004), TEMPEST has also been used for simulating metrology (Tadros, Neureuther, & Guerrieri, 1991) and alignment systems (Deng, Pistor, & Neureuther, 2001; Yin et al., 2000). The FDTD method has also found use in modeling microscopy modalities. The earliest work on this subject began with the simulation of near-field imaging modalities such as the near-field scanning optical microscope (NSOM) (Furukawa & Kawata, 1996; Krug, Sanchez, & Xie, 2002; Simpson & Hanna, 2001; Symons, Whites, & Lodder, 2003; Vasilyeva & Taflove, 1998a, 1998b). Far-field microscopy modeling was later tackled by the incorporation of ray principles and diffraction formulas from optics into the solution algorithm (Capoglu et al., 2011; Hollmann et al., 2004; Török et al., 2008). Upon suitable modification of the optical far-field data, different modalities such as differential-contrast microscopy (DIC) (Munro & Török, 2005), phase-contrast microscopy (Tanev, Tuchin, et al., 2008), and confocal microscopy (Salski & Gwarek, 2009a; Simon & DiMarzio, 2007) can be handled. A novel algorithm based on the extended Nijboer–Zernike (ENZ) theory of diffraction (see Section 3.4) was coupled with FDTD for photomask imaging (Janssen et al., 2008; van Haver et al., 2008) and the imaging of general 3D objects (van Haver, Braat, Janssen, Janssen, & Pereira, 2009). A variant of the FDTD method, called the pseudo-spectral time-domain (PSTD) method (Liu, 1997), is particularly suited to the analysis of scattering from optically large structures, such as macroscopic biological tissue (Tseng, 2007).

FE methods have found less usage than FD methods in simulating optical imaging systems. As mentioned above, this is primarily due to the difficulty of implementation. On the other hand, if geometrical flexibility is a crucial requirement in the spatial discretization of the scattering object, a FE method may be a viable option. An explicit time-domain FE scheme has been used to simulate the optical imaging of cylindrical dielectric objects on silicon surfaces (Wojcik, Vaughan, & Galbraith, 1987), integrated-circuit line features (Wojcik et al., 1991a), alignment marks (Wojcik et al., 1991b), and phase-shifting masks (Wojcik, John Mould, Ferguson, Martino, & Low, 1994). Frequency-domain FE methods were applied to various complex diffracting structures (Lichtenberg & Gallagher, 1994; Wei et al., 2007). The ease of mesh refinement in FE discretization has led to adaptive approaches to the modeling of optical diffractive devices (Bao, Chen, & Wu, 2005).

Among all the rigorous numerical solution methods for optical scattering, FD and FE methods clearly dominate. Still, there may be certain situations where an integral-equation method is the best choice for solving an optical scattering problem. For example, if the near field in the vicinity of

the scattering structure is very intense, absorbing boundary conditions for FD or FE methods may get too cumbersome or numerically intensive. In this case, the implicit radiation condition provided by an integral equation becomes a handy feature. However, integral-equation methods can only be effectively used in optical problems where the scattering structure is homogeneous or piecewise homogeneous. Consequently, their usage has been limited to the numerical simulation of optical scattering from relatively simple diffractive structures. Some examples are dielectric-coated gratings (Botten, 1978; Kleemann, Mitreiter, & Wyrowski, 1996; Maystre, 1978), diffractive lenses (Prather, Mirotznik, & Mait, 1997), and integrated-circuit structures (Aguilar & Mendez, 1994; Aguilar et al., 2002; Marx, 2007).

It is a trade-off between the accuracy requirements and the resource constraints that determines the numerical method best suited to a problem. For guidance, one can consult benchmark studies that compare multiple numerical methods for a given application (Besbes et al., 2007; Erdmann, Fuhner, Shao, & Evanschitzky, 2009; Nikolaev & Erdmann, 2003; Vallius & Turunen, 2006; Wojcik, Mould, Marx, & Davidson, 1992).

3.3 Collection

After the scattering from the sample is calculated, the scattered field should be propagated to the image plane to complete the imaging process. These two steps are commonly performed by an *objective*. As shown diagrammatically in Figure 1, the first task of the objective is to collect the portion of the light scattered from the sample that falls within its entrance pupil. The entrance pupil is defined as the image of the limiting aperture stop as seen from the object side. Among all the aperture stops in the system, the one that limits the angular extent of the rays emanating from the object is the limiting aperture (Born & Wolf, 1999). In general, the entrance pupil can be located at any finite distance from the sample. However, it is more advantageous to place the entrance pupil at infinity whenever possible. Such a system is referred to as *object-side telecentric* (Born & Wolf, 1999). In a telecentric system, the size of the blurred image does not change as the object moves out of focus. This is a very desirable feature in both microscopy and photolithography. In the former, telecentricity makes it easier to manually focus on a sample, as the image size does not change with defocus. In the latter, telecentricity is much more crucial, for it ensures that the image of the photomask is much less sensitive to positioning tolerances. In addition to its advantages, telecentricity poses its own difficulties. First of all, it should be remembered that the limiting aperture stop should be

located at the back focal plane of the optical system preceding it. If the lateral dimension of a certain lens in the optical setup is not large enough, that lens could act as the limiting aperture and destroy telecentricity. For this reason, telecentricity usually requires that the lenses have much larger lateral dimensions than the sample, resulting in more expensive equipment. This is usually not a big issue for microscopes. However, telecentricity places a more stringent constraint on photolithography projection lenses in terms of both price and performance. Almost all modern microscopes and photolithography projection lenses are telecentric on the object side. We will assume object-side telecentricity in the following analysis.

The collection geometry for an object-side telecentric system is shown in Figure 9. Since the entrance pupil is at infinity, the scattering object can be regarded as a point at O, and the scattered wavefront W_s is spherical. The conventional spherical coordinates (r, θ, ϕ) are defined with respect to O and the z axis chosen to coincide with the optical axis. The *far zone* (also called the *Fraunhofer* or *radiation* zone) is defined as the region $r \gg d^2/\lambda$ where d is the maximum dimension of the sample and λ is the wavelength. In the far zone, the radial dependence of the field can be factored out, and the wavefront W_s is completely specified by its angular profile (Harrington, 2001; Stratton, 2007):

$$\boldsymbol{E}_s(r, \theta, \phi) = \boldsymbol{E}_s(\theta, \phi)\frac{e^{-jkr}}{r}. \tag{40}$$

Here, $k = nk_0$ is the wavenumber in the homogeneous space between the object and the entrance pupil, and n is the refractive index of the same homogeneous space. The vector quantity $\boldsymbol{E}_s(\theta, \phi)$ in (40) is called the *strength factor* of the ray associated with the far-zone direction (θ, ϕ)

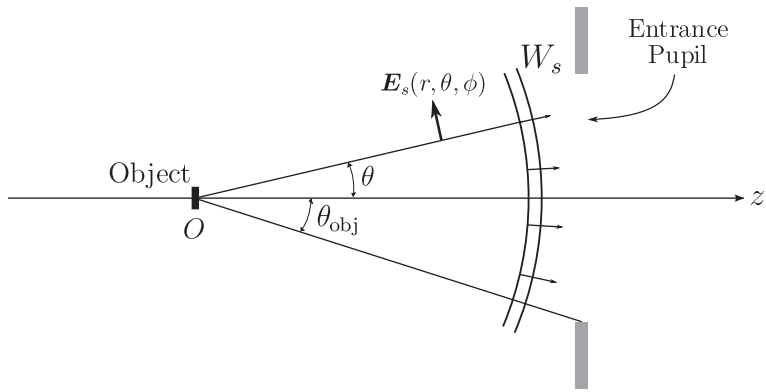

Figure 9 The collection geometry for a telecentric system.

(Kline & Kay, 1979; Richards & Wolf, 1959). The collection step therefore reduces to numerically calculating the strength factor $\boldsymbol{E}_{\mathrm{s}}(\theta, \phi)$ at various observation directions. Depending on the scattering geometry, this may be accomplished in several different ways. One common property of almost all numerical collection methods is that the near-field information is used to obtain the far-zone field using certain theorems of electromagnetics. The term *near-field-to-far-field transform* (NFFFT) is commonly used for such an algorithm that computes the far field from the near field. All NFFFT algorithms rely on either spatial Fourier analysis or a Green's-function formalism. We will examine these two cases separately.

3.3.1 Fourier Analysis

In certain cases, the strength factor $\boldsymbol{E}_{\mathrm{s}}(\theta, \phi)$ can be found using the spatial Fourier transform of the near field around the scatterer. This near field should be given on an infinite planar surface S between the scatterer and the entrance pupil, as shown in Figure 10. Let us define the origin O on S and denote the vector electric field on this plane as $\boldsymbol{E}(x, y)$. The far-zone field $\boldsymbol{E}_{\mathrm{s}}(r, \theta, \phi)$ on the wavefront W_{s} at the entrance pupil can be found by expanding $\boldsymbol{E}(x, y)$ into its plane-wave (or angular) spectrum, and propagating it to very large r using the steepest-descent method (Smith, 1997). Let us define the plane-wave spectrum of the 2D electric-field distribution $\boldsymbol{E}(x, y)$ as the following Fourier-transform operation

$$\widetilde{\mathbf{E}}(k_x, k_y) = \iint_{-\infty}^{\infty} \boldsymbol{E}(x, y) \mathrm{e}^{j(k_x x + k_y y)} \, \mathrm{d}x \mathrm{d}y \tag{41}$$

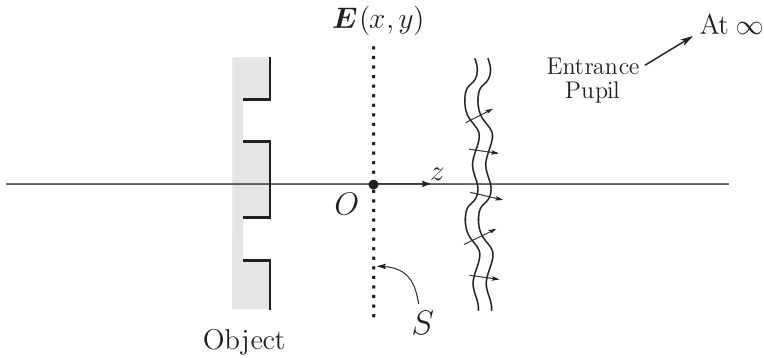

Figure 10 The collection geometry for a Fourier-analysis-based near-field-to-far-field transform. The near fields should be given on an infinite planar surface S between the scatterer and the entrance pupil.

with the inverse transform (or the plane-wave representation)

$$\boldsymbol{E}(x, y) = \frac{1}{(2\pi)^2} \iint_{-\infty}^{\infty} \widetilde{\mathbf{E}}(k_x, k_y) \mathrm{e}^{-j(k_x x + k_y y)} \, \mathrm{d}k_x \mathrm{d}k_y. \tag{42}$$

It is understood in (41) and (42) that the Fourier transform is applied to the Cartesian components of the vector integrands separately. The representation (42) for the vector field $\boldsymbol{E}(x, y)$ is an infinite summation of plane waves (propagating and evanescent) whose lateral propagation coefficients are (k_x, k_y). This plane-wave representation can be used to extrapolate the electric field to the region above the plane S, i.e., $z > 0$. The following field satisfies the boundary condition (42) at $z = 0$ and Maxwell's equations (5)–(8) in the region $z > 0$:

$$\boldsymbol{E}(x, y, z) = \frac{1}{(2\pi)^2} \iint_{-\infty}^{\infty} \widetilde{\mathbf{E}}(k_x, k_y) \mathrm{e}^{-j(k_x x + k_y y + k_z z)} \, \mathrm{d}k_x \mathrm{d}k_y, \tag{43}$$

with $k_z = (k^2 - k_x^2 - k_y^2)^{1/2}$ (Smith, 1997). Here, k is the wavenumber ω/c in the homogeneous space between S and the entrance pupil. Only the plane-wave components in (42) with $k_x^2 + k_y^2 < k^2$ will contribute to the far field; because a complex k_z represents an evanescent plane wave decaying exponentially in z. Now, propagating the plane waves with $k_x^2 + k_y^2 < k^2$ into the space $z > 0$ and using the steepest-descent method at the far zone, one arrives at the following formula for the strength factor $E_s(\theta, \phi)$ (Smith, 1997):

$$\boldsymbol{E}_{\mathrm{s}}(\theta, \phi) = E_\theta(\theta, \phi)\hat{\boldsymbol{\theta}} + E_\phi(\theta, \phi)\hat{\boldsymbol{\phi}}, \tag{44}$$

where the θ and ϕ components are given by

$$E_\theta(\theta, \phi) = \frac{jk}{2\pi} \left[\widetilde{E}_x(\alpha, \beta) \cos\phi + \widetilde{E}_y(\alpha, \beta) \sin\phi \right], \tag{45}$$

$$E_\phi(\theta, \phi) = \frac{jk}{2\pi} \cos\theta \left[-\widetilde{E}_x(\alpha, \beta) \sin\phi + \widetilde{E}_y(\alpha, \beta) \cos\phi \right], \tag{46}$$

in which $\widetilde{E}_x(k_x, k_y)$ and $\widetilde{E}_y(k_x, k_y)$ are the x and y components of the plane-wave spectrum $\widetilde{\mathbf{E}}(k_x, k_y)$ in (41), and the definition

$$(\alpha, \beta) = (k \cos\phi \sin\theta, k \sin\phi \sin\theta) \tag{47}$$

has been introduced for brevity. This expression can be put into a more compact vectorial form. The electric-field divergence equation (7) in free

space ($\rho = 0$) applied to (43) will quickly reveal that $\widetilde{\mathbf{E}}(k_x, k_y)$ is transverse to the propagation vector $\bar{k} = k_x\hat{x} + k_y\hat{y} + k_z\hat{z}$. It follows that the vector $\widetilde{\mathbf{E}}(\alpha, \beta)$ only possesses $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\phi}}$ components. Expanding $\widetilde{E}_x$ and $\widetilde{E}_y$ in (45) and (46) in terms of $\widetilde{E}_\theta$ and $\widetilde{E}_\phi$, it is readily found that

$$\boldsymbol{E}_s(\theta, \phi) = \frac{jk\cos\theta}{2\pi}\widetilde{\mathbf{E}}(\alpha, \beta). \tag{48}$$

We will now show how the expressions (44)–(48) can be used in different scattering geometries to calculate the strength factor $\boldsymbol{E}_s(\theta, \phi)$ numerically.

The simplest scattering geometry results when the sample is a *phase object*, represented by a complex transmittance $T(x, y)$ (see Section 3.2). This concept can be generalized to vector fields by assuming that the complex transmittance acts on the tangential component of the incident field. The normal component is then determined by the transversality property $\bar{k} \cdot \widetilde{\mathbf{E}}(k_x, k_y) = 0$. The phase-transmittance property suggests that the plane S in Figure 10 be defined immediately above the sample, opposite the illumination side. If the incident field is a unit-amplitude plane wave

$$\boldsymbol{E}_i(x, y, z) = \hat{\boldsymbol{a}}\mathrm{e}^{-jk(\gamma_x x + \gamma_y y + \gamma_z z)}, \tag{49}$$

the tangential component of the transmitted field on the plane S is given by

$$\begin{aligned}\boldsymbol{E}_+^{\parallel}(x, y) &= E_x(x, y, 0^+)\hat{\boldsymbol{x}} + E_y(x, y, 0^+)\hat{\boldsymbol{y}} \\ &= \mathrm{e}^{-jk(\gamma_x x + \gamma_y y)}\, T(x, y)(a_x\hat{\boldsymbol{x}} + a_y\hat{\boldsymbol{y}}),\end{aligned} \tag{50}$$

where a_x, a_y are the x and y components of the incident unit polarization vector $\hat{\boldsymbol{a}}$. The plane-wave spectrum of the vector field in (50) is given by

$$\widetilde{\mathbf{E}}_+^{\parallel}(k_x, k_y) = \widetilde{T}(k_x - k\gamma_x, k_y - k\gamma_y)(a_x\hat{\boldsymbol{x}} + a_y\hat{\boldsymbol{y}}), \tag{51}$$

where $\widetilde{T}(k_x, k_y)$ is the 2D Fourier transform of the complex transmittance $T(x, y)$ according to (41). The frequency-shift property of the Fourier transform was used in deriving the above expression from (50). Substituting (51) in (45), (46), we obtain the following expression for the θ and ϕ components of the strength factor $\boldsymbol{E}_s(\theta, \phi)$:

$$\begin{aligned}E_\theta(\theta, \phi) = \frac{jk}{2\pi}&\widetilde{T}(\alpha - k\gamma_x, \beta - k\gamma_y) \\ &\times \left[a_x\cos\phi + a_y\sin\phi\right],\end{aligned} \tag{52}$$

$$E_\phi(\theta,\phi) = \frac{jk}{2\pi}\widetilde{T}(\alpha - k\gamma_x, \beta - k\gamma_y) \times \cos\theta\left[-a_x\sin\phi + a_y\cos\phi\right]. \tag{53}$$

If the object is not sufficiently thin, the model can be improved by calculating the response $\boldsymbol{E}_+^0(x, y)$ of the object to a normally incident plane wave by some rigorous method and obtaining the responses to other incidences perturbatively from $\boldsymbol{E}_+^0(x, y)$. As mentioned in Section 3.2, this approach is viable only if the illumination and collection NAs are very small (Yeung et al., 1993). Such an assumption is often valid in photolithography, but not microscopy (Totzeck, 2006) (see Figure 14). Let us assume that $\boldsymbol{E}_+^0(x, y)$ is the response to an x-polarized unit-amplitude plane wave. The corresponding response to a y-polarized plane wave can be added to the final result in the general case and is omitted here for brevity. Under the perturbation assumption, the tangential response to an x-polarized off-axis plane wave becomes

$$\boldsymbol{E}_+^{\parallel}(x, y) = (E_{x_+}^0(x, y)\hat{\boldsymbol{x}} + E_{y_+}^0(x, y)\hat{\boldsymbol{y}})\mathrm{e}^{-jk(\gamma_x x + \gamma_y y)}. \tag{54}$$

With this assumption, (52) and (53) become

$$E_\theta(\theta,\phi) = \frac{jk}{2\pi}\left[\widetilde{E}_{x_+}^0(\alpha - k\gamma_x, \beta - k\gamma_y)\cos\phi + \widetilde{E}_{y_+}^0(\alpha - k\gamma_x, \beta - k\gamma_y)\sin\phi\right], \tag{55}$$

$$E_\phi(\theta,\phi) = \frac{jk}{2\pi}\cos\theta\left[-\widetilde{E}_{x_+}^0(\alpha - k\gamma_x, \beta - k\gamma_y)\sin\phi + \widetilde{E}_{y_+}^0(\alpha - k\gamma_x, \beta - k\gamma_y)\cos\phi\right], \tag{56}$$

where $\widetilde{E}_{x_+}^0(k_x, k_y)$ and $\widetilde{E}_{y_+}^0(k_x, k_y)$ are the 2D Fourier transforms of $E_{x_+}^0(x, y)$ and $E_{y_+}^0(x, y)$.

When neither the thin-film assumption in (50) nor the perturbation assumption in (54) is valid, the scattered electric field $\boldsymbol{E}(x, y)$ on the plane S and its plane-wave spectrum $\widetilde{\mathbf{E}}(k_x, k_y)$ should be calculated using rigorous numerical methods detailed in Section 3.2 for a given incident beam. If a modal method is employed (see Section 3.2.1), the plane-wave spectrum $\widetilde{\mathbf{E}}(k_x, k_y)$ is obtained *directly*, eliminating any need for 2D Fourier transform. Since the $z = 0^+$ plane in Figure 7 is between the scattering structure and the entrance pupil, it can be chosen as the collection plane S in Figure 10.

The second term on the right-hand side of (29) is the scattered electric field on S for TE illumination:

$$E_y(x) = \sum_{p=-\infty}^{\infty} R_p \mathrm{e}^{-j\beta_p x}, \tag{57}$$

with the equally spaced Floquet wavenumbers $\beta_p = k\sin\theta_i - p(2\pi/d)$, in which d is the lateral period of the structure and θ_i is the longitudinal incidence angle of the excitation plane wave. Comparing (57) with (42), it is seen that the reflection coefficients R_p play the role of the plane-wave spectral coefficients $\widetilde{\mathbf{E}}(k_x, k_y)$,while β_p play the role of k_x. One difference is that R_p is a finite set of numbers unlike $\widetilde{\mathbf{E}}(k_x, k_y)$ because of the periodicity of the structures in Figure 7. Another difference is that R_p is defined for a 2D geometry that is invariant in y. If the scattering structure is periodic in both x and y with periods d_x and d_y, the scattered field should be expressed as a doubly infinite sum of vector-valued Floquet modes R_{pq} (Maystre & Neviere, 1978):

$$\boldsymbol{E}(x, y) = \sum_{p=-\infty}^{\infty}\sum_{q=-\infty}^{\infty} \boldsymbol{R}_{pq} \mathrm{e}^{-j(\beta_p x + \beta_q y)}, \tag{58}$$

where now the Floquet wavenumbers

$$\begin{aligned} \beta_p &= k\cos\phi_i \sin\theta_i - p(2\pi/d_x), \\ \beta_q &= k\sin\phi_i \sin\theta_i - q(2\pi/d_y) \end{aligned} \tag{59}$$

play the roles of k_x and k_y, respectively. Here, ϕ_i and θ_i are the spherical incidence angles of the incident plane wave. Comparing this expression with (42), the plane-wave spectrum $\widetilde{\mathbf{E}}(k_x, k_y)$ can be written in terms of R_{pq} as follows:

$$\widetilde{\mathbf{E}}(k_x, k_y) = (2\pi)^2 \sum_{p=-\infty}^{\infty}\sum_{q=-\infty}^{\infty} \boldsymbol{R}_{pq}\delta(k_x - \beta_p)\delta(k_y - \beta_q), \tag{60}$$

where $\delta(\cdot)$ is the Dirac delta function. Substituting this expression in (48), the strength factor $\boldsymbol{E}_s(\theta, \phi)$ is obtained as

$$\boldsymbol{E}_s(\theta, \phi) = (jk2\pi\cos\theta)\sum_{p}\sum_{q} \boldsymbol{R}_{pq}\delta(\alpha - \beta_p)\delta(\beta - \beta_q). \tag{61}$$

Carrying the term $\cos\theta = (1 - (\alpha/k)^2 - (\beta/k)^2)^{1/2}$ inside the summation and using the properties of the delta function, this can be written as

$$\boldsymbol{E}_s(\theta, \phi) = (jk2\pi) \sum_p \sum_q c_{pq} \boldsymbol{R}_{pq} \delta(\alpha - \beta_p) \delta(\beta - \beta_q), \tag{62}$$

in which c_{pq} is the dimensionless cosine parameter

$$c_{pq} = \sqrt{1 - (\beta_p/k)^2 - (\beta_q/k)^2}. \tag{63}$$

It is seen that the far-zone field is nonzero only at a discrete set of directions. This is a direct result of the periodicity of the system, and the discreteness of the Floquet wavenumbers β_p, β_q. Second, the finite range of the variables $\alpha = k\cos\phi\sin\theta$, $\beta = k\sin\phi\sin\theta$ between 0 and $k\sin\theta_{obj}$ (where $\sin\theta_{obj}$ is the collection NA in Figure 9) only allows for a finite number of observation directions to be collected by the objective. It is easy to see that any Floquet mode $\boldsymbol{R}_{pq}$ with p index higher than a maximum value p_{max} will not be collected by the objective, where

$$p_{max} = \frac{d\sin\theta_{obj}}{\lambda}. \tag{64}$$

The same concept applies to the q index. For the best reconstruction of the scattered field at the image plane, p_{max} should be maximized. This can be accomplished by reducing the wavelength λ or increasing the collection NA $\sin\theta_{obj}$. On the other extreme, if p_{max} is less than 1, only the homogeneous zeroth-order mode ($p = 0$) propagates to the objective, resulting in a uniform image (see Section 3.4, specifically Figure 16).

For finite numerical methods (see Section 3.2.2), it is not the diffracted orders $\boldsymbol{R}_{pq}$ but the electromagnetic field $\boldsymbol{E}(x, y)$ that is directly available. If an integral-equation method is used, the collection algorithm is usually based on a Green's-function formalism (see Section 3.3.2). Collection algorithms based on Fourier analysis are more widely used in conjunction with differential-equation methods such as FDTD or FEM. One difficulty associated with using Fourier analysis in differential-equation methods is that the computational grids are always bounded in space. Since Fourier analysis requires that the near field be given on an infinite planar surface, the range of applications for Fourier-based collection algorithms is severely limited. One broad category of problems that falls within this range is those that employ *periodic boundary conditions*. In such problems, both the simulation geometry and the incident beam are constrained to be periodic along a certain direction or two orthogonal directions. The electromagnetic field

on an infinite lateral plane (parallel to the direction of periodicity) is therefore determined completely by the electromagnetic field in the finite grid. This allows the use of Fourier analysis for the collection of the far-zone field. It is sufficient to consider the simpler 2D case in order to illustrate the concept of periodic boundary conditions. Some examples of 2D periodic structures are shown in Figure 7. Let us assume a unit-amplitude TE plane-wave incident from angle θ_i with respect to the z axis, with y component

$$E_y^i(x, z) = \mathrm{e}^{-jk(\gamma_x x - \gamma_z z)}, \tag{65}$$

in which $\gamma_x = \sin\theta_i$ and $\gamma_z = \cos\theta_i$. The spatial period of the structure is d; therefore, the Floquet theorem (Peterson et al., 1998; Taflove & Hagness, 2005) says that the electromagnetic field obeys the following pseudo-periodicity condition everywhere:

$$E_y(x + d, z) = E_y(x, z)\mathrm{e}^{-jk\gamma_x d}. \tag{66}$$

The linear phase term $\exp(-jk\gamma_x d)$ is enforced by the incident plane wave and becomes unity for normal incidence. The incorporation of periodic boundary conditions to a frequency-domain finite-element scheme is pretty straightforward. Assuming that the lateral dimension of the FEM grid is equal to the spatial period d, the pseudo-periodicity condition (66) is imposed at the opposite boundaries of the grid:

$$\{E, H\}_{\text{right}} = \{E, H\}_{\text{left}}\mathrm{e}^{-jk\gamma_x d}. \tag{67}$$

If the periodicity is along two directions, the same condition applies between the "back" and "front" boundaries as well, with $\gamma_y y$ instead of $\gamma_x x$. There is no extra conceptual difficulty introduced by the pseudo-periodicity condition (67), although there may be numerical issues regarding the efficient construction of the resulting matrix (Mias, Webb, & Ferrari, 1999; Sukumar & Pask, 2009). There are much larger implementation obstacles for time-domain methods such as FDTD, which will not be reviewed here. For a good review on periodic boundary conditions in FDTD, the reader may consult Taflove and Hagness (2005). A newer method for enforcing periodic boundary conditions in FDTD can be found in Lee and Smith (2006). Let us assume that the vector field $\boldsymbol{E}(x, y)$ scattered from a 3D object has been computed on the plane S (see Figure 10) using some finite numerical method at equally spaced spatial points $m\Delta x$ and $n\Delta y$, resulting in the discrete array $\boldsymbol{E}[m, n]$:

$$\begin{aligned} \boldsymbol{E}[m, n] = \boldsymbol{E}(m\Delta x, n\Delta y), \quad & m = 0 \ldots M - 1, \\ & n = 0 \ldots N - 1 \end{aligned} \tag{68}$$

and the entire periods in both x and y are covered by the sampling:

$$M\Delta x = d_x, \quad N\Delta y = d_y. \tag{69}$$

We will now describe how the vector amplitudes of the Floquet modes in (58) can be obtained by using two-dimensional discrete Fourier transform (DFT). Expressing the results in terms of DFT is always advantageous, since there exists a very efficient algorithm for the evaluation of the DFT called the fast Fourier transform (FFT). We first define the phase-shifted sampled array $\bar{\boldsymbol{E}}[m, n]$ as follows:

$$\bar{\boldsymbol{E}}[m, n] = \boldsymbol{E}[m, n]\mathrm{e}^{jk\sin\theta_i(d_x\cos\phi_i\frac{m}{M}+d_y\sin\phi_i\frac{n}{N})}. \tag{70}$$

The phase shift above depends on the direction of incidence (θ_i, ϕ_i) of the excitation plane wave. This shift removes the phase condition (66) imposed by the incident plane wave and simplifies the resulting expression considerably. The 2D DFT of this modified array is conventionally defined as (Bracewell, 1986; Oppenheim et al., 1999)

$$\tilde{\bar{\mathbf{E}}}[p, q] = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\bar{\boldsymbol{E}}[m, n]\mathrm{e}^{-j2\pi(\frac{pm}{M}+\frac{qn}{N})}. \tag{71}$$

It can be shown (see Appendix B) that the DFT array $\tilde{\bar{\mathbf{E}}}[p, q]$ is related to the Floquet modes $\boldsymbol{R}_{pq}$ as follows:

$$\tilde{\bar{\mathbf{E}}}[p, q] = MN\sum_{r=-\infty}^{\infty}\sum_{s=-\infty}^{\infty}\boldsymbol{R}_{p+rM,q+sN}. \tag{72}$$

Equation (72) expresses the results of the 2D DFT operation on the phase-shifted sampled field array $\bar{\boldsymbol{E}}[m, n]$ in terms of the Floquet modes $\boldsymbol{R}_{pq}$ of the original, *continuous* field $\boldsymbol{E}(x, y)$. From (72), we can immediately draw some key conclusions. First of all, the DFT array $\tilde{\bar{\mathbf{E}}}[p, q]$ is seen to be equal to an infinite summation of shifted copies of $\boldsymbol{R}_{pq}$ in both p and q indices. In other words, $\tilde{\bar{\mathbf{E}}}[p, q]$ is an aliased version of R_{pq}. In order for $\tilde{\bar{\mathbf{E}}}[p, q]$ to represent R_{pq} faithfully, the shifted copies of R_{pq} should not overlap. This requires that the shifting periods M and N be larger than the effective widths W_p and W_q of R_{pq} in the p and q indices:

$$M > W_p, \quad N > W_q, \tag{73}$$

which, using (69), can also be written in terms of the sampling periods as

$$\Delta x < \frac{d_x}{W_p}, \quad \Delta y < \frac{d_y}{W_q}. \tag{74}$$

If M or N is too small, shifted copies of $\boldsymbol{R}_{pq}$ overlap, and $\boldsymbol{R}_{pq}$ cannot be recovered fully from $\widetilde{\widetilde{\mathbf{E}}}[p, q]$ If both M and N are large enough so that neighboring replicas of $\boldsymbol{R}_{pq}$ do not overlap, $\boldsymbol{R}_{pq}$ can be recovered from $\widetilde{\widetilde{\mathbf{E}}}[p, q]$ using the relationship (72):

$$\boldsymbol{R}_{pq} = \frac{1}{MN}\widetilde{\widetilde{\mathbf{E}}}[p, q] \tag{75}$$

for a range of p, q values around $p = q = 0$. Some simple estimates for the effective widths W_p and W_q of $\boldsymbol{R}_{pq}$ can be made in certain circumstances. For example, if the plane S is *far enough* above the periodic structure, the evanescent Floquet modes in (58) become negligible, and it is only necessary to consider the propagating modes. Using the propagation conditions $|\beta_p| < k$ and $|\beta_q| < k$, the following expressions are obtained for W_p and W_q:

$$W_p = 2\frac{d_x}{\lambda}, \quad W_q = 2\frac{d_y}{\lambda}. \tag{76}$$

Substituting these expressions into (74), we obtain

$$\Delta x < \frac{\lambda}{2}, \quad \Delta y < \frac{\lambda}{2}. \tag{77}$$

This is, in effect, a discrete version of the celebrated Nyquist sampling theorem for bandlimited signals (Oppenheim et al., 1999). It states that if only propagating modes are present on the plane S, the spatial sampling periods need only be smaller than half the wavelength.

Fourier-analysis-based NFFFTs have the advantage that they do not require separate treatment for multilayered media. This is because the sampling plane S lies above the scattering structure and any stratification over which it might be situated. However, these methods have their own difficulties. The most important of these is the requirement that the field be specified on a laterally infinite plane S. It was shown in the foregoing discussion that this is a surmountable problem if the scattering structure is periodic in space. A Fourier-analysis-based NFFFT might still be feasible even for non-periodic structures if both the scattered field and the reflection of the incident field from the interfaces are bounded in space. Then,

the collection surface S can in principle be made large enough to cover them. Nevertheless, the preferred method for non-periodic structures remains the Green's-function formalism explained in the next subsection.

3.3.2 Green's-Function Formalism

If the scattering medium is not periodic, the Fourier-analysis-based approaches of the previous subsection have very limited use. The preferred method for a non-periodic scatterer is usually based on a Green's-function formalism, which operates by applying certain integral theorems to the near fields on a closed surface surrounding the scatterer and computes the far field by the finite approximation of these integrals. The geometry of a Green's-function NFFFT is shown in Figure 11(a). The scattering (or radiating) structure A is enclosed in a closed surface S, on which the tangential and normal components of the electric and magnetic fields are assumed to be known. The tangential components are denoted with a subscript t, whereas the normal components are denoted with n. If an integral-equation method is used in the scattering step, the result is usually the tangential components of the electromagnetic field on the surface of the scattering structure A. Therefore, the surface S can be assumed to coincide with the surface of A, and the following discussion is still valid. Without much loss

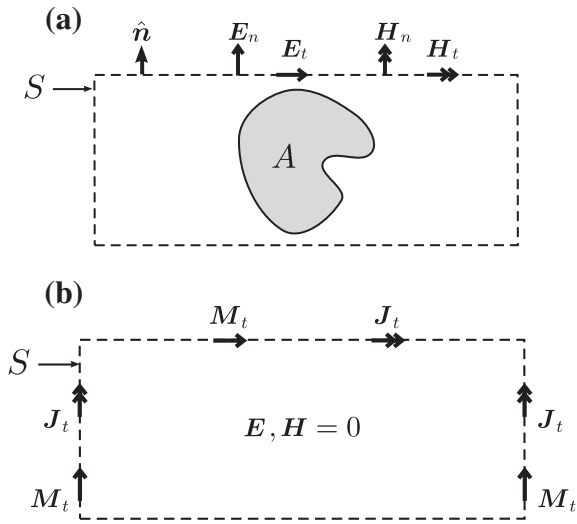

Figure 11 Pictorial description of a Green's-function near-field-to-far-field transform (NFFFT). (a) The geometry of the NFFFT. The scattering structure A is enclosed in a closed surface S. (b) Equivalent surface currents on S, with the interior of S removed.

of generality, we will assume that a differential-equation method has been employed in the scattering step, and the surface S is some arbitrary surface surrounding A. The most prominent method for obtaining the radiated far field in terms of the near fields on S is the *vector-potential* formulation, wherein the near fields are first converted to equivalent electric and magnetic surface currents on S using the *surface equivalence theorem* (Balanis, 1989; Chen, 1989; Harrington, 2001). The derivation of this theorem is quite lengthy, so it will not be reproduced here. The interested reader may consult the references. However, the result of the theorem is very simple and is shown in Figure 11(b). The equivalent electric and magnetic surface currents $\boldsymbol{M}_t$ and $\boldsymbol{J}_t$ are now radiating in free space, without the scatterer A inside the NFFFT surface S. The removal of A is justified because the equivalent surface currents $\boldsymbol{M}_t$ and $\boldsymbol{J}_t$ create a null field inside S. The relationship between the equivalent surface currents $\boldsymbol{M}_t$ and $\boldsymbol{J}_t$ and the electromagnetic field on S is quite simple:

$$\boldsymbol{M}_t = \boldsymbol{E} \times \hat{\boldsymbol{n}} = \boldsymbol{E}_t \times \hat{\boldsymbol{n}}, \tag{78}$$

$$\boldsymbol{J}_t = \hat{\boldsymbol{n}} \times \boldsymbol{H} = \hat{\boldsymbol{n}} \times \boldsymbol{H}_t, \tag{79}$$

in which $\hat{\boldsymbol{n}}$ is the outward normal unit vector shown in Figure 11(a). Notice that only the *tangential* components of the electric and magnetic fields are needed for this formulation. Once the equivalent currents are placed on S and the interior region of S is filled with free space (or the homogeneous material outside S), the fields radiated by these currents can be found using a variety of methods. In the vector-potential method, the currents are first inserted into certain integrals that yield intermediate quantities called the vector potentials. Among several slightly different conventions for their definitions, we will follow that of (Balanis, 1989). The vector potentials are obtained from the surface currents as follows:

$$\boldsymbol{A}(\boldsymbol{r}) = \frac{\mu_0}{4\pi} \iint_S \boldsymbol{J}_t(\boldsymbol{r}') \frac{\mathrm{e}^{-jk|\boldsymbol{r}-\boldsymbol{r}'|}}{|\boldsymbol{r}-\boldsymbol{r}'|} \mathrm{d}\boldsymbol{r}', \tag{80}$$

$$\boldsymbol{F}(\boldsymbol{r}) = \frac{\epsilon_0}{4\pi} \iint_S \boldsymbol{M}_t(\boldsymbol{r}') \frac{\mathrm{e}^{-jk|\boldsymbol{r}-\boldsymbol{r}'|}}{|\boldsymbol{r}-\boldsymbol{r}'|} \mathrm{d}\boldsymbol{r}', \tag{81}$$

in which $\boldsymbol{A}(\boldsymbol{r})$ and $\boldsymbol{F}(\boldsymbol{r})$ are called the electric and magnetic vector potentials, respectively. The primed coordinates $\boldsymbol{r}'$ represent the source points on S, while the unprimed coordinates $\boldsymbol{r}$ represent the observation points. The

electric and magnetic fields at the observation point $\boldsymbol{r}$ result from the following differentiation operations on the vector potentials:

$$\boldsymbol{E}(\boldsymbol{r}) = -j\omega\left[\boldsymbol{A} + \frac{1}{k^2}\nabla(\nabla\cdot\boldsymbol{A})\right] - \frac{1}{\epsilon_0}\nabla\times\boldsymbol{F}, \tag{82}$$

$$\boldsymbol{H}(\boldsymbol{r}) = -j\omega\left[\boldsymbol{F} + \frac{1}{k^2}\nabla(\nabla\cdot\boldsymbol{F})\right] + \frac{1}{\mu_0}\nabla\times\boldsymbol{A}. \tag{83}$$

In the near field, the evaluation of (80)–(83) can be extremely complicated. However, considerable simplification occurs when the observation point approaches infinity ($|\boldsymbol{r}| \to \infty$), which is the far-zone region we are interested in. In the far zone, the $|\boldsymbol{r} - \boldsymbol{r}'|$ term in the exponential in (80) and (81) can be approximated as $(r - \hat{\boldsymbol{r}}\cdot\boldsymbol{r}')$, where $\hat{\boldsymbol{r}} = \boldsymbol{r}/r = (\cos\phi\sin\theta, \sin\phi\sin\theta, \cos\theta)$ is the unit vector in the direction of observation, while the $|\boldsymbol{r} - \boldsymbol{r}'|$ term in the denominator can be approximated as r. This results in the following far-zone expressions for the vector potentials:

$$\boldsymbol{A}(\boldsymbol{r}) = \mu_0\frac{\mathrm{e}^{-jkr}}{4\pi r}\iint_S \boldsymbol{J}_t(\boldsymbol{r}')\mathrm{e}^{jk\hat{\boldsymbol{r}}\cdot\boldsymbol{r}'}\,\mathrm{d}\boldsymbol{r}' \tag{84}$$

$$\boldsymbol{F}(\boldsymbol{r}) = \epsilon_0\frac{\mathrm{e}^{-jkr}}{4\pi r}\iint_S \boldsymbol{M}_t(\boldsymbol{r}')\mathrm{e}^{jk\hat{\boldsymbol{r}}\cdot\boldsymbol{r}'}\,\mathrm{d}\boldsymbol{r}'. \tag{85}$$

As a result of the far-zone approximation, the r dependence in (84) and (85) is completely factored out, and the surface integrals only depend on the observation angles θ, ϕ. The differentiation relations (82) and (83) also assume simpler forms for large r if the terms that decay faster than $1/r$ are neglected. Expanding the ∇ operator in the spherical coordinates (r, θ, ϕ) and neglecting $1/r^2$ and higher terms, (82) and (83) simplify to

$$E_r = 0, \tag{86}$$

$$E_\theta = -j\omega(A_\theta + \eta_0 F_\phi), \tag{87}$$

$$E_\phi = -j\omega(A_\phi - \eta_0 F_\theta), \tag{88}$$

$$H_r = 0, \tag{89}$$

$$H_\theta = \frac{j\omega}{\eta_0}(A_\phi - \eta_0 F_\theta), \tag{90}$$

$$H_\phi = -\frac{j\omega}{\eta_0}(A_\theta + \eta_0 F_\phi), \tag{91}$$

where $\eta_0 = (\mu_0/\epsilon_0)^{1/2}$ is the wave impedance of free space. The far-zone electric and magnetic fields are transverse ($\hat{\boldsymbol{r}} \cdot \boldsymbol{E} = 0$, $\hat{\boldsymbol{r}} \cdot \boldsymbol{H} = 0$) and orthogonal to each other ($\eta_0 \boldsymbol{H} = \hat{\boldsymbol{r}} \times \boldsymbol{E}$).

Within the context of finite numerical methods, the term "near-field-to-far-field transform" is usually reserved for the implementation of the formulas (84)–(91). The frequency-domain NFFFT described above was first incorporated into the finite-difference time-domain (FDTD) method by Umashankar and Taflove (1982) and Taflove and Umashankar (1983). A time-domain version of the vector-potential NFFFT in three dimensions was developed later (Luebbers, Kunz, Schneider, & Hunsberger, 1991; Yee, Ingham, & Shlager, 1991). For a good review of these methods, the reader is referred to Taflove and Hagness (2005).

Despite the assumption in the beginning of this section that the collection system is telecentric on the object side and the entrance pupil lies at infinity (see Figure 9), we digress briefly to introduce another numerical collection algorithm that is extremely convenient for non-telecentric cases. Commonly referred to as the *Stratton–Chu* formulation, it operates directly on the tangential *and* normal fields on S [see Figure 11(a)] without any intermediate vector potentials. Its derivation, based on vector versions of Green's identities, proceeds very similarly to that of the surface equivalence principle. For details, the reader may consult (Stratton, 2007; Stratton & Chu, 1939; Török, Munro, & Kriezis, 2006). Although the derivation is quite lengthy, the result is pretty simple. The electric and magnetic fields anywhere outside the closed surface S in Figure 11(a) are given by

$$\boldsymbol{E}(\boldsymbol{r}) = \iint_S \big[(\hat{\boldsymbol{n}} \cdot \boldsymbol{E})\nabla' G(\boldsymbol{r}, \boldsymbol{r}') - j\omega\mu_0(\hat{\boldsymbol{n}} \times \boldsymbol{H})G(\boldsymbol{r}, \boldsymbol{r}') + (\hat{\boldsymbol{n}} \times \boldsymbol{E}) \times \nabla' G(\boldsymbol{r}, \boldsymbol{r}')\big]\, \mathrm{d}S', \tag{92}$$

$$\boldsymbol{H}(\boldsymbol{r}) = \iint_S \big[(\hat{\boldsymbol{n}} \cdot \boldsymbol{H})\nabla' G(\boldsymbol{r}, \boldsymbol{r}') + j\omega\epsilon_0(\hat{\boldsymbol{n}} \times \boldsymbol{E})G(\boldsymbol{r}, \boldsymbol{r}') + (\hat{\boldsymbol{n}} \times \boldsymbol{H}) \times \nabla' G(\boldsymbol{r}, \boldsymbol{r}')\big]\, \mathrm{d}S'. \tag{93}$$

Here, $G(\boldsymbol{r}, \boldsymbol{r}') = \exp(-jk|\boldsymbol{r} - \boldsymbol{r}'|)/4\pi|\boldsymbol{r} - \boldsymbol{r}'|$ is the free-space Green's function, and the gradient operator ∇' operates on the source coordinate $\boldsymbol{r}'$. The Stratton–Chu formula (92) and (93) for the field outside S is actually a combination of the vector-potential formulas (80)–(83), with the differentiation operator ∇ carried inside the surface integrals. If one is interested in the asymptotic far-zone field that decays as $1/r$, it is more efficient to use

the vector-potential formulas (84)–(91) instead of the asymptotic form of the Stratton–Chu formula (92) and (93), because the latter requires extra operations on the normal field components. In the near field, however, the Stratton–Chu formulation has more superior numerical properties. Since the ∇' operator only acts on the Green's function $G(\boldsymbol{r}, \boldsymbol{r}')$, it can be evaluated in closed form:

$$\nabla' G(\boldsymbol{r}, \boldsymbol{r}') = jk \frac{\mathrm{e}^{-jk|\boldsymbol{r}-\boldsymbol{r}'|}}{4\pi|\boldsymbol{r}-\boldsymbol{r}'|}\left[1 - \frac{j}{k|\boldsymbol{r}-\boldsymbol{r}'|}\right]\hat{\boldsymbol{r}}_d, \tag{94}$$

in which $\hat{\boldsymbol{r}}_d$ is the unit vector $(\boldsymbol{r}-\boldsymbol{r}')/|\boldsymbol{r}-\boldsymbol{r}'|$. If the wavelengths present in the illumination are much smaller than the distance between the scattering object and the entrance pupil, then the second term in square brackets in (94) can be neglected. Note that this expression depends both on the source coordinate $\boldsymbol{r}'$ and the observation coordinate $\boldsymbol{r}$, therefore it does not result in a full decoupling of the r dependence unlike (84) and (85). Nevertheless, the differentiation operator ∇' is analytically evaluated, which simplifies the numerical implementation of (92) and (93) significantly. The presence of second-order differentiation operators in the vector-potential formulation (82) and (83) renders the numerical implementation of the near-field-to-near-field transform more tedious and impractical. The price paid by choosing the more convenient Stratton–Chu formulation is the extra computation and/or storage requirements imposed by the normal components $(\hat{\boldsymbol{n}} \cdot \boldsymbol{E})$, $(\hat{\boldsymbol{n}} \cdot \boldsymbol{H})$ of the field on S. The Stratton–Chu formulation has been used to calculate the collected light in coherent microscopes (Munro & Török, 2005; Török et al., 2008) and photolithography systems (Janssen et al., 2008; van Haver et al., 2009) where the entrance pupil is assumed to be situated at a finite distance.

Near-field-to-far-field transforms based on Green's-function formalisms pose a difficulty when the scattering object A in Figure 11(a) is radiating in a planar multilayered medium. In this case, the equivalent surface currents $\boldsymbol{J}_t$, $\boldsymbol{M}_t$ in Figure 11(b) do not radiate in free space; therefore, the free-space Green's function $G(r) = \exp(-jkr)/4\pi r$ cannot be used. Instead, the appropriate Green's functions associated with the multilayered medium should be used in (80) and (81) as well as (92) and (93). In the near field, obtaining exact expressions for these Green's functions can be an extremely complicated task (Felsen & Marcuvitz, 1994; Michalski & Mosig, 1997). In the far zone, however, closed-form analytical expressions of these Green's functions may be found. Frequency-domain NFFFT algorithms for the

FDTD analysis of multilayered media were introduced in Demarest, Huang, and Plumb (1996) and Martin and Pettersson (2001). A direct time-domain FDTD NFFFT was later developed for a three-layered medium (Capoglu, 2007; Capoglu & Smith, 2006). A more general and stable formulation of the frequency-domain NFFFT was developed by Capoglu, Taflove, and Backman (2012). Spatial Fourier-transform methods have also been used to obtain the far-zone field in FDTD (Janssen et al., 2008; van Haver et al., 2009) and finite-element simulations (Wei et al., 2007). The transmission-line formulation used in Capoglu and Smith (2006) and Capoglu (2007) and the spatial Fourier-transform method operate on the same basic principles.

In imaging applications, the far field (86)–(91) has to be calculated at multiple observation directions (θ, ϕ) in order to construct the final image. The choice of these observation directions is more obvious in the Fourier-based NFFFT of the previous subsection. If the scattering is calculated using a modal method, the Floquet modes $\boldsymbol{R}_{pq}$ in (58) contain all the necessary information regarding the far-zone scattered field. For a finite method applied to a periodic structure, the 2D discrete Fourier-transform operation of the phase-shifted sampled field $\bar{\boldsymbol{E}}[m, n]$ in (70) was shown to contain sufficient information regarding the far-zone scattered field, provided the sampling is fine enough to satisfy (74) or (77). In a Green's-function-based NFFFT, however, the choice of the observation directions is not immediately obvious. It is evident that a discrete arrangement of observation directions (similar to that of the incoherent plane waves constituting Köhler illumination in Figure 5) is needed. Two different arrangements of observation directions are shown in Figure 12. In Figure 12(a), the direction cosines $(s_x, s_y) = (\cos\phi \sin\theta, \sin\phi \sin\theta)$ are equally spaced, resulting in a Cartesian distribution of observation directions in the (s_x, s_y) space. The loss of rotational symmetry in ϕ can be mitigated by increasing the number of points. Alternatively, a rotationally symmetric arrangement can be obtained by parametrizing the region inside the circle $s = (s_x^2 + s_y^2)^{1/2} < \sin\theta_{\text{obj}}$ by the polar coordinates (ρ, ϕ), such that

$$s_x = \rho \cos\phi, \quad s_y = \rho \sin\phi \tag{95}$$

with the ranges

$$-\sin\theta_{\text{obj}} < \rho < \sin\theta_{\text{obj}}, \quad -\pi/2 < \phi < \pi/2. \tag{96}$$

Applying Gaussian quadrature in ρ (Bochkanov & Bystritsky, 2008; Press et al., 1992) and maintaining equal spacing in ϕ, the discrete arrangement in Figure 12(b) is obtained. Note that the rotational symmetry is preserved,

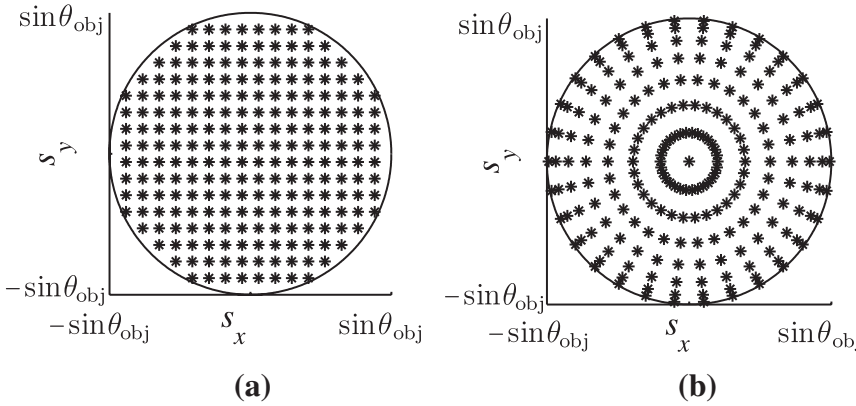

Figure 12 Two types of discrete arrangements for far-zone observation directions in a numerical imaging application. (a) Equal spacing of direction cosines (s_x, s_y) = ($\cos\phi\sin\theta$, $\sin\phi\sin\theta$). (b) Polar representation (s_x, s_y) = ($\rho\cos\phi$, $\rho\sin\phi$), followed by Gaussian quadrature in $-\sin\theta_{obj} < \rho < \sin\theta_{obj}$ and equal spacing of $-\pi/2 < \phi < \pi/2$.

but there is an inhomogeneity in the density of points inside the collection numerical aperture. In Section 3.4, the respective advantages of the arrangements in Figure 12(a) and (b) will be seen more clearly.

For the Cartesian arrangement in Figure 12(a), there is an upper limit for the spacings Δs_x, Δs_y of the direction cosines if the resulting image is to be constructed accurately. Here, we will merely note this limit and defer its derivation until Section 3.4. Consider the scatterer A in Figure 11(a). It is obvious that the scattered electromagnetic field will be stronger near the scatterer A and will gradually decay to zero away from it. Let us define an area of dimensions W_x and W_y around the scatterer A, outside which the scattered electromagnetic field can be assumed negligible. An area having dimensions several times the dimensions of the scatterer L_x and L_y will usually be sufficient. Given the dimensions W_x, W_y of the "nonzero-field" area, the condition for the image to be constructed without loss of information is

$$\Delta s_x < \frac{2\pi}{kW_x}, \quad \Delta s_y < \frac{2\pi}{kW_y}. \tag{97}$$

This implies that a larger scatterer requires a finer sampling of the far-zone electromagnetic field. In a sense, this relation is dual to (77), which describes the condition for the reconstruction of the far-zone field from the sampled near field.

There is a subtle complication that arises in the collection step when either the incident beam or the reflection of this beam from the planar multilayers (if applicable) falls within the angular collection range of the objective. The former case may happen in transmission-mode microscopy or photolithography, where the illumination and scattering happen on opposite sides of the object. The latter case will happen in reflection-mode microscopy, if part of the beam reflected from the layer interfaces is within the collection numerical aperture. This is usually less of a problem for a Fourier-based collection scheme because it is the *total field* that is observed on the planar surface S of Figure 10, including the incident or reflected beams. The real problem arises when a Green's-function-based scheme is used with near-field information on a closed surface S as in Figure 11(a). Almost invariably, the near field on S is only the *scattered* field. The incident field is calculated only as an excitation term either inside the scattering regions (called the pure-scattered field-formalism—used both in FDTD (Taflove & Hagness, 2005) and FEM (Jin, 2002)) or inside a fictitious surface surrounding the scatterer (called the total-field/scattered-field (TF/SF) formalism—used mostly in FDTD). In the TF/SF formalism, the fictitious surface should be inside the NFFFT surface S. Otherwise, the imbalance between the magnitudes of the incident and scattered fields will cause larger numerical errors in the scattered field. For this reason, the incident or reflected beam should be treated separately from the scattered field and propagated individually through the collection and refocusing system. This will be discussed further at the end of Section 3.4.

3.4 Refocusing

Since we are only concerned with real images that can be projected on a recording medium, the final step of the imaging process involves the refocusing of the rays collected from the scatterer onto an image plane. The collection and refocusing steps in Figure 1 are reproduced schematically in Figure 13 for convenience. The entrance and exit pupils of the system are images of each other with respect to the collection-refocusing optics in the middle. The direction-cosine variables (s_x, s_y) and (s'_x, s'_y) are used to parametrize the entrance and exit pupils. The object and the image are centered around O and O', and the angles subtended by the entrance and exit pupils at O and O' are denoted by θ_{obj} and θ_{img}. The refractive indices of the object and image spaces are n and n', respectively. Allowing arbitrary n and n' can be useful for modeling liquid-immersion lenses. Two Cartesian coordinate systems are defined with respect to the origins O and O', having common z and

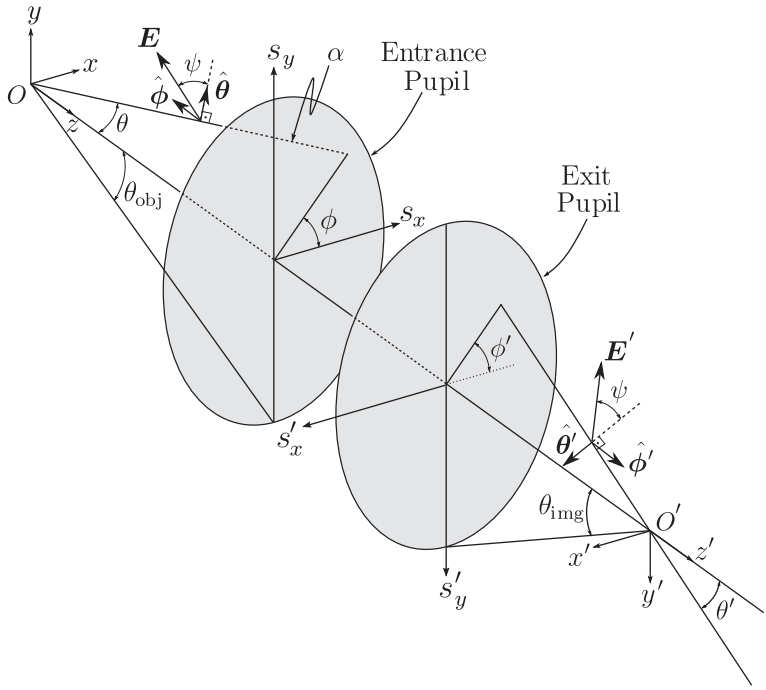

Figure 13 An illustration of the general geometry of the collection and refocusing optics.

z' axes along the optical axis OO'. The x, y and x', y' axes are anti-parallel, i.e., $x' = -x$ and $y' = -y$. In the following, unprimed and primed coordinates are used to denote variables in the object and image spaces, respectively.

In Section 3.3, the assumption of telecentricity was made in the object space, meaning that the entrance pupil is at infinity. In Figure 13, a ray $\boldsymbol{\alpha}$ is shown entering the entrance pupil at angles (θ, ϕ). This ray traverses the collection-refocusing system, leaving the exit pupil at angles (θ', ϕ'). Assuming that the collection-refocusing system is rotationally symmetric, the ray stays on the meridional plane (defined by the ray $\boldsymbol{\alpha}$ and the line OO'). This requires that the azimuthal angles are equal: $\phi' = \phi$. We will only consider a subclass of optical systems that satisfy the *Abbe sine condition* (Barrett & Myers, 2004; Born & Wolf, 1999; Kingslake, 1978) between the sines of the ray angles θ and θ' at the entrance and exit pupils:

$$\frac{n \sin\theta}{n' \sin\theta'} = M = \frac{\text{NA}_{\text{obj}}}{\text{NA}_{\text{img}}}, \tag{98}$$

where M is a constant that is a characteristic of the collection-refocusing system. In (98), $\mathrm{NA}_{\mathrm{obj}}$ and $\mathrm{NA}_{\mathrm{img}}$ are the *collection* and *imaging numerical apertures* defined as $\mathrm{NA}_{\mathrm{obj}} = n \sin\theta_{\mathrm{obj}}$ and $\mathrm{NA}_{\mathrm{img}} = n' \sin\theta_{\mathrm{img}}$, respectively. Up to the first order in off-axis distances, this constant is equal to the negative of the *lateral magnification* of the imaging system (Born & Wolf, 1999). The negative sign is a consequence of the fact that the imaging geometry in Figure 13 always results in an *inverted* image. The constant M will be called "magnification" in the following, bearing in mind that the actual lateral magnification is $(-M)$. For notational convenience, we define another parameter M' representing the angular de-magnification:

$$M' = \frac{n'}{n} M = \frac{\sin\theta}{\sin\theta'}. \tag{99}$$

The Abbe sine condition (98) ensures that aberrations that depend on the first power of the off-axis distance of an object point (called circular coma) are absent. In other words, it is the required condition for the sharp imaging of points at small off-axis distances. This condition is usually satisfied in well-corrected optical imaging systems.

Two opposite situations regarding the lateral magnification M are encountered in photolithography and microscopy, as illustrated in Figure 14 (Totzeck, 2006). In photolithography [Figure 14(a)], a de-magnified image of the mask is projected on the photoresist by a projection lens, so $M < 1$ and $\mathrm{NA}_{\mathrm{obj}} < \mathrm{NA}_{\mathrm{img}}$. De-magnification in photolithography is usually specified in terms of the "reduction ratio," defined as the inverse of M and notated as $(1/M)$:1. For example, a projection lens with $M = 0.2$ is said to have a reduction ratio of 5:1. The nonlinear response of photoresists in modern photolithography demands that the projection lenses be corrected for every conceivable aberration well beyond the requirement of simple diffraction limited resolution. In microscopy, however, requirements for aberration correction are less stringent, and most of the challenge lies with the higher NA of the objective lens. The structure of a general microscopy system is shown in Figure 14(b). Most modern microscope objectives are infinite-conjugate, meaning that the optimum sample position for best aberration correction is at the front focal plane of the objective, resulting in an image at infinity. This image is brought to a finite position by the tube lens, as shown in Figure 14(b). The image-side numerical aperture $\mathrm{NA}_{\mathrm{img}}$ of a microscope is equal to the object-side numerical aperture $\mathrm{NA}_{\mathrm{obj}}$ divided by the magnification, which can be as high as $M = 100$. This results in very small incidence angles for the rays in the image space, which can be

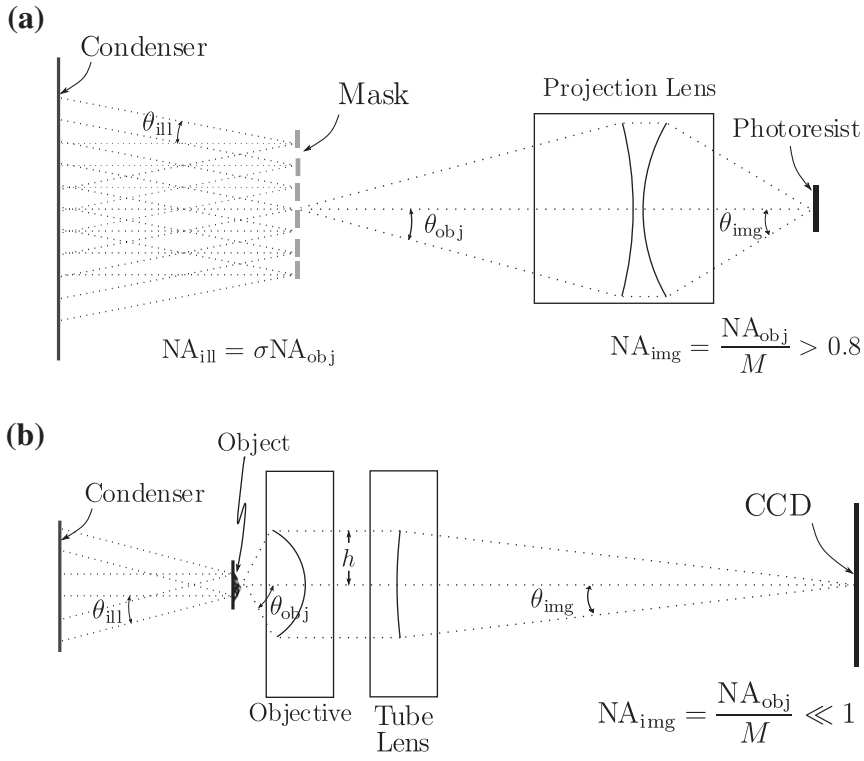

Figure 14 Comparison of the collection and refocusing geometries in photolithography and microscopy. (a) In photolithography, a de-magnified image of the mask is projected on the photoresist. Typical values are $\sigma = 0.3$–0.8, M = 0.1–0.25. (b) In microscopy, a magnified image of the object is projected on the CCD by the objective-tube lens combination. Typical values are $M = 10$–100.

handled without much error by inexpensive tube lenses that do not require much aberration correction. On the other hand, the microscope objective is usually an expensive, well-corrected optical component. This is because the maximum object-side ray angle θ_{obj} has to be pretty large for good imaging resolution. If both the objective and the tube lens satisfy the Abbe sine condition (98), the magnification M can also be expressed in terms of the focal lengths f_1, f_2 of the objective and the tube lens. Let us denote the height of the marginal ray between the objective and the tube lens by h [see Figure 14(b)]. It can be shown that the Abbe sine condition for this ray takes the form (Born & Wolf, 1999)

$$h = f_1 \sin\theta_{\mathrm{obj}} = f_2 \sin\theta_{\mathrm{img}}. \tag{100}$$

Using (98), the magnification M is then equal to

$$M = \frac{n}{n'}\frac{f_2}{f_1}. \tag{101}$$

The objective/tube lens arrangement in Figure 14(b) provides a way of altering the magnification M of the system by changing the focal lengths of either the objective or the tube lens. Changing the focal length f_2 of the tube lens simply makes the image bigger or smaller with no change in resolution. Changing the focal length f_1 of the objective amounts to zooming into or out of the image with lower or higher resolution. In many microscopes, one cycles through different objectives with different focal lengths (f_1), effectively changing the magnification (10×, 100×, etc.)

In order to construct the field distribution at the image plane, it is necessary to know the properties of all the rays α that leave the exit pupil. The azimuthal angles ϕ, ϕ' at the entrance and exit pupils are identical. The polar exit angle θ' of the rays is given by the Abbe sine condition (98). The strength factor $\boldsymbol{E}'_{\mathrm{s}}(\theta', \phi')$ of the ray at the exit pupil still needs to be found (see (40) for the definition of the strength factor.) Let us start with the polarization of $\boldsymbol{E}'_{\mathrm{s}}(\theta', \phi')$. From the laws of geometrical optics, $\boldsymbol{E}'_{\mathrm{s}}(\theta', \phi')$ lies in a plane perpendicular to the ray. A good approximation for the polarization can be obtained by making the same assumption as in Section 3.1.2, wherein it was argued (in reference to Figure 6) that the angle ψ between the electric-field vector and the meridional plane remains constant as the ray α traverses the system (Born & Wolf, 1999; Richards & Wolf, 1959). This requires that the angles of incidence at each refracting surface be small. In highly corrected optical components with multiple lenses, the deviation of a ray at each surface is minimal; therefore, the above assumption is valid. With this assumption, the strength factors $\boldsymbol{E}_{\mathrm{s}}(\theta, \phi)$ and $\boldsymbol{E}'_{\mathrm{s}}(\theta', \phi')$ of the ray α at the entrance and exit pupils make the same angle ψ with the meridional plane, as shown in Figure 13. The magnitude of $\boldsymbol{E}'_{\mathrm{s}}(\theta', \phi')$ follows from the intensity law of geometrical optics (Flagello, Milster, & Rosenbluth, 1996). Let us track an infinitesimal tube of rays containing the ray α in the object and image spaces. These rays emanate from the object-side origin O and converge at the image-side origin O', see Figure 13. If the aberrations of the collection-refocusing system are small, the principal radii of curvature of the geometrical-optics wavefront in the image space are both approximately equal to the distance r' from O' (Wolf, 1959). The light intensities on the ray α in the object and image spaces are, from (9),

$$I_1 = n|\boldsymbol{E}_{\mathrm{s}}|^2/(\eta_0 r^2),\ I_2 = n'|\boldsymbol{E}'_{\mathrm{s}}|^2/(\eta_0 (r')^2), \tag{102}$$

in which r and r' are arbitrary distances from O and O', respectively. The infinitesimal areas on the spherical wavefronts intersected by the tubes of rays are

$$\mathrm{d}S_1 = r^2 \sin\theta \,\mathrm{d}\theta \,\mathrm{d}\phi, \quad \mathrm{d}S_2 = (r')^2 \sin\theta' \,\mathrm{d}\theta' \,\mathrm{d}\phi'. \tag{103}$$

Assuming that the absorptive, reflective, and refractive losses in the collection-refocusing system are negligible, conservation of energy dictates that the total powers crossing $\mathrm{d}S_1$ and $\mathrm{d}S_2$ are equal. Since the total power crossing an infinitesimal area $\mathrm{d}S$ is (intensity $\times$ $\mathrm{d}S$), this is equivalent to the intensity law of geometrical optics:

$$I_1 \,\mathrm{d}S_1 = I_2 \,\mathrm{d}S_2. \tag{104}$$

We therefore have

$$|\boldsymbol{E}'_{\mathrm{s}}| = |\boldsymbol{E}_{\mathrm{s}}| \sqrt{\frac{n \sin\theta \,\mathrm{d}\theta \mathrm{d}\phi}{n' \sin\theta' \,\mathrm{d}\theta' \mathrm{d}\phi'}}. \tag{105}$$

From the Abbe sine condition, $n \sin\theta = M(n' \sin\theta')$. Using the chain rule, one can write

$$\frac{\mathrm{d}\theta}{\mathrm{d}\theta'} = \frac{\cos\theta'}{\cos\theta} \frac{\mathrm{d}(\sin\theta)}{\mathrm{d}(\sin\theta')} = M \frac{n' \cos\theta'}{n \cos\theta}. \tag{106}$$

Also noting that $\mathrm{d}\phi' = \mathrm{d}\phi$, (105) becomes

$$|\boldsymbol{E}'_{\mathrm{s}}| = M \sqrt{\frac{n' \cos\theta'}{n \cos\theta}} |\boldsymbol{E}_{\mathrm{s}}|. \tag{107}$$

If $\boldsymbol{E}_{\mathrm{s}}(\theta, \phi)$ and $\boldsymbol{E}'_{\mathrm{s}}(\theta', \phi')$ are expressed in spherical coordinates centered around O and O', respectively, a quick inspection of Figure 13 reveals that the $\hat{\boldsymbol{\theta}}'$ and $\hat{\boldsymbol{\phi}}'$ components of $\boldsymbol{E}'_{\mathrm{s}}(\theta', \phi')$ are given by

$$\begin{aligned} \boldsymbol{E}'_{\mathrm{s}}(\theta', \phi') \cdot \hat{\boldsymbol{\theta}}' &= -M \sqrt{\frac{n' \cos\theta'}{n \cos\theta}} \boldsymbol{E}_{\mathrm{s}}(\theta, \phi) \cdot \hat{\boldsymbol{\theta}}, \\ \boldsymbol{E}'_{\mathrm{s}}(\theta', \phi') \cdot \hat{\boldsymbol{\phi}}' &= -M \sqrt{\frac{n' \cos\theta'}{n \cos\theta}} \boldsymbol{E}_{\mathrm{s}}(\theta, \phi) \cdot \hat{\boldsymbol{\phi}}. \end{aligned} \tag{108}$$

Now, both the directions and the strength factors of the rays leaving the exit pupil are determined and we are ready to construct the field at the image plane. The final step of the imaging process requires a connection between the geometrical-optics field determined by the rays at the exit

pupil and the electromagnetic field at the image plane $z' = 0$. This can be achieved by use of vectorial diffraction theory (Kline & Kay, 1979; Török et al., 2008; Wolf, 1959). The vector field at the image plane (x', y') is given by the *Debye–Wolf integral*:

$$\boldsymbol{E}_{\text{img}}(x', y') = \frac{jk'}{2\pi} \iint_{\Omega_{\text{img}}} \boldsymbol{E}'_{\text{s}}(s'_x, s'_y) \mathrm{e}^{-jk'(s'_x x' + s'_y y')} \mathrm{d}\Omega, \tag{109}$$

where $k' = n' k_0$ is the wavenumber in the image space. A change of variables is made from the angle variables (θ', ϕ') to the direction-cosine variables (s'_x, s'_y):

$$(s'_x, s'_y) = (\cos\phi' \sin\theta', \sin\phi' \sin\theta'). \tag{110}$$

In (109), Ω_{img} is the solid angle bounded by θ_{img}, and $\mathrm{d}\Omega = \mathrm{d}s'_x \mathrm{d}s'_y / s'_z = \mathrm{d}s'_x \mathrm{d}s'_y / \cos\theta'$. It is straightforward to show that (Wolf, 1959) the Debye–Wolf integral in (109) is the "inverse" of the vectorial far-field expression (48). The Debye–Wolf integral in (109) can also be regarded as an infinite summation of plane waves incident from a spectrum of directions (θ', ϕ'). For this reason, it is also called the *angular-spectrum* or the *plane-wave* representation of the image field.

The range of validity of the Debye–Wolf integral (109) warrants some discussion. If the exit pupil is at infinity, then the refocusing system is *image-side telecentric*, and the Debye–Wolf integral (109) strictly applies (Sheppard, 2007). However, if a certain geometrical condition is satisfied, (109) is also applicable for an exit pupil at a finite position. In Figure 15, a spherical wavefront W_f passing through the center C of the exit pupil is shown converging toward the focal point O' at a distance d from the pupil. Let the radius of the exit pupil be a and the maximum distance between W_f and the exit pupil be denoted by $\Delta_{\max}$. To a good approximation, $\Delta_{\max}$ is equal to $a^2/2d$. The *Fresnel number* N_{F} is a dimensionless quantity defined as $\Delta_{\max}$ divided by half of the wavelength λ' in the image space:

$$N_{\text{F}} = \frac{\Delta_{\max}}{\lambda'/2} \approx \frac{a^2}{\lambda' d}. \tag{111}$$

The Fresnel number is approximately equal to the number of Fresnel zones that fill the aperture when viewed from the focal point O' (Born & Wolf, 1999). It can be shown (Li & Wolf, 1982; Wolf & Li, 1981) that the

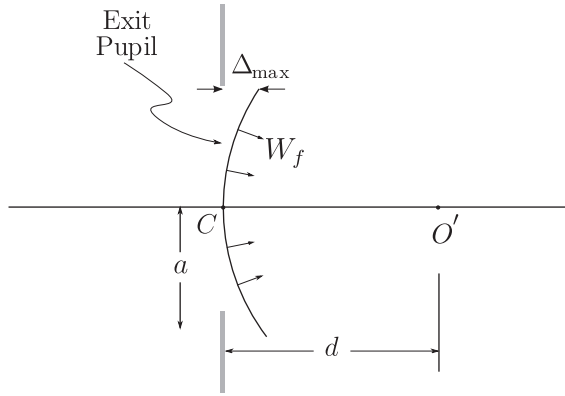

Figure 15 The geometrical parameters used in the definition of the Fresnel number N_F.

required condition for the validity of the Debye–Wolf integral (109) is that the Fresnel number is very large:

$$N_{\mathrm{F}} \gg 1. \tag{112}$$

For visible-range optical imaging systems employed in microscopy, photolithography, metrology, inspection, and alignment, the exit-pupil radius a is well on the order of centimeters, so a/λ' is on the order of 10^4. The ratio a/d of the exit-pupil radius to the pupil distance is equal to $\tan\theta_{\mathrm{img}}$, which may range from 10^{-2} to infinity, depending on the magnification M. Therefore, it can safely be assumed that the Debye–Wolf integral (109) is a very accurate representation of the electromagnetic field in the image space for a wide range of optical systems.

If the image space is homogeneous, the Debye–Wolf integral in (109) gives the final result for the image field. If there is a non-trivial scattering topography in the image space such as a CCD or a photoresist, the integral (109) for the image field should be considered only as an incident field. The calculation methods detailed in Section 3.2 should then be used to compute the scattered field resulting from this incident field. Since plane-wave incidence is usually the easiest illumination scheme to handle, the angular-spectrum interpretation of the Debye–Wolf integral becomes quite handy in many cases. The incident field (109) is a coherent illumination beam (see Section 3.1.1) that can be written as the sum of plane-wave components

$$\mathrm{d}\boldsymbol{E}_{\mathrm{img}}(x', y') = \frac{jk'}{2\pi}\boldsymbol{E}'_{\mathrm{s}}(s'_x, s'_y)\mathrm{e}^{-jk'(s'_x x' + s'_y y')}\,\mathrm{d}\Omega. \tag{113}$$

If the image space is simply a stack of laterally infinite planar layers, each plane wave (113) can be propagated into this medium using standard Fresnel refraction formulas (Bernard & Urbach, 1991; Flagello et al., 1996; Tang et al., 2005; Török & Varga, 1997; Török, Varga, Laczik, & Booker, 1995; van de Nes, Billy, Pereira, & Braat, 2004). For more complex topographies, advanced numerical methods such as the waveguide method (Tanabe, 1992), the differential method (Kirchauer & Selberherr, 1997), the C-method (Yeung, 1990), the finite-element method (Matsuzawa, Moniwa, Hasegawa, & Sunami, 1987; Urbach & Bernard, 1989), the finite-difference time-domain method (Gamelin, Guerrieri, & Neureuther, 1989), and the integral-equation method (Yeung & Barouch, 1997) can be used to obtain the field distribution in the image space.

The Debye–Wolf integral (109) can be generalized to include the aberrations of the collection-refocusing optics by the inclusion of an additional phase factor in the exponential kernel of the integral:

$$\boldsymbol{E}_{\mathrm{img}}(x', y') = \frac{jk'}{2\pi} \iint_{\Omega_{\mathrm{img}}} \boldsymbol{E}'_{\mathrm{s}}(s'_x, s'_y) \mathrm{e}^{-jk'\left[s'_x x' + s'_y y' + \Phi(s'_x, s'_y)\right]} \, \mathrm{d}\Omega, \tag{114}$$

where the *aberration function* $\Phi(s'_x, s'_y)$ is a measure of the deviation of the wavefront from perfect spherical shape (Wolf, 1959). Regarding the aberration function as a small perturbation, the validity condition (112) can still be assumed to hold for (114). If the image space is homogeneous, the generalized Debye–Wolf integral (114) gives the final field distribution. Let us now discuss the numerical evaluation of the generalized Debye–Wolf integral (114) for a homogeneous image space, regarding the original Equation (109) as a special case. We will discuss the cases of periodic and non-periodic scatterers separately.

3.4.1 Periodic Scatterers

We assume that, at the end of the collection step, the strength factor $\boldsymbol{E}_{\mathrm{s}}(\theta, \phi)$ at the far zone has been found in the form (62). The Floquet modes $\boldsymbol{R}_{pq}$ may have either been provided directly by a modal method or by discrete Fourier transform (DFT) of near-field values provided by a finite method (see Section 3.3.1). Substituting the definitions of α, β in (47) and β_p, β_q in (59), the strength factor in (62) becomes

$$\begin{aligned} \boldsymbol{E}_{\mathrm{s}}(s_x, s_y) = (jk2\pi) \sum_p \sum_q c_{pq} \boldsymbol{R}_{pq} \\ \times \delta(ks_x - ks_{x_i} + p2\pi/d_x) \\ \times \delta(ks_y - ks_{y_i} + q2\pi/d_y), \end{aligned} \tag{115}$$

where k is the wavenumber in the object space. A change of variables is made from the angle variables (θ, ϕ) to the direction cosines (s_x, s_y) at the entrance pupil:

$$(s_x, s_y) = (\alpha/k, \beta/k) = (\cos\phi \sin\theta, \sin\phi \sin\theta) \tag{116}$$

and the direction cosines of the incident plane wave in the object space are defined as

$$(s_{x_i}, s_{y_i}) = (\cos\phi_i \sin\theta_i, \sin\phi_i \sin\theta_i). \tag{117}$$

The Abbe sine condition states that the direction cosines (s'_x, s'_y) of a ray at the exit pupil are $1/M'$ times the direction cosines (s_x, s_y) of the same ray at the entrance pupil, where M' is given by (99). Substituting (115) with $(s_x, s_y) = (M's'_x, M's'_y)$ into (108) and using the scaling property $\delta(ax) = \delta(x)/|a|$ of the Dirac delta function, we obtain the strength factor at the exit pupil:

$$\begin{aligned} \boldsymbol{E}'_s(s'_x, s'_y) = \frac{j2\pi}{M'k'} \sum_p \sum_q (n' c_{pq} c'_{pq}/n)^{1/2} \boldsymbol{R}'_{pq} \\ \times \delta\left(s'_x - s'_{x_i} + p2\pi/(M'k\, d_x)\right) \\ \times \delta\left(s'_y - s'_{y_i} + q2\pi/(M'k\, d_y)\right), \end{aligned} \tag{118}$$

where

$$(s'_{x_i}, s'_{y_i}) = (\cos\phi_i \sin\theta_i/M', \sin\phi_i \sin\theta_i/M') \tag{119}$$

and the cosine parameters c'_{pq} are defined the same way as in (63), with (β_p, β_q) replaced by $(\beta_p/M', \beta_q/M')$. The $\hat{\boldsymbol{\theta}}'$ and $\hat{\boldsymbol{\phi}}'$ components of the vector amplitude $\boldsymbol{R}'_{pq}$ are

$$\begin{aligned} \boldsymbol{R}'_{pq\theta'} &= -\boldsymbol{R}_{pq\theta}, \\ \boldsymbol{R}'_{pq\phi'} &= -\boldsymbol{R}_{pq\phi}. \end{aligned} \tag{120}$$

The image field is obtained by substituting (118) into the Debye–Wolf integral (114):

$$\boldsymbol{E}_{\text{img}}(x', y') = \mathrm{e}^{-jk'(s'_{x_i}x' + s'_{y_i}y')} \sum_p \sum_q \bar{\boldsymbol{R}}'_{pq} \mathrm{e}^{j\frac{2\pi}{M}\left(p\frac{x'}{d_x} + q\frac{y'}{d_y}\right)}, \tag{121}$$

with the modified vector Floquet mode $\bar{\boldsymbol{R}}'_{pq}$ defined by

$$\bar{\boldsymbol{R}}'_{pq} = -\sqrt{\frac{n'c_{pq}}{nc'_{pq}}}\frac{1}{M'}\boldsymbol{R}'_{pq}\mathrm{e}^{-jk'\Phi\left(s'_{x_i}-\frac{p2\pi}{M'kd_x},s'_{y_i}-\frac{q2\pi}{M'kd_y}\right)}. \tag{122}$$

The phase factor in front of the summations in (121) is also present in the object-space field distribution (58). It is enforced by the plane wave incident on the periodic scatterer. An interesting consequence of (121) and (122) is that not only the image is inverted, but the *polarization* of the electromagnetic field in the image is inverted as well. This is seen more clearly if we assume $M' = 1$ and $\Phi = 0$ in (122), which gives $\bar{\boldsymbol{R}}'_{pq} = -\boldsymbol{R}'_{pq}$. This result is intuitively satisfying, since it implies a *vector inversion* of the electromagnetic field as a generalization of the classical image inversion of geometrical optics.

If the image field (121) is to be evaluated at a few (x', y') positions, it can be calculated directly by a brute-force method. If a whole region is of interest, then a DFT-based evaluation is more efficient. It will now be shown that, using DFT (and its efficient computation by the fast Fourier transform, FFT), Equation (121) for the image field can be evaluated at a discrete rectangular grid of (x', y') points with an arbitrary spacing in x' and y'. First, it is important to remember that the p and q indices in (121) belong to a finite set. They are the indices of the scattered Floquet modes $\boldsymbol{R}_{pq}$ that fall within the entrance pupil and subsequently leave the exit pupil. The range of indices in (121) is thus defined by the following condition:

$$\left(s'_{x_i} - p\frac{2\pi}{M'kd_x}\right)^2 + \left(s'_{y_i} - q\frac{2\pi}{M'kd_y}\right)^2 < \sin^2\theta_{\mathrm{img}}. \tag{123}$$

This condition is shown geometrically in Figure 16. A rectangular grid of points (s'_{x_p}, s'_{y_q}) are represented by the indices p and q. The origin $(p, q) = (0, 0)$ of the indices corresponds to (s'_{x_i}, s'_{y_i}). Only the direction cosines that fall within a circle of radius $\sin\theta_{\mathrm{img}}$ are counted in (121), because of the condition (123). Let $p_{\min}$ and $p_{\max}$ denote the minimum and maximum permissible indices in p, and $q_{\min}$, $q_{\max}$ the corresponding indices for q. Let us write the summation term in (121) as a double summation over a rectangular region of indices limited by $p_{\min}$, $p_{\max}$, $q_{\min}$, and $q_{\max}$, with the implicit assumption that $\bar{\boldsymbol{R}}'_{pq}$ vanishes outside the range set by (123):

$$\sum_{p=p_{\min}}^{p_{\max}}\sum_{q=q_{\min}}^{q_{\max}} \bar{\boldsymbol{R}}'_{pq}\mathrm{e}^{j\frac{2\pi}{M}\left(p\frac{x'}{d_x}+q\frac{y'}{d_y}\right)}, \tag{124}$$

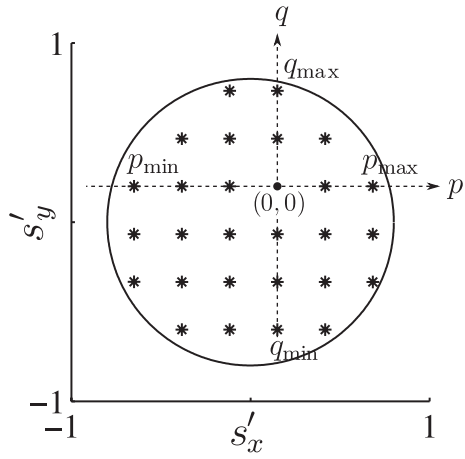

Figure 16 The admissible Floquet modes for a periodic scatterer. Only those that fall within the exit pupil contribute to the image.

which becomes, after shifting the indices by $p_{\min}$ and $q_{\min}$,

$$e^{j\frac{2\pi}{M}\left(p_{\min}\frac{x'}{d_x}+q_{\min}\frac{y'}{d_y}\right)} \times \sum_{p=0}^{p_{\max}-p_{\min}} \sum_{q=0}^{q_{\max}-q_{\min}} \bar{\boldsymbol{R}}'_{p+p_{\min},q+q_{\min}} e^{j\frac{2\pi}{M}\left(p\frac{x'}{d_x}+q\frac{y'}{d_y}\right)}. \tag{125}$$

The summation term above will take a true DFT form if it is sampled at a discrete set of points. Consider the sampling

$$(x', y') = (m\Delta x, n\Delta y), \quad \begin{matrix} m = 0 \ldots P-1, \\ n = 0 \ldots Q-1 \end{matrix} \tag{126}$$

in which Δx and Δy are chosen such that the sampling covers the entire magnified periods in both x' and y':

$$P\Delta x = Md_x, \quad Q\Delta y = Md_y. \tag{127}$$

Substituting (126) in (125), the summation term becomes

$$\sum_{p=0}^{p_{\max}-p_{\min}} \sum_{q=0}^{q_{\max}-q_{\min}} \bar{\boldsymbol{R}}'_{p+p_{\min},q+q_{\min}} e^{j2\pi\left(\frac{pm}{P}+\frac{qn}{Q}\right)}. \tag{128}$$

This expression is almost in the same form as DFT and can be evaluated using the same efficient FFT algorithms (Oppenheim et al., 1999). In

fact, when divided by PQ, it is called the $(P \times Q)$-point *inverse DFT* of $\bar{\boldsymbol{R}}'_{p+p_{\min},q+q_{\min}}$. For an efficient FFT operation, P and Q should be chosen to have small multiple prime factors. Once the inverse-DFT expression (128) is computed, the image field (121) is directly obtained at the sampled image positions $(x', y') = (m\Delta x, n\Delta y)$. This sampling can be made arbitrarily fine by increasing P and Q.

3.4.2 Non-periodic Scatterers

For a non-periodic scatterer, the far-zone field is evaluated at a discrete set of observation directions. Two different arrangements for this discrete set are shown in Figure 12. If the rectangular arrangement of Figure 12(a) is chosen, then the direction cosines (s'_x, s'_y) are also distributed in a rectangular grid inside a circle of radius $\sin\theta_{\text{img}}$. The relationship between the object-side strength factor $\boldsymbol{E}_{\text{s}}(s_x, s_y)$ and the image-side strength factor $\boldsymbol{E}'_{\text{s}}(s'_x, s'_y)$ is given by (108). Once $\boldsymbol{E}'_{\text{s}}(s'_x, s'_y)$ is determined, the Debye–Wolf integral (114) for the image field can then be evaluated numerically using DFT. A quick comparison shows that (114) is in the same form as the inverse-Fourier-transform relation (42), except a trivial change of variables $(s'_x, s'_y) \to (k's'_x, k's'_y)$. Since the double integral only covers the region $(s'^2_x + s'^2_y)^{1/2} \leqslant \sin\theta_{\text{img}}$, we can extend the limits of the integral from $-\infty$ to ∞ and assign $\boldsymbol{E}'_{\text{s}}(s'_x, s'_y) = 0$ for $(s'^2_x + s'^2_y)^{1/2} > \sin\theta_{\text{img}}$. Using the 2D Fourier relation (41), we can invert the relation (114) to obtain

$$\begin{aligned}\boldsymbol{E}'_{\text{s}}(s'_x, s'_y) = {} & \frac{(-jk')\cos\theta'}{2\pi} \mathrm{e}^{jk'\Phi(s'_x,s'_y)} \\ & \times \iint_{x,y} \boldsymbol{E}_{\text{img}}(x', y') \mathrm{e}^{jk'(s'_x x' + s'_y y')} \,\mathrm{d}x'\mathrm{d}y'.\end{aligned} \tag{129}$$

Let the direction cosines inside the imaging cone be discretized by dividing the rectangular region defined by $-\sin\theta_{\text{img}} < s'_x < \sin\theta_{\text{img}}$ and $-\sin\theta_{\text{img}} < s'_y < \sin\theta_{\text{img}}$ into $P \times Q$ rectangular patches, and choosing (s'_{x_p}, s'_{y_q}) to be at the center of each patch:

$$\begin{aligned} s'_{x_p} &= s'_{x_0} + p\Delta s'_x, \quad p = 0 \ldots P-1, \\ s'_{y_q} &= s'_{y_0} + q\Delta s'_y, \quad q = 0 \ldots Q-1, \end{aligned} \tag{130}$$

where

$$\begin{aligned} s'_{x_0} &= -\sin\theta_{\text{img}}(1 - 1/P), \\ s'_{y_0} &= -\sin\theta_{\text{img}}(1 - 1/Q) \end{aligned} \tag{131}$$

and the sampling periods $\Delta s'_x$ and $\Delta s'_y$ are

$$\begin{aligned}\Delta s'_x &= 2\sin\theta_{\text{img}}/P,\\ \Delta s'_y &= 2\sin\theta_{\text{img}}/Q.\end{aligned} \tag{132}$$

Let us define an auxiliary variable $\boldsymbol{G}'(s'_x, s'_y)$ that combines the strength factor $\boldsymbol{E}'_{\text{s}}(s'_x, s'_y)$ with the $\cos\theta'$ factor and the aberration phase factor $\mathrm{e}^{-jk'\Phi(s'_x,s'_y)}$ in (129) as follows:

$$\boldsymbol{G}'(s'_x, s'_y) \triangleq \frac{\boldsymbol{E}'_{\text{s}}(s'_x, s'_y)}{\cos\theta'}\mathrm{e}^{-jk'\Phi(s'_x,s'_y)}. \tag{133}$$

This definition facilitates the direct application of 2D DFT to the numerical evaluation of (114). The 2D array resulting from the sampling of the auxiliary variable $\boldsymbol{G}'(s'_x, s'_y)$ according to (130) is defined as

$$\boldsymbol{G}'[p, q] = \boldsymbol{G}'(s'_{x_p}, s'_{y_q}). \tag{134}$$

Now, it will be shown how the 2D DFT of the array $\boldsymbol{G}'[p, q]$ can be related to the continuous image field $\boldsymbol{E}_{\text{img}}(x', y')$ in (114). Adopting the same convention as in Section 3.3, we define 2D DFT of $\boldsymbol{G}'[p, q]$ as

$$\boldsymbol{E}[m, n] = \sum_{p=0}^{N_p-1}\sum_{q=0}^{N_q-1}\boldsymbol{G}'[p, q]\mathrm{e}^{-j2\pi\left(\frac{pm}{N_p}+\frac{qn}{N_q}\right)}. \tag{135}$$

The DFT lengths N_p and N_q are greater than or equal to P and Q, respectively. If they are greater, then $\boldsymbol{G}'[p, q]$ is zero-padded up to the required length. It is shown in Appendix C that $\boldsymbol{E}[m, n]$ is a sampled and periodically replicated (aliased) version of the continuous image field $\boldsymbol{E}_{\text{img}}(x', y')$:

$$\begin{aligned}\boldsymbol{E}[m, n] = {} & \frac{2\pi}{(jk')\Delta s'_x\Delta s'_y}\\ & \times\sum_{r=-\infty}^{\infty}\sum_{s=-\infty}^{\infty}\mathrm{e}^{j2\pi\left(\frac{s'_{x0}}{\Delta s'_x N_p}(m+rN_p)+\frac{s'_{y0}}{\Delta s'_y N_q}(n+sN_q)\right)}\\ & \times\boldsymbol{E}_{\text{img}}\left(m\Delta_x + rD'_x, n\Delta_y + sD'_y\right),\end{aligned} \tag{136}$$

in which the spatial sampling periods Δ_x, Δ_y and the aliasing periods D'_x, D'_y are defined as

$$\Delta_x = \frac{2\pi}{k'\Delta s'_x N_p}, \quad \Delta_y = \frac{2\pi}{k'\Delta s'_y N_q}, \tag{137}$$

$$D'_x = N_p \Delta_x = \frac{2\pi}{k' \Delta s'_x}, \quad D'_y = N_q \Delta_y = \frac{2\pi}{k' \Delta s'_y}. \tag{138}$$

If the shifted replicas of $\boldsymbol{E}_{\text{img}}(m\Delta_x, n\Delta_y)$ in (136) do not overlap, then $\boldsymbol{E}_{\text{img}}(m\Delta_x, n\Delta_y)$ can be retrieved from (136) as follows:

$$\boldsymbol{E}_{\text{img}}(m\Delta_x, n\Delta_y) \approx \frac{(jk')\Delta s'_x \Delta s'_y \mathrm{e}^{-j2\pi\left(\frac{s'_{x0}}{\Delta s'_x N_p} m + \frac{s'_{y0}}{\Delta s'_y N_q} n\right)}}{2\pi} \boldsymbol{E}[m, n]. \tag{139}$$

for a range of m and n values centered around $m = n = 0$ over which $\boldsymbol{E}_{\text{img}}(m\Delta_x, n\Delta_y)$ is nonzero. The condition for this retrieval is that the aliasing periods D'_x, D'_y are greater than the x' and y' dimensions W'_x, W'_y of the "nonzero-field area" over which the amplitude of the electromagnetic field $\boldsymbol{E}_{\text{img}}(x', y')$ is non-negligible:

$$D'_x > W'_x, \quad D'_y > W'_y. \tag{140}$$

Barring diffraction effects on the order of $\sim\lambda$, the dimensions W'_x, W'_y are equal to the corresponding dimensions at the object side multiplied by the magnification M of the system. At the end of Section 3.3, these object-side dimensions were defined as W_x and W_y. It follows that the dimensions W'_x, W'_y of the "nonzero-field area" at the image space are given by MW_x and MW_y. Using the definitions of the aliasing periods D'_x, D'_y in (138), the non-aliasing condition (140) becomes

$$\Delta s'_x < \frac{2\pi}{k' M W_x}, \quad \Delta s'_y < \frac{2\pi}{k' M W_y}. \tag{141}$$

The Abbe sine condition (98) also relates the sampling periods $(\Delta s_x, \Delta s_y)$, $(\Delta s'_x, \Delta s'_y)$ at the entrance and exit pupils linearly through $M' = n'M/n$. Using this relationship and $k'/k = n'/n$ in (141), we obtain the sampling relations (97) given at the end of Section 3.3, reproduced here for convenience:

$$\Delta s_x < \frac{2\pi}{k W_x}, \quad \Delta s_y < \frac{2\pi}{k W_y}. \tag{142}$$

This condition places an upper limit on the distances Δs_x and Δs_y between the direction cosines of the angles at which the far-zone field is collected (see Figure 12(a)).

Assuming that the retrieval (139) is accurate, the electric field $\boldsymbol{E}_{\text{img}}(x', y')$ is now known at discrete spatial positions $(m\Delta_x, n\Delta_y)$. We

know that the vector field $\boldsymbol{E}_{\text{img}}(x', y')$ is spatially bandlimited, since $\boldsymbol{E}'_{\text{s}}(s'_x, s'_y)$ is only nonzero inside $(s'^2_x + s'^2_y)^{1/2} \leqslant \sin\theta_{\text{img}}$. The spatial bandwidth of the field is therefore $2k' \sin\theta_{\text{img}}$. From the Nyquist sampling theorem, this bandlimited field is completely determined by its sampled version $\boldsymbol{E}_{\text{img}}(m\Delta, n\Delta)$ if the sampling period is smaller than or equal to $\Delta = 2\pi/(2k' \sin\theta_{\text{img}}) = \lambda'/(2\sin\theta_{\text{img}})$, which corresponds to the traditional definition of the "diffraction limit." For the minimum allowable DFT lengths $N_p = P$ and $N_q = Q$ in (135), it follows from (132) that both sampling periods in (137) are equal to this limit, and the continuous field $\boldsymbol{E}_{\text{img}}(x', y')$ is represented by the least possible number of sampling points. In order to evaluate the continuous field $\boldsymbol{E}_{\text{img}}(x', y')$ at higher spatial precision, one can simply increase the DFT lengths N_p and N_q.

A second way to numerically evaluate (114) follows from a generalization of the Nijboer–Zernike aberration theory (Born & Wolf, 1999), called the extended Nijboer–Zernike (ENZ) theory by its developers (Braat, Dirksen, & Janssen, 2002; Braat, Dirksen, Janssen, & van de Nes, 2003; Braat, Dirksen, Janssen, van Haver, & van de Nes, 2005; Janssen, 2002). The important features of the ENZ method are the evaluation of the series expansions in the original Nijboer–Zernike method with more terms, and the redefinition of the aberration function $\Phi(s'_x, s'_y)$ as a complex quantity, thereby accounting for the variations of the vector amplitude $\boldsymbol{E}'_{\text{s}}(s'_x, s'_y)$ on the wavefront. In the ENZ method, one starts by expanding the aberration function $\Phi(s'_x, s'_y)$ into a series in the form

$$\Phi(s'_x, s'_y) = \sum_{n,m} \alpha_{nm} R_n^m(s') \cos(m\phi'), \tag{143}$$

in which $s' = (s'^2_x + s'^2_y)^{1/2}$ and ϕ' is the azimuthal angle in the image space. In (143), $R_n^m(s')$ are the Zernike circle polynomials (Born & Wolf, 1999), and different terms in (143) are orthogonal to each other inside the unit circle. The coefficients α_{nm} can therefore be obtained by integrating $\Phi(s'_x, s'_y) R_n^m(s') \cos(m\phi')$ over the unit circle and making use of the orthogonality property. Because polar coordinates (s', ϕ') are used in (143), the polar arrangement of (s_x, s_y) in Figure 12(b) at the collection step is more suitable for this integration. Next, the generalized Debye–Wolf integral (114) is expanded into an infinite series in $\Phi(s'_x, s'_y)$, resulting in definite integrals involving the products of exponential and Bessel functions. These integrals can be evaluated *off-line*, and lookup tables can be generated for repeated use in the future. Using these lookup tables and the coefficients α_{nm} of the aberration function $\Phi(s'_x, s'_y)$, the generalized Debye–Wolf integral (114) can be

calculated to a desired degree of accuracy. Any additional phase term in the aberration function $\Phi(s'_x, s'_y)$ requires only the recalculation of the coefficients α_{nm}. At this point, the use of Zernike circle polynomials in (143) becomes a real advantage. If the additional phase term for the aberration is expressed in terms of Zernike polynomials, only the α_{nm} terms that have the same indices as those polynomials will be affected. As a result, the effects of primary aberrations (spherical aberration, coma, astigmatism, etc.) on the imaging performance can be investigated extremely efficiently, without any need to carry out two-dimensional numerical integration or numerical quadrature for each different aberration. The ENZ method has also been generalized to planar multilayered structures (Braat, van Haver, Janssen, & Pereira, 2009).

The ENZ formulation does not suffer from the aliasing artifacts encountered in the DFT-based formulation. It is inherently geared toward synthesizing the images of non-periodic structures. In van Haver et al. (2008, 2009) and Janssen et al. (2008), the image-space field distributions (also called *aerial images* in photolithography) of non-periodic masks were computed using the ENZ method. On the other hand, convergence and range of validity issues are of greater importance in the ENZ method because of the heavy use of series expansions. The relative theoretical complexity and the difficulty of constructing the lookup tables is another disadvantage.

It should be remembered that the integral expression (114) for the image field is created by a coherent illumination beam. If Köhler illumination is employed (see Section 3.1.2), this coherent illumination beam is one of the plane-wave components in Figure 5. In order to obtain the total intensity at the image space, the image intensities corresponding to every incidence direction and polarization in Figure 5 should be added. As mentioned in Section 2, the image intensity due to a single plane wave is proportional to the absolute square of $\boldsymbol{E}_{\mathrm{img}}(x', y')$ in (114).

One subtle point that needs to be addressed with regard to non-periodic scatterers is the presence of planar material layers in the object space. This issue was touched upon at the end of Section 3.3. The NFFFT surface S in Figure 11(b), only collects the *scattered field* and not the incident or reflected beam. Therefore, the contribution to the image by the latter needs to be calculated separately. Let the incident or reflected beam be a plane wave with direction cosines (s_{x_i}, s_{y_i}) in the object space. Other coherent beams can be expressed as a sum of plane waves (see Section 3.1.1). Notationally, this plane wave can be regarded as the zeroth-order Floquet mode in (58):

$$\boldsymbol{E}_i(x, y) = \boldsymbol{R}_{00} \mathrm{e}^{-jk(s_{x_i} x + s_{y_i} y)}. \tag{144}$$

With this notation, the results for the image field of periodic scatterers in Section 3.4.1 are immediately applicable. Only considering the $(p, q) = (0, 0)$ mode in (121), the image field due to the incident or reflected plane wave is found as

$$\boldsymbol{E}_{\text{img}}(x', y') = -\sqrt{\frac{n' \cos\theta_i}{n \cos\theta_i'}} \frac{1}{M'} \boldsymbol{R}'_{00} \mathrm{e}^{-jk'\Phi(s'_{x_i}, s'_{y_i})} \mathrm{e}^{-jk'(s'_{x_i}x' + s'_{y_i}y')}, \quad (145)$$

in which $\boldsymbol{R}'_{00}$ is given by (120). Using (145), each incident or reflected plane wave can be propagated to the image space and added coherently to the image field (139) scattered from the object.

We conclude this section with a brief discussion of broadband imaging. Almost all of the formulation presented for the illumination, collection, and refocusing steps has been for harmonic time dependence $\exp(j\omega t)$, which corresponds to full temporal coherence and an infinitesimally narrow bandwidth. All the results so far can be immediately generalized to a broadband source. It was mentioned in Section 2 that a large class of optical sources can be modeled as statistically stationary in time. The total intensity at any position (x', y') at the image plane is the integral of the power-spectral density $S_{\text{img}}(x', y'; \omega)$ across all frequencies. Therefore, from the power-spectral density relation (12), the total intensity at the image plane can be found by repeating the entire numerical imaging formulation for every frequency component present in the source, and adding the *image intensities* corresponding to each frequency component. As mentioned in Section 2, the frequency response $H(\omega)$ can be evaluated at multiple frequencies in a single simulation run if a broadband method such as FDTD is used to calculate the time-domain scattering response. Another technical point related to broadband simulation is the effect of different excitation wavelengths on the numerical calculation of the image. If the discrete arrangement of the direction cosines $\Delta s'_x$, $\Delta s'_y$ defined by (130) is kept fixed at every frequency, the spatial sampling periods Δ_x, Δ_y as well as the aliasing periods D'_x, D'_y in (137) and (138) for the sampled image field (136) scale linearly with the wavelength λ. This complicates the direct summation of the image field in the image plane, since the field is evaluated at different spatial points at different wavelengths. To avoid this complication, it is advisable to *scale* the direction cosines Δs_x, Δs_y by the wavelength λ in the collection step. This implies that a different set of observation directions is recorded for each wavelength in the near-field-to-far-field transform (NFFFT) step (see Section 3.3). An additional advantage of scaling the direction cosines

with the wavelength is that the no-aliasing condition (142) becomes only dependent on the spatial extent of the object-side field distribution, and independent of the wavelength.

4. IMPLEMENTATION EXAMPLES

In this section, we will present some example results obtained using an optical imaging simulation software package that features some of the methods reviewed in previous sections. The simulation package is based on an in-house implementation of the three-dimensional finite-difference time-domain (FDTD) method. It has been used to generate original results for previous publications (Capoglu & Smith, 2006, 2008; Capoglu et al., 2008, 2011), and has been thoroughly tested and verified. The simulation package, named Angora, is freely available under the GNU Public License (Capoglu, 2012).

In all the examples, the object and image spaces are both free space, i.e., $n' = n = 1$. The object and image-space wavenumbers are equal ($k' = k = k_0$), and from (99), the angular de-magnification M' is equal to the magnification M.

For our first simulation example, we consider the numerical microscope image of a thin structure in the form of the letters "N" and "U" embossed on an infinite dielectric substrate. Two-dimensional cross-sections of the simulation geometry across the xy and xz planes are shown in Figure 17(a). The embossed structure is a dielectric material with refractive index $n = 1.38$ and dimensions $12\ \mu\text{m} \times 6\ \mu\text{m} \times 0.27\ \mu\text{m}$. It is situated on a glass half plane with refractive index $n = 1.5$. This could represent an idealized situation where the embossed structure is on a glass slide, if the illumination is focused on the top surface of the slide and the thickness of the slide is much larger than the focal depth of the illumination. Even the smallest-NA illumination has a finite focal depth in practice, so this condition is very easy to fulfill. When these conditions are fulfilled, the scattered beam due to the bottom surface of the slide is far out of focus and spread over a large area; therefore, it can be neglected. The structure is illuminated normally by a y-polarized plane wave that has a modulated-Gaussian time waveform $\sin(2\pi f_0 t)\exp(-t^2/2\tau^2)$ with $f_0 = 5.89 \times 10^{14}$ and $\tau = 2.13 \times 10^{-15}$ s. The -20 dB wavelengths of this beam in free space are 400 and 700 nm. This approximates a polarized illumination beam with very low illumination NA (see Section 3.1.1). Incoherent and unpolarized illumination can be achieved by repeating this simulation for multiple incidence directions and

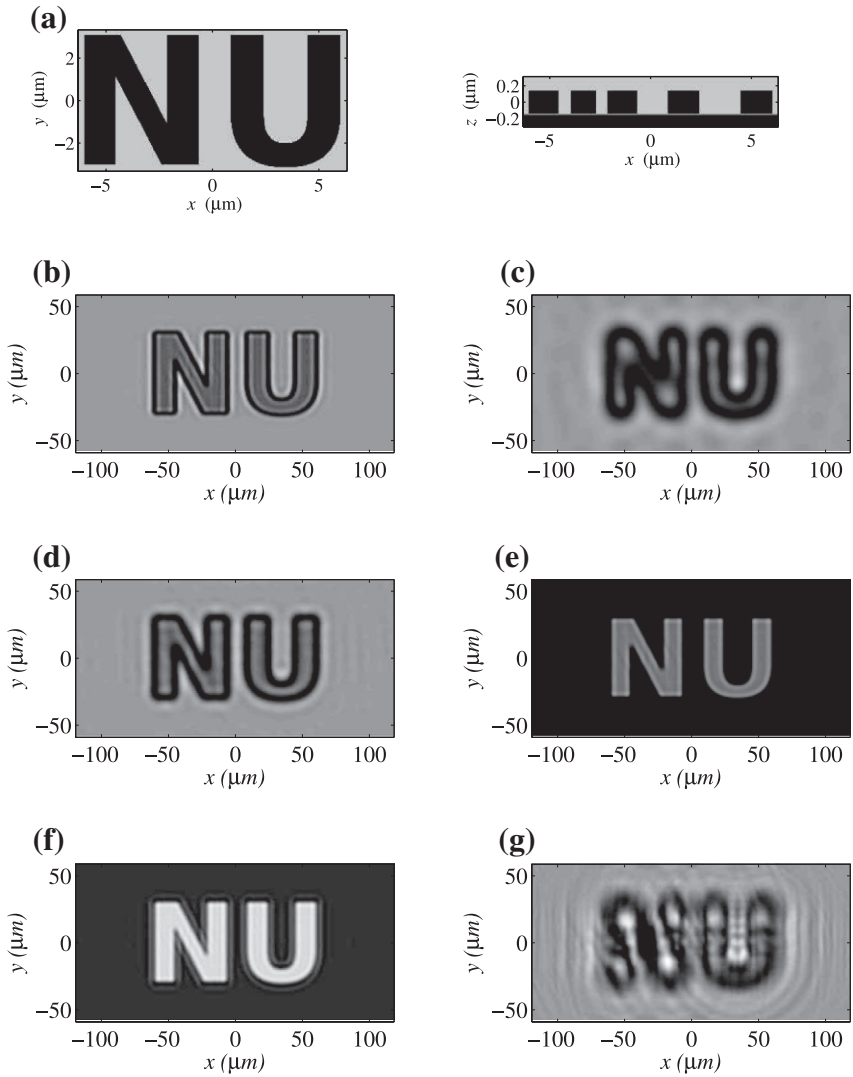

Figure 17 An example computational imaging simulation. The letters "N" and "U" are embossed on a glass substrate and imaged at magnification $M = 10$. (a) The *xy* and *xz* sections of the FDTD grid. (b) Bright-field image for $NA_{obj} = 0.9$. (c) Bright-field image for $NA_{obj} = 0.2$. (d) Bright-field image for $NA_{obj} = 0.4$. (e) Dark-field image. (f) Phase-contrast image. (g) Image of the off-focus plane $z = 6$ μm.

polarizations (see Section 3.1.2, especially the discussion involving Figure 5). The scattered near fields are computed in an FDTD grid with the following parameters: grid dimensions $12.635\ \mu m \times 6.65\ \mu m \times 0.5985\ \mu m$, grid

spacing $\Delta x = \Delta y = \Delta z = \Delta = 13.33$ nm, time step $\Delta t = (0.98/\sqrt{3})\Delta/c$ The computational grid is truncated by 10-cell thick convolution perfectly matched layer (CPML) (Roden & Gedney, 2000). The illumination plane wave is sourced into the FDTD grid using the total-field/scattered-field (TF/SF) formulation (Taflove & Hagness, 2005) with multilayer capabilities (Capoglu & Smith, 2008). The TF/SF boundary is placed eight cells away from the PML boundary. The scattered field is collected on a surface four cells away from the PML, and transformed to the far field using a frequency-domain vector-potential near-field-to-far-field transform (NFFFT) algorithm (Taflove & Hagness, 2005) for a two-layered medium (Capoglu et al., 2012). The far field is calculated at a set of observation directions (θ, ϕ) arranged as in Figure 12(a), with equally spaced direction cosines (s_x, s_y) within a collection numerical aperture $\mathrm{NA}_{\mathrm{obj}} = 0.9$. The spacings of the direction cosines are $\Delta s_x = 0.0167$ and $\Delta s_y = 0.0333$. A smaller spacing is necessary in s_x because the structure is wider in the x direction [see Section 3.4.2, Equation (142)]. At each observation direction, the far field is calculated at 7 wavelengths between 400 and 700 nm with equal spacing in $k = 2\pi/\lambda$. It is assumed that the microscope has magnification $M = 10$ and is free of aberrations ($\Phi = 0$). The sampled field distribution $\boldsymbol{E}_{\mathrm{img}}(m\Delta_x, n\Delta_y)$ at the image plane is calculated using the DFT-based refocusing algorithm of Section 3.4.2, described by Equations (134)–(139) with $\boldsymbol{E}'_{\mathrm{s}}(s'_{x_p}, s'_{y_q})$ given by (108). The continuous field $\boldsymbol{E}_{\mathrm{img}}(x', y')$ is oversampled with $N_p = N_q = 256$ in (135) for a smooth intensity image. Since the scattering geometry is two layered, the plane wave reflected from the air–glass interface has to be propagated to the image space using (144) and (145). Unless otherwise noted, all intensity spectra are normalized by the intensity spectrum at a pixel corresponding to the glass region. In Figure 17(b), the mean of the normalized intensity across all wavelengths (called the *bright-field image*) for a collection NA of 0.9 is shown in grayscale with black and white corresponding to 0 and 1.85, respectively. In Figure 17(c) and (d), the same image is shown for collection NAs of 0.2 and 0.4 within the same grayscale limits. The blurring of the image due to the increased diffraction limit is immediately apparent. In Figure 17(e), the plane wave reflected from the glass slide is subtracted from the image, resulting in a modality similar to *dark-field microscopy*. If the reflected plane wave is phase-shifted by 90° instead of being subtracted from the total image, the image in Figure 17(f) is obtained. This is very similar to the procedure followed in *phase-contrast microscopy*. In both Figure 17(e) and (f), the collection NA is 0.9 and the spectra are normalized by the same glass spectrum used to

normalize the previous figures. However, the grayscale limits are 0 and 3.5 because of the higher intensity contrast. It is seen that the phase-contrast image yields better contrast than the dark-field image. Finally, in Figure 17(g), the normalized bright-field image of the off-focus plane $z = 6\ \mu\text{m}$ at the object space is shown for a collection NA of 0.9. The distortion of the image due to the lack of sharp focus is clearly visible.

In our second implementation example (Capoglu et al., 2011), we compare the numerically calculated "microscope in a computer" images and spectra of polystyrene latex beads to experimental results obtained from an actual microscope coupled to a spectrometer and a CCD camera. In the actual experiment, two different sizes of polystyrene latex beads (2.1 μm and 4.3 μm diameter, Thermo Scientific) are placed on a glass slide with refractive index $n = 1.5$. The refractive index of the latex beads is frequency-dependent, but is guaranteed to be between 1.59 and 1.61 in the wavelength range 486–589 nm according to the bead specifications. The illumination is passed through a diffuser to smoothen the inhomogeneity of the white-light xenon lamp, and projected on the sample using a Köhler setup with illumination numerical aperture $\text{NA}_{\text{ill}} = 0.2$. The magnification of the microscope is $M = 40$ and the collection numerical aperture is $\text{NA}_{\text{obj}} = 0.6$. The image of the sample is projected on a spectrograph with a 10 μm slit width coupled with a CCD camera. The spectrograph records the spectra at a column of pixels, resulting in a 2D data array. A 3D spectroscopic image is acquired by scanning the slit of the spectrograph over the image with a 10 μm step. More details on the optical setup can be found in (Liu, Li, Kim, & Backman, 2005). For the numerical simulation of this microscope, we used the FDTD package mentioned in the beginning of the section. The parameters for the FDTD simulation are as follows: grid dimensions $5\ \mu\text{m} \times 5\ \mu\text{m} \times 5\ \mu\text{m}$ with spacing $\Delta x = \Delta y = \Delta z = 31$ nm, time step $\Delta t = (0.98/\sqrt{3})\Delta x/c$. The grid is terminated with 10-cell thick convolution PML. In the simulations, a fixed refractive-index value of $n = 1.61$ is chosen for the polystyrene beads as a first approximation. Exploiting the rotational symmetry of the bead and the microscope setup, 204 incident plane waves (102×2 polarizations) are distributed in the Cartesian arrangement of Figure 5 only within the first quadrant of the circle of illumination. The final image is synthesized by rotating the resulting image intensity by 0°, 90°, 180°, and 270°, and adding the intensities together. Each incident plane wave has a sine-modulated Gaussian waveform, with −20-dB wavelengths at 486 and 589 nm. The scattered light is collected at a set of observation directions arranged with

equally spaced 50×50 direction cosines (s_x, s_y) inside a collection numerical aperture $\mathrm{NA}_{\mathrm{obj}} = 0.6$. The spacing in s_x and s_y is uniform, and equal to $\Delta s_x = \Delta s_y = 0.024$. The far field at each direction is calculated at 30 wavelengths between 486 and 589 nm spaced linearly in $k = 2\pi/\lambda$. The sampled field distribution $\boldsymbol{E}_{\mathrm{img}}(m\Delta_x, n\Delta_y)$ at the image plane is calculated with no oversampling (at 50×50 points) using the refocusing algorithm (134)–(139) with $\boldsymbol{E}'_{\mathrm{s}}(s'_{x_p}, s'_{y_q})$ given by (108). Both the measured/simulated spectra are normalized by the measured/simulated spectrum at a glass pixel. The grayscale plots on the left and center columns of Figure 18 show the simulated and measured bright-field images, respectively. The plots on the right column show the measured and simulated spectra between

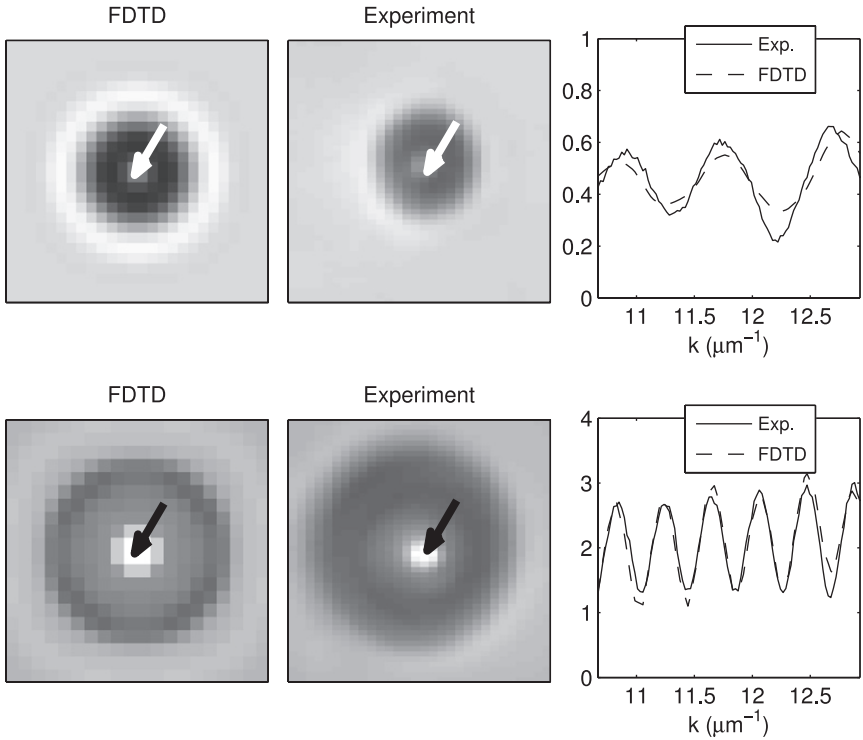

Figure 18 Comparison of microscopy images and pixel spectra obtained via an FDTD "microscope in a computer" software package and an actual spectroscopic microscope. The bright-field images from FDTD and experiment are shown in the left and center columns, while the spectra from the center pixel are shown in the right column. (Top) 2.1 μm bead. (Bottom) 4.3 μm bead. (*Source:* Capoglu et al. (2011), © 2011 The Optical Society.)

486 and 589 nm at the pixels annotated by arrows. Since the precise focal plane position in the measurement is unknown, FDTD image at the optimum focusing depth is chosen for each comparison. The optimum focal positions at the object space are $z = 1.58\ \mu\text{m}$ for the 2.1 μm bead and $z = 1.44\ \mu\text{m}$ for the 4.3 μm bead. At these optimum focal positions, the root-mean-square errors in the spectra are 11.4% and 8.2%, respectively. The error is primarily caused by the dispersive nature of the beads and the variations in their sizes. It should be noted that the same multilayer techniques used for the previous example have also been used here. The bottom surface of the glass slide on which the beads is placed is far out of focus; therefore, the reflection from that interface is spread over a large area with much reduced intensity at the top interface. For this reason, the glass slide is modeled as a two-layered space.

Our final example will demonstrate the potential of the imaging algorithms reviewed in this paper for modeling spatial features in the nanometer scale. We consider a cell model placed on a glass slide ($n = 1.5$), and calculate its microscope images using the same FDTD package used for the previous examples. The surface profile of the cell model is that of a human buccal cell from inside the cheek, measured using atomic-force microscopy (AFM) and read into an FDTD grid. The measured AFM surface profile of the cell is shown in grayscale in Figure 19(a). The maximum height value, represented by the brightest shade of gray, is 990 nm. The interior of the cell model is filled with a homogeneous material of refractive index $n = 1.38$—a value guided by previous cell refractometry studies (Beuthan, Minet, Helfmann, Herrig, & Muller, 1996; Lue et al., 2006). The buccal cell is approximately of dimensions 80 μm × 80 μm × 1 μm. The parameters of the FDTD grid are as follows: 85 μm × 85 μm × 1.45 μm grid with spacing $\Delta x = \Delta y = \Delta z = 25$ nm, time step $\Delta t = \left(0.98/\sqrt{3}\right) \Delta x/c$, five-cell thick convolution PML. The illumination is a normally incident plane wave with a sine-modulated Gaussian waveform whose –20-dB wavelengths are 400 and 700 nm, respectively. Two orthogonal polarizations are sent separately, and the resulting intensities are added for the final image. The far-zone field is calculated at directions arranged as in Figure 12(a), with direction-cosine spacing $\Delta s_x = \Delta s_y = 0.0048$ inside a collection numerical aperture of $\text{NA}_{\text{obj}} = 0.6$. For each observation direction, 10 wavelengths (spaced linearly in k) are recorded between 400 and 700 nm. The final intensity spectrum is normalized by the spectrum of a glass pixel, resulting in a normalized spectroscopic reflectance image. In Figure 19(b),

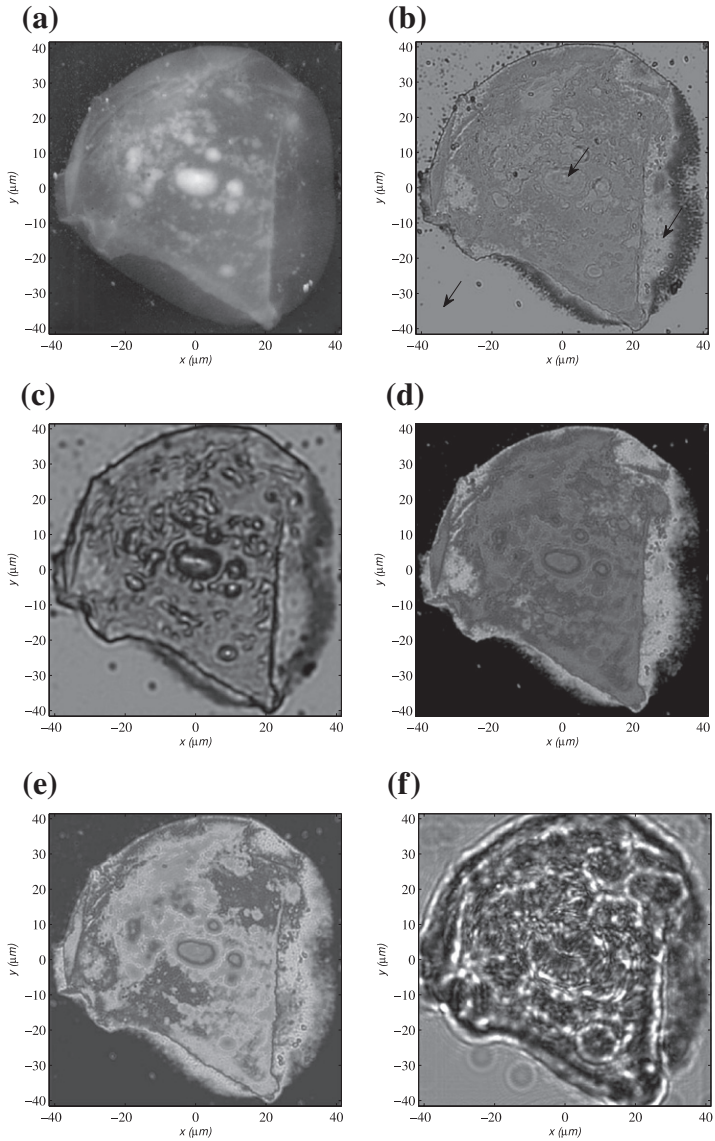

Figure 19 Simulated microscope images of a buccal (cheek) cell model at magnification $M = 1$. (a) The measured AFM profile of the cell. The maximum height is 990 nm. (b) Bright-field image for $NA_{obj} = 0.6$. The spectra at the pixels annotated by arrows are plotted in Figure 20. (c) Bright-field image for $NA_{obj} = 0.2$. (d) Dark-field image. (e) Phase-contrast image. (f) Image of the off-focus plane $z = 20\ \mu m$. [Grayscale values are 137 nm→990 nm for (a); 0 →2 for (b), (c), (f); 0 →6 for (d) and (e).]

the simulated bright-field reflectance image at $NA_{obj} = 0.6$ is shown in grayscale. The minimum and maximum brightness values correspond to 0 and 2, respectively. In Figure 19(c), the bright-field image for a reduced collection NA of 0.2 is shown. The dark-field and phase-contrast images are shown in Figure 19(d) and (e), with brightness values between 0 and 6. The effect of focusing is demonstrated in Figure 19(f), where the bright-field image of the off-focus plane $z = 20$ μm is shown at the same grayscale level as Figure 19(b)–(f). It should be noted that there is a spectrum associated with every recorded pixel, and Figure 19(b)–(f) all represent certain averages over these spectra. The normalized reflectance spectra at the three pixels (nucleus, cytoplasm, glass) annotated by the arrows in Figure 19(b) are shown in Figure 20. The spectrum at the pixel outside the cell is almost unity, since the spectra are normalized by the spectrum of a glass pixel. The deviation from unity is because the AFM measurements are not flat and noise-free outside the cell area. Finally, we would like to note that the partially coherent Köhler illumination described in Section 3.1.2 has *not* been implemented in this example. However, the angular-shift invariance approximation mentioned in Section 3.1.1 and Section 3.2 can safely be made for this example. The lateral dimensions of the buccal cell are almost two orders of magnitude larger than the axial dimension. Under the angular-shift invariance, the response to the oblique plane waves within a finite illumination numerical aperture NA_{ill} can be approximated as angularly shifted versions of the response to the normally incident plane wave.

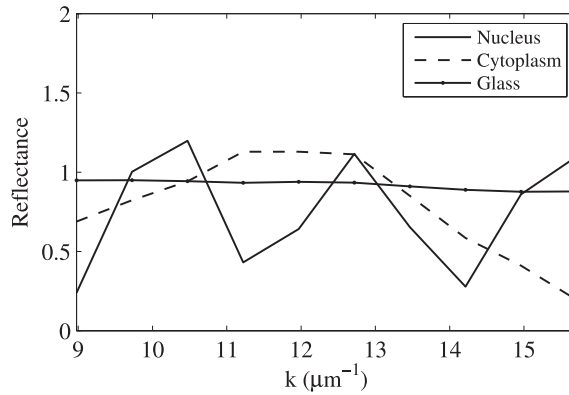

Figure 20 Reflectance spectra between 400 and 700 nm at three pixels in the simulated microscope image of the buccal-cell model. The three pixels, annotated by arrows in Figure 19(b), fall within the nuclear, cytoplasmic, and glass regions.

The last implementation example clearly demonstrates the power of numerical computation for simulating optical microscopy modalities. Although microscopy simulation has been previously applied to integrated-circuit inspection and mark alignment, computational resources have only recently caught up with the possibility of applying the same principles to large and extremely complex biological media (Hollmann et al., 2004; Starosta & Dunn, 2009). The continued increase in the speed and availability of computing resources will soon make it possible to simulate even the most complicated biological samples within manageable time limits. Although a substantially large scatterer such as the buccal-cell model in Figure 19 is still beyond the range of a personal workstation, this is bound to be overcome in the near future.

5. SUMMARY

In this tutorial and review paper, a comprehensive account of the theoretical principles behind the numerical electromagnetic modeling of optical imaging systems and a thorough literature survey of the entire subject is presented. In short, this virtual system can be called a "microscope in a computer." The underlying principles, however, are applicable to any optical imaging modality, including those used in photolithography, metrology, inspection, and alignment. The optical imaging system is decomposed into four self-contained subcomponents (illumination, scattering, collection, and refocusing), and each of these subcomponents is mathematically analyzed. Approximate numerical methods used in the modeling of each subcomponent are explained in appropriate detail. Relevant practical applications are cited whenever applicable. The algorithms reviewed in the paper are illustrated via several implementation examples involving the simulated microscopy images of nanoscale structures. The paper will hopefully constitute a useful starting point for those interested in modeling optical imaging systems from a rigorous electromagnetics point of view. A distinct feature of this paper is the extra attention paid to the issues of discretization and signal processing. This is a key issue in finite methods, where the electromagnetic field is only given at a finite set of spatial and temporal points.

ACKNOWLEDGMENTS

This work was supported by the NIH Grants R01EB003682 and R01CA128641. The numerical simulations in Section 4 have been made possible by a supercomputing grant on the Quest high performance computing system at Northwestern University. The authors would also like to thank Samantha Dale Strasser and the NUANCE facility for providing the atomic-force microscopy measurements.

APPENDIX A. DERIVATION OF (18)

The mutual coherence function $J^*(x_d; y_d)$ resulting from the finite collection of plane waves in Figure 5 is

$$J^*(x_d; y_d) = \Delta s_x \Delta s_y \sum_{m,n} P(s_{x_m}, s_{y_n}) \mathrm{e}^{-jk(s_{x_m} x_d + s_{y_n} y_d)}, \tag{146}$$

where both m and n range from $-\infty$ to ∞. The expression (15) for the original mutual coherence function $J(x_d; y_d)$ is in the form of a Fourier transform and can be inverted to yield the following for $P(s_x, s_y)$:

$$P(s_x, s_y) = \frac{k^2}{(2\pi)^2} \iint_{-\infty}^{\infty} J(x'_d; y'_d) \mathrm{e}^{jk(s_x x'_d + s_y y'_d)} \, \mathrm{d}x'_d \mathrm{d}y'_d. \tag{147}$$

Primed coordinates are used to avoid confusion in what follows. Using (147), (146) becomes

$$\begin{aligned} J^*(x_d; y_d) = & \iint_{-\infty}^{\infty} \mathrm{d}x'_d \mathrm{d}y'_d J(x'_d; y'_d) \\ & \times \left[\frac{k^2 \Delta s_x \Delta s_y}{(2\pi)^2} \sum_{m,n} \mathrm{e}^{jk(s_{x_m}(x'_d - x_d) + s_{y_n}(y'_d - y_d))} \right]. \end{aligned} \tag{148}$$

Substituting $s_{x_m} = m\Delta s_x$ and $s_{y_n} = n\Delta s_y$, the expression in square brackets becomes (Oppenheim, Willsky, & Nawab, 1997)

$$\sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \delta\left(x'_d - x_d - r\frac{2\pi}{k\Delta s_x}\right) \delta\left(y'_d - y_d - s\frac{2\pi}{k\Delta s_y}\right), \tag{149}$$

which, when substituted into (148), yields the desired relation (18).

APPENDIX B. DERIVATION OF (72)

The original, periodic, continuous vector field $E(x, y)$ can be written using the Floquet expansion (58) as follows:

$$\boldsymbol{E}(x, y) = \sum_{a=-\infty}^{\infty} \sum_{b=-\infty}^{\infty} \boldsymbol{R}_{ab} \mathrm{e}^{-j(\beta_a x + \beta_b y)}, \tag{150}$$

where the integral indices a, b are used to avoid later confusion with the Fourier indices p, q. Sampling this function at $m\Delta x$ and $n\Delta y$, where Δx

and Δy are given by (69), and applying the phase shift as defined in (70), one obtains

$$\bar{\boldsymbol{E}}[m,n] = \sum_{a=-\infty}^{\infty}\sum_{b=-\infty}^{\infty} \boldsymbol{R}_{ab} \mathrm{e}^{-j(\beta_a m\Delta x+\beta_b n\Delta y)} \times \mathrm{e}^{jk\sin\theta_i\left(d_x\cos\phi_i\frac{m}{M}+d_y\sin\phi_i\frac{n}{N}\right)}. \tag{151}$$

Substituting the expressions for the Floquet wavenumbers $\beta_a = k\cos\phi_i\sin\theta_i - a(2\pi/d_x)$ and $\beta_b = k\sin\phi_i\sin\theta_i - b(2\pi/d_y)$, and the sampling relations (69), the above expression simplifies to

$$\bar{\boldsymbol{E}}[m,n] = \sum_{a=-\infty}^{\infty}\sum_{b=-\infty}^{\infty} \boldsymbol{R}_{ab} \mathrm{e}^{j2\pi\left(\frac{am}{M}+\frac{bn}{N}\right)}. \tag{152}$$

Now, the DFT of this array is given by

$$\begin{aligned}\tilde{\bar{\mathbf{E}}}[p,q] &= \sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\bar{\boldsymbol{E}}[m,n]\mathrm{e}^{-j2\pi\left(\frac{pm}{M}+\frac{qn}{N}\right)} \\ &= \sum_a\sum_b \boldsymbol{R}_{ab}\left[\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\mathrm{e}^{-j2\pi\left(\frac{(p-a)m}{M}+\frac{(q-b)n}{N}\right)}\right].\end{aligned} \tag{153}$$

The expression in square brackets is equal to

$$MN\sum_{r=-\infty}^{\infty}\sum_{s=-\infty}^{\infty}\delta[a,p+rM]\delta[b,q+sN], \tag{154}$$

where $\delta[\cdot,\cdot]$ is the Kronecker delta symbol. Substituting (154) in (153), we obtain the desired relation (72).

APPENDIX C. DERIVATION OF (136)

The sampled auxiliary variable $\boldsymbol{G}'[p,q]$ of (134) is, from (129)–(132),

$$\boldsymbol{G}'[p,q] = \frac{-jk'}{2\pi}\int\int_{-\infty}^{\infty}\mathrm{d}x'\mathrm{d}y'\boldsymbol{E}_{\mathrm{img}}(x',y') \times \mathrm{e}^{jk'\left(s'_{x_0}x'+s'_{y_0}y'\right)}\mathrm{e}^{jk'\left(p\Delta s'_x x'+q\Delta s'_y y'\right)}. \tag{155}$$

The DFT of $\boldsymbol{G}'[p, q]$ is, from (135),

$$\begin{aligned}\boldsymbol{E}[m, n] &= \sum_{p=0}^{N_p-1}\sum_{q=0}^{N_q-1} \boldsymbol{G}'[p, q]\mathrm{e}^{-j2\pi\left(\frac{pm}{N_p}+\frac{qn}{N_q}\right)} \\ &= \frac{-jk'}{2\pi}\int\int_{-\infty}^{\infty} \mathrm{d}x'\mathrm{d}y'\boldsymbol{E}_{\mathrm{img}}(x', y')\mathrm{e}^{jk'\left(s'_{x_0}x'+s'_{y_0}y'\right)} \\ &\quad\times \sum_{p=0}^{N_p-1}\sum_{q=0}^{N_q-1} \mathrm{e}^{j\left[\left(k'\Delta s'_x x'-\frac{2\pi m}{N_p}\right)p+\left(k'\Delta s'_y y'-\frac{2\pi n}{N_q}\right)q\right]}.\end{aligned} \tag{156}$$

The summations in p and q can be extended to infinity, since $\boldsymbol{G}'[p, q]$ is only nonzero for a finite number of p and q values. The resulting infinite summation is equal to an infinite series of Dirac delta functions (Oppenheim et al., 1997)

$$\begin{aligned}&\sum_{p=-\infty}^{\infty}\sum_{q=-\infty}^{\infty} \mathrm{e}^{j\left[\left(k'\Delta s'_x x'-\frac{2\pi m}{N_p}\right)p+\left(k'\Delta s'_y y'-\frac{2\pi n}{N_q}\right)q\right]} \\ &= (2\pi)^2\sum_{r=-\infty}^{\infty}\sum_{s=-\infty}^{\infty} \delta\left(k'\Delta s'_x x'-\frac{2\pi m}{N_p}-2\pi r\right) \\ &\quad\times\delta\left(k'\Delta s'_y y'-\frac{2\pi n}{N_q}-2\pi s\right).\end{aligned} \tag{157}$$

Using the scaling property of the Dirac delta function, this becomes

$$\frac{(2\pi)^2}{(k')^2\Delta s'_x\Delta s'_y}\sum_{r=-\infty}^{\infty}\sum_{s=-\infty}^{\infty} \delta(x'-m\Delta_x-rD'_x)\times\delta(y'-n\Delta_x-sD'_y), \tag{158}$$

in which the spatial sampling periods Δ_x, Δ_y and the aliasing periods D'_x, D'_y are given by (137) and (138). Substituting this expression into (156) and using the sifting property of the delta function to evaluate the integral, the desired relation (136) is obtained.

REFERENCES

Adam, K., & Neureuther, A. R. (2002). Methodology for accurate and rapid simulation of large arbitrary 2D layouts of advanced photomasks. In G. T. Dao, & B. J. Grenon (Eds.), *21st Annual Bacus symposium on photomask technology, Pts 1 and 2. Proceedings of SPIE* (Vol. 4562, pp. 1051–1067).

Aguilar, J. F., & Mendez, E. R. (1994). Imaging optically thick objects in optical scanning microscopy: Perfectly conducting surfaces. *Journal of the Optical Society of America A – Optics Image Science and Vision, 11*(1), 155–167.

Aguilar, J. F., Mendez, E. R., & Maradudin, A. A. (2002). Imaging thick objects in optical scanning microscopy: Penetrable surfaces and patterned-layer structures. *Journal of Optical Technology, 69*(6), 422–427.

Arridge, S. R., & Hebden, J. C. (1997). Optical imaging in medicine: II. Modelling and reconstruction. *Physics in Medicine and Biology, 42*(5), 841–853.

Balanis, C. A. (1989). *Advanced Engineering Electromagnetics*. New York: Wiley.

Bao, G., Chen, Z. M., & Wu, H. J. (2005). Adaptive finite-element method for diffraction gratings.. *Journal of the Optical Society of America A – Optics Image Science and Vision, 22*(6), 1106–1114.

Barouch, E., Cole, D. C., Hollerbach, U., & Orszag, S. A. (1993). Vector aerial image with off-axis illumination. In J. D. Cuthbert (Ed.), *Optical/laser microlithography, Pts 1 and 2. Proceedings of SPIE* (Vol. 1927, pp. 686–708).

Barrett, H., & Myers, K. (2004). *Foundations of Image Science*. Hoboken, NJ: Wiley-Interscience.

Berenger, J.-P. (1994). A perfectly matched layer for the absorption of electromagnetic waves. *Journal of Computational Physics, 114*(2), 185–200.

Bernard, D. A., & Urbach, H. P. (1991). Thin-film interference effects in photolithography for finite numerical apertures.. *Journal of the Optical Society of America A – Optics Image Science and Vision, 8*(1), 123–133.

Besbes, M., Hugonin, J. P., Lalanne, P., van Haver, S., Janssen, O. T. A., Nugrowati, A. M., et al. (2007). Numerical analysis of a slit-groove diffraction problem. *Journal of the European Optical Society – Rapid Publications, 2*, 07022-1–17.

Beuthan, J., Minet, O., Helfmann, J., Herrig, M., & Muller, G. (1996). The spatial variation of the refractive index in biological cells. *Physics in Medicine and Biology, 41*(3), 369–382.

Bochkanov, S., Bystritsky, V. (2008). *Computation of Gauss–Legendre quadrature rule nodes and weights*. Alglib.net – Web resource. <http://www.alglib.net/integral/gq/glegendre.php> Accessed 08, 2008.

Booth, M. J., Juskaitis, R., & Wilson, T. (2008). Spectral confocal reflection microscopy using a white light source.. *Journal of the European Optical Society – Rapid Publications, 3*, 08026-1–08026-6.

Born, M., & Wolf, E. (1999). *Principles of optics: Electromagnetic theory of propagation, interference and diffraction of light* (7th ed.). Cambridge: Cambridge University Press.

Botten, L. C. (1978). A new formalism for transmission gratings. *Optica Acta, 25*(6), 481–499.

Botten, I. C., Craig, M. S., McPhedran, R. C., Adams, J. L., & Andrewartha, J. R. (1981). The dielectric lamellar diffraction grating. *Optica Acta, 28*(3), 413–428.

Braat, J., Dirksen, P., & Janssen, A. J. E. M. (2002). Assessment of an extended Nijboer–Zernike approach for the computation of optical point-spread functions. *Journal of the Optical Society of America A – Optics Image Science and Vision, 19*(5), 858–870.

Braat, J. J. M., Dirksen, P., Janssen, A. J. E. M., & van de Nes, A. S. (2003). Extended Nijboer–Zernike representation of the vector field in the focal region of an aberrated high-aperture optical system. *Journal of the Optical Society of America A – Optics Image Science and Vision, 20*(12), 2281–2292.

Braat, J. J. M., Dirksen, P., Janssen, A. J. E. M., van Haver, S., & van de Nes, A. S. (2005). Extended Nijboer–Zernike approach to aberration and birefringence retrieval in a high-numerical-aperture optical system. *Journal of the Optical Society of America A – Optics Image Science and Vision, 22*(12), 2635–2650.

Braat, J. J. M., van Haver, S., Janssen, A. J. E. M., & Pereira, S. F. (2009). Image formation in a multilayer using the extended Nijboer–Zernike theory. *Journal of the European Optical Society – Rapid Publications, 4*, 09048–09112.

Bracewell, R. N. (1986). *The Fourier transform and its applications* (2nd ed.). Boston: McGraw Hill.

Brukman, M., Deng, Y. F., & Neureuther, A., 2000. Simulation of EUV multilayer mirror buried defects. In E. A. Dobisz (Ed.), *Emerging lithographic technologies IV. Proceedings of SPIE* (Vol. 3997, pp. 799–806).

Burckhardt, C. B. (1966). Diffraction of a plane wave at a sinusoidally stratified dielectric grating. *Journal of the Optical Society of America A – Optics Image Science and Vision, 56*(11), 1502–1509.

Capoglu, I. R. (2007). *Techniques for handling multilayered media in the FDTD method*. Ph.D. thesis, Georgia Institute of Technology, Atlanta, GA. <http://goo.gl/7gGN7>.

Capoglu, I. R. (2012). *Angora: A free software package for finite-difference time-domain (FDTD) electromagnetic simulation*. <http://www.angorafdtd.org> Accessed 2012.

Capoglu, I. R., & Smith, G. S. (2006). A direct time-domain FDTD near-field-to-far-field transform in the presence of an infinite grounded dielectric slab. *IEEE Transactions on Antennas and Propagation, 54*(12), 3805–3814.

Capoglu, I. R., & Smith, G. S. (2008). A total-field/scattered-field plane-wave source for the FDTD analysis of layered media. *IEEE Transactions on Antennas and Propagation, 56*(1), 158–169.

Capoglu, I. R., Taflove, A., & Backman, V. (2008). Generation of an incident focused light pulse in FDTD. *Optics Letters, 16*(23), 19208–19220.

Capoglu, I. R., Taflove, A., & Backman, V. (2012). A frequency-domain near-field-to-far-field transform for planar layered media. *IEEE Transactions on Antennas and Propagation, 60*(4), 1878–1885.

Capoglu, I. R., White, C. A., Rogers, J. D., Subramanian, H., Taflove, A., & Backman, V. (2011). Numerical simulation of partially-coherent broadband optical imaging using the FDTD method. *Optics Letters, 36*(9), 1596–1598.

Chandezon, J., Dupuis, M. T., Cornet, G., & Maystre, D. (1982). Multicoated gratings: A differential formalism applicable in the entire optical region. *Journal of the Optical Society of America, 72*(7), 839–846.

Chateau, N., & Hugonin, J. P. (1994). Algorithm for the rigorous coupled-wave analysis of grating diffraction. *Journal of the Optical Society of America A – Optics Image Science and Vision, 11*(4), 1321–1331.

Chen, K.-M. (1989). A mathematical formulation of the equivalence principle. *IEEE Transactions on Microwave Theory and Techniques, 37*(10), 1576–1581.

Chew, W. C. (1990). *Waves and fields in inhomogeneous media*. New York: Van Nostrand Reinhold.

Choi, K., Chon, J. W. M., Gu, M., & Lee, B. (2007). Characterization of a subwavelength-scale 3D void structure using the FDTD-based confocal laser scanning microscopic image mapping technique. *Optics Express, 15*(17), 10767–10781.

Christov, I. P. (1986). Propagation of partially coherent light pulses. *Optica Acta, 33*(1), 63–72.

Chu, R. S., & Kong, J. A. (1977). Modal theory of spatially periodic media. *IEEE Transactions on Microwave Theory and Techniques, 25*(1), 18–24.

Coen, S., Chau, A. H. L., Leonhardt, R., Harvey, J. D., Knight, J. C., Wadsworth, W. J., et al. (2002). Supercontinuum generation by stimulated Raman scattering and parametric four-wave mixing in photonic crystal fibers.. *Journal of the Optical Society of America B – Optical Physics, 19*(4), 753–764.

Coifman, R., Rokhlin, V., & Wandzura, S. (1993). The fast multipole method for the wave equation: A pedestrian prescription. *IEEE Antennas and Propagation Magazine, 35*(3), 7–12.

Cole, D. C., Barouch, E., Conrad, E. W., & Yeung, M. (2001). Using advanced simulation to aid microlithography development. *Proceedings of the IEEE, 89*(8), 1194–1215.

Cole, D. C., Barouch, E., Hollerbach, U., & Orszag, S. A. (1992a). Derivation and simulation of higher numerical aperture scalar aerial images. *Japanese Journal of Applied Physics, Part 1, 31*(12B), 4110–4119.

Cole, D. C., Barouch, E., Hollerbach, U., & Orszag, S. A. (1992b). Extending scalar aerial image calculations to higher numerical apertures. *Journal of Vacuum Science and Technology B – Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena, 10*(6), 3037–3041.

Cotter, N. P. K., Preist, T. W., & Sambles, J. R. (1995). Scattering-matrix approach to multilayer diffraction. *Journal of the Optical Society of America A – Optics Image Science and Vision, 12*(5), 1097–1103.

Davidson, M. P. (1999). Developments in optical modeling methods for metrology. In Singh, B. (Ed.), *Metrology, inspection, and process control for microlithography XIII, Pts 1 and 2. Proceedings of SPIE* (Vol. 3677, pp. 866–875).

Davidson, D. B., & Ziolkowski, R. W. (1994). Body-of-revolution finite-difference time-domain modeling of space-time focusing by a three-dimensional lens. *Journal of the Optical Society of America A – Optics Image Science and Vision, 11*(4), 1471–1490.

Demarest, K., Huang, Z., & Plumb, R. (1996). An FDTD near-to-far-zone transformation for scatterers buried in stratified grounds. *IEEE Transactions on Antennas and Propagation, 44*(8), 1150–1157.

Deng, Y. F., Pistor, T., & Neureuther, A. R., 2001. Rigorous electromagnetic simulation applied to alignment systems. In: Progler, C. J. (Ed.), *Optical microlithography XIV, Pts 1 and 2. Proceedings of SPIE* (Vol. 4346, pp. 1533–1540).

Erdmann, A., & Evanschitzky, P. (2007). Rigorous electromagnetic field mask modeling and related lithographic effects in the low k_1\(k_1\) and ultrahigh numerical aperture regime. In *Journal of Micro/Nanolithography, MEMS and MOEMS. Proceedings of SPIE* (Vol. 6, pp. 031002-1–031002-16).

Erdmann, A., Evanschitzky, P., Citarella, G., Fuhner, T., & De Bisschop, P. (2006). Rigorous mask modeling using waveguide and FDTD methods: An assessment for typical Hyper NA imaging problems. In M. Hoga (Ed.), *Photomask and next generation lithography mask technology XIII, Pts 1 and 2. Proceedings of SPIE* (Vol. 6283, p. 28319).

Erdmann, A., Fuhner, T., Shao, F., & Evanschitzky, P. (2009). Lithography simulation: Modeling techniques and selected applications. In: *Modeling aspects in optical metrology II. Proceedings of SPIE* (Vol. 7390, pp. 739002-1–739002-17).

Felsen, L. B., & Marcuvitz, N. (1994). *Radiation and scattering of waves*. Piscataway, NJ: IEEE Press.

Flagello, D. G., Milster, T., & Rosenbluth, A. E. (1996). Theory of high-NA imaging in homogeneous thin films. *Journal of the Optical Society of America A – Optics Image Science and Vision, 13*(1), 53–64.

Furukawa, H., & Kawata, S. (1996). Analysis of image formation in a near-field scanning optical microscope: Effects of multiple scattering. *Optics Communications, 132*(1–2), 170–178.

Gale, D. M., Pether, M. I., & Dainty, J. C. (1996). Linnik microscope imaging of integrated circuit structures. *Applied Optics, 35*(1), 131–148.

Gamelin, J., Guerrieri, R., & Neureuther, A. R. (1989). Exploration of scattering from topography with massively parallel computers. *Journal of Vacuum Science and Technology B – Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena,* 7(6), 1984–1990.

Gaylord, T. K., & Moharam, M. G. (1982). Planar dielectric grating diffraction theories. *Applied Physics B – Photophysics and Laser Chemistry, B28*(1), 1–14.

Goodman, J. W. (1996). *Introduction to Fourier optics* (2nd ed.). New York: McGraw-Hill.

Goodman, J. W. (2000). *Statistical optics*. New York, NY: Wiley.

Goodman, D. S., & Rosenbluth, A. E. (1988). Condenser aberrations in Kohler illumination. In *Optical/laser microlithography* (922nd ed.). *Proceedings of SPIE* (Vol. 922, pp. 108–134).

Gu, M., & Sheppard, C. J. R. (1995). Three-dimensional image formation in con-focal microscopy under ultra-short-laser-pulse illumination. *Journal of Modern Optics, 42*(4), 747–762.

Guerrieri, R., Tadros, K. H., Gamelin, J., & Neureuther, A. R. (1991). Massively parallel algorithms for scattering in optical lithography. *IEEE Transactions on Computer – Aided Design of Integrated Circuits and Systems, 10*(9), 1091–1100.

Hansen, R. (1985). Focal region characteristics of focused array antennas. *IEEE Transactions on Antennas and Propagation, 33*(12), 1328–1337.

Harrington, R. F. (2001). *Time-harmonic electromagnetic fields*. New York: Wiley-IEEE Press.

Haykin, S. (2001). *Communication systems* (4th ed.). Hoboken, NJ: Wiley.

Hollmann, J. L., Dunn, A. K., & DiMarzio, C. A. (2004). Computational microscopy in embryo imaging. *Optics Letters, 29*(19), 2267–2269.

Hopkins, H. H. (1951). The concept of partial coherence in optics. *Proceedings of the Royal Society of London Series A – Mathematical and Physical Sciences, 208*(1093), 263–277.

Hopkins, H. H. (1953). On the diffraction theory of optical images. *Proceedings of the Royal Society of London Series A – Mathematical and Physical Sciences, 217*(1130), 408–432.

Huttunen, J., & Turunen, J. (1995). Phase images of grooves in a perfectly conducting surface. *Optics Communications, 119*(5–6), 485–490.

Ibragimov, E. (1995). Focusing of ultrashort laser pulses by the combination of diffractive and refractive elements. *Applied Optics, 34*(31), 7280–7285.

Ichikawa, H., Masuda, K., & Ueda, T. (2009). Analysis of micro-Fresnel lenses with local grating theory and its comparison with fully electromagnetic methods. *Journal of the Optical Society of America A – Optics Image Science and Vision, 26*(9), 1938–1944.

Ishimaru, A. (1999). *Wave propagation and scattering in random media*. New York: Wiley-IEEE Press.

Janssen, A. J. E. M. (2002). Extended Nijboer–Zernike approach for the computation of optical point-spread functions. *Journal of the Optical Society of America A – Optics Image Science and Vision, 19*(5), 849–857.

Janssen, O. T. A., van Haver, S., Janssen, A. J. E. M., Braat, J. J. M., Urbach, P., & Pereira, S. F. (2008). Extended Nijboer–Zernike (ENZ) based mask imaging: Efficient coupling of electromagnetic field solvers and the ENZ imaging algorithm. In H. J. Levinson, & M. V. Dusa (Eds.), *Optical microlithography XXI, Pts 1-3. Proceedings of SPIE* (Vol. 6924, pp. 692410-1–692410-9).

Jin, J. (2002). *The finite element method in electromagnetics*. New York, NY: John Wiley & Sons.

Judkins, J. B., Haggans, C. W., & Ziolkowski, R. W. (1996). Two-dimensional finite-difference time-domain simulation for rewritable optical disk surface structure design. *Applied Optics, 35*(14), 2477–2487.

Judkins, J. B., & Ziolkowski, R. W. (1995). Finite-difference time-domain modeling of non-perfectly conducting metallic thin-film gratings. *Journal of the Optical Society of America A – Optics Image Science and Vision, 12*(9), 1974–1983.

Kaspar, F. G. (1973). Diffraction by thick, periodically stratified gratings with complex dielectric constant. *Journal of the Optical Society of America, 63*(1), 37–45.

Kempe, M., Stamm, U., Wilhelmi, B., & Rudolph, W. (1992). Spatial and temporal transformation of femtosecond laser pulses by lenses and lens systems. *Journal of the Optical Society of America B – Optical Physics, 9*(7), 1158–1165.

Kingslake, R. (1978). *Lens design fundamentals*. Boston: Academic Press.

Kintner, E. C. (1978). Method for the calculation of partially coherent imagery. *Applied Optics, 17*(17), 2747–2753.

Kirchauer, H., & Selberherr, S. (1997). Rigorous three-dimensional photoresist exposure and development simulation over nonplanar topography. *IEEE Transactions on Computer – Aided Design of Integrated Circuits and Systems, 16*(12), 1431–1438.

Kleemann, B. H., Mitreiter, A., & Wyrowski, F. (1996). Integral equation method with parametrization of grating profile – Theory and experiments. *Journal of Modern Optics, 43*(7), 1323–1349.

Kline, M., & Kay, I. W. (1979). *Electromagnetic theory and geometrical optics*. Huntington, NY: Krieger Pub. Co.

Knop, K. (1978). Rigorous diffraction theory for transmission phase gratings with deep rectangular grooves. *Journal of the Optical Society of America A – Optics Image Science and Vision, 68*(9), 1206–1210.

Kogelnik, H. (1969). Coupled wave theory for thick hologram gratings. *Bell System Technical Journal, 48*(9), 2909–2947.

Kotlyar, V. V., & Nesterenko, D. V. (2001). Analysis of light diffraction by binary micro-optics using a combination of boundary element method and finite element method. In D. A. Zimnyakov (Ed.), *Saratov fall meeting 2000: Coherent optics of ordered and random media. Proceedings of SPIE* (Vol. 4242, pp. 125–132).

Krug, J. T., Sanchez, E. J., & Xie, X. S. (2002). Design of near-field optical probes with optimal field enhancement by finite difference time domain electromagnetic simulation. *Journal of Chemical Physics, 116*(24), 10895–10901.

Kundu, N. N., Mathur, B. P., & Gupta, S. N. (1986). Computation of optical image profile for lithography and linewidth measurement. *Microelectronics Journal, 17*(6), 14–23.

Lajunen, H., Vahimaa, P., & Tervo, J. (2005). Theory of spatially and spectrally partially coherent pulses. *Journal of the Optical Society of America A – Optics Image Science and Vision, 22*(8), 1536–1545.

Lam, M. C. (2005). *Fast simulation methods for non-planar phase and multilayer defects in DUV and EUV photomasks for lithography*. Ph.D. thesis, University of California, Berkeley.

Lam, M. C., & Neureuther, A. R. (2004). Fast simulation methods for defective EUV mask blank inspection. In W. Staud, & J. T. Weed (Eds.), *24th Annual Bacus symposium on photomask technology, Pt 1 and 2. Proceedings of SPIE* (Vol. 5567, pp. 741–750).

Lee, W., & Degertekin, E. L. (2004). Rigorous coupled-wave analysis of multilayered grating structures. *Journal of Lightwave Technology, 22*(10), 2359–2363.

Lee, J. F., Lee, R., & Cangellaris, A. (1997). Time-domain finite-element methods. *IEEE Transactions on Antennas and Propagation, 45*(3), 430–442.

Lee, R. T., & Smith, G. S. (2006). An alternative approach for implementing periodic boundary conditions in the FDTD method using multiple unit cells. *IEEE Transactions on Antennas and Propagation, 54*(2), 698–705.

Li, L. F. (1997). New formulation of the Fourier modal method for crossed surface-relief gratings. *Journal of the Optical Society of America A – Optics Image Science and Vision, 14*(10), 2758–2767.

Li, L. F. (1999). Oblique-coordinate-system-based Chandezon method for modeling one-dimensionally periodic, multilayer, inhomogeneous, anisotropic gratings. *Journal of the Optical Society of America A – Optics Image Science and Vision, 16*(10), 2521–2531.

Li, L. F., Chandezon, J., Granet, G., & Plumey, J. P. (1999). Rigorous and efficient grating-analysis method made easy for optical engineers. *Applied Optics, 38*(2), 304–313.

Li, Y., & Wolf, E. (1982). Focal shift in focused truncated Gaussian beams. *Optics Communications, 42*(3), 151–156.

Lichtenberg, B., & Gallagher, N. C. (1994). Numerical modeling of diffractive devices using the finite element method. *Optical Engineering, 33*(11), 3518–3526.

Liu, Q. H. 1997. The pseudospectral time-domain (PSTD) method: A new algorithm for solutions of Maxwell's equations. In *Antennas and propagation society international symposium, 1997, IEEE, 1997 Digest* (Vol. 1, pp. 122–125).

Liu, Y., Li, X., Kim, Y. L., & Backman, V. (2005). Elastic backscattering spectroscopic microscopy. *Optics Letters, 30*(18), 2445–2447.

Lucas, K., Yuan, C. M., & Strojwas, A. (1992). Rigorous and practical vector model for phase-shifting masks in optical lithography. In J. D. Cuthbert (Ed.), *Optical/laser microlithography V, Pts 1 and 2. Proceedings of SPIE* (Vol. 1674, pp. 252–263).

Lucas, K. D., Tanabe, H., & Strojwas, A. J. (1996). Efficient and rigorous three-dimensional model for optical lithography simulation. *Journal of the Optical Society of America A – Optics Image Science and Vision, 13*(11), 2187–2199.

Lue, N., Popescu, G., Ikeda, T., Dasari, R. R., Badizadegan, K., & Feld, M. S. (2006). Live cell refractometry using microfluidic devices. *Optics Letters, 31*(18), 2759–2761.

Luebbers, R., Kunz, K., Schneider, M., & Hunsberger, F. (1991). A finite-difference time-domain near zone to far zone transformation. *IEEE Transactions on Antennas and Propagation, 39*(4), 429–433.

Mack, C. A. (2007). *Fundamental principles of optical lithography: The science of microfabrication*. Hoboken, NJ: Wiley.

Magnusson, R., & Gaylord, T. K. (1978). Equivalence of multiwave coupled-wave theory and modal theory for periodic-media diffraction. *Journal of the Optical Society of America A – Optics Image Science and Vision, 68*(12), 1777–1779.

Mandel, L. (1963). Intensity fluctuations of partially polarized light. *Proceedings of the Physical Society of London, 81*(524), 1104–1114.

Martin, T., & Pettersson, L. (2001). FDTD time domain near-to-far-zone transformation above a lossy dielectric half-space. *Applied Computational Electromagnetics Society Journal, 16*(1), 45–52.

Marx, E. (2007). Images of strips on and trenches in substrates. *Applied Optics, 46*(23), 5571–5587.

Matsuzawa, T., Moniwa, A., Hasegawa, N., & Sunami, H. (1987). Two-dimensional simulation of photolithography on reflective stepped substrate. *IEEE Transactions on Computer – Aided Design of Integrated Circuits and Systems, 6*(3), 446–451.

Maystre, D. (1978). A new general integral theory for dielectric coated gratings. *Journal of the Optical Society of America, 68*(4), 490–495.

Maystre, D., & Neviere, M. (1978). Electromagnetic theory of crossed gratings. *Journal of Optics, 9*(5), 301–306.

Mias, C., Webb, J. P., & Ferrari, R. L. (1999). Finite element modelling of electromagnetic waves in doubly and triply periodic structures. *IEE Proceedings – Optoelectronics, 146*(2), 111–118.

Michalski, K. A., & Mosig, J. R. (1997). Multilayered media Green's functions in integral equation formulations. *IEEE Transactions on Antennas and Propagation, 45*(3), 508–519.

Mirotznik, M. S., Prather, D. W., & Mait, J. N. (1996). A hybrid finite element boundary element method for the analysis of diffractive elements. *Journal of Modern Optics, 43*(7), 1309–1321.

Moharam, M. G., & Gaylord, T. K. (1981). Rigorous coupled-wave analysis of planar-grating diffraction. *Journal of the Optical Society of America, 71*(7), 811–818.

Moharam, M. G., & Gaylord, T. K. (1982). Diffraction analysis of dielectric surface-relief gratings. *Journal of the Optical Society of America, 72*(10), 1385–1392.

Moharam, M. G., & Gaylord, T. K. (1983). Three-dimensional vector coupled-wave analysis of planar-grating diffraction. *Journal of the Optical Society of America, 73*(9), 1105–1112.

Moharam, M. G., & Gaylord, T. K. (1986). Rigorous coupled-wave analysis of metallic surface-relief gratings. *Journal of the Optical Society of America A – Optics Image Science and Vision, 3*(11), 1780–1787.

Moharam, M. G., Grann, E. B., Pommet, D. A., & Gaylord, T. K. (1995). Formulation for stable and efficient implementation of the rigorous coupled-wave analysis of binary gratings. *Journal of the Optical Society of America A – Optics Image Science and Vision, 12*(5), 1068–1076.

Munro, P. R. T., & Török, P. (2005). Vectorial, high numerical aperture study of Nomarski's differential interference contrast microscope. *Optics Express, 13*(18), 6833–6847.

Neureuther, A. R. (1992). Simulation of optical lithography and inspection. *Microelectronic Engineering, 17*(1–4), 377–384.

Neureuther, A., 2008. If it moves, simulate it! In H. J. Levinson, & M.V. Dusa (Eds.), *Optical microlithography XXI, Pts 1–3. Proceedings of SPIE* (Vol. 6924, pp. 692402-1–692402-15).

Neviere, M., Vincent, P., & Petit, R. (1974). Theory of conducting gratings and their applications to optics. *Nouvelle Revue D Optique, 5*(2), 65–77.

Nikolaev, N. I., & Erdmann, A. (2003). Rigorous simulation of alignment for microlithography. *Journal of Microlithography, Microfabrication, and Microsystems, 2*(3), 220–226.

Nolte, A., Pawley, J. B., & Höring, L. (2006). Non-laser light sources for three-dimensional microscopy. In J. B. Pawley (Ed.), *Handbook of biological confocal microscopy* (3rd ed.). New York: Springer.

Noponen, E., & Turunen, J. (1994). Eigenmode method for electromagnetic synthesis of diffractive elements with three-dimensional profiles. *Journal of the Optical Society of America A – Optics Image Science and Vision, 11*(9), 2494–2502.

Nyyssonen, D. (1982). Theory of optical edge detection and imaging of thick layers. *Journal of the Optical Society of America, 72*(10), 1425–1436.

Nyyssonen, D., & Kirk, C. P. (1988). Optical microscope imaging of lines patterned in thick layers with variable edge geometry: Theory. *Journal of the Optical Society of America A – Optics Image Science and Vision, 5*(8), 1270–1280.

Oppenheim, A.V., Schafer, R.W., & Buck, J. R. (1999). *Discrete-time signal processing* (2nd ed.). Upper Saddle River, N.J: Prentice Hall.

Oppenheim, A. V., Willsky, A. S., & Nawab, S. H. (1997). *Signals and systems*. Upper Saddle River, NJ: Prentice Hall.

Papoulis, A. (1991). *Probability, random variables, and stochastic processes*. New York: McGraw-Hill.

Peterson, A. F., Ray, S. L., & Mittra, R. (1998). *Computational methods for electromagnetics*. New York: IEEE Press.

Pistor, T.V. (2001). *Electromagnetic simulation and modeling with applications in lithography*. Ph.D. thesis, University of California, Berkeley.

Pistor, T. V., Adam, K., & Neureuther, A. (1998). Rigorous simulation of mask corner effects in extreme ultraviolet lithography. *Journal of Vacuum Science and Technology B – Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena, 16*(6), 3449–3455.

Pistor, T. V., & Neureuther, A. (1999a). Extreme ultraviolet mask defect simulation. *Journal of Vacuum Science and Technology B – Microelectronics and Nanometer Structures Processing, Measurement, and Phenomena, 17*(6), 3019–3023.

Pistor, T.V., & Neureuther, A. R. (1999b). Calculating aerial images from EUV masks. In Y. Vladimirsky (Ed.), *Emerging lithographic technologies III, Pts 1 and 2. Proceedings of SPIE* (Vol. 3676, pp. 679–696).

Popov, E., & Bonod, N. (2007). Differential theory of diffraction in cylindrical coordinates. *Physica Status Solidi B – Basic Solid State Physics, 244*(10), 3463–3478.

Popov, E., Neviere, M., Gralak, B., & Tayeb, G. (2001). The classical different method, the rigorous coupled wave theory, and the modal method: Comparative analysis of convergence properties in staircase approximation. In *Physics, theory, and applications of periodic structures in optics. Proceedings of SPIE* (Vol. 4438, pp. 12–18).

Popov, E., & Neviere, M. (2000). Grating theory: New equations in Fourier space leading to fast converging results for TM polarization. *Journal of the Optical Society of America A – Optics Image Science and Vision, 17*(10), 1773–1784.

Prather, D. W., Mirotznik, M. S., & Mait, J. N. (1997). Boundary integral methods applied to the analysis of diffractive optical elements. *Journal of the Optical Society of America A – Optics Image Science and Vision, 14*(1), 34–43.

Prather, D.W., Shi, S.Y., & Sonstroem, J. (2002). Electromagnetic analysis of finite-thickness diffractive elements. *Optical Engineering, 41*(8), 1792–1796.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W.T. (1992). *Numerical recipes in C: The art of scientific computing* (2nd ed.). Cambridge: Cambridge University Press.

Rafler, S., Schuster, T., Frenner, K., Osten, W., & Seifert, U. (2008). Improvements on the simulation of microscopic images for the defect detection of nanostructures. In J. A. Allgair, & C. J. Raymond (Eds.), *Metrology, inspection, and process control for microlithography XXII. Proceedings of SPIE* (Vol. 6922, p. 692215).

Richards, B., & Wolf, E. (1959). Electromagnetic diffraction in optical systems. II. Structure of the image field in an aplanatic system. *Proceedings of the Royal Society of London Series A – Mathematical and Physical Sciences, 253*(1274), 358–379.

Roden, J. A., & Gedney, S. D. (2000). Convolution PML (CPML): An efficient FDTD implementation of the CFD-PML for arbitrary media. *Microwave and Optical Technology Letters, 27*(5), 334–339.

Salski, B., & Gwarek, W. (2008). Hybrid FDTD-Fresnel modeling of microscope imaging. In *International conference on recent advances in microwave theory and applications, 2008. MICROWAVE 2008, Jaipur, India* (pp. 398–399).

Salski, B., & Gwarek, W. (2009a). Hybrid FDTD-Fresnel modeling of the scanning confocal microscopy. In *Scanning microscopy. Proceedings of SPIE* (Vol. 7378, pp. 737826-1–737826-6).

Salski, B., Celuch, M., & Gwarek, W. (2010). FDTD for nanoscale and optical problems. *IEEE Microwave Magazine, 11*(2), 50–59.

Salski, B., & Gwarek, W. (2009b). Hybrid finite-difference time-domain Fresnel modeling of microscopy imaging. *Applied Optics, 48*(11), 2133–2138.

Sheppard, C. J. R. (2007). The optics of microscopy. *Journal of Optics A – Pure and Applied Optics, 9*(6), S1–S6.

Sheridan, J. T., & Sheppard, C. J. R. (1993). Coherent imaging of periodic thick fine isolated structures. *Journal of the Optical Society of America A – Optics Image Science and Vision, 10*(4), 614–632.

Sierra, H., DiMarzio, C. A., & Brooks, D. H. (2008). Modeling images of phase information for three-dimensional objects. In J. A. Conchello, C. J. Cogswell, T. Wilson, & T. G. Brown (Eds.), *Three-dimensional and multidimensional microscopy: Image acquisition and processing XV. Proceedings of SPIE* (Vol. 6861, pp. 68610A-1–68610A-9).

Simon, B., & DiMarzio, C. A. (2007). Simulation of a theta line-scanning confocal microscope. *Journal of Biomedical Optics, 12*(6), 064020-1–064020-9.

Simpson, S. H., & Hanna, S. (2001). Analysis of the effects arising from the near-field optical microscopy of homogeneous dielectric slabs. *Optics Communications, 196*(1–6), 17–31.

Smith, G. S. (1997). *An introduction to classical electromagnetic radiation*. New York, NY: Cambridge University Press.

Smith, M. D., & Mack, C. A. (2003). Methods for benchmarking photolithography simulators. In A. Yen (Ed.), *Optical microlithography XVI, Pts 1–3. Proceedings of SPIE* (Vol. 5040, pp. 57–68).

Starosta, M. S., & Dunn, A. K. (2009). Three-dimensional computation of focused beam propagation through multiple biological cells. *Optics Express, 17*(15), 12455–12469.

Stratton, J. A. (2007). *Electromagnetic theory*. Hoboken, NJ: Wiley.

Stratton, J. A., & Chu, L. J. (1939). Diffraction theory of electromagnetic waves. *Physical Review A – Atomic, Molecular and Optical Physics, 56*, 99–107.

Subramanian, H., Pradhan, P., Liu, Y., Capoglu, I. R., Li, X., Rogers, J. D., et al. (2008). Optical methodology for detecting histologically unapparent nanoscale consequences of genetic alterations in biological cells. *Proceedings of the National Academy of Sciences of the United States of America, 105*(51), 20118–20123.

Subramanian, H., Pradhan, P., Liu, Y., Capoglu, I. R., Rogers, J. D., Roy, H. K., et al. (2009). Partial-wave microscopic spectroscopy detects subwavelength refractive index fluctuations: An application to cancer diagnosis. *Optics Letters, 34*(4), 518–520.

Sukumar, N., & Pask, J. E. (2009). Classical and enriched finite element formulations for Bloch-periodic boundary conditions. *International Journal for Numerical Methods in Engineering, 77*(8), 1121–1138.

Symons, W. C., Whites, K. W., & Lodder, R. A. (2003). Theoretical and experimental characterization of a near-field scanning microwave (NSMM). *IEEE Transactions on Microwave Theory and Techniques, 51*(1), 91–99.

Tadros, K., Neureuther, A. R., & Guerrieri, R. (1991). Understanding metrology of polysilicon gates through reflectance measurements and simulation. In W. H. Arnold (Ed.), *Integrated circuit metrology, inspection, and process control V. Proceedings of SPIE* (Vol. 1464, pp. 177–186).

Taflove, A. (1980). Application of the finite-difference time-domain method to sinusoidal steady-state electromagnetic-penetration problems. *IEEE Transactions on Electromagnetic Compatibility, 22*(3), 191–202.

Taflove, A., & Hagness, S. C. (2005). *Computational electrodynamics: The finite-difference time-domain method* (3rd ed.). Boston: Artech House.

Taflove, A., & Umashankar, K. (1983). Radar cross section of general three-dimensional scatterers. *IEEE Transactions on Electromagnetic Compatibility, 25*(4), 433–440.

Tanabe, H. (1992). Modeling of optical images in resists by vector potentials. In *Optical/laser microlithography V. Proceedings of SPIE* (Vol. 1674, pp. 637–649).

Tanev, S., Pond, J., Paddon, P., & Tuchin, V. V. (2008). Optical phase contrast microscope imaging: A FDTD modeling approach. In V. V. Tuchin (Ed.), *Optical technologies in biophysics and medicine IX. Proceedings of SPIE* (Vol. 6791, pp. 7910E-1–7910E-11).

Tanev, S., Tuchin, V. V., & Pond, J. (2008). Simulation and modeling of optical phase contrast microscope cellular nanobioimaging. In *15th International school on quantum electronics; laser physics and applications. Proceedings of SPIE* (Vol. 7027, pp. 702716-1–702716-8).

Tanev, S., Sun, W. B., Pond, J., Tuchin, V. V., & Zharov, V. P. (2009). Flow cytometry with gold nanoparticles and their clusters as scattering contrast agents: FDTD simulation of light-cell interaction. *Journal of Biophotonics, 2*(8–9), 505–520.

Tanev, S., Tuchin, V. V., & Paddon, P. (2006). Cell membrane and gold nanoparticles effects on optical immersion experiments with noncancerous and cancerous cells: finite-difference time-domain modeling. *Journal of Biomedical Optics, 11*(6), 064037-1–064037-6.

Tang, X. G., Gao, F. H., Guo, Y. K., Du, J. L., Liu, S. J., & Gao, F. (2005). Analysis and simulation of diffractive imaging field in thick film photoresist by using angular spectrum theory. *Optics Communications, 244*(1–6), 123–130.

Tayeb, G., & Petit, R. (1984). On the numerical study of deep conducting lamellar diffraction gratings. *Optica Acta, 31*(12), 1361–1365.

Teixeira, F. L. (2008). Time-domain finite-difference and finite-element methods for Maxwell equations in complex media. *IEEE Transactions on Antennas and Propagation, 56*(8), 2150–2166.

Teixeira, F., & Chew, W. (1999). Lattice electromagnetic theory from a topological viewpoint. *Journal of Mathematical Physics, 40*(1), 169–187.

Tervo, J., Setälä, T., & Friberg, A. T. (2004). Theory of partially coherent electromagnetic fields in the space-frequency domain. *Journal of the Optical Society of America A – Optics Image Science and Vision, 21*(11), 2205–2215.

Thompson, L. F. Willson, C. G. & Bowden, M. J., (Eds.), (1994). *Introduction to microlithography*. Washington, DC: American Chemical Society.

Tirapu Azpiroz, J. (2004). *Analysis and modeling of photomask near-fields in sub-wavelength deep ultraviolet lithography*. Ph.D. thesis, University of California, Los Angeles.

Török, P., Munro, P. R. T., & Kriezis, E. E. (2006). Rigorous near- to far-field transformation for vectorial diffraction calculations and its numerical implementation. *Journal of the Optical Society of America A – Optics Image Science and Vision, 23*(3), 713–722.

Török, P., Munro, P. R. T., & Kriezis, E. E. (2008). High numerical aperture vectorial imaging in coherent optical microscopes. *Optics Express, 16*(2), 507–523.

Török, P., & Varga, P. (1997). Electromagnetic diffraction of light focused through a stratified medium. *Applied Optics, 36*(11), 2305–2312.

Török, P., Varga, P., Laczik, Z., & Booker, G. R. (1995). Electromagnetic diffraction of light focused through a planar interface between materials of mismatched refractive indices: An integral representation. *Journal of the Optical Society of America A – Optics Image Science and Vision, 12*(2), 325–332.

Totzeck, M. (2001). Numerical simulation of high-NA quantitative polarization microscopy and corresponding near-fields. *Optik, 112*(9), 399–406.

Totzeck, M. (2006). Some similarities and dissimilarities of imaging simulation for optical microscopy and lithography. In W. Osten (Ed.), *The 5th international workshop on automatic processing of fringe patterns. Fringe 2005* (pp. 267–274). Berlin, Heidelberg, Germany: Springer.

Tseng, S. H. (2007). Virtual optical experiment: Characterizing the coherent effects of light scattering through macroscopic random media. *Japanese Journal of Applied Physics, Part 1, 46*(12), 7966–7969.

Umashankar, K., & Taflove, A. (1982). A novel method to analyze electromagnetic scattering of complex objects. *IEEE Transactions on Electromagnetic Compatibility, 24*(4), 397–405.

Urbach, H. P., & Bernard, D. A. (1989). Modeling latent-image formation in photolithography, using the Helmholtz equation. *Journal of the Optical Society of America A – Optics Image Science and Vision, 6*(9), 1343–1356.

Vallius, T. (2002). Comparing the Fourier modal method with the C method: Analysis of conducting multilevel gratings in TM polarization. *Journal of the Optical Society of America A – Optics Image Science and Vision, 19*(8), 1555–1562.

Vallius, T., & Turunen, J. (2006). Electromagnetic models for the analysis and design of complex diffractive microstructures. In Y. L. Sheng, S. L. Zhuang, & Y. M. Zhang (Eds.), *ICO20: Optical information processing, Pts 1 and 2. Proceedings of SPIE* (Vol. 6027, pp. 602704-1–602704-4).

van de Nes, A. S., Billy, L., Pereira, S. F., & Braat, J. J. M. (2004). Calculation of the vectorial field distribution in a stratified focal region of a high numerical aperture imaging system. *Optics Express, 12*(7), 1281–1293.

van Haver, S., Braat, J. J. M., Janssen, A. J. E. M., Janssen, O. T. A., & Pereira, S. F. (2009). Vectorial aerial-image computations of three-dimensional objects based on the extended Nijboer–Zernike theory. *Journal of the Optical Society of America A – Optics Image Science and Vision, 26*(5), 1221–1234.

van Haver, S., Janssen, O. T. A., Braat, J. J. M., Janssen, A. J. E. M., Urbach, H. P., & Pereira, S. F. (2008). General imaging of advanced 3D mask objects based on the fully-vectorial extended Nijboer–Zernike (ENZ) theory. In H. J. Levinson, & M. V. Dusa (Eds.), *Optical microlithography XXI, Pts 1–3. Proceedings of SPIE* (Vol. 6924, pp. 69240U-1–69240U-8).

Vasilyeva, E., & Taflove, A. (1998a). 3-D FDTD image analysis in transmission illumination mode of scanning near-field optical microscopy*Antennas and propagation society international symposium, 1998* (pp. 1800–1803) (pp.1800–1803). GA, USA: IEEE.

Vasilyeva, E., & Taflove, A. (1998b). Three-dimensional modeling of amplitude-object imaging in scanning near-field optical microscopy. *Optics Letters, 23*(15), 1155–1157.

Veetil, S. P., Schimmel, H., Wyrowski, F., & Vijayan, C. (2006). Wave optical modelling of focusing of an ultra short pulse. *Journal of Modern Optics, 53*(15), 2187–2194.

Wang, L. G., Lin, Q., Chen, H., & Zhu, S.Y. (2003). Propagation of partially coherent pulsed beams in the spatiotemporal domain. *Physical Review E: Statistical, Nonlinear, and Soft Matter Physics, 67*(5, Part 2), 0566131–0566137.

Wei, X. H., Wachters, A. J. H., & Urbach, H. P. (2007). Finite-element model for three-dimensional optical scattering problems. *Journal of the Optical Society of America A – Optics Image Science and Vision, 24*(3), 866–881.

Winton, S. C., Kosmas, P., & Rappaport, C. M. (2005). FDTD simulation of TE and TM plane waves at nonzero incidence in arbitrary layered media. *IEEE Transactions on Antennas and Propagation, 53*(5), 1721–1728.

Wojcik, G. L., Vaughan, D. K., & Galbraith, L. K. (1987). Calculation of light scatter from structures on silicon surfaces. In *Lasers in microlithography. Proceedings of SPIE* (Vol. 774, pp. 21–31).

Wojcik, G. L., Mould, J., Monteverde, R. J., Prochazka, J. J., & Frank, J. R. J. (1991a). Numerical simulation of thick line width measurements by reflected light. In *Integrated circuit metrology, inspection and process control V. Proceedings of SPIE* (Vol. 1464, pp. 187–203).

Wojcik, G. L., Vaughan, D. K., Mould, J. J., Leon, F. A., Qian, Q.-D., & Lutz, M. A. (1991b). Laser alignment modeling using rigorous numerical simulations. In V. Pol (Ed.), *Optical/laser microlithography IV. Proceedings of SPIE* (Vol. 1463, pp. 292–303).

Wojcik, G. L., Mould, J., Marx, E., & Davidson, M. P. (1992). Numerical reference models for optical metrology simulation. In M. T. Postek (Ed.), *Integrated circuit metrology, inspection, and process control VI. Proceedings of SPIE* (Vol. 1673, pp. 70–82).

Wojcik, G. L., John Mould, J., Ferguson, R. A., Martino, R. M., & Low, K. K. (1994). Some image modeling issues for I-line, 5X phase-shifting masks. In *Optical/laser microlithography VII. Proceedings of SPIE* (Vol. 2197, pp. 455–465).

Wolf, E. (1959). Electromagnetic diffraction in optical systems. I. An integral representation of the image field. *Proceedings of the Royal Society of London Series A – Mathematical and Physical Sciences, 253*(1274), 349–357.

Wolf, E., & Li, Y. (1981). Conditions for the validity of the Debye integral-representation of focused fields. *Optics Communications, 39*(4), 205–210.

Wong, A. K. (1994). *Rigorous three-dimensional time-domain finite-difference electromagnetic simulation*. Ph.D. thesis, University of California, Berkeley.

Wong, A. K., Guerrieri, R., & Neureuther, A. R. (1995). Massively-parallel electromagnetic simulation for photolithographic applications. *IEEE Transactions on Computer – Aided Design of Integrated Circuits and Systems, 14*(10), 1231–1240.

Wong, A. K., & Neureuther, A. R. (1994). Mask topography effects in projection printing of phase-shifting masks. *IEEE Transactions on Electron Devices, 41*(6), 895–902.

Wong, A. K., & Neureuther, A. R. (1995). Rigorous three-dimensional time-domain finite-difference electromagnetic simulation for photolithographic applications. *IEEE Transactions on Semiconductor Manufacturing, 8*(4), 419–431.

Wu, R. B., & Itoh, T. (1997). Hybrid finite-difference time-domain modeling of curved surfaces using tetrahedral edge elements. *IEEE Transactions on Antennas and Propagation, 45*(8), 1302–1309.

Xu, Y.-L. (1995). Electromagnetic scattering by an aggregate of spheres. *Applied Optics, 34*(21), 4573–4588.

Yang, S.-H., Milster, T., Park, J. R., & Zhang, J. (2010). High-numerical-aperture image simulation using Babinet's principle. *Journal of the Optical Society of America A – Optics Image Science and Vision, 27*(5), 1012–1023.

Yang, S.-H., Milster, T., Zhang, J., & Chen, T. (2010). Characteristics of evanescent polarization imaging. *Journal of Modern Optics, 57*(9), 783–797.

Yang, S., Taflove, A., & Backman, V. (2011). Experimental confirmation at visible light wavelengths of the backscattering enhancement phenomenon of the photonic nanojet. *Optics Express, 19*(8), 7084–7093.

Yee, K. S. (1966). Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media. *IEEE Transactions on Antennas and Propagation, 14*(3), 302–307.

Yee, K. S., Ingham, D., & Shlager, K. (1991). Time-domain extrapolation to the far field based on FDTD calculations. *IEEE Transactions on Antennas and Propagation, 39*(3), 410–413.

Yeh, C., Colak, S., & Barber, P. (1982). Scattering of sharply focused beams by arbitrarily shaped dielectric particles: An exact solution. *Applied Optics, 21*(24), 4426–4433.

Yeung, M. S. (1988). Modeling high numerical aperture optical lithography. In *Optical/laser microlithography. Proceedings of SPIE* (Vol. 922, pp. 149–167).

Yeung, M. S. (1990). Photolithography simulation on nonplanar substrates. In V. Pol (Ed.), *Optical/laser microlithography III. Proceedings of SPIE* (Vol. 1264, pp. 309–321).

Yeung, M. S., & Barouch, E. (1997). Three-dimensional nonplanar lithography simulation using a periodic fast multipole method. In Fuller, G. E. (Ed.), *Optical microlithography X. Proceedings of SPIE* (Vol. 3051, pp. 509–521).

Yeung, M. S., & Barouch, E. (1999). Application of the hybrid finite-difference time-domain method to modeling curved surfaces in three-dimensional lithography simulation. In *Optical microlithography XII, Pts 1 and 2. Proceedings of SPIE* (Vol. 3679, pp. 1093–1103).

Yeung, M. S., & Neureuther, A. R. (1995). Three-dimensional reflective-notching simulation using multipole accelerated physical-optics approximation. In *Optical/laser microlithography VII. Proceedings of SPIE* (Vol. 2440, pp. 395–409).

Yeung, M. S., Lee, D., Lee, R., & Neureuther, A. R. (1993). Extension of the Hopkins theory of partially coherent imaging to include thin-film interference effects. In *Optical/laser microlithography VI. Proceedings of SPIE* (Vol. 1927, pp. 452–463).

Yin, X. M., Wong, A., Wheeler, D., Williams, G., Lehner, E., Zach, F., et al. (2000). Sub-wavelength alignment mark signal analysis of advanced memory products. In N. T. Sullivan (Ed.), *Metrology, inspection, and process control for microlithography XIV. Proceedings of SPIE* (Vol. 3998, pp. 449–459).

Yuan, C. M. (1992). Efficient light scattering modeling for alignment, metrology, and resist exposure in photolithography. *IEEE Transactions on Electron Devices, 39*(7), 1588–1598.

Yuan, C. M., & Strojwas, A. J. (1991a). Modeling optical equipment for wafer alignment and line-width measurement. *IEEE Transactions on Semiconductor Manufacturing, 4*(2), 99–110.

Yuan, C. M., & Strojwas, A. J. (1991b). Modeling optical microscope images of integrated-circuit structures. *Journal of the Optical Society of America A – Optics Image Science and Vision, 8*(5), 778–790.

Zhang, J., Kim, Y., Yang, S.-H., & Milster, T. D. (2010). Illumination artifacts in hyper-NA vector imaging. *Journal of the Optical Society of America A – Optics Image Science and Vision, 27*(10), 2272–2284.

Zhang, L., & Seideman, T. (2010). Rigorous formulation of oblique incidence scattering from dispersive media. *Physical Review B: Condensed Matter and Materials Physics, 82*(15), 155117-1–155117-15.