Early View publication on wileyonlinelibrary.com
(issue and page numbers not yet assigned;
citable using Digital Object Identifier – DOI)

LASER & PHOTONICS REVIEWS

Laser Photonics Rev., 1–39 (2011) / **DOI** 10.1002/lpor.201100009

Abstract Diffraction is a natural phenomenon, which occurs when waves propagate or encounter an obstacle. Diffraction is also a fundamental aspect of modern optics: all imaging systems are diffraction systems. However, like a coin has two sides, diffraction also leads to some unfavorable effects, such as an increase in the size of a beam during propagation, and a limited minimal beam size after focusing. To overcome these disadvantages set by diffraction, many techniques have been developed by various groups, including apodization techniques to reduce the divergence of a laser beam and increase the resolution, and time reversal, STED microscopy, super lenses and optical antennas to obtain resolution down to nano-scale. This review concentrates on the diffraction of electromagnetic waves, and the ways to overcome beam divergence and the diffraction limit.

Fighting against diffraction: apodization and near field diffraction structures

Haifeng Wang[1,*], Colin J. R. Sheppard[2], Koustuban Ravi[3], Seng Tiong Ho[3], and Guillaume Vienne[1]

1. Diffraction, the positive and negative aspects of it

Diffraction is a natural phenomenon: it occurs when waves encounter an obstacle. This phenomenon has been widely exploited in imaging systems, where a circular aperture is most commonly used. Here we would like to concentrate on the diffraction of electromagnetic waves. Diffraction of light by a small aperture can generate a focusing effect, and the earliest type of camera had an imaging system consisting of only a pinhole [1–5]. A pinhole is even believed to have been used by the ancient Egyptians as a magnification device in making tiny works of art [6]. The optimum design of a pinhole lens has been defined as [5]

$$d^2 = 3.8\lambda l, \tag{1}$$

where d is the diameter of the pinhole, λ is the wavelength and l is the focus position measured from the centre of the pinhole. Now, suppose the diameter d is 1.0 mm and the wavelength of light is 0.0005 mm, then the focus position is $l = 526.315$ mm, which is a long distance. The focusing effect is very poor, because the focusing angle is $\theta = d/(2l) = 0.00095$, the small relative aperture resulting in poor illuminance when it is used in imaging. However, this ancient tool also has its application in modern optics, for example, pinhole arrays have been used in the focusing of soft x-rays [7–9] and even diffraction limited resolution can be achieved by using a pinhole array [10]. The field diffracted by a pinhole is usually calculated by a diffraction integral [11], and the diffraction of a pinhole array is the summation of the Fresnel diffraction of all the individual pinholes.

Suppose an electromagnetic wave is incident on to an aperture in a black screen. Its distribution on the aperture is $E(x',y',0)$, and the field outside the aperture is zero, the electric field diffraction pattern at a point (x,y,z), as shown in Fig. 1, is given by:

$$E(x,y,z) = \frac{z}{i\lambda}\iint E(x',y',0)\frac{e^{ikr}}{r^2}dx'dy', \tag{2}$$

where $r = \sqrt{(x-x')^2+(y-y')^2+(z-z')^2}$, i is the imaginary unit, x' and y' are the coordinates in the plane of the aperture, and $k = 2\pi/\lambda$. The Fresnel approximation of Eq. (2) is given by Eq. (3):

$$E(x,y,z) = \frac{e^{ikz}}{i\lambda z}\iint E(x',y',0)\,e^{\frac{ik}{2z}[(x-x')^2+(y-y')^2]}dx'dy'. \tag{3}$$

To specify the relative distance to the diffraction aperture, the Fresnel number $F = a^2/\lambda z$ is defined, where a is the radius of the diffraction aperture, the condition for applying Fresnel diffraction theory being that the Fresnel number $F \geqslant 1$. The Fresnel number at the focal point of a pinhole lens is $F = 0.95$ [5].

[1] Advanced Concept Group, Data Storage Institute (DSI), Agency for Science, Technology and Research, DSI Building, 5 Engineering Drive 1, 117608 Singapore [2] Division of Bioengineering, National University of Singapore, 7 Engineering Drive 1, 117574 Singapore [3] Nanophotonics and Electronicphotonic Integration Group, Data Storage Institute (DSI), Agency for Science, Technology and Research, DSI Building, 5 Engineering Drive 1, 117608 Singapore
[*] Corresponding author: e-mail: WANG_Haifeng@dsi.a-star.edu.sg

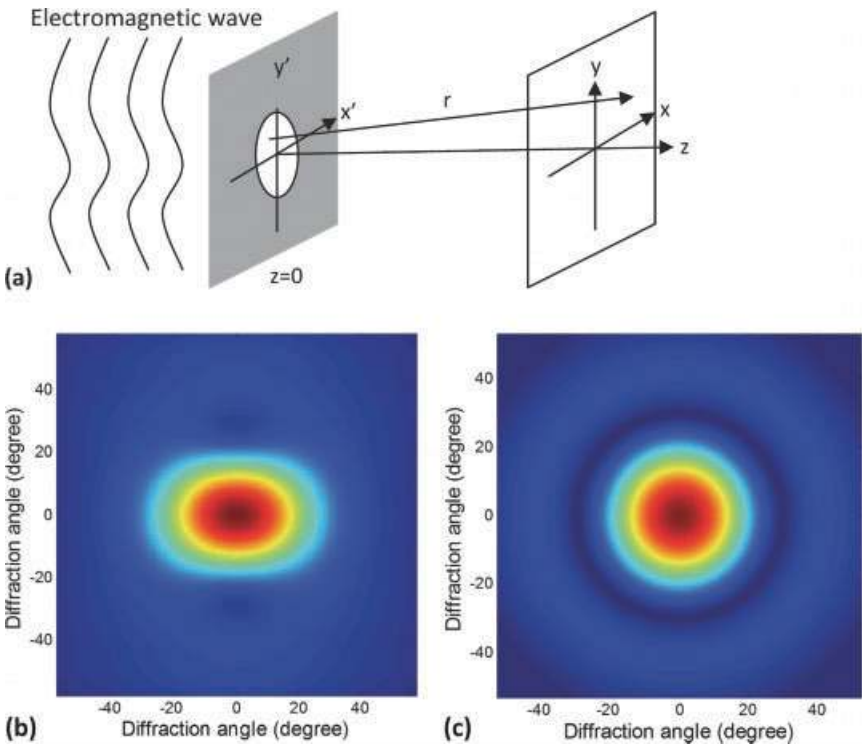

Figure 1 (online color at: www.lpr-journal.org) (a) Diffraction geometry, electromagnetic wave with wavelength of $\lambda = 400$ nm is incident onto a circular aperture in a black screen, the diffracted electromagnetic field is received by a plane located at position z. (b) Diffraction pattern of an aperture with $D = 2.44\lambda$ with an incident plane wave producing the Fresnel diffraction image at $F = 0.95$ position (627 nm from the exit of the aperture). (c) The Fraunhofer image in the far field.

However, the optical efficiency of pinhole lenses is very low: for a 1.0 millimetre diameter aperture, natural diffraction would only focus light at over 0.5 meter away, and the light condensing efficiency is very low. This makes it unsuitable to be widely used in many real applications. Most modern optical systems, like magnifying lenses, microscopes and telescopes, have a big aperture, high condensing efficiency and controllable focal length. These systems combine the diffraction effect of an aperture and the refracting effect of the curved surfaces of transparent materials. The curved surfaces can pull the focus from infinitely far away to a position close to the aperture. The diffraction pattern of an aperture at a far away distance, with Fresnel number $F \ll 1$, is represented by Fraunhofer diffraction, which is a further approximation of Fresnel diffraction [11]:

$$E(x,y,z) = \frac{e^{ikz}\, e^{\frac{ik}{2z}\cdot(x^2+y^2)}}{i\lambda z} \iint E(x',y',0)\, e^{\frac{-ik}{z}[xx'+yy']} dx'\, dy' \tag{4}$$

However, the field distribution far away is more conveniently expressed in terms of the angular spectrum. Integrating Eq. (4) for a circular aperture we get

$$E(\theta) = \frac{e^{ikz} e^{\frac{ikr_0^2}{2z}}}{i\lambda z} 2\pi r_0^2 \frac{J_1(kr_0 \sin\theta)}{kr_0 \sin\theta}, \tag{5}$$

where θ is the angle relative to the z axis, r_0 is the radius of the aperture, and J_1 is a first order Bessel function.

The intensity distribution of the diffraction pattern is

$$I(\theta) = |E(\theta)|^2 = \left(\frac{\pi r_0^2}{\lambda z}\right)^2 \left[\frac{2J_1(kr_0 \sin\theta)}{kr_0 \sin\theta}\right]^2. \tag{6}$$

The first zero of the intensity occurs at $kr_0 \sin\theta = 3.832$, $\sin\theta = 1.22\lambda/D$, where $D = 2r_0$. This means that when plane waves are incident on to a circular aperture with diameter D, the divergence angle caused by this aperture in the far field is

$$\theta = \arcsin(1.22\lambda/D). \tag{7}$$

It is clear that the smaller the aperture is, the larger the divergence angle will be, and when $D = 2.44\lambda$, the far field divergence angle is 30 degrees. The divergence angle caused by the aperture also represents the angular resolution of such an aperture in the far field, which is defined by the Rayleigh criterion. To have a better understanding of the diffraction by an aperture in the Fresnel diffraction region and that in the Fraunhofer diffraction region, the diffraction patterns of an aperture with diameter of $D = 2.44\lambda$ are plotted in Fig. 1b and Fig. 1c for each case. The diffraction pattern of Fresnel diffraction is calculated with Finite Difference Time Domain (FDTD) method: a horizontally polarized plane wave with wavelength of 400 nm is incident on to an aperture inside a 280 nm thick gold film, the radius of the aperture is 488 nm, and the magnitude of the electric field in the image plane ($F = 0.95$, 627 nm from the exit side of the gold film) is plotted, where the magnitude of the electric field is calculated as $\left(\left|E_x^2\right| + \left|E_y^2\right| + \left|E_z^2\right|\right)^{1/2}$. The magnitude of the electric field versus diffraction angle in the far field is calculated with Eq. (5), as shown in Fig. 1c. The Fraunhofer diffraction pattern has its first dark ring at 30 degrees, but in the Fresnel diffraction region the image does not have a distinct diffraction ring, as shown in Fig. 1b, the pattern being elliptical in shape with a longer horizontal axis. The zero field position can be found along the short axis direction, which also occurs at around 30 degrees. Now we see how polarization affects the diffraction pattern in the Fresnel diffraction region when the radius of the diffraction

aperture is comparable with the wavelength of light. However, when the radius of the diffraction aperture is much larger than light wavelength, the scalar diffraction Eq. (3) is applicable, which will give a circularly symmetric field distribution. Equation (5) represents the field far away from the aperture. By putting a condensing lens immediately behind this aperture, the image of this aperture from far away is pulled near to the aperture, the intensity distribution of this image being expressed as

$$I(r) = I_0 \left[\frac{2J_1(kr\sin\theta)}{kr\sin\theta} \right]^2, \tag{8}$$

where I_0 is a constant, $\sin\theta$ is the numerical aperture of the condensing lens, and r is the coordinate in the radial direction of its focal plane. The condensing lens makes the focusing angle much larger than that of natural diffraction by an aperture, the large focusing angle resulting in fast divergence of the beam after the focal point, which is unfavorable for some applications.

For a microscope objective lens with numerical aperture (NA) of 0.9 and focal length $f = 2$ mm, the pupil aperture radius is 1.8 mm, and if the incident laser wavelength $\lambda = 0.0005$ mm, then the Fresnel number $F = 3240$, which is much bigger than 1. When the Fresnel number is small, a focal shift occurs [12]: the focus moves closer to the lens because diffraction by the aperture generates a weak focusing effect. This effect can also be explained by the fact that the NA increases for points closer to the lens, and a balance with defocus is reached at the focus [13]. In practice the aperture should be placed in the front focal plane of the lens, so that the Fresnel number is infinite for any radius [14]. The condensed light spot in the focal region is usually called the point spread function, which has a size given by a full-width at half-intensity-maximum (FWHM) approximately equal to $\lambda/2\mathrm{NA}$, which is also the resolution limit of this condensing lens, in free space. The resolution limit actually describes the minimum resolvable distance between two images generated by two point sources at infinitely far distance. The fields from these two point sources are actually two plane waves with a certain crossing angle. Their images are two identical point spread functions of the same system. Suppose the two point sources are incoherent with respect to each other, and are of equal intensity. Then to distinguish the two images, the minimum distance between them corresponds to the case when the two intensity profiles have their half maximum cross each other, resulting in a flat-top image, the distance between the two images is just the full-width at half-intensity-maximum of its point spread function. For a dry objective with NA < 1, the highest resolution in the far field is limited to $\lambda/2$. In fact, when light is incident on to a subwavelenth or spatially discrete object, waves with spatial frequency higher than that of the incident light are generated, which propagate along the surface of the object. Such waves attenuate exponentially away from the object surface, and are the so-called evanescent waves. Evanescent waves satisfy the following relationship:

$$E(z) = E(0)\, e^{ik_z z}, \tag{9}$$

where $k_z^2 = k_0^2 - k_e^2$, $k_0 = 2\pi/\lambda$, $k_e = 2\pi/\lambda_e$ and λ_e is the transverse wavelength of the evanescent wave, with $\lambda_e < \lambda$. These waves carry information of the object features with spatial frequency higher than that corresponding to the light wavelength, and decay exponentially so that they are not collected by an objective lens in the far field. Therefore, when an objective lens with collection angle of $\theta = \arcsin(\mathrm{NA})$ is used, the far-field resolution is limited to $\lambda/2\mathrm{NA}$.

Now we see that diffraction is a fundamental property of all imaging systems. The pinhole camera uses the diffraction pattern of its aperture in the Fresnel diffraction region, where the Fresnel number is near to 1.0. Modern imaging systems like condensing glasses, telescopes and microscopes use the diffraction pattern of their apertures at an infinitely far distance (Fraunhofer diffraction region). However, diffraction sets a limit to far field imaging, i. e., high resolution evanescent waves from an object cannot be detected by such systems. The highest resolution that is achievable from far field imaging is decided by the size of its point spread function: $\lambda/2\mathrm{NA}$. Diffraction also causes beams to diverge: the smaller the diffraction aperture is, the faster the beam will diverge: for a beam condensed by a focusing lens, the smaller the condensed spot is, and the faster the beam will diverge. The following sections will centre on solutions to overcome the unfavourable aspects of diffraction: beam divergence and the diffraction limit, and are organized in the following way. In Sect. 2, we concentrate on the far field apodization technique, which can be used to reduce beam divergence (2.1) in different ways: a pure phase apodizer for elimination of beam divergence in the focal region of a focusing lens (2.1.1), and an amplitude type apodizer for generating nondiffracting beams (2.1.2). The application of the apodization technique to obtain super-resolution is addressed in Sect. 2.2, and can be realized in different ways: generating a super-resolution focusing spot (2.2.1), obtaining super-resolution imaging through illuminating the object with a fringe structured pattern (2.2.2) and obtaining super-resolution imaging through using two beams, which overlap and interact with fluorescence material to generate an effectively small imaging area (2.2.3). In Sect. 3, we concentrate on near field diffraction structures, which can be used to reduce the divergence of beam from a tiny aperture (3.1) and obtain super-resolution through different ways (3.2): use of a diffraction structure to turn evanescent waves into propagating waves (3.2.1), use of a superlens to realize near field super-resolution image reconstruction (3.2.2), and the use of an optical antenna to realize super-resolution light-focusing (3.2.3). And Sect. 4 is the conclusion and outlook.

2. Far field apodization technique

Apodization literally means "removing the foot". In optics, it was initially used to reduce the diffraction edge effect of an image from telescope. Here, apodization has a more generalized meaning, it refers to approaches that purposely change the input intensity of an optical system in order to modify light distribution in the focal region, on the object or

in the image plane. From this point of view, stimulated emission depletion (STED) microscopy, which uses two beams with predefined profiles which overlap and interact with fluorescent materials to reduce the effective imaging area, can be taken as a kind of apodization. The structured illumination imaging technique that projects a grating or fringe pattern on the object is also a kind of apodization. This chapter will first address the use of apodization techniques to reduce the divergence of light beam in the focal region and generate nondiffracting beams, then address the use of apodization techniques to obtain super-resolution: super-resolution focusing through reducing the size of the point spread function, superresolution imaging through structured illumination, and fluorescence switching.

2.1. Reduction of beam divergence

In free space, all light beams with limited size diverge during propagation. The divergence of the light can be increased when it goes through a concave lens or after focusing by a convex lens; it can also be increased through diffraction by a small aperture. Nevertheless, the diffraction effect can also be used to reduce the divergence of light beams, for example, with specially designed apodizers or apertures.

2.1.1. Apodizer design for elimination of beam divergence

A collimated laser beam diverges during propagation due to its limited size. The smaller the beam size is, the faster it diverges: for a Gaussian beam with beam waist radius of ω, its divergence angle can be approximated as $\theta \approx \lambda/(\pi\omega)$ [15]. When a beam is focused to a subwavelength scale, strong defocusing will cause it to diverge rapidly away from the focal plane, resulting in a very short depth of focus [16]. The fact that an annular aperture can greatly increase the depth of focus has been known for many years [17–22]. In 1952, Toraldo di Francia proposed to split the aperture of a focusing lens into a multiple annular ring structure, and by modulating the amplitude and phase of each ring on the lens pupil [23, 24], the divergence of the focused beam can be reduced, the reduction of beam divergence being accompanied by lower intensity at the focus [25–29].

In 1987, Durnin et al. proposed a 'diffraction-free' beam, which has the characteristics of intensity and spot size invariance along the optical axis [30]. The scalar solution of such beam is a zero-order Bessel function of the first kind, which rigorously exists only in infinite free space. As was actually stated several years earlier, 'A wave with zero-order Bessel-function radial distribution propagates without change' [31]. Any realization of such beams in an experimental setup requires a finite aperture, which limits the propagation distance of the beam. An approximation of the diffraction-free beam can be experimentally realized by placing an extremely narrow annular aperture in the lens pupil [32–34], and as a result, the intensity of this beam decreases and the beam size increases gradually away from the focal plane. A diffraction-free beam with limited propagation distance can also be realized by focusing light with an axicon lens [34]. The axial intensity of the beam generated by an axicon lens usually increases with its propagation distance, making it difficult to realize subwavelength focusing [35, 36]. To eliminate the divergence of a subwavelength focused light beam with high optical efficiency, one needs to use a pure phase apodizer to modulate only the phase on the aperture of a focusing lens [33, 35, 36]. A detailed review of diffraction free beams will be addressed in Sect. 2.1.2: the current section will concentrate on the design of a phase adpodizer for eliminating beam divergence.

Design of the phase apodizer

The structure of a binary apodizer is shown in Fig. 2, consisting of a multiple annular ring structure, the phase difference between adjacent rings being π. This kind of binary apodizer can consist of a transparent substrate with annular grooves or bumps: the depth of the grooves or bumps is given by $\lambda/[2(n-1)]$, where λ is the wavelength and n is the refractive index of the substrate. To realize a combined superresolution and nondiffracting effect, the apodizer has to be placed at the entrance pupil of an objective lens, as shown in Fig. 2.

Apodizer design based on scalar focusing method

When a collimated laser beam traverses a multi-belt annular apodizer and is then focused by the lens, as is shown in Fig. 2, the normalized amplitude distribution in the focal region in the paraxial approximation can be simplified

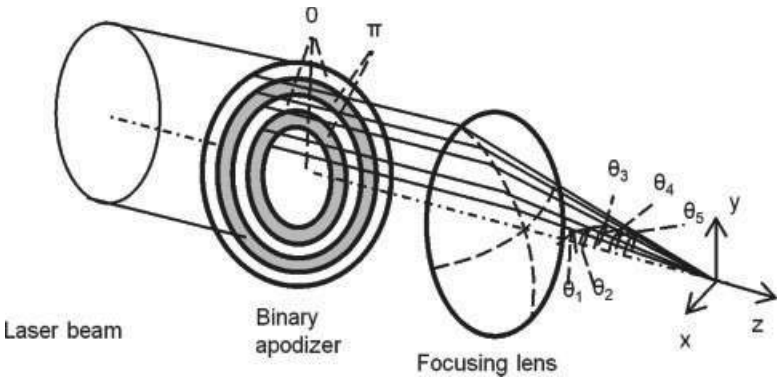

Figure 2 Schematic configuration of a system for generating super-resolution and nondiffracting light beam, laser beam undergoes a concentric binary apodizer and then focused by a lens, super-resolution and nondiffracting beam is generated in the focal region of the lens.

as [37, 38]

$$G(\rho,u) = 2\sum_{j=1}^{N} \exp(i\varphi_j) \int_{r_{j-1}}^{r_j} rJ_0(\rho r)g(r) \times \exp\left[-(1/2)iur^2\right] dr, \quad (10)$$

where $J_0(\rho,u)$ is the zero-order Bessel function of the first kind, r is is the radial coordinate of the objective lens pupil plane and $g(r)$ is the amplitude distribution in the radial direction of the lens pupil plane. ρ and u are normalized radial and axial coordinates, respectively:

$$\rho = (2\pi/\lambda)(\mathrm{NA})R, \quad (11)$$

$$u = (2\pi/\lambda)(\mathrm{NA})^2 Z, \quad (12)$$

where R and Z are the genuine radial and axial coordinates. NA is the numerical aperture of the objective lens. The phase of each belt on the pupil plane is $\varphi_j, (j = 1,2,\ldots N)$, and the radius for each belt is $r_j, (j = 1,2,\ldots N)$, where $r_{j-1} < r_j < r_N = 1$. For a 3-belt pure phase apodizer, $r_1 = b$, $r_2 = a$, $r_3 = 1$, and for simplicity we choose $g(r) = 1$, representing uniform illumination across the aperture, so that the axial amplitude distribution is

$$\begin{aligned} G(0,u) &= 2\sum_{j=1}^{3} \exp(i\varphi_j) \int_{r_{j-1}}^{r_j} r\exp[-1/2iur^2]dr \\ &= \frac{2}{iu}\{\exp(-1/2iu) - 1 \\ &\quad - 2\left[\exp\left(-1/2iua^2\right) - \exp\left(-1/2iub^2\right)\right]\}, \end{aligned} \quad (13)$$

and the axial intensity distribution is

$$\begin{aligned} I(0,u) &= |G(0,u)|^2 \\ &= \left\{10 + 16\sin\left[\frac{u}{4}\left(a^2-b^2\right)\right]\sin\frac{u}{4}\cos\left[\frac{u}{4}\left(1-a^2-b^2\right)\right]\right. \\ &\quad \left. - 8\cos\left[\frac{u}{2}\left(a^2-b^2\right)\right] - 2\cos\frac{u}{2}\right\}\frac{4}{u^2}. \end{aligned} \quad (14)$$

To find the value of a and b pairs towards obtaining constant axial intensity, we need to solve a second order differential equation from Eq. (14)

$$\frac{\partial^2 I(0,u)}{\partial u^2}\Big|_{u=0} = 0, \quad (15)$$

and then the relationship between a and b is obtained as

$$-0.5 - 2(a^2-b^2)^4 - (1-a^2)^4 + (1-b^2)^4 + a^8 - b^8 = 0. \quad (16)$$

Equation (16) is solved in a numerical way, with $0 \leqslant b \leqslant 1$, $0 \leqslant a \leqslant 1$ and $b < a$. When b increases from 0 to 1, the corresponding positive real value a is obtained and plotted in Fig. 3. It can be seen that when $b < 0.4$, for each value of b there are two positive real value of a that satisfy Eq. (16), and therefore two curves are plotted. When $b \geqslant 0.4$, for each value of b there is only one positive real value of a that satisfies Eq. (16). For the case when the numerical aperture of the lens is 0.85, the axial intensity distribution for some pairs of a and b taken from curve 1 and curve 2 are plotted in Fig. 4. The solid curve corresponds to the system without the apodizer, and the one with $a = 0.927$ and $b = 0.22$ is taken from curve 2 in Fig. 3, the rest being from curve 1 in Fig. 3.

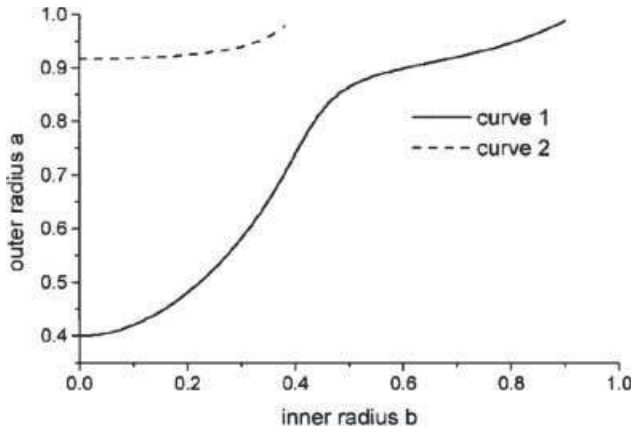

Figure 3 Relationship between outer radius a and inner radius b of an optimized 3-belt apodizer towards obtaining constant axial intensity when it is applied to the aperture of a focusing lens.

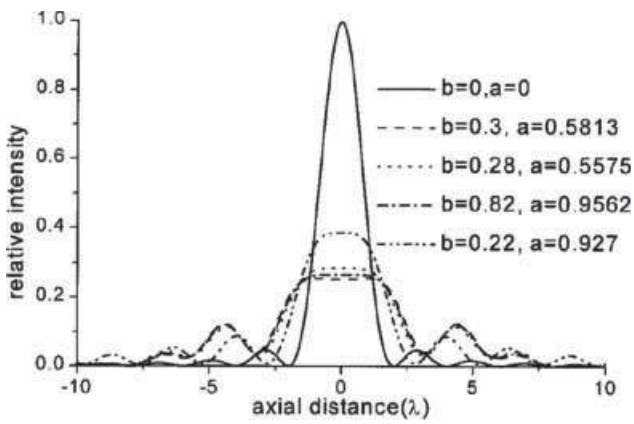

Figure 4 Axial intensity in the focal region of a focusing lens with NA = 0.85 when optimized 3-belt apodizers with different pairs of a and bare applied to its aperture.

For the optimized value of a and b for eliminating beam divergence, the radial behavior of the beam can be investigated by looking at the intensity beam profile in the focal plane, which is given as

$$\begin{aligned} I(\rho,0) &= |G(\rho,0)|^2 \\ &= \left|\frac{2}{\rho}\{J_1(\rho) - 2[aJ_1(a\rho) - bJ_1(b\rho)]\}\right|^2. \end{aligned} \quad (17)$$

The corresponding radial behavior for the pairs of a and b used in Fig. 4 is plotted in Fig. 5. Now we see that not all the

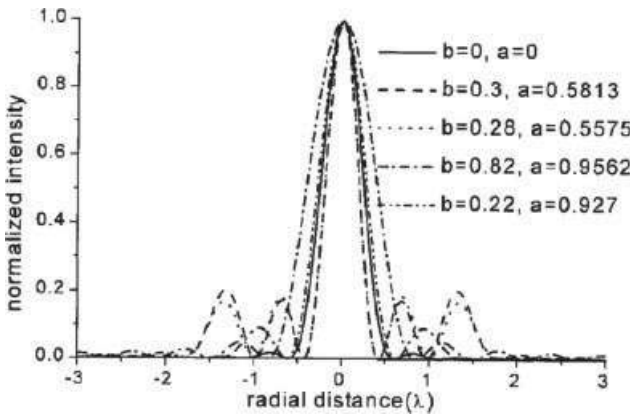

Figure 5 Radial intensity profiles on the focal plane of an NA = 0.85 lens when optimized 3-belt apodizers with different pairs of outer radius a and inner radius b are applied to the aperture of the lens.

pairs of a and b for optimized axial intensity can result in a superresolution light spot. The solid curve corresponds to a system without the apodizer. Only two of the selected pairs of a and b can result in a smaller beam spot than that obtained without the apodizer, i. e. ($b = 0.28$, $a = 0.5575$) and ($b = 0.3$, $a = 0.5813$). Therefore, this optimization process also includes beam spot size comparison. In fact, we can also ensure that the fourth derivative of the axial intensity is zero, by taking the values $a = 0.2864$ and $b = 0.8248$ [39].

Apodizer design based on vector focusing method

When the numerical aperture of the optical lens is above 0.6, a vector focusing method is preferable in the design of the apodizer. When plane-polarized light is refracted by a focusing lens, cross components of polarization are introduced upon focusing. These degrade the focused spot. For example, for a Bessel beam generated from focused plane-polarized light, the focal spot splits into two when the NA is greater than about 0.92 ($\sim$ 66° [40,41].

Supposing the field on the exit pupil of the focusing lens is linearly polarized in the X direction, the field in the focal region is given as [16,42–44]

$$E_x = -iA\left(I_0 + I_2\cos 2\varphi\right), \tag{18}$$

$$E_y = -iAI_2\sin 2\varphi, \tag{19}$$

$$E_z = -2AI_1\cos\varphi, \tag{20}$$

where

$$I_0 = \int_0^\alpha \sqrt{\cos\theta}\sin\theta(1+\cos\theta)J_0(kr\sin\theta) \times \exp(ikz\cos\theta)d\theta, \tag{21}$$

$$I_1 = \int_0^\alpha \sqrt{\cos\theta}\sin^2\theta J_1(kr\sin\theta)\exp(ikz\cos\theta)d\theta, \tag{22}$$

$$I_2 = \int_0^\alpha \sqrt{\cos\theta}\sin\theta(1-\cos\theta)J_2(kr\sin\theta) \times \exp(ikz\cos\theta)d\theta, \tag{23}$$

in which $\alpha = \arcsin(\mathrm{NA})$ denotes the largest focusing angle, φ is the azimuthal angle, $k = 2\pi/\lambda$, J_0, J_1 and J_2 are zero order, first order and second order, first kind Bessel functions. The constant $A = \pi l_0 f/\lambda$, $l_0 = 1$ for uniform illumination, and f is the focal length.

A linearly polarized plane wave goes through a multi-belt binary optics element, and then is focused by an aplanatic lens with NA of 0.85. The transmission of the aperture is $T(\theta)$. The optimization of the binary optics element can be achieved by making the intensity on the optical axis to be constant within a certain range. Because E_z and E_y are zero on the optical axis, the optimization can be achieved by looking only at the E_x field intensity on the optical axis. The strength of the electric field at an axial point with a distance z from the focal plane is given as [45]

$$F(z) = \int_0^\alpha \sqrt{\cos\theta}(1+\cos\theta)T(\theta)\exp(ikz\cos\theta)\sin\theta\, d\theta. \tag{24}$$

For an n-belt binary optics element, ($n = 1,2,3,\ldots$), the radius of each belt is r_j, ($j = 1,2,\ldots,n$), and $r_{j-1} < r_j < r_n = 1$, $r_0 = 0$, and the corresponding focusing angle is $\alpha_j = \arcsin(r_j\mathrm{NA})$, so that $r_j = \sin\alpha_j/\mathrm{NA}$. By making the transmission coefficient within each belt to be 1.0, the transmission function $T(\theta)$ can be expressed as a function of the belt order, and is given as $T_j = (-1)^{j+1}$. Thus Eq. (24) can be further expressed as:

$$\begin{aligned} F(z) &= \sum_{j=1}^{n} T(j)\int_{\alpha_{j-1}}^{\alpha_j}(\cos\theta)^{1/2}(1+\cos\theta) \\ &\quad\times\exp(ikz\cos\theta)\sin\theta d\theta \\ &= \sum_{j=1}^{n}(-1)^{j+1}\left(f(\alpha_j,z)-f(\alpha_{j-1},z)\right), \end{aligned} \tag{25}$$

where

$$\begin{aligned} f(\alpha_j,z) = &\Big(\exp(ikz)(3-4ikz) \\ &+ i\exp(ikz\cos\alpha_j)(\cos\alpha_j)^{1/2} \\ &\times(3i+2kz+2kz\cos\alpha_j)\Big)\Big/\left(2k^2z^2\right) \\ &+\sqrt{i\pi kz}(3i+2kz) \\ &\times\left(\mathrm{Erfi}\left(\sqrt{ikz}\right)-\mathrm{Erfi}\left(\sqrt{ikz\cos\alpha_j}\right)\right) \\ &\Big/\left(4k^3z^3\right) \end{aligned}$$

and $\mathrm{Erfi}(x)$ is the imaginary error function. Equation (25) describes the field on the optical axis generated by a system using the multi-belt π phase binary optical element.

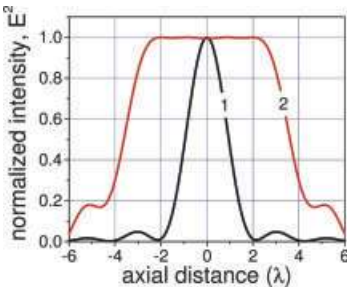

Figure 6 (online color at: www.lpr-journal.org) Axial intensity in the focal region of the NA = 0.85 lens. original system "1" and that with seven-belt optimized binary apodizer "2".

With this expression and the relation $\alpha_j = \arcsin(r_j \mathrm{NA})$, it is easy to find a series of values r_j to obtain the expected axial intensity. A nondiffracting beam can be obtained by optimizing the radius (r_j) of each belt towards obtaining a constant axial intensity, but to scale the size of the nondiffracting beam to sub-wavelength and smaller than the diffraction limit, the spot size has to be taken into consideration in the optimization process. For example, the axial intensity (as shown in Fig. 6) is made constant within an appreciable range by using a seven-belt ($r_1 = 0.0896, r_2 = 0.2852, r_3 = 0.4869, r_4 = 0.6136, r_5 = 0.6755, r_6 = 0.7688, r_7 = 1$) phase element, the FWHM of the total intensity profile in the focal plane is 0.53λ. The diffraction limit of this objective lens is 0.59λ, so that the beam size is about 9% smaller than the diffraction limit. The image of the beam in the focal region before and after using the binary apodizer is shown in Fig. 7. It is clear that for the original system, the beam diverges rapidly away from the focal plane, while for the system with binary apodizer, the beam does not diverge within a range of 5 wavelengths.

Binary apodizer design for radially polarized light

Radially polarized light has its polarization direction pointing outward in all transversal directions from its beam center. It was observed many years ago in the output of different types of laser [46]. Radially polarized light can also be generated from linearly-polarized light by a variety of different optical methods [47, 48]. When such a beam is focused, a small longitudinally-polarized focal spot is produced [49]. However, this is degraded by shoulders of radially-polarized light in the focal distribution. Using a ring aperture to obstruct the central part of the incident beam can result in a smaller light spot than achievable by focusing plane-polarized light [49–54], due to the suppression of the radial field component contributed by the lower aperture field, thus effectively reducing the cross-polarization effect. However, such a beam diverges when it is out of focus, and the obstruction results in low optical efficiency. By replacing the ring aperture with a binary apodizer, and increasing the numerical aperture of the focusing lens to 0.95, a superresolution needle of (nondiffracting) longitudinally polarized light can be achieved [55].

The electric fields in the focal region for illumination of the high aperture lens with a radially polarized Bessel-Gaussian beam is expressed as [55–58],

$$E_r(r,z) = A\int_0^{\alpha} \sqrt{\cos\theta}\sin 2\theta\, \ell(\theta) J_1(kr\sin\theta) e^{ikz\cos\theta} d\theta, \tag{26}$$

$$E_z(r,z) = 2iA\int_0^{\alpha} \sqrt{\cos\theta}\sin^2\theta\, \ell(\theta) J_0(kr\sin\theta)\, e^{ikz\cos\theta} d\theta, \tag{27}$$

where $\alpha = \arcsin(\mathrm{NA})$ and NA is the numerical aperture, and the function $\ell(\theta)$ describes the amplitude distribution of the Bessel-Gaussian beam, which is given by

$$\ell(\theta) = \exp\left[-\beta^2\left(\frac{\sin\theta}{\sin\alpha}\right)^2\right] J_1\left(2\gamma\frac{\sin\theta}{\sin\alpha}\right), \tag{28}$$

where β and γ are parameters that are taken as unity in this configuration. The numerical aperture of the focusing lens is $\mathrm{NA} = 0.95$ ($\alpha \approx 71.8°$. The corresponding field distribution is shown in Fig. 8.

As is shown in Fig. 8a, the peak of the radial component of the intensity $|E_r^2|$ is about 30% of that of the longitudinal intensity $|E_z^2|$. This strong cross-polarization effect makes the beam size as big as FWHM = 0.68 λ, which is larger than the diffraction limit for this focusing lens $\lambda/(2\mathrm{NA}) = 0.526\lambda$. And this beam diverges rapidly away from the focus, as is shown in Fig. 8b. By applying a binary apodizer to the exit pupil of the focusing lens, a superresolution and nondiffracting beam can be realized [55].

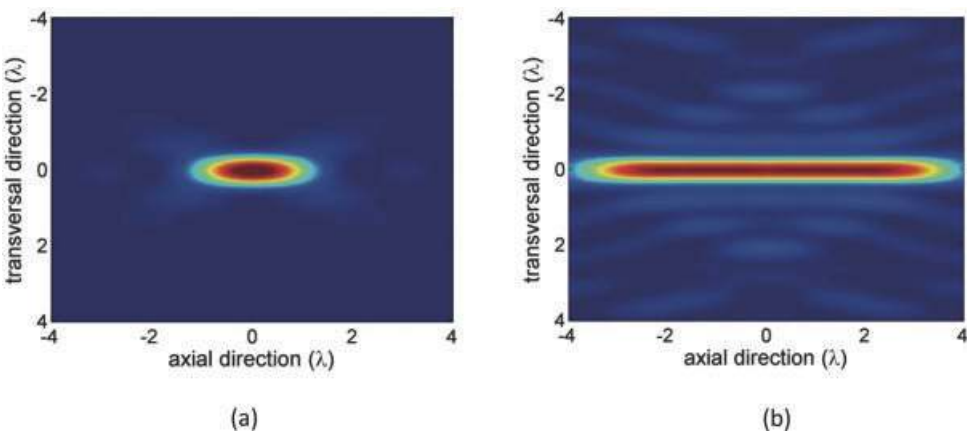

Figure 7 (online color at: www.lpr-journal.org) Intensity image in the focal region of the NA = 0.85 lens. (a) Without binary apodizer. (b) With binary apodizer.

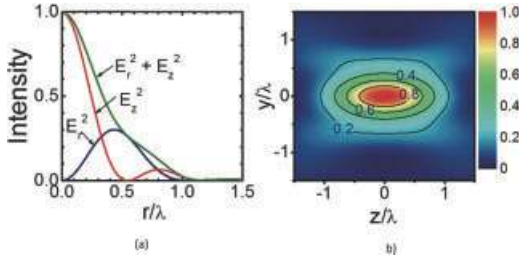

Figure 8 (online color at: www.lpr-journal.org) Intensity in the focal region of a NA = 0.95 lens illuminated with a radially-polarized Bessel-Gaussian beam. (a) Radial component $|E_r^2|$, longitudinal component $|E_z^2|$ and the total intensity $|E_r^2| + |E_z^2|$ on the focal plane. (b) Contour plot of the total intensity distribution on the y-z cross-section.

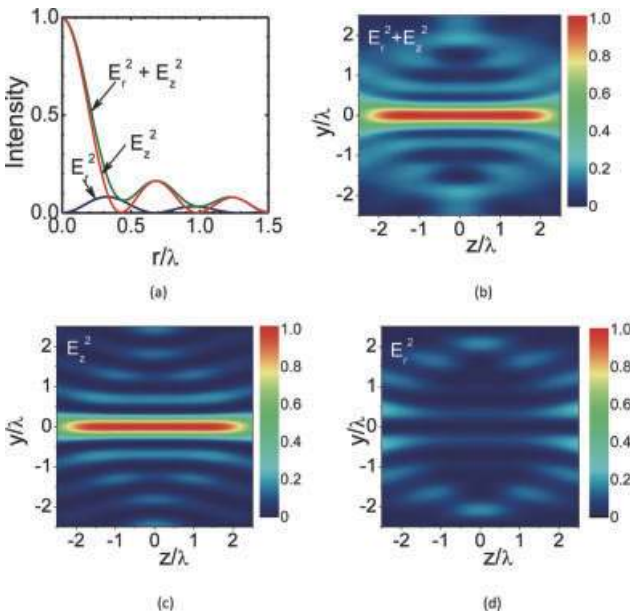

Figure 9 (online color at: www.lpr-journal.org) Intensity profiles on the focal plane and contour plots of the intensity distributions in the yz-plane after using the binary apodizer. (a) Intensity profile on the focal plane. (b) The total intensity distribution. (c) The longitudinal components. (d) The radial component.

For example, when a five belt binary apodizer ($r_1 = 0.091, r_2 = 0.391, r_3 = 0.592, r_4 = 0.768, r_5 = 1$) is applied, the radial intensity $|E_r^2|$ can be reduced to around 8% of that of the longitudinal intensity $|E_z^2|$, and the beam size, i. e. the FWHM of the $|E_r^2| + |E_z^2|$ profile is only 0.43λ, which is about 18% smaller than the diffraction limit of the optical system, as is shown in Fig. 9a. A nondiffracting effect is also achieved, as is shown in Fig. 9b. The total intensity beam size is constant within a range of 4 wavelengths, and also the longitudinally-polarized intensity dominates, as can be seen in Fig. 9c. The radially-polarized intensity is quite low, as is indicated in Fig. 9d: this beam is substantially longitudinally polarized. The function of the binary apodizer is like a polarization filter, diffracting the radially polarized light away from the center of the beam.

Discussion and conclusion

In conclusion, the divergence of a focused laser beam can be reduced by using annular apodizers, where the narrow ring slit serves as an angular spectrum extender. Different angular spectra of light rays from the ring slit are focused at different positions along the optical axis, which form a longer axial light spot, and therefore the divergence of light beam in the focal region is reduced. However, to eliminate beam divergence within a specified range with high optical efficiency, phase apodizers have to be used: the phase apodizers can diffract light and generate multiple focal points along the optical axis, where the defocusing spherical aberration of the neighboring light spots have opposite signs, which can totally offset each other when their intervals are properly adjusted through apodizer design. The parameters of the apodizers depend much on the field distribution and polarization state of the incident light. A longitudinally-polarized beam that propagates a few wavelengths without divergence can be generated by tightly focusing radially polarized light after going through a binary apodizer. For a low numerical aperture focusing system (NA < 0.6) a scalar design method can be applied, but when NA > 0.6, the vector design method is preferred for taking different polarization states of light into account. Apodization techniques can reduce the divergence of a laser beam, and therefore a new class of "nondiffracting beams" is coined, which will be addressed in the next section.

2.1.2. Nondiffracting beams

The propagation of light without transverse spreading may be seen as "a theorist's dream", but a very useful one when transposed to the constrained world of experiments. Like the plane wave, a nondiffracting beam is a concept that would require a source of infinite extent and energy. Naturally this condition cannot be met in experiments where nondiffracting beams are an approximation to the ideal solution. As a result, practical nondiffracting beams do not totally eliminate beam spreading, but they are able very significantly to mitigate it. Several techniques have been successfully implemented to generate these beams. We will first briefly describe various nondiffracting beams, such as the Bessel beam, with emphasis on their properties and generation. We will discuss the underlying physics but for more quantitative details the reader could for example refer to the excellent reviews in [59, 60]. We will also present a few applications of nondiffracting beams, most of them still emerging. The properties of these beams can offer new opportunities to several areas of optics such as laser material processing, and biological imaging.

Properties and generation of "nondiffracting" beams

Lasers have an extraordinary ability to concentrate energy in space. This is best achieved when a laser operates in its fundamental mode (typically the mode denoted TEM_{00}). In its simplest form the emitted beam is then described as a

Gaussian beam, which corresponds to a so-called diffraction-limited beam. This name emphasizes the limits imposed by diffraction. Diffraction affects the collimation of a beam, so that it is never exactly "pencil-like" and spreads while propagating. For a focused beam, diffraction can very severely limit the depth of field. Indeed, when focusing a beam of light, one faces a trade-off between the beam waist size at focus and the distance along the optical axis over which the beam waist size remains close to its value at focus. For Gaussian beams this is quantified by the Rayleigh range Z_R, which is the distance over which the beam increases its cross-sectional area by a factor of two, given by $Z_R = \pi w_0^2/\lambda$, where w_0 is the beam waist size and λ is the wavelength. In theory it is possible to go beyond the limits of diffraction with the so-called "nondiffracting beams". Really the beams are not nondiffracting, but only appear to be so: the energy in the central lobe of a Bessel beam spreads upon propagation, but energy from the strong outer rings diffracts inwards, attaining a dynamic equilibrium. In practical implementations, where the energy and cross-section of the beam are limited, nondiffracting beams are not fully immune to beam spreading. Nevertheless, they can conserve a small beam central lobe size over a propagation distance far beyond the Rayleigh range. But the total energy is spread over a much bigger region than the central lobe.

The Helmholtz equation governs the phenomenon of diffraction:

$$(\nabla^2 + \mathbf{k}^2)\psi(\mathbf{r},\mathbf{k}) = 0\,, \tag{29}$$

where ∇^2 is the Laplacian, $\mathbf{k}$ is the wave number, $\mathbf{r}$ is the position vector, and $\psi(\mathbf{r},\mathbf{k})$ is the electromagnetic field component. The Bessel beam is usually attributed to Durnin, who pointed out that this equation admits a class of diffraction-free mode solutions [61, 62]. The plane wave was already known to be non-diffractive, but it does not correspond to a beam of light, as its energy is not concentrated in space. In his experimental report, Durnin generated an approximation of a zero-order Bessel beam and demonstrated its exceptionally low spreading. This non-diffractive behavior also characterizes higher-order Bessel beams, which take their name from the expression of their intensity profile, which is the square of a Bessel function of the first kind of order l, noted J_l. Their electric field is given in cylindrical coordinate (r,φ,z) by:

$$E(r,\varphi,z,t) = E_0 \exp[i(-\omega t + k_{||} z \pm l\varphi)] J_l(k_\perp r)\,, \tag{30}$$

where z is the position along the optical axis, t is the time, ω is the angular frequency, $k_{||} = (2\pi/\lambda)\cos\theta$ is the longitudinal wavevector, $k_\perp = (2\pi/\lambda)\sin\theta$ is the transverse wavevector, and θ is a fixed angle. Actually Eq. (30) was given by Stratton in 1941, as the elementary wave functions in cylindrical coordinates [63]. He went on also to give the fields for vectorial solutions. His solutions apply for the electromagnetic modes of a cylindrical waveguide. Sheppard discussed the limiting case of free space propagation, and described how the Bessel beam "propagates without change" [64]. This paper also introduces the Bessel-Gauss beam, which is a finite-energy solution with weak diffractive spreading. A vectorial Bessel beam solution was also given by Sheppard, and the transverse electric and magnetic modes mentioned [65].

The non-diffractive nature of a Bessel beam is evident from its mathematical formulation. Its intensity distribution $J_l^2(k_\perp r)$ does not depend on z, the position along the optical axis, which means that it is propagation-invariant. The non-diffractive nature of Bessel beams can also be understood in the following way [60]. We first note that any beam can be described as a sum of plane waves. This is the so-called plane wave expansion. If each plane wave has its wavevector lying on a cone, all wavevectors have the same longitudinal component $k_{||}$ and form the same angle θ with $k_{||}$, while the transverse components $k_\perp$ lie on a circle, as illustrated in Fig. 10a,b1. The Fourier transform of a circle, Fig. 10b1,

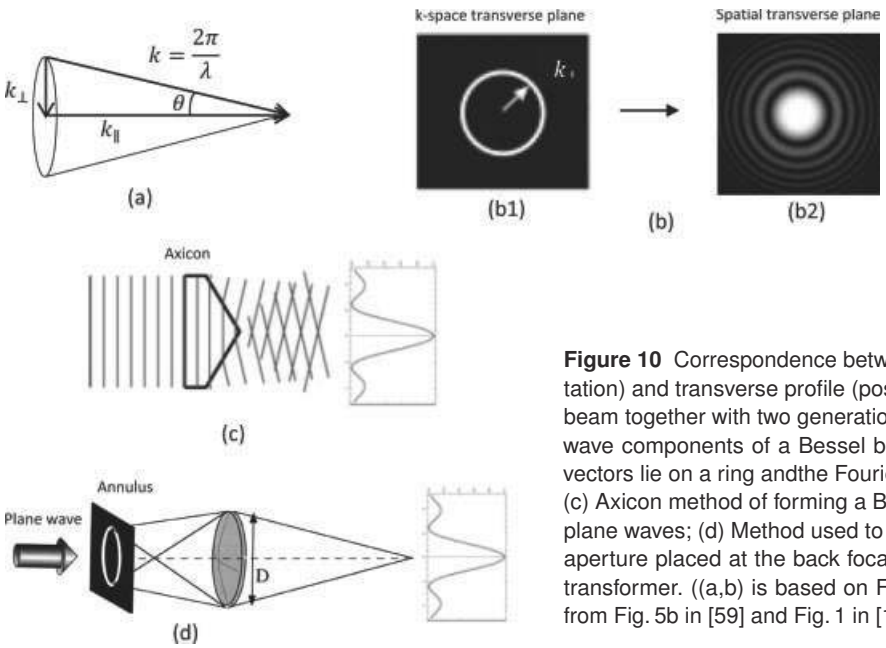

Figure 10 Correspondence between angular spectrum (k-space representation) and transverse profile (positional space representation) of a Bessel beam together with two generation methods: (a) The k-vectors of the plane wave components of a Bessel beam lie on a cone; (b) The transverse k-vectors lie on a ring andthe Fourier transform of a ring is a Bessel function; (c) Axicon method of forming a Bessel beam showing the superposition of plane waves; (d) Method used to generate Bessel beams using an annular aperture placed at the back focal plane of a lens which acts as a Fourier transformer. ((a,b) is based on Fig. 11 in [60]. (c) and (d) are reproduced from Fig. 5b in [59] and Fig. 1 in [102], respectively.

is a Bessel function, Fig. 10b2. The non-diffractive nature of Bessel beams is evident from this wavevector geometry: moving along a distance dz along the optical axis, each plane wave accrues the same phase shift $k_{||}z$. Since no dephasing is introduced between the plane-wave components, the beam, which is the addition (or interference) of these plane waves, remains invariant while propagating along z.

Another remarkable property of Bessel beams is that of reconstruction: when a beam is obstructed by an object (locally changing the phase or the intensity of the beam), it is able to reconstruct behind the shadow of the object.[1] The field is disturbed locally just behind the obstruction, but energy is redistributed among the rings to re-establish a Bessel profile over a distance behind the obstruction appropriately termed the "healing distance". Actually the beam does not really reconstruct, but appears to do so: the energy in the central lobe is continuously replenished from the outer rings. Using Fig. 10a to construct the shadow, it is clear that the healing distance decreases with increasing θ , as explained in detail in [59]. As we will see below, this property is put to good use in many applications.

Another noticeable property of Bessel beams of order l is that they carry an orbital angular momentum $\pm lh$, where h is the Planck constant. This is related to the vortex phase term $\exp(\pm il\varphi)$ in Eq.(30), which leads to a twisted phase front. Beams carrying angular momentum have led to many exciting developments [66].

Although Bessel beams are by far the most common class of nondiffracting beams, a few other classes of nondiffracting beams have been proposed and recently demonstrated, including Mathieu and Airy beams. First we mention that the two-dimensional analog of the Bessel beam is the cosine beam: two interfering plane waves generate a set of cosine fringes that are propagation invariant. But perhaps we cannot consider this as a beam as it is not localized. Mathieu beams can be considered as elliptical generalisations of Bessel beams (in fact they can be seen as a superposition of Bessel beams [67]). In addition to the parameter l, the order of the mode, which quantifies the orbital angular momentum as for Bessel beams, a parameter usually denoted q accounts for the "ellipticity" of the Mathieu beam [68]. The other class of nondiffracting beams, which were predicted more than 30 years ago [69], are the Airy beams. They are "freely accelerating" and have the peculiarity that they appear not to propagate in a straight line but along a parabolic path. But although the central lobe moves in a nonlinear fashion, it is well known that the center of gravity of a beam must travel in a straight line as a consequence of conservation of momentum. Only recently have Airy beams been experimentally demonstrated [70]. They are generated by Fourier transformation of a cubic phase. In fact they are closely related to the cubic phase mask method for depth of field enhancement [71]. Temporal Airy beams can also show invariant propagation in time: in other words they are non-dispersive waves. Recently Chong et al. have combined the spatial invariance of a Bessel beam (nondiffracting) and the temporal invariance of an Airy beam (non-dispersive) to create a linear light bullet, which neither spreads in space nor in time [72, 73]. Such a light bullet is illustrated in Fig. 11, showing the two stages of its synthesis, as well as its spatial and temporal evolution. Advantages of this type of light bullets compared to their nonlinear counterpart include ease of generation, flexible choice of the bullet's energy, and propagation independent of the medium's dispersion. It should however be noted that the intensity in the central lobe of the beam is traded off for its nondiffracting property. The more rings are present in the generated Bessel beam, the lower is the beam spreading, but the lower is the peak intensity (in the central lobe).

In this section we are considering "nondiffracting" beams in the context of linear optics. However, it should be noted that at high optical intensities a balance between self-focusing and diffraction can effectively supress beam spreading. Typically, self-focusing occurs through the optical Kerr effect, where the index of refraction of a medium changes in relation to the intensity of the light propagating through it. Such a balance results in optical spatial solitons [74]. When plasma defocusing is also present, filamentation occurs [75]. It was recently shown that this phenomenon could induce condensation in air and it was speculated that it may allow triggering rain on demand [76]. Although these non-linear optical techniques do provide a way to fight against diffraction, they are relatively complex from both conceptual and experimental viewpoints and we refer interested readers to the key papers cited in this paragraph for details.

Generation of non-diffractive beams

The "nondiffracting" beam generated by Durnin was a zero-order Bessel beam [61]. The generation method drew on the fact that the Fourier transform of a ring in k-space (spatial frequency space) is a Bessel function, see Fig. 10b. Figure 10d shows the setup used, where an annulus was imaged by a lens, which acted as a Fourier transformer. It should be noted that this setup had already been used to increase the depth of field prior to Durnin's work (see [77] and references therein). First proposed by Airy [17], the J_0 amplitude variation for a thin annulus was described by Rayleigh [78]. Welford described the poor imaging of an extended object using an annulus, due to the strong side lobes, and this effect was demonstrated experimentally in the image of a razor blade edge [79]. Durnin's simple setup provided a convincing demonstration of nondiffracting beams, but did poorly in two regards: the intensity of the central lobe along the propagation axis was seen to oscillate widely; in addition, this method was inherently energy-inefficient, since most of the beam was blocked by using an aperture, which had to be made thin to produce a good approximation of a Bessel

[1] In fact this property is reminiscent of the so-called "Poisson point", a controversy between Poisson, supporter of the corpuscular theory of light, and Fresnel, whose work brilliantly contributed to establishing the wave nature of light. To object to Fresnel's diffraction theory, Poisson pointed out that it predicted the presence of a bright spot behind an opaque disk illuminated by a beam of light. In 1818, Arago proved experimentally that the beam indeed reconstructed behind the opaque disk to form this bright spot. This result significantly helped tilting the balance in favor of the wave theory of light.

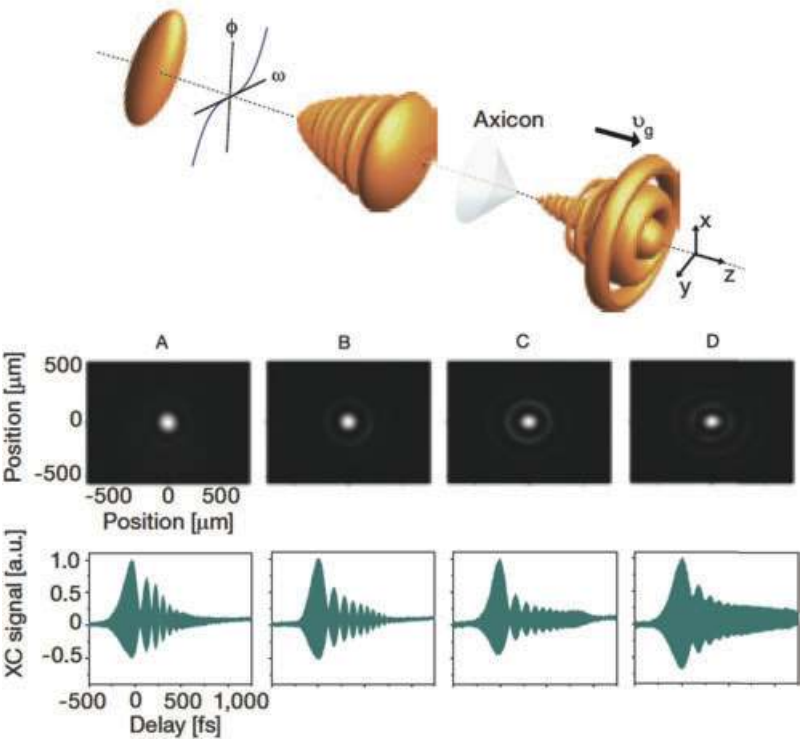

Figure 11 (online color at: www.lpr-journal.org) (Top) Schematic to generate Airy-Bessel wave packets. (Bottom) Propagation of an Airy-Bessel wave packet. A initial spatial and temporal profile, B profiles after propagation through 3.3 LR and 1.8 Ld, C 5.4 LR and 3.6 Ld, D 7.5 LR and 5.4 Ld. (LR = diffraction length of a beam with a diameter of 180 mm, Ld = dispersion length of a 100 fs pulse). (After [73]. Reproduced with permission from the Optical Society of America.)

beam. The first drawback could later be alleviated by combining the aperture with a resonant cavity [80, 81]. However, to shape a beam without wasting energy, it is best to modulate its phase rather that its intensity. This can be done with a special lens, conical in shape, termed an axicon [34, 82]. Refractive axicons provide a simple method to obtain the desired superposition of plane waves to generate a zero-order Bessel beam: see Fig. 10c. The J_0 behavior of an axicon was described by Fujiwara [83]. Diffractive axicons [84] serve the same function as refractive ones, but are generally more compact and introduce less chirping on intense optical pulses [85]. It was shown many years ago that spherical aberration (a quartic phase) can produce an axicon-like behavior over a limited range [86] (Remember the Airy beam is produced from a cubic phase.) This approach has been generalized to a general phase power greater than two, and also been applied to pulse shaping [87]. Illuminating an axicon with a Laguerre-Gaussian beam, one can also generate higher-order Bessel beams. The most flexible method to shape the phase of a beam is to use a spatial light modulator and produce a computer-generated hologram. This method is very flexible, as it allows imprinting any phase pattern on a beam, and has been successfully used to generate both Bessel and Mathieu beams, as well as to enable the first demonstration of Airy beams [70].

New methods to generate non-diffractive beams have been reported recently. Guided optics provides an alignment-free and compact means to generate Bessel beams. In the first demonstration a micro-axicon was formed on the output facet of an optical fiber [88]. More recently Ramachandran and Ghalmi have generated Bessel beams by forming a Bragg grating within a multimode fiber [89]. They showed that the generated beams could have a depth of focus that was 32 times larger than for a Gaussian beam of similar waist. Kim et al. also proposed and demonstrated an all-fiber Bessel beam generated by a method analogous to the free-space method used by Durnin [90]. It consisted of a hollow fiber with a ring-shaped core spliced to a coreless silica fiber with a polymer lens deposited on the end-facet. This design has recently been modified and simplified [91]. Zhan has also proposed an astute and simple way to generate evanescent Bessel beams with a high degree of confinement [92]. This method was recently demonstrated experimentally [93]. It consists of illuminating a metallic surface with a tightly focussed radially polarized beam. The radial polarization ensures that the beam is entirely TM with respect to the metallic surface. Due to the angular selection of plasmon excitation, the metallic surface effectively acts as an axicon, and an evanescent beam is generated.

Applications

Lasers have an extremely wide range of applications. In many of them, nondiffracting beams can lend a helping hand. It is beyond the scope of this review to give an exhaustive overview of the applications of non-diffractive beams. A recent review article gives an extended account of this topic, including optical manipulation and non-linear optics, which we will not cover here [59]. In the following we illustrate how the "rod of light" (i. e. large depth of focus) or reconstruction properties of Bessel beams are benefiting two major applications of laser and photonics.

Material processing

Non-diffractive beams have proved useful in microstructuring transparent materials without the requirement of scan-

ning the workpiece in the depth direction. For example Amako et al. used multishot subpicosecond pulses from an amplified titanium sapphire laser in combination with an axicon to produce long through-holes in silica. This involved two steps: structural modification and hydrofluoric acid etching [85]. The holes had a diameter in the order of 100 µm for a length in the order of millimeters. Recently, Bhuyan et al. also used Bessel beams for machining through-holes in glass, albeit on a much finer scale [94]. Their process was also more straightforward, using a single pulse with peak intensity above the breakdown intensity of the material for ablation, so that no etching was required. They produced nanochannels in borosilicate glass with diameters of only a few hundred nanometers and with aspect ratios in excess of 100. They attributed the channel wall parallelism to the fundamental stationarity of Bessel beams, which allows them to resist transverse beam breakup at ablation-level intensities.

Tsampoula et al. made use of Bessel beams for machining a different type of material: a cell membrane [95]. Many approaches exist for penetrating a cell membrane, but one of them has gained prominence. It makes use of a laser pulses together with multiphoton absorption to punch a hole through the cell membrane. This technique is particularly useful for delivering foreign DNA to a cell, a process called transfection. Tsampoula et al. applied Bessel beams to transfection, and showed two major benefits. The large depth of focus of Bessel beams served to eliminate the requirement to precisely focus the beam upon the membrane surface, which is a painstaking task and would preclude automation. In addition, the self-reconstruction ability of Bessel beams allowed performing this procedure even in the presence of an obstructing turbid layer.

Biological imaging

Nondiffracting beams have been proposed to benefit biological imaging, in particular through optical coherence tomography and microscopy. The main benefit is to allow for imaging deeper under the sample surface than is possible with the more standard Gaussian beams. The strong side lobes of the Bessel beam can be reduced by using a confocal imaging system [79].

Optical coherence tomography (OCT) is a kind of optical radar mainly used to image biological tissues. It is an attractive alternative to biopsy, which is highly accurate but invasive. Light backscattered from a sample is measured at scanned lateral positions. The axial resolution is obtained not from time of flight, which would be much too short to be measured accurately, but from interferences, which are highly spatially selective if a broadband source is used. Ding et al. proposed to use an axicon in the sample arm of an OCT interferometer in order to increase the depth of focus into the sample [96]. This is significant because in this way no dynamical focusing is required to obtain good lateral resolution deep inside a sample. Their demonstration showed a lateral resolution of 10 µm or better over a 6 mm depth (a Gaussian beam would result in a depth of only 0.25 mm for the same lateral resolution), thereby overcoming the usual trade-off between lateral resolution and focusing depth when conventional optical elements are used. Moreover, the intensity of the beam was approximately constant along this depth. It should however be noted that this intensity, being evenly distributed, is lower than the intensity at the focus of a conventional lens. Nevertheless, Lee and Rolland demonstrated that a high sensitivity can be obtained across a depth of several millimetres [97]. They also pushed the idea of Ding et al. one step further by demonstrating the advantage of Bessel beams over Gaussian beams for OCT on a biological sample.

Fahrbach et al. recently demonstrated a microscope with self-reconstructing beams (MISERB), and showed that the scanned Bessel beam did not only reduce scattering artefacts, but also increased the image quality and penetration depth in dense media [98]. The authors studied the performance of their new type of light-sheet microscopy with three different specimens. Most significantly in a piece of human skin, which is a highly inhomogeneous medium, a comparison of Gaussian and Bessel beam illumination showed that less scattering artefacts (and therefore more sample's details) as well as a longer penetration depth could be obtained with Bessel beams, see Fig. 12. Furthermore, Betzig et al. recently showed that Bessel beam plane illumination in conjunction with structured illumination and/or two-photon excitation was particularly suited to fast and/or high resolution 3D microscopic imaging [99].

Bessel beams, thanks to their "rod-of-light" geometry, are able to produce in-depth images with a simple 2D scan, whereas 3D imaging techniques such as confocal and two–photon microscopy would require a much longer acquisition time and additional data processing [100]. In projection tomography, which does not rely on "optically sectioning" the specimen, nondiffracting beams have the potential to image much greater depths than confocal microscopy [101, 102].

In fact, besides applications in reducing beam divergence and generating Bessel beams, the apodization technique can also be used to achieve superresolution focusing. This can be done through increasing the relative ratio of the high spatial frequency field to the low spatial frequency field [33]. For example, a Bessel beam generated using an annular aperture actually reduces the central low spatial frequency field, which also reduces the point spread function of the system, resulting in higher resolution than light focusing without an annular aperture [33]. The binary apodizers used in reducing beam divergence in the focal region can also reduce the point spread function of the system, which is achieved through interference between the fields from different belts of the apodizer.

2.2. Superresolution apodization

Usually, far field superresolution can be achieved through two ways. One way is through reducing the size of the PSF of the focusing system. To do this, one can use apodizers to change the amplitude or phase on the pupil of the focusing system, or use two beams to control the excitation of fluorescence material to form an effectively smaller PSF,

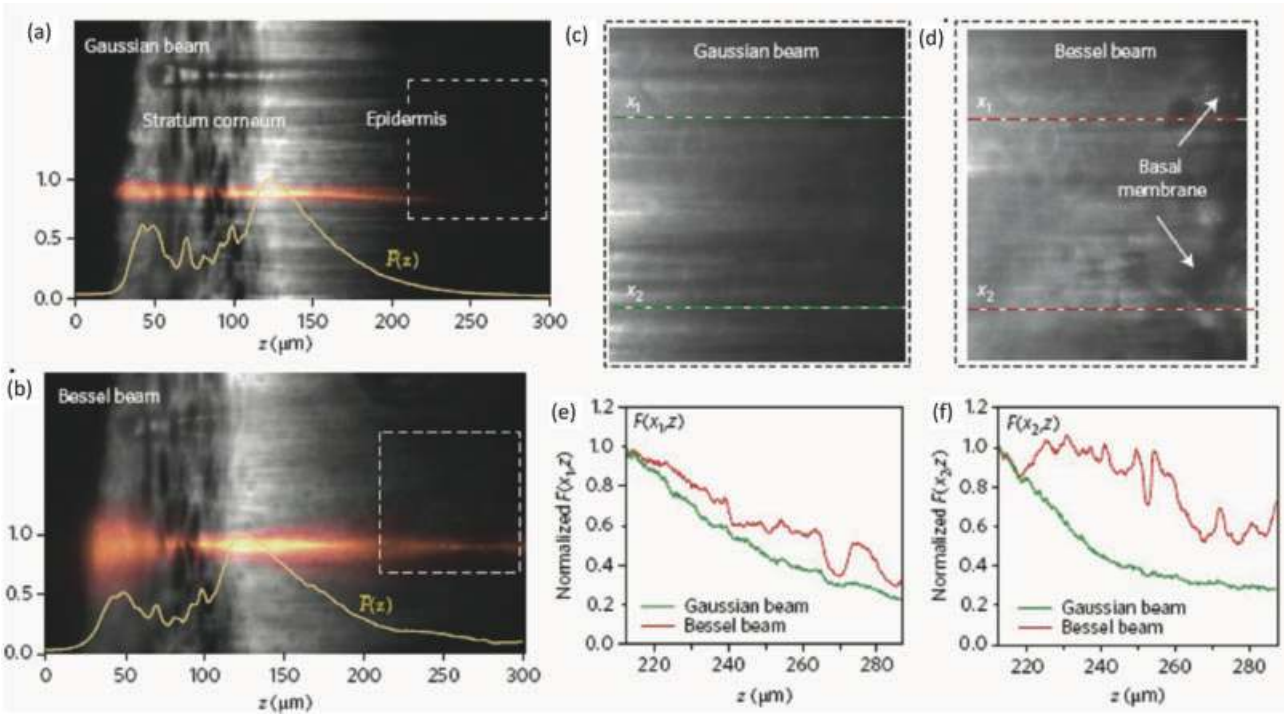

Figure 12 (online color at: www.lpr-journal.org) Maximum-selection images of human skin. Illumination by a conventional beam (a) or a self-reconstructing beam (b). The beams illuminate the skin from lef to right. Images from the Gaussian and Bessel beams at a single position are overlayed in orange-hot colors. Averaged intensity linescans show an exponential decay through the epidermis. (c,d) Part of the epidermis close to the basal membrane. magnified and autoscaled (boxes with dashed outline in a,b), revealing single cells only for Bessel beam illumination. (e,f) Line scans F(x,z) normalized to $F(x, z = 0)$ for $x = x_1, x_2$ (indicated by dashed lines in c,d), showing the strong increase in contrast for the Bessel beam illumination. (After Fig. 5 in [98], reproduced with permission from Macmillan Publishers Ltd.)

like in stimulated emission depletion (STED) microscopy. The other way is to increase the band width of the imaging system, like in structured illumination microscopy, by generating grating or fringe patterns to illuminate an object. The band width can be increased by up to two times, which corresponds to a two-fold increase in the resolution.

2.2.1. Superresolution apodization through pupil masks

The use of pupil masks (apodizers) to attain a focal spot smaller than the classical limit was first proposed by Toraldo di Francia [23, 24] and in the same year by Boivin [103]. It is found that these pupils can also achieve either increased depth of focus (as discussed in 2.1.3) or axial superresolution (i. e. decreased depth of focus) [27, 29, 104]. Both of these behaviors have potential applications. In fact, it is even possible to generate an axial minimum in the focal plane, corresponding to a bifocal effect. A general method to compare the performance of different filters with real (positive or negative) amplitude transmittance is to introduce performance parameters in terms of moments of the pupil [27]. For now, we consider scalar, paraxial systems. These parameters include S, the Strehl ratio compared with an unobscured pupil, which is related to the zero order moment of the pupil. This is maximized at unity for an unobscured pupil, which is called the Luneberg apodization

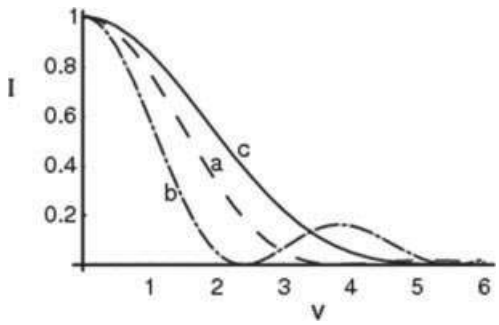

Figure 13 The image of a point object for different lens pupils: (a) unobstructed pupil (Airy disk), (b) a narrow annular pupil (Bessel beam), and (c) a pupil weighted for minimum second moment.

condition [105]. There have been several extensive reviews of the basic designs [106–108]. Figure 13 shows the point spread functions for two special cases. The Bessel beam gives a small central spot, but there are strong side lobes and, in fact, $S = 0$. Also shown is a pupil apodized for minimum second moment width [107, 109, 110].

One popular class of designs consists of an array of rings of different amplitude transmittance. Again, as the central lobe of the focused spot is reduced, S decreases and the side lobes become stronger. For binary phase masks the

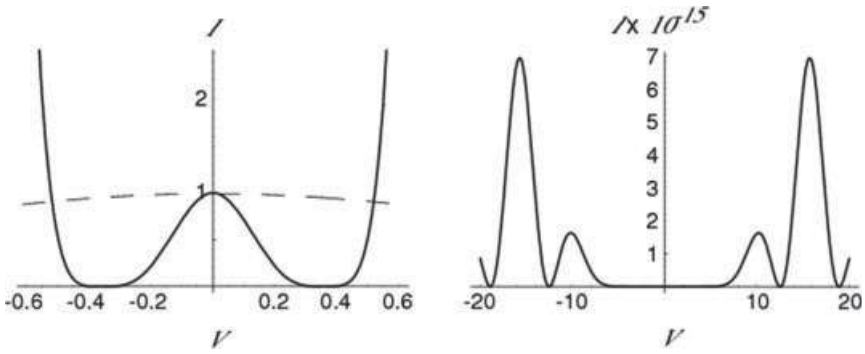

Figure 14 The intensity variation in the focal plane for a focusing system with a pupil consisting of 11 equally spaced delta functions, with coefficients as given by Zheludev [130]. The result for a slit pupil of the same width is shown dashed.

total power in a cross-section of the focal spot is independent of the filter. The simplest case of array of rings has just two elements [111–113]. In fact it has been shown that the 2-zone binary phase filter gives the highest value of S for a given spot size [114]. A high value of S is important for some applications, e. g. collection optics. But for many applications a more important parameter is F, the intensity at the focus compared with the integrated power or intensity (these are not the same for nonparaxial sysyems) in the focal plane [27, 115]. Maximum F is achieved for an amplitude filter, called a leaky filter, as this absorbs some energy that would otherwise appear in the side lobes [126–128]. 2-zone leaky filters can give simultaneous transverse and axial superresolution. 3-zone filters [114, 118–121] can also give simultaneous transverse and axial superresolution, with higher F than for a 2-zone filter.

The performance parameters based on pupil moments has also been extended to the nonparaxial regime (for scalar and various different vectorial cases) [122–125], and various designs have also been presented [45, 55, 126]. If superresolving masks are used in confocal system, there is more flexibility in the design, and side-lobe strength can be reduced [127, 128].

Recently, superresolving filters have been rediscovered under the name of super-oscillations [129]. Super-oscillations, as do any super-resolving filter, suffer from a number of practical limitations. First is the tolerance on the strengths and frequencies of the components. Second is the efficiency. Figure 14 plots the intensity associated with a function described by Zheludev [130]. The intensity (squared modulus) has been replotted in terms of the normalized optical coordinate v, so that the classical focus of a slit pupil of the same width has a first zero at $v = \pi$. As described by Zheludev, the super-oscillating function is much narrower: its first zero is at about 0.36. (Actually it does not go exactly to zero, an indication of the sensitivity to errors.) But the strength of the highest side lobe is 7×10^{15}, so a very small amount of the energy ends up in the focal spot. The third limitation is the field of view: the intensity is less than that at the focus for a width of 1.04, so an image formed with this pupil would only contain about two pixels. A similar filter could be designed to have 8 closely spaced zeros, which would have increased the field of view to about 8 pixels. The fourth limitation is axial behavior [27, 104]. Figure 15 shows the intensity along the axis, plotted against the axial optical coordinate $u = [8\pi n \sin^2(\alpha/2)]/\lambda$. At least in this case there is a 3D maximum in intensity at the focal point. But the intensity is less than that at the focus for a range of u of only 0.16. To put it into perspective, this compares with the distance along the axis between the first zeros for a circular aperture of about 25. So the focus has to be set with very high accuracy. The outer side lobes go up to an intensity of 5×10^{15}. The focus is surrounded by an extremely bright "corral". This particular example is not as badly behaved as some: the focal spot can actually be situated at a minimum in axial intensity [104]. But it demonstrates the difficulty in designing practical superresolving masks.

Another point that should be made is that the performance of superresolving pupils is fundamentally different from that of supergain antennas [24]. This is because supergain antennas have an aperture of a particular *width* that is used to produce a beam with an *angular* extent. So the size of the antenna can always be increased to make a narrower beam. On the other hand, for a focusing system the *angular* aperture is used to produce a focal spot with a resultant *width*. The angular aperture cannot be increased indefinitely.

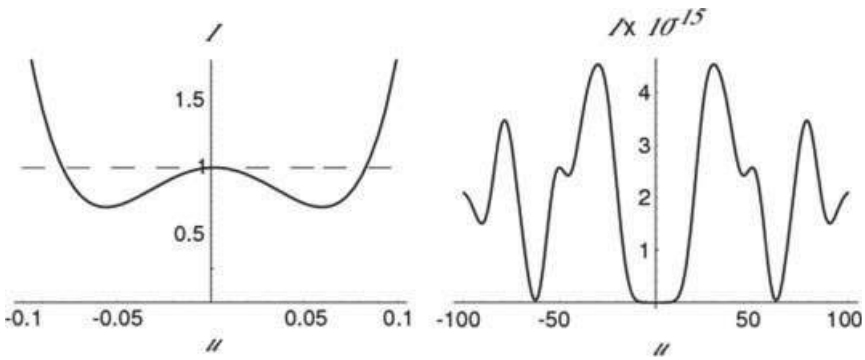

Figure 15 The intensity variation along the axis for a focusing system with a pupil consisting of 11 equally spaced delta functions, with coefficients as given by Zheludev [130]. The result for a slit pupil of the same width is shown dashed.

Moreover, the huge side lobes that are present with superresolving pupils can be made evanescent for the supergain antenna case, so that they appear in the reactive field of the antenna. This is a result of the interchange of the roles of angles and distances between the two techniques.

Superresolving filters can decrease the size of the focal spot produced by a lens, but the spatial frequency bandwidth is unchanged. In the next section, we discuss structured illumination, which allows the spatial frequency cut-off to be increased by a factor of two.

2.2.2. Structured illumination

A classical coherent imaging system has a spatial frequency cut-off of $(n\sin\alpha)/\lambda$, where the numerical aperture is $n\sin\alpha$ and λ is the wavelength. Here n is the refractive index of the immersion medium and α is the semi-angular aperture of the imaging lens. It has been known since the time of Abbe that incoherent imaging, as in fluorescence, has a spatial frequency cut-off of $(2n\sin\alpha)/\lambda$, i. e. twice that for the coherent case. (Here we neglect the Stokes' shift of the fluorescent light). Although it is not completely fair to compare this with the coherent figure, nevertheless a grating of weak contrast with appropriate period is imaged by the incoherent system while it is not visible using the coherent one.

Also since the time of Abbe, it has been known that oblique illumination in a bright field imaging system increases the cut-off frequency. For a weakly scattering object the cut-off can be increased to $(2n\sin\alpha)/\lambda$, also an improvement of a factor of two. But until more recently it was not known how to combine these two approaches. Now we know that this combination can be achieved using structured illumination of a fluorescent object. Structured illumination in the wider sense encompasses several techniques. This not only includes of projection of a grating or fringe pattern on to the object [131, 132], which is the technique most commonly referred to as structured illumination, but also confocal microscopy, where the object is illuminated with a single spot of light and a pinhole is used before detection [53]. In these cases the cut-off can now be $(4n\sin\alpha)/\lambda$ (for equal illumination and collection apertures), a factor of four increase over the classical coherent case [131–133]. If the refractive index is 1.5, for the limiting case of $\sin\alpha = 1$, the cut-off thus corresponds to a period of $\lambda/6$. Using solid-immersion lens technology, for luminescent imaging in silicon a period of $\lambda/14$ could be imaged [134, 135]. Structured illumination is a modulation-demodulation scheme, where the demodulation can be performed either optically using a grating, or digitally. While the cut-off is the same for confocal imaging or for structured illumination using a grating, the spatial frequency response within the pass band is better in the structured illumination case, and inverse filtering can be performed during the demodulation process: resolution close to 100 nm has been achieved using visible light [136]. Figure 16 shows the point spread function for various different systems to illustrate the possibilities of confocal imaging or structured illumination with inverse filtering. Here $v = 2\pi rn\sin\alpha/\lambda$, where r is the cylindrical radial coordinate.

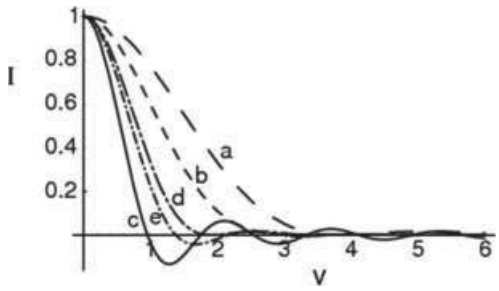

Figure 16 The image of a point object for different systems: (a) conventional system (Airy disk), (b) confocal system, (c) a superresolved system with cut-off $4n\sin a/\lambda$ and constant-valued OTF, (d) as in (c) with OTF weighted as in the conventional OTF, and (e) as in (c) with apodization to minimize the second moment width of I^2.

A similar improvement in resolution to that in structured illumination can be achieved using two-photon fluorescence microscopy, as a result of nonlinear effects [137, 138]. But this improvement is determined relative to the incident, longer-wavelength. Using confocal two-photon fluorescence, the cut-off is $(4n\sin\alpha)/\lambda$ in terms of the fluorescent wavelength. The resolution improvement associated with nonlinear effects has been long known in microlithography and optical data storage [139]. Recently other nonlinear effects have been combined with scanned imaging or structured illumination. STED (stimulated emission depletion) microscopy has achieved resolution in the order of 28nm [140] and saturated structured illumination microscopy has achieved 50 nm [141].

The methods described above have the great advantage over near field methods in that they can be applied to thick objects. In principle it is possible to illuminate the sample with a full spherical beam [142], which can be approximated by using two opposing microscope objectives [143, 144]. This technique can be performed in either confocal (4Pi microscopy) or conventional fluorescence mode (I5M) [144, 145]. Recent results have achieved 100 nm isotropic 3D resolution [146], while combination with nonlinear effects from stimulated emission depletion (STED) has achieved 45 nm 3D resolution, or $\lambda/16$ [147]. Detailed discussion of STED will be addressed in the next section.

2.2.3. Superresolution based on fluorescence switching

Fluorescence switching refers to controlling the emission of fluorescent markers by the use of light at different wavelengths. Typically, one beam is used to switch on and another to switch them off. When the beams are used to selectively switch on or switch off markers, superresolution imaging can be achieved in fluorescence microscopy. Two broad approaches for achieving superresolution using fluorescence switching exist. In the first approach, illumination profiles of beams are shaped through apodization schemes to reduce the effective PSF to subdiffraction dimensions in techniques

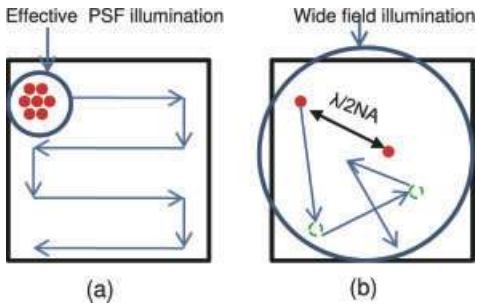

Figure 17 (online color at: www.lpr-journal.org) (a) Targeted Readout: Super resolution is achieved by reducing the PSF to sub-diffraction dimensions by controlling the on and off states of fluorophores. The detected fluorescence is thus an aggregate of the emission from all molecules contained within this illuminated area as shown by the numerous red dots within the demarcated illumination spot. In order to image the entire sample, the beam is scanned as shown by the blue arrows. (b) Stochastic readout: A wide field illumination scheme is used. Controlling the activation and excitation intensities ensures that only a single molecule emits within a diffraction limited region. For example, if N markers exist within a diffraction limited region, the activation intensity is set to $I_{\text{activation}}/N$, such that the chance of more than one molecule being excited diminishes significantly. The emitter is then localized by computationally fitting the recorded emission pattern. The excitation beam then bleaches the sample. The stochastic nature of the process requires several activation localization and excitation cycles to regenerate the complete image. The green circles indicate markers which are currently not emitting but will emit eventually upon repeated activation and excitation due to the stochastic nature of the process.

such as STED [140, 147, 148] as shown in Fig. 17a. Since this approach directly reduces the size of the PSF, it can be used to image most samples with minimal change to the instrumental setup. In the second approach, low intensity wide field illumination is used to activate and localize few markers with separations greater than the diffraction limit randomly within the illuminated region as shown in Fig. 17b. Due to the stochastic nature of this approach, several iterations of activation and localization have to be performed to reconstruct the complete image. Photoactivation Loacalization Microscopy (PALM) and Stochastic Optical Reconstruction Microscopy (STORM) fall in this category [149]. The stochastic nature of these methods requires several activation-localization cycles to obtain high resolutions. This translates to relatively slow image acquisition speed. In the subsequent portions of this section, the ideas behind various fluorescence switching approaches, their pros and cons, applications and recent developments are discussed.

STED microscopy uses two different light frequencies to excite and de-excite a fluorophore (fluorescent emitter). The effective fluorophore excitation area is reduced by overlapping the intensity pattern of the two light frequencies. For example, when a Gaussian beam is used for excitation and a Bessel-Gaussian beam, called the STED beam, is used for de-excitation, then the effective PSF is much smaller, as is shown in Fig. 18 [140, 148–150]. The optical setup of the system is identical to that of a confocal microscope [151] with an additional STED beam. The fluorophore molecular energy states describing the mechanism of this method are described in Fig. 19. The ground state of the unexcited fluorophore is denoted by S_0 and the excited state by S_1.

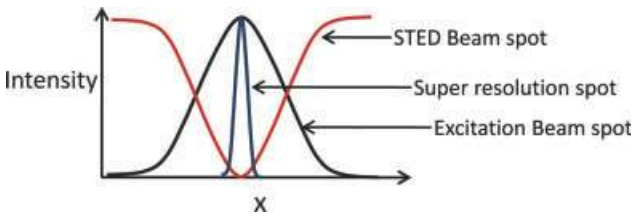

Figure 18 (online color at: www.lpr-journal.org) Schematic diagram of how the excitation spot is effectively reduced in size by STED microscopy. A diffraction limited excitation beam excites fluorophores within the region. A doughnut shaped STED beam with a null at the center of the excitation peak then serves to depopulate the excited fluorophores along the periphery of the excited region, allowing only the fluorophores at the null of the doughnut to fluoresce. Thus, the effective PSF is reduced as represented by the blue curve.

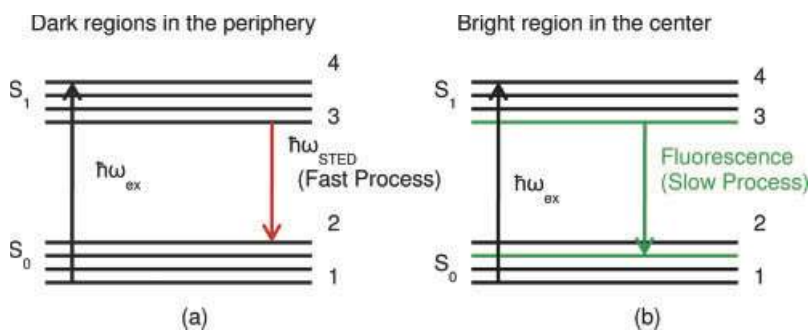

Figure 19 (online color at: www.lpr-journal.org) Molecular transitions depicting the concept of STED: (a) In the periphery of the excitation beam shown in Fig. 18, molecules are excited from state 1 to 4 by the excitation beam represented by $\hbar\omega_{\text{ex}}$. The molecules rapidly decay from state 4 to state 3 in around 0.2 ps. The STED beam resonates at the energy difference of states 3 and 2, which depopulate state 3, leaving this region dark. (b) At the centre of the excitation beam shown in Fig. 18, the STED beam contains a null, therefore the molecules excited by the excitation beam relax to state 3 and fluoresce over a nanosecond scale.

Each state consists of many closely spaced energy levels as shown in Fig. 19. When we excite the fluorophore with a photon of energy $\hbar\omega_{ex}$, molecules are excited to state 4 from state 1. They quickly relax non-radiatively to state 3 over a time scale less than 0.5 ps [152], which is significantly smaller than the fluorescence lifetime. Thus the population of state 3 builds up rapidly. Then over a longer timescale of a few nanoseconds, photons are emitted and molecules relax to state 2 or other nearby states. Finally, they relax to state 1 through quick non-radiative transitions. Now, consider that the flurophore is subject to perturbations by two beams of light at different frequencies – one at the excitation frequency of ω_{ex} (the state 1 to state 4 transition) and the other at a similar frequency as the fluorophore emission frequency given by ω_{STED} as shown in Fig. 19a. When appropriate intensities, time delays and pulse widths for the excitation (ω_{ex}) and stimulated emission beam (ω_{STED}) are used, the fluorophores excited by the beam resonant at ω_{ex} are rapidly brought back to the ground state by the beam resonant at ω_{STED} through stimulated emission. Stimulated emission, being a faster process thus depletes state 3 before the onset of spontaneous emission. Regions illuminated by both beams hence do not fluoresce and are 'dark' regions. Regions illuminated by the excitation beam alone are however able to fluoresce through spontaneous emission and are the 'bright' regions as shown in Fig. 19b.

If the STED beam shape is engineered so that it spatially overlaps the excitation beam only along the periphery, leaving a null at its center as shown schematically in Fig. 18, the effective PSF can be reduced. Beams of complex shapes may be designed using binary phase plates [152–154].

As one increases the intensity of the STED beam, the size of the effective PSF becomes smaller. The resolution limit or mimimum distance resolvable by STED is approximated by the relation $\lambda/\left(2\,\mathrm{NA}\sqrt{1+I_{max}/I_{sat}}\right)$ [155, 156]. Here, I_{max} is the peak intensity of the STED beam, I_{sat} is the minimum intensity required to deplete state 3 and NA is the numerical aperture of the focusing lens. Since we would like state 3 to be depleted before it can fluoresce through spontaneous emission, I_{sat} is estimated by setting the condition that the rate of stimulated emissions for the transition between state 3 and state 2 just exceeds the spontaneous emission rate between the two states. Due to the fast relaxation of the fluorescent markers from state 2 to state 1, the rate of stimulated emission can be approximated by $\sigma_{32}N_3I$, where σ_{32} corresponds to the gain cross-section of the marker and N_3 is the number of markers in state 3. The spontaneous emission rate is given by $\tau_{rad}^{-1}N_3$. Thus, I_{sat} is estimated by $\sigma_{32}N_3I_{sat}\left(\hbar\omega_{STED}\right)^{-1} = \tau_{rad}^{-1}N_3 \Rightarrow I_{sat} = \hbar\omega_{STED}\left(\sigma_{32}\tau_{rad}\right)^{-1}$. Typical values for the gain cross-section and the fluorescent lifetime are $10^{-16}\,\mathrm{cm}^2$ and 1–5 ns, which necessitates STED beam intensities in the $10\,\mathrm{MWcm}^{-2}$ range for effective operation. For most implementations of STED microscopy, a Ti:Sapphire mode-locked laser and an optical parametric amplifier were used to produce the excitation and STED beams respectively. However, the STED concept has been demonstrated using laser diodes [157] hence promising low cost implementation of this method.

The expression for the resolution yielded by the use of STED microscopy, indicates that as $I_{max}/I_{sat} \to \infty$, the resolution becomes arbitrarily high. However, this is perhaps not realistic due to the effects of photo-bleaching at high intensities (Fig. 20) and the finite transition time between state 2 and state 1. Additionally, with increasing intensities the rate of upward transition from state 2 to state 3 by absorption of photons from the STED beam can become comparable to the rate of non-radiative decay from state 2 to state 1. Thus the efficacy of the depletion process could reduce. For optimal operation of STED microscopy, the use of STED beam pulses of several picoseconds and short excitation pulses, few picoseconds in duration are recommended. This ensures that excitation occurs very fast, followed by efficient depletion of the fluorescent states. Such an arrangement where the fluorescent lifetime is much larger than the STED pulse duration would ensure that regions excited by the STED beam are switched off before they can fluoresce.

While STED has demonstrated high resolutions in three dimensions, isotropic resolution < 50 nm has been achieved after several improvements. The first implementation of

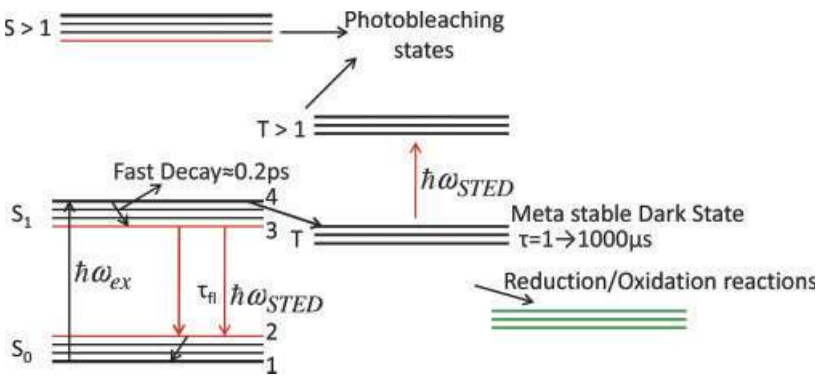

Figure 20 (online color at: www.lpr-journal.org) Photobleaching reaction mechanism: The higher bright states ($S > 1$) and metastable states ($T > 1$) are the nsnal starting points for photobleaching reactions. The long lifetime of several microseconds of the metastable state (T) increases the probability of a molecule in this state to be promoted to $T > 1$ by the STED beam. A way to alleviate this problem is to use ROXS fluorophores which depopulate the metastable state through reduction/oxidation reactions.

STED used a spatial overlap of STED and excitation beams and produced only a 1.3 improvement in lateral resolution and almost no improvement in axial resolution [158]. Eventually, the STED concept was married with 4Pi confocal techniques to produce significant improvements in the axial resolution [159]. 4Pi techniques coherently add two wavefronts at the focal point to realize a sharper point spread function. In conjunction with STED, 4Pi-STED produced an axial resolution of 33 nm or 23 times smaller than the wavelength [160]. Modifying the polarization of the excitation beam and detection scheme to enable maximum overlap with the dipoles of the organic dye molecules [161] led to lateral resolutions of 16 nm or a $\lambda/45$ resolution [161]. In order to obtain an isotropic resolution of less than 50 nm, two STED lasers called STEDxy and STEDz were orthogonally polarized and used in conjunction with the 4Pi method to yield constructive and destructive interference respectively at the focal point in a method called isoSTED [147]. Resolution improvements were further obtained by using a 4Pi interferometric detection scheme to improve the collection efficiency [162]. Other demonstrations involved intertwining two-photon microscopy and STED concepts, although this does not yield signficant advantages over single photon techniques in terms of resolution due to the longer wavelengths used in two photon absorption [163]. However two-photon STED could be more suited to 3D imaging because of lesser scattering by samples at longer wavelengths. With these gradual improvements, STED has now been able to demonstrate sub-10 nm resolutions. For instance, STED microscopy of nitrogen vacancy (NV) centers in diamond with resolutions of 5.8 nm with an $I_{max} = 8.6\,\mathrm{GWcm^{-2}}$ [164] was demonstrated.

Despite milestone achievements, STED has some limitations. One key limitation lies in the requirement for fluorophore transitions compatible with the employed excitation and STED beam wavelengths. Fluorophore transition states are however susceptible to ambient conditions [165]. Secondly, the high intensities used in STED may damage specimens and also increase the likelihood of photobleaching. Photo-bleaching refers to the formation of free radicals due to the sustained illumination of the dye. Starting points for photobleaching reactions are higher metastable states as shown in and the probability of these states being occupied increases with intensity. Figure 20 depicts the mechanism of the photobleaching process. There has been plenty of work to address these challenges.

To reduce the intensities used, a method called Ground State Depletion (GSD) microscopy was proposed [166] employing the same concept as STED microscopy but differing in the mechanism of depleting the fluorescent states. In fact, various targeted read out methods such as GSD and STED which overcome the diffraction limit by spatially switching fluorescence on or off fall have been categorized under Reversible Saturable Optical Fluorescence Transitions (RESOLFT) microscopy. GSD microscopy utilizes dyes with a metastable dark state with a lifetime of several µs, as shown in Fig. 21, to deplete the fluorescent states. When a high intensity excitation beam called the pump beam is applied with a doughnut-like spatial beam profile similar to the STED beam, fluorescent markers in these regions are excited from state 1 to state 4 followed by rapid decay to state 3. Simultaneously, a fraction of them make a non-radiative inter state crossing to the metastable state. This fraction increases with increasing intensity of the pump beam. The longer lifetime of the metastable state translates to an I_{sat} in the range of a few $\mathrm{kWcm^{-2}}$. Regions in the null of this pump beam are excited with a lower intensity probe beam at the same wavelength which leaves more molecules in state 3 and less in the metastable state. Thus, the central region near the null of the pump beam exhibits more fluorescence than the peripheral regions. Resolutions of 8 nm have been demonstrated with GSD at 910 times the saturation intensity (in the $100\,\mathrm{MWcm^{-2}}$ range) in NV centers, which is a much smaller intensity requirement compared to STED for similar resolutions [167, 168]. Other methods to alleviate intensity requirements include switching from dark to bright states

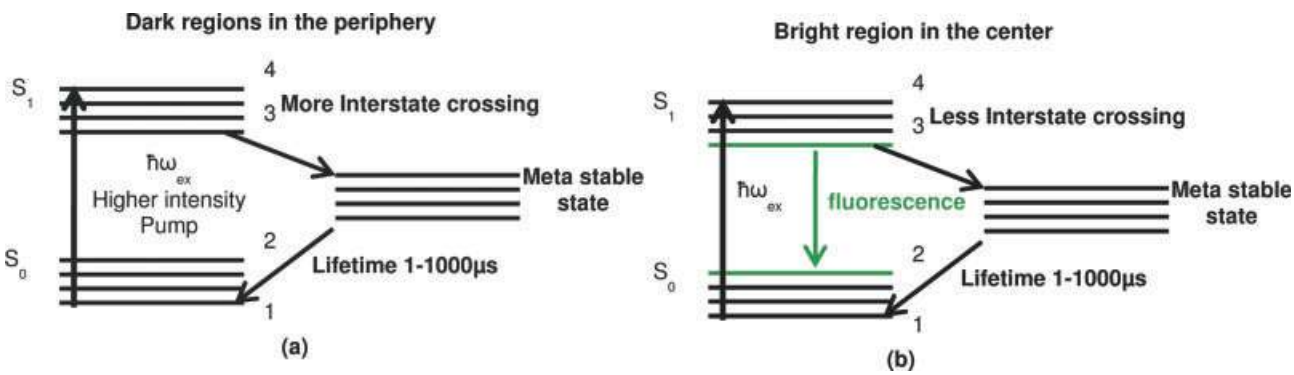

Figure 21 (online color at: www.lpr-journal.org) Transitions in Ground State Depletion (GSD) Microscopy: (a) In the periphery, long lifetime metastable states are populated by the use of a higher intensity pump beam with a doughnut shape (similar to a STED beam) represented by $\hbar\omega_{ex}$ The higher intensity ensures that a larger number of molecules are transferred to the metastable states, thus leaving lesser number of molecules to fluoresce in the peripheral regions. (b) Close to the null of the higher intensity pump beam, a lower intensity probe beam is used which means more molecules are available in this region for fluorescent emission with a lesser number of interstate crossings to the metastable state. Thus quenching of the emission along the periphery by enabling interstate crossing to the metastable states causes a reduced PSF.

with photo-switchable proteins as markers [169]. For example, light of one wavelength switches these proteins from a dark to a 'bright state' and light of a different wavelength is able to switch bright states back to dark. This switching may represent different states of isomerization of the protein.

Since GSD relies on a metastable state, it is more susceptible to photo-bleaching as molecules in the metastable state may be excited to higher energy states. To address the problem of photobleaching, metastable states were allowed to completely relax before being illuminated again. This technique can increase the stability of the fluorophore and was called Triplet Relaxation (T-REX) STED [170]. It allows usage of higher intensities and has demonstrated lateral resolutions of 15–20 nm. Besides, there has been conscious effort towards engineering more photostable fluorophores. For instance, using electron transfer (reduction or oxidation), long-lived metastable states are depopulated in Reducing and Oxidizing System (ROXS) fluorophores by the introduction of states closely coupled to the metastable states as shown in Fig. 20 [171]. Other efforts towards fluorophore engineering involve surrounding a donor fluorophore with a few acceptor fluorophores which effectively quench the donor emission in the outer regions of the confocal focus, thereby improving resolution [172]. A good review dealing with the engineering of fluorescent markers for super resolution methods is provided in [173].

The above solutions to challenges posed by STED were significantly motivated by the prospect of using STED as a tool for biological studies. These studies encompass investigations into the finer structural details of cell organelles, and imaging of living cells to understand cellular processes. For instance STED has been successfully employed to record new observations in neuroscience. For instance, insights into the dynamics of synaptic vesicles which were previously not clear were obtained through STED-based super resolution studies [174]. Synapse refers to the junction through which neurons interact to control different functions such as muscular movement. Synaptic vesicles contain neurotransmitters which are the main channel of exercising this control. The work in [174] provides conclusive evidence to decipher how the synaptic vesicles are regenerated after the process of exocytosis (process when the vesicles are divulged from the neuron). Structural details of the dendritic spines in neurons [175] and arrangements of neurotransmitters [176] during synapse, which were previously not resolvable, have been resolved by STED microscopy.

Since cell damage is quite prevalent when fluorophores photo-bleach and damage the cells [177], photoswitchable proteins with lower saturation intensities have been used. For instance, the expression of Green Fluorescent Protein (GFP) found in jellyfish in cells by genetic encoding provided an alternative to the use of fluorescence dyes [178]. STED has been demonstrated by GFP tagging to yield images of the endoplasmic reticulum of cells at 70 nm lateral resolution [179]. Furthermore, immunofluorescence studies with STED have also been performed, yielding axial resolutions of 50 nm. Such studies rely on the bonding of antigens to antibodies and super resolution could provide further insights in complex structures which cannot be resolved by standard light microscopy [180]. Another challenge lies in imaging regions of highly convoluted structures which lie farther than the scattering length from the sample surface. In such cases, the scattering of the excitation beam causes poorer than expected resolution. One solution to reduce this scattering used two photon STED microscopy employing photons of longer wavelength. This method was used to image dendritic spines of neurons located within brain slices [181].

While single super resolution images of live cells reveal structural details not resolvable by confocal microscopy [182, 183], images rendered in quick succession could be useful to the study of molecular dynamics and intracellular processes [184]. The theoretical limit to scanning speeds in STED microscopy corresponds to the time it takes for excited fluorophores to return to the ground state, which is in the order of nanoseconds. In practice, the scanning method places limits on the scanning speed. Beam scanning methods with parallel recording of information from pixels spaced at least half a wavelength apart can significantly enhance scanning speeds. Using parallel focal points, previously unresolved details of live mitochondria (power units of the cell) structure were revealed using STED microscopy with a frame refresh time of a few seconds, but millisecond recording of images is possible with more extensive parallel focusing. The study of intracellular processes with two-color imaging with red and green channels was demonstrated with STED at resolutions of 25 nm for green and 60 nm for red channels, to study protein-protein interactions [185]. Two-color STED microscopy has been used to reveal details into the interaction of proteins found in mitochondria called voltage-dependent anion selective channels (VDAC) proteins [186].

Concomitantly with STED microscopy, another class of super resolution methods based on stochastic read out have emerged, which rely on accurate localization of fluorescent emitters rather than reducing the effective size of the PSF. These methods include Photoactivation Localization Microscopy (PALM), Stochastic Optical Reconstruction Microscopy (STORM) and their variants. Contrary to STED, STORM and PALM are wide field methods where conditions are chosen such that only one or a few markers are switched on within a diffraction-limited region. Since it is not possible to determine *a priori* as to which molecule will be switched on, the process is inherently random and hence termed Stochastic Readout. The fluorescence from the excited emitter is collected and its spatial distribution is fitted with a Gaussian function to localize its position. This step is referred to as single molecule localization and its accuracy increases with the number of photons detected. However, one must note that detection of a greater number of photons entails more time and lowers frame acquisition speed. Since the tags are switched on randomly and many markers are present within a diffraction limited region, several iterations (100 to 10^5 depending on the density of tagged molecules) of single molecule localization are performed and distilled to yield a single image of the sample. In each iteration a different set of molecules are activated and localized. This operating principle of PALM/STORM is outlined in Fig. 22.

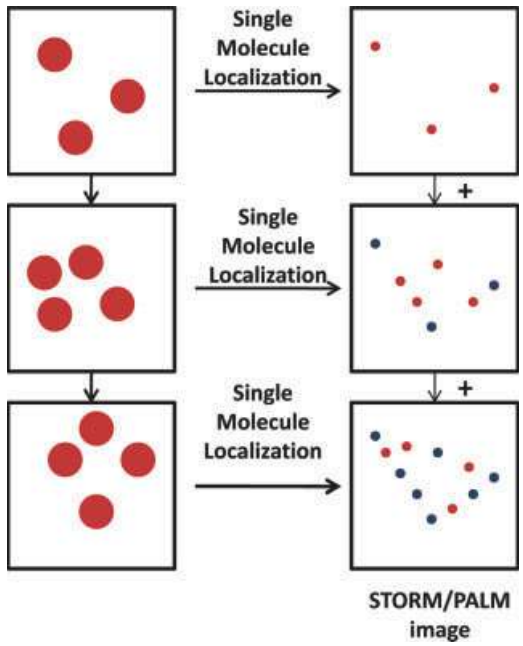

Figure 22 (online color at: www.lpr-journal.org) Formation of a PALM/STORM image: Each box in the left hand column represents a separate activation and excitation cycle. In each cycle only few molecules are switched on. Single molecule localization is then performed to identify the spatial locations of the individual emitters. The newly localized molecules (shown as little red dots in the right column) are superposed with the already localized molecules (shown as little blue dots in the right column) to result in the aggregated STORM image as shown at the bottom of the right hand column. Several activation cycles are necessary to reconstruct the entire image and the number of cycles depends greatly on the density of tagged molecules.

One main challenge in stochastic readout is to ensure that only few fluorescent tags are excited in the illuminated region. This is achieved by using a low activation intensity to switch on the tags. For instance, if N tagged molecules are expected to be present within a diffraction limited region, an intensity I_{sat}/N is used, where I_{sat} is the saturation intensity of the marker. The intensities used in stochastic readout are significantly smaller than those used in RESOLFT and are usually in the W/cm^2 range.

In addition to sparse activation, low background emission is necessary to localize molecules accurately. In PALM, this is partially achieved by using photo activated fluorescent proteins (PA-FPs) as markers. An advantage of PA-FPs is that they have almost no background emission unless they are activated by a beam of specific wavelength referred to as the activation beam. A second beam called the excitation beam at a different wavelength is then used to excite fluorescence and switch off the emitter. Accuracy of localization also depends on robust methods to fit the detected emission pattern. Several algorithms to fit the detected emission patterns, check for symmetry and distinguish between multiple peaks from closely spaced emitters and discard spurious results have been developed to assist in single molecule localization [187]. While in principle it appears that one can improve accuracy by increasing the number of emitted photons by increasing the activation intensity, in reality increasing the activation intensity indefinitely would result in the activation of greater number of fluorescent tags within the diffraction limited region, thus counteracting any accuracy improvement. Additionally, increasing the number of detected photons increases the image acquisition time. In order to ensure that a sufficient resolution is attained with a reasonable image acquisition time, several parameters, such as the intensites of the activation and excitation beam, the fluorescent tags used and data processing algorithms, would have to be optimized depending on the density of fluorescent tags.

Due to the need for low background, stochastic readout methods such as PALM and STORM predominantly use a Total Internal Reflection Fluorescence microscopy scheme (TIRF) [187, 188]. This limits the imaging region to thin samples in many cases. For example, the first demonstration of PALM for thin samples by Betzig produced images with resolutions of 25 nm with an image generation time of 2 to 12 hours for approximately one million localized molecules [188]. Independently, Hess [189] devised an almost identical method with a similar name of Fluorescence PALM (FPALM). FPALM differs from PALM in that it uses a high numerical aperture objective lens rather than a TIRF scheme which allows it to image thicker samples. STORM is governed by the same general concept as PALM and FPALM. It differs in that it uses a synthetic fluorescent tag of Cy5-Cy3 instead of PA-FPs. STORM was first demonstrated by Rust [190] using a Cy5-Cy3 dye system resulting in 20 nm resolution two dimensional images. More recently, three dimensional STORM images with a 20 nm lateral resolution and 60 nm axial resolution were demonstrated using a cylindrical lens [191, 192].

The above mentioned requisites of sparse activation and low background limit the viable candidates for use as markers. Various approaches [193] have been demonstrated to render a wider class of fluorescent tags compatible with stochastic readout techniques. For example, the issue of sparse activation has been addressed through demonstrations of multi-color STORM imaging using dyes with different emission and excitation wavelengths [194] and multicolor PALM imaging using photoactivable mCherry PA-FPs which demonstrate high photostability and high photoactivation contrast [195]. Additionally in [196] it is shown that the dyes like Cy5 can be reversibly switched for hundreds of cycles without the need for activator fluorophores such as Cy3 in the vicinity, albeit at intensities two orders of magnitude larger. This work reported 20 nm resolution images and the technique was termed Direct STORM or dSTORM. In other approaches, the concept of ground state depletion was married with stochastic read out techniques in a method known as Ground State Depletion followed by Individual Molecule return or GDSIM [197]. This method uses a single beam of light to switch on the fluorophores. However, a fraction of the activated molecules enter a dark metastable state which has a lifetime more than 10^7 times greater than the spontaneous emission lifetime [197]. Thus within a diffraction limited region, one can adjust the excitation intensity

Table 1 A comparison of targeted readout methods such as STED and GSD with stochastic readout methods such as PALM/, STORM, the pros and cons of each method give them relative advantages for different applications.

Targeted Readout	Stochastic Readout
Reduces effective PSF to subdiffraction scale	Wide field method which localizes single molecule emission within a diffracted limited region.
Requires higher intensities	Operates at lower intensities
Due to the collective emission of molecules within targeted region, fast scanning is possible and is hence suited to imaging dynamic processes.	Due to collection of many photons from a single molecule and requirement of multiple activation cycles, it is slower and more suited to tracking single molecule dynamics.
Produces the same resolution for a given set of intensities irrespective of density of fluorescent markers.	The intensity of excitation and activation beams have to be tuned to the local marker density.
More sophisticated optical setup	Simple optical setup
More suited to 3D imaging as a confocal imaging scheme is utilized and offers superior axial resolution	Predominantly employs a TIRF scheme (although 3D imaging is possible), thus limiting the thickness of samples.

to ensure that only a few emitters are switched on while the majority reside in metastable dark states. Subsequently, an image is formed after several single molecule localization steps. Similarly, stochastic read out with quantum dots as the fluorescent emitter has been demonstrated resulting in a 12 nm image resolution [198]. This method relies on the fact that quantum dots exhibit a stochastic blue shift in their emission wavelength with continuous excitation. It is hence possible for just a single molecule to fluoresce within a certain wavelength window within the diffraction limited region. Thus, by scanning different wavelength windows and performing several single molecule localization steps, the image can be reconstructed.

The requirement for several activation-localization cycles to capture a single image and the need to detect a large number of photons from a single emitter places limitations on PALM/STORM in imaging dynamic processes on a millisecond scale. However, these methods could be used for the visualization of dynamics with timescales in the minute regime. For instance, the dynamics of adhesion complexes were resolved at 60 nm by PALM in [199]. Since stochastic readout methods localize single molecule emissions, they are more suited to single molecule tracking. Recently, tracking of membrane protein trajectories at 20 frames per second has also been demonstrated using PALM and has been named single particle tracking PALM (sptPALM) [200].

In summary, microscopy techniques exploiting the reversible switching of fluorescent markers have been demonstrated using targeted and stochastic readout techniques, enabling nanoscopic imaging of living cells in both cases. Neither of these methods trumps the other on all grounds and their relative advantages are summarized in Table 1. The most attractive application of these imaging techniques lies in biological studies such as investigations into structural details of cells or mechanisms of biochemical reactions. There have been quite a few investigations in this regard. A great review of open biological questions ranging from cellular architecture to protein assembly dynamics of HIV, which necessitate super resolution techniques are provided in [184]. Since PALM/STORM benefit from being relatively inexpensive due to simple optical setups, they may be more suited to imaging quasi-static samples. However, the large number of frames and high photon count necessary to obtain a single image of the sample may make them less suited to imaging fast dynamic processes. On the contrary, STED despite its more complicated optical setup is capable of high speed 3-D imaging and could be instrumental in visualizing dynamic processes. The commercialization of STED and PALM/STORM microscopes promises to reveal the contributions of these methods in the coming years.

3. Handling of evanescent waves with near field diffraction structures

Near field diffraction structures can also be used to reduce the divergence of light beam and obtain superresolution. For example, when a plane wave is incident on to a grating structure, it can generate different orders of diffraction, including propagating fields with large diffraction angles and evanescent fields. Inversely, when light incident from a certain angle that matches a diffraction angle of the grating, part of the light can be diffracted to propagate normal to the grating direction. This also happens to the evanescent waves that match the evanescent field diffracted by the grating, i. e. evanescent waves can also be diffracted to propagating waves. The principle based on such inversed processes can be used to reduce the divergence of laser beams from subwavelength apertures, and turn evanescent waves into propagating waves for superresolution imaging. A superlens, which takes advantage of negative refraction effect, is actually an ideal diffraction system, an ideal lens that can realize near field to near field image transfer. Optical antennas can also be considered as diffraction structures, which diffract light waves to evanescent waves and part of which are coupled to surface plasmon waves. These surface plasmon waves are further localized to a superresolution light spot much smaller than the light wavelength.

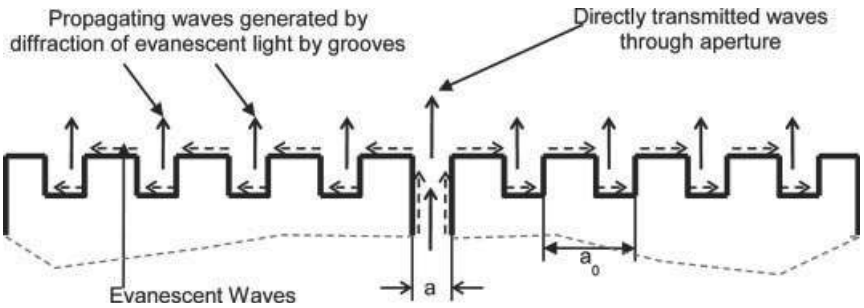

Figure 23 Evanescent waves propagate along the surface of the corrugation structnre and are diffracted to free space propagating waves, these propagating waves together with the propagating waves directly transmitted from the apertnre form a large beam spot, the large beam spot results in low beam divergence.

3.1. Reduction of beam divergence with near field diffraction structures

Light from subwavelength apertures suffers severe divergence [201–220]. To reduce the divergence of the beam from a subwavelength aperture, for instance a single mode laser diode, a multimode waveguide can be used to effectively reduce the beam divergence and realize focusing with a short free working distance [202]. Such a multimode waveguide can also serve as a micron-sized lens [203]. Besides beam divergence, the transmission of the aperture is also very low: for an aperture with radius of r in a perfect conductor, its transmission is proportional to $(r/\lambda)^4$ [201], where λ is the wavelength of the incident light. It is clear that a shorter wavelength will result in higher transmission. The problem of low transmission can be resolved through fabricating concentric periodic corrugation structures surrounding the subwavelength aperture, which can couple the incident propagating waves into evanescent waves [204–211]. These evanescent waves have much shorter transversal wavelength, and therefore, the transmission is enhanced. The relationship between the incident waves and the evanescent waves is shown in Eq. (31), where k_e represents the transversal wave number of evanescent waves, k_{incident} is the transversal wave number of the incident waves, a is the period of the corrugation and m is the order of diffraction by the periodic corrugation structure:

$$k_e = k_{\text{incident}} + m\frac{2\pi}{a}. \tag{31}$$

An optimum period for corrugations is found to be between 0.8λ and 0.95λ, regardless of the size of the diffraction aperture. When metals are used as the corrugation structure, surface plasmon waves will also be excited, corresponding to part of the evanescent waves. The wave number of surface plasmon waves is described in Eq. (32), where ε_m and ε_d are the permittivities of the metal and its immediate dielectric material:

$$k_{\text{sp}} = \frac{2\pi}{\lambda}(\varepsilon_m\varepsilon_d)^{1/2}(\varepsilon_m+\varepsilon_d)^{-1/2}. \tag{32}$$

The behavior of such a concentric periodic structure is like that of a lens, which can condense light and can also act as a beam collimator. The collimation capability of such a periodic structure has been explored by fabricating it on the exit side of a subwavelength aperture [211–214], which can effectively reduce the divergence of light.

The principle of reducing near-field beam divergence is schematically described in Fig. 23. The propagating light waves consist of two parts: one part is that directly transmitted from the aperture, and the other part is that converted from evanescent waves, they together form a larger beam spot in the near field [211–214]. As is discussed in Sect. 1, a larger beam spot will result in a smaller divergence angle. Corrugation structures can effectively enhance light emission and reduce the beam divergence from laser diodes [215–218], and based on the same principle, the emission direction from a photonic crystal can be very well controlled by making the lattice constant slightly smaller than half of the light wavelength [219, 220].

3.2. Superresolution using near-field diffraction structures

Superresolution can be achieved through generating a light spot with a size smaller than that determined by far field diffraction limit, i. e., $\lambda/2$. Using such a smaller light spot as a light source to illuminate a sample, only the illuminated area of the sample is seen by a detector, and therefore, the spot size actually determines the resolving power of the system, one efficient way of generating a superresolution light spot is through optical antennas. Another way to obtain superresolution is through increasing the bandwidth of an imaging system, which is limited by the existence of evanescent waves. To extend this band width, one efficient way is to turn evanescent waves into propagating waves.

3.2.1. Turning evanescent waves into propagating waves

Evanescent waves carry information of the high spatial frequency features of an object. However, they attenuate exponentially away from the object surface, which makes it difficult to achieve high resolution imaging from the far field. A superlens, which is expected to use a block of material with negative refractive index ($n < 0$), can theoretically achieve arbitrary high resolution. In practice, the original idea is applicable to the transformation from near field to near field [221–261]. The recent far field superlens concept introduces a diffraction structure to the original superlens, which is able to turn evanescent waves into propagating waves, and makes possible far field nano-imaging [262–266]. The schematic principle of turning evanescent waves into propagating waves is shown in Fig. 24. Light with wavelength of

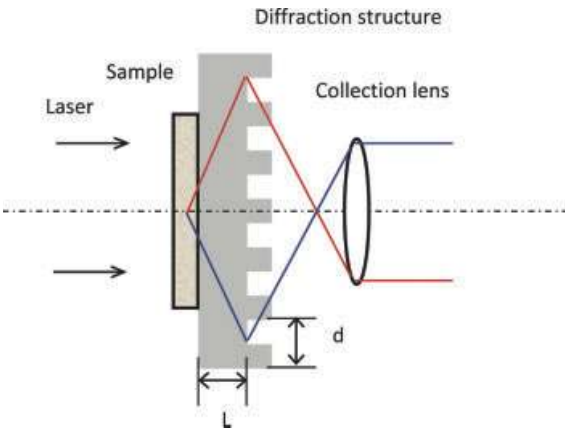

Figure 24 (online color at: www.lpr-journal.org) Schematic for turning evanescent waves into propagating waves using diffraction structure, laser is incident onto a sample, generating evanescent waves with wave vector $k_e \gg 2\pi/\lambda$. Such evanescent waves propagate a distance L and are then diffracted by a corrugation with period of d, the diffracted light is collected by a lens.

λ is incident on to a sample, generating evanescent waves with wave vector $k_e \gg 2\pi/\lambda$. Such evanescent waves propagate a distance L and are then diffracted by a corrugation with period of d. If the material filling the space between the sample and the corrugation is a superlens material of silver, this evanescent wave can be enhanced, with optimum distance L of about 35 nm [262]. If it is a non-superlens material, the distance is required to be as small as possible, because evanescent waves attenuate during propagation. The diffracted beam propagating in free space has a wave number that satisfies

$$k_p = k_e + 2\pi m/d, \tag{33}$$

where m is the diffraction order, which has the inverse sign to k_e , $-2\pi/\lambda < k_p < 2\pi/\lambda$. Then for arbitrary evanescent wave number k_e, we can always find the proper m and d to turn it into a propagating wave. Now the resolving power of the system is decided by the diffraction efficiency and period of the corrugation structure.

A hyperlens is another solution for turning evanescent waves into propagating waves [267–282]. In a cylinder consisting of isotropic media, as shown in Fig. 25a, the wave vector along the radial direction k_r and that along the tangential direction k_θ satisfy the relation

$$k_r^2 + k_\theta^2 = \varepsilon\omega^2/c^2, \tag{34}$$

where $k_\theta r = m$, and m is an integer, representing the order of a mode that exists in the cylinder. As r decreases, k_θ increases and k_r decreases. However, when $k_\theta > \varepsilon^{1/2}\omega/c, k_r$ becomes imaginary, which means that the electromagnetic field is exponentially decreasing towards the center, as can be seen from Fig. 25b, for the 20th mode. When the radius is smaller than a certain value, the intensity becomes extremely low. However, the situation can be changed by using an anisotropic material as shown in Fig. 25c, which consists

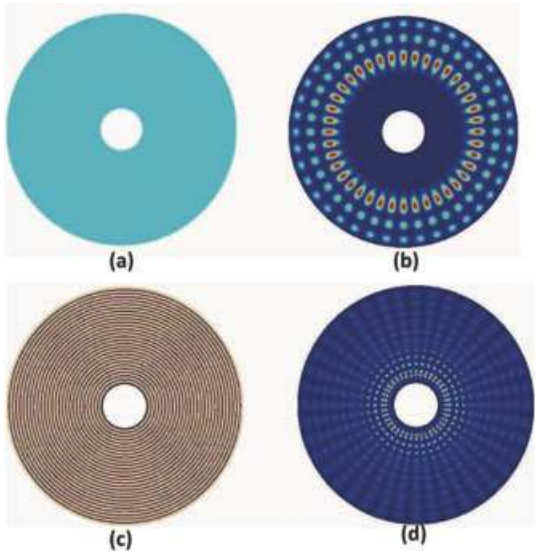

Figure 25 (online color at: www.lpr-journal.org) (a) Top view of a hollow core cylinder with inner radius of $r = \lambda$ made from uniform dielectric $\varepsilon = 1.5$ (average of ε_m and ε_d). (b) Calculated light intensity for the $m = 20$ angular momentum state, in false color representation where red denotes high intensity and blne corresponds to low intensity. (c) The hyperlens made of 50 alternating layers of metal (dark regions) with $\varepsilon_m = -2$ and dielectric (grey regions) with $\varepsilon_d = 5$. The outer radius is 2.2 µm and the inner radius is 250 nm. (d) Corresponding intensity for the $m = 20$ mode of (c). (Reproduced from [267] with permission from the Optical Society of America.)

of concentric metallic layers alternating with dielectric layers. When the layer thickness is much smaller than the light wavelength, this material can be treated as an effective medium with $\varepsilon_\theta = (\varepsilon_m + \varepsilon_d)/2$ and $\varepsilon_r = 2\varepsilon_m\varepsilon_d/(\varepsilon_m + \varepsilon_d)$, where ε_m and ε_d denote the dielectric permittivities of the metal and dielectric layers, respectively. Suitable selection of ε_m and ε_d can make $\varepsilon_\theta > 0$ and $\varepsilon_r < 0$. The radial and tangential wave vectors of TM modes satisfy the hyperbolic dispersion relation

$$k_r^2/\varepsilon_\theta - k_\theta^2/\varepsilon_r = \omega^2/c^2, \tag{35}$$

which allows for both k_r and k_θ to increase towards the center of the cylinder, and therefore, the light field does not decrease towards the center of the cylinder, i. e. a light field with arbitrary high tangential wave vector (evanescent wave) can propagate out from the center, and its wave vector decreases during this propagation until it comes to free space with a wave number much smaller than $2\pi/\lambda$. However, the dielectric permittivities are ill-defined at the center and a close approximation can only be realized when $r \geqslant \lambda$, as shown in Fig. 25d, which shows the 20th mode intensity distribution in a hollow core anisotropic cylinder made of 50 alternating layers of metal (dark regions) with $\varepsilon_m = -2$ and dielectric (grey region) with $\varepsilon_m = 5$. This hyperlens allows a highest tangential wave vector of $k_\theta = m/\lambda$, the 20th mode corresponding to $k_\theta = 20/\lambda$, which has a maximum resolving power of about $\lambda/6.4$. The process of turning

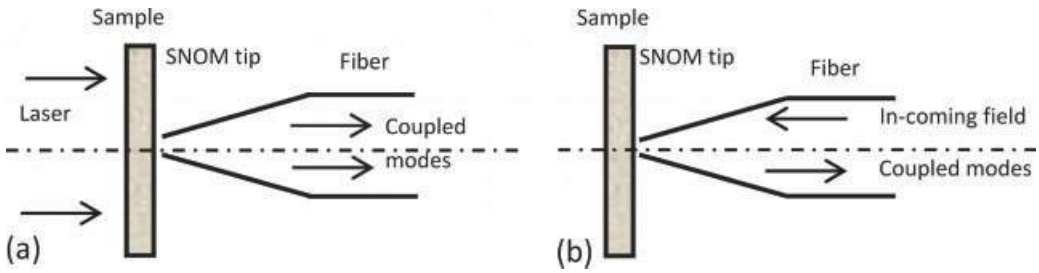

Figure 26 (online color at: www.lpr-journal.org) Schematic diagram of turning evanescent waves into propagating waves using SNOM. (a) Transmission mode, light is incident onto the sample, it generates propagating waves and evanescent waves. Evanescent waves can couple into the SNOM tip through the open aperture and become propagating field. (b) Reflection mode, light comes from the SNOM tip and is incident onto the sample, the field reflected from the sample is coupled back into the SNOM tip and become propagating waves.

evanescent waves into propagating waves is invertible, i. e. propagating waves can also be turned into evanescent waves using this device [283], and 3D planar imaging capability is highly expected from such a hyperlens [284, 285].

In fact, all scanning near-field optical microscopes (SNOMs) [286–291] and apertureless SNOMs [292–296] realize superresolution through turning evanescent waves into propagating waves, where the conversion is through the diffraction of the tiny aperture or the tip. The schematic principle of SNOM is shown in Fig. 26 and can operate in transmission mode and reflection mode, as is shown in Fig. 26a and Fig. 26b. For the transmission mode, when the light is incident on to the sample, it generates propagating waves and evanescent waves. Both the transmission field and evanescent waves can couple into the SNOM tip through the open aperture. Suppose the aperture has a diameter d, then the maximum wave vector generated by this aperture is $2.44\pi/d$, the wave vector of the field propagating inside the SNOM can be approximated as

$$k_p = k_e - 2.44\pi/d \tag{36}$$

where k_p is the guided mode of the fiber, its wave vector range depends on the supported mode of the fiber. The resolution limit of the SNOM is determined by the size of the tip aperture, as $k_e = k_p + 2.44\pi/d$. For the reflection mode, the field coming from the fiber is turned into evanescent waves at the SNOM tip, the maximum wave vector of the evanescent waves being $2.44\pi/d$ in this case, and the wave vector of the field propagating inside the SNOM can be approximated as

$$k_p = k_e - 4.88\pi/d. \tag{37}$$

Therefore the maximum resolution of light is determined by $k_e = k_p + 4.88\pi/d$, which is about twice the resolution of that of the transmission mode. A smaller diameter of aperture will result in higher resolution, but this will lead to low optical efficiency, especially when it is smaller than the size that can allow a single mode to go through. This actually limits the resolution of an aperture-type SNOM.

Apertureless SNOM, which takes advantage of the scattering between its sharp tip and the sample under investigation, is not subject to the transmission issues that are encountered in a normal SNOM, and therefore higher resolution can be achieved, because there is no optical constraint to the size of its tip. The principle of turning evanescent waves into propagating waves for apertureless SNOM is similar to that of a reflection mode SNOM, the difference being that the strong localized light near the tip is achieved through surface plasmon resonance [292–296].

A solid immersion lens (SIL) system is another solution to turning evanescent waves into propagating waves [297, 298]. This is because the material of the SIL has a refractive index n larger than 1.0, which allows the propagation of light with a wave number n times of that in free space, i. e. $2\pi n/\lambda$. As is shown in Fig. 27a, light is incident on to the sample, which generates evanescent waves propagating along the sample surface (red line). Part of the evanescent waves couple to the SIL lens and propagate inside the lens with a wave number $2\pi n \sin\theta/\lambda$ (the blue line, θ is the maximum far field collection angle). When it crosses the spherical interface, its wave number is reduced to $2\pi \sin\theta/\lambda$, and now the originally evanescent wave is changed to a free space propagating wave. Detecting this wave from the far field can achieve higher resolution. The resolution limit of transmission mode SIL lens is $\lambda/(n\sin\theta)$, where θ is the far-field collection angle. When the reflection mode is used, i. e. light is first focused into the SIL lens on to the sample and then collected by the lens, this is actually working in a confocal mode, as is shown in Fig. 27b. Light is first focused into the SIL lens, when a light ray at angle θ crosses the spherical interface, its wave number increasing n times from $2\pi \sin\theta/\lambda$ to $2\pi n \sin\theta/\lambda$. The light ray further propagates inside the SIL lens to the flat side of the SIL lens, becoming an evanescent wave between the gap of the SIL lens and the sample. It interacts with the sample, couples back to the SIL lens and then propagates out to free space. This confocal setup will have higher resolution of $\lambda/(2n\sin\theta)$ [298]. When the SIL lens works in a confocal scanning mode, its resolution increases to between $\lambda/(3n\sin\theta)$ and $\lambda/(4n\sin\theta)$ [297, 299–304].

In conclusion, evanescent waves can be turned into propagating waves using a corrugated structure, a hyperlens, a scanning near field optical microscope or a solid immersion lens. The process of converting evanescent waves to propagating waves, and propagating waves to evanescent waves is associated with most near-field nano-imaging, light beaming, and optical antennas, as will be discussed in following sections.

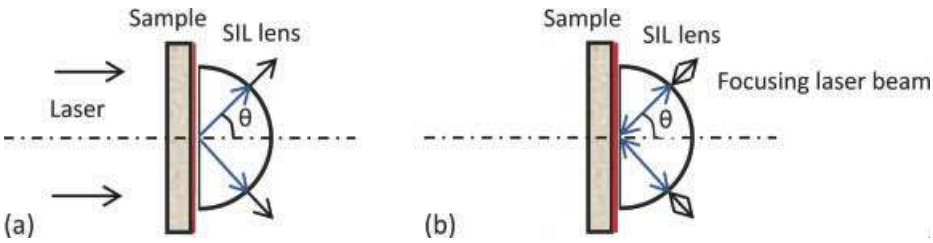

Figure 27 (online color at: www.lpr-journal.org) Schematic diagram of turning evanescent waves into propagating waves using a SIL lens. (a) Transmission mode, light is incident on to the sample, which generates evanescent waves propagating along the sample surface (red line). Part of the evanescent waves couple to the SIL lens and propagate inside the lens with a wave number $2\pi n \sin\theta/\lambda$ (the blue line, θ is the maximum far field collection angle). When it crosses the spherical interface, its wave number is reduced to $2\pi \sin\theta/\lambda$, and now the originally evanescent wave is changed to a free space propagating wave. (b) Reflection mode, light is first focused into the SIL lens, when a light ray at angle θ crosses the spherical interface, its wave number increasing n times from $2\pi \sin\theta/\lambda$ to $2\pi n \sin\theta/\lambda$. The light ray further propagates inside the SIL lens to the flat side of the SIL lens, becoming an evanescent wave between the gap of the SIL lens and the sample. It interacts with the sample, couples back to the SIL lens and then propagates out to free space.

3.2.2. Superlens

In the past decade, there has been much interests and research activities in the realization of negative refractive-index materials (NIMs), due to their unusual negative refraction effect, which enables the realization of a "perfect lens" [221, 231]. A perfect lens can image an object without the loss of any spatial resolution [231]. A NIM has negative refractive index ($n < 0$) for the propagation of electromagnetic waves, and is thus drastically different from normal materials which have positive refractive indices. Such a material does not exist in nature. It can be shown that the refractive index n of a material effectively becomes negative ($n < 0$) when its electric permittivity ε and magnetic permeability μ are both negative ($\varepsilon < 0$ and $\mu < 0$) [221]. NIM can only be realized through the fabrication of nanostructures embedded in suitable materials. Such materials are called metamaterials [221, 226, 231, 305–313]. In practice, an ideal perfect lens cannot be realized due to loss and dispersion, but one can in principle approach a near-perfect lens by engineering the right metamaterial [314]. A lens with a perfect lens-like property is also often called a sub-diffraction-limited super-focusing lens (superlens) [242]. While NIM can be used to realize negative refraction and a perfect lens, it is not the only way. Alternatively, negative refraction and a near-perfect lens may be realized by exciting a surface plasmon resonance in metal films, which can result in a negative permittivity ε (but μ is still positive), called a Negative Dielectric Permittivity Material (NDM) [242]. However, NDM can only show these effects for the TM (or p) polarization. Likewise, a Negative Magnetic Permeability Material (NMM) with negative μ but positive ε could be used to realize negative refraction and a near perfect lens for the TE (or s) polarization.

A NIM will focus light even when it is in the form of a flat slab with parallel surfaces. This is because the light inside an $n < 0$ medium makes a negative angle with the surface normal and undergoes "negative refraction" (see Fig. 28) [242, 307, 311–313]. As shown in Fig. 29, this enables an object in air to form a first image inside a flat slab of NIM material without the need for any surface curvature

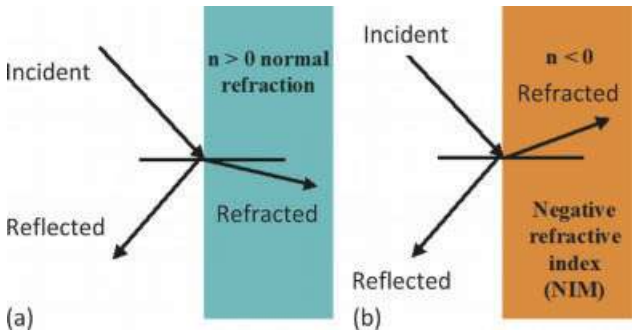

Figure 28 (online color at: www.lpr-journal.org) (a) Normal refraction case, refracted light ray makes a positive angle with the surface normal; (b) Refraction for NIM, refracted light ray makes a negative angle with the surface normal.

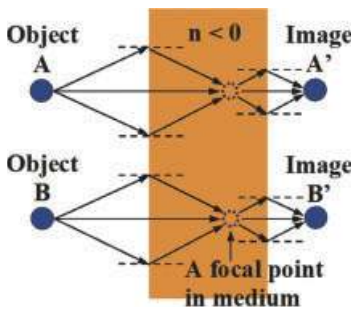

Figure 29 (online color at: www.lpr-journal.org) Perfect focusing for an object through a material with $n < 0$, an object at point A in air form a first image inside the flat slab of NIM material, the image further propagates to the air interface at the other side of the slab, the beam bends inwards when it goes from the material back into air and forms an image at point A'. As the object moves "downwards" from point A to point B, the image also moves downwards from point A' to point B', which is in the same direction as the object movement, and the magnification is always unity.

at the air-material interface. The focused beam further propagates to the air interface at the other side of the slab. Again, at this interface, the beam bends inwards when it goes from

the material back into air. This enables a second image to be formed at the other side of the NIM "flat lens" [221].

Even more surprisingly, a NIM medium can cancel the decay of evanescent waves which carry the high spatial frequency information from an object [231]. As these evanescent waves are lost in normal imaging, a normal lens is limited to an imaging resolution that is limited to at best half a wavelength of light, often referred to as diffraction-limited resolution. If these evanescent waves could somehow be amplified and recovered then they could contribute to the formation of image and remove the image resolution limitation. This means that an entire object surface can be imaged on the other side of the NIM slab with perfect resolution as pointed out by Pendry [231]. Also when $\varepsilon = \mu = -1$, the optical reflection at the lens-air interface becomes zero [231]. One interesting property of the NIM lens is that as the object moves "downwards" from point A to point B in Fig. 29, the image also moves downwards from point A′ to point B′, which is in the same direction as the object movement, and the magnification is always unity. The "geometrical optics" of a perfect lens for which the lens has an effective refractive index of "$-n$" and the surrounding material has refractive index of "$+n$" is shown in Fig. 30. In this case, ray tracing shows that the distance from object to lens front surface (L_{obj1}) is exactly equal to the first focal length ($L_{foc1} = L_{\text{obj1}}$) and the distance from the 1st focal length to the back surface (2nd "object" distance) (L_{obj2}) is equal to the 2nd focal length (L_{foc2}) [221, 231].

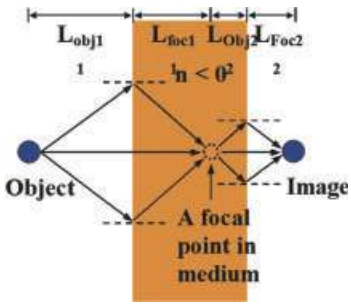

Figure 30 (online color at: www.lpr-journal.org) Perfect lens or superlens optics: distance from object to lens front surface (L_{obj1}) is exactly equal to the first focal length ($L_{\text{foc1}} = L_{\text{obj1}}$) and the distance from the 1st focal length to the back surface (2nd "object" distance) (L_{obj1}) is equal to the 2nd focal length ($L_{\text{foc2}} = L_{\text{obj2}}$).

As mentioned above, this perfect-lens effect can also be achieved in NDMs and NMMs, but only for one of the two polarizations [231–233, 236, 237, 242, 315–318]. In the case of NDM, only the TM (or p) polarization will form the focused image and thus the imaging power efficiency will be lower. A NDM can be realized in practice using metals, as the real part of the dielectric constant of metals can be negative. In a causal medium (medium in which wave propagation obeys causality) with atomic resonances, the permittivity, permeability, and refractive index of the medium must vary with wavelength and cannot be a wavelength-independent constant. This is referred to as optical dispersion. In an optically dispersive medium, the imaginary part of the complex permittivity and permeability cannot be zero due to causality. This means that in any realistic scheme to realize a NIM, NDM, or NMM, there must be optical loss and dispersion. The optical loss and dispersion in a NIM, NDM or NMM realized with use of realistic materials will smear out the image sharpness but sub-wavelength resolution is still achievable [226, 232, 233, 236, 237, 262, 315–323].

Thus, a "perfect lens" based on NIM, MDM, or NMM realized with realistic materials will not be perfect, but can still achieve sub-diffraction-limited super-focusing, and such lenses are referred to as superlenses [242]. Such a super-focusing effect has been demonstrated experimentally with use of a thin film of silver at its surface plasmon resonance (SPR) frequency. Working at or near the SPR frequency is essential for propagating and amplifying the normally-evanescent waves [232, 233, 236, 237, 242, 315–318]. The demonstrated structure involved a pre-etched chrome (Cr) pattern with tens of nanometers feature size on quartz as the nano-object. The object is imaged through a 35 nm thick silver NDM lens with sub-diffraction-limited resolution [242].

Physics of NIMs

To understand the basic physics involved in a superlens based on NIM/NDM/NMM, first let us write down the main two Maxwell Equations in a dielectric or magnetic material, which are given by:

$$\nabla \times \bar{e}(r,t) = -\mu \frac{\partial \bar{h}}{\partial t}, \tag{38}$$

$$\nabla \times \bar{h}(r,t) = \varepsilon \frac{\partial \bar{e}}{\partial t}. \tag{39}$$

Let us assume monochromatic light. For ease of calculation, we employ the complex representation for the fields so that:

$$\bar{\mathbf{e}}(r,t) = \mathrm{Re}\left[\overline{E_A}(\bar{r})e^{-i\omega.t}\right], \tag{40}$$

$$\bar{\mathbf{h}}(r,t) = \mathrm{Re}\left[\overline{H_A}(\bar{r}).e^{-i\omega.t}\right]. \tag{41}$$

Let us further assume plane waves so that:

$$\overline{E_A}(\bar{r}) = \overline{E_A}e^{i\overline{K_{\mathrm{eA}}}.\bar{r}}, \tag{42}$$

$$\overline{H_A}(\bar{r}) = \overline{H_A}e^{i\overline{K_{\mathrm{hA}}}.\bar{r}}, \tag{43}$$

where $\overline{K_{\mathrm{eA}}}$ and $\overline{K_{\mathrm{hA}}}$ are the **K**-vectors for the electric field and magnetic field respectively. Assuming monochromatic plane waves, the Maxwell equations can then be transformed to:

$$\overline{K_{\mathrm{eA}}} \times \overline{E_A} = \mu\omega\overline{H_A}, \tag{44}$$

$$\bar{K}_{\mathrm{hA}} \times \bar{H}_A = -\varepsilon.\omega.\bar{E}_A. \tag{45}$$

Assuming a plane wave propagating across a material boundary from a medium 1 with permittivity ε_1 and permeability μ_1 to a medium 2 with permittivity ε_2 and permeability μ_2, the following boundary conditions can be obtained from Maxwell's equations:

$$E_{1t} = E_{2t}; \tag{46}$$

$$H_{1t} = H_{2t}\,; \tag{47}$$

$$\varepsilon_1 E_{1n} = \varepsilon_2 E_{2n}\,; \tag{48}$$

$$\mu_1 H_{1n} = \mu_2 H_{2n}\,. \tag{49}$$

The complex Poynting vector and the time-averaged (averaged over one optical cycle) real Poynting vector are given by:

$$\bar{S}(\bar{r}) = \bar{E}(\bar{r}) \times \bar{H}(\bar{r})\,, \qquad \langle s(\bar{r},t)\rangle = \frac{1}{2}\left[\overline{S(\bar{r})}\right]. \tag{50}$$

Since the Poynting vector is not dependent on the sign of either ε or μ, we see that in a NIM, the Poynting vector allows the usual right-hand rule with direction of energy flow given by $E \times H$. The same is true for a NDM or NMM. From Eq. (45), one can see that if $\varepsilon < 0$, the **K**-vector $\overline{K_{\mathrm{hA}}}$ for the magnetic field in a NIM will be pointing in a direction opposite from that of the Poynting vector (this direction of **K** makes sense for NIM only as for NDM and NMM, $K^2 = \mu\varepsilon\omega^2$ means that the magnitude of the **K**-vector given by **K** becomes complex as only ε or μ is negative, and **E** and **H** are 90 degree out of phase). Thus, the phase of the field will appear to flow in a direction opposite to the direction of energy propagation. If $\mu < 0$, then the **K**-vector $\overline{K_{\mathrm{eA}}}$ for the electric field will be pointing in a direction opposite to that of the Poynting vector. Detailed investigations of wave packets propagating into a NIM can be found in [319, 320].

We can see from Eq. (48) that for a boundary with NIM or NDM, if $\varepsilon_1 > 0$ and $\varepsilon_2 < 0$, the normal components of the electric fields across the boundary will have opposite signs. Likewise, the normal components of the magnetic fields across a boundary with NIM or NMM will also change sign via Eq. (49). A representative figure of a monochromatic plane wave impinging on a material interface from medium 1 to medium 2 is shown in Fig. 31 for the case of a transverse magnetic (TM or p polarized) wave, where subscripts "I", "R", and "T", label the incident, reflected, and transmitted fields of their **K**-vectors, respectively. The solid line in the upper half of the figure denotes the transmitted fields for the case of a normal ($\varepsilon > 0$) medium and the dotted line denotes the transmitted fields for the case of a NIM ($\varepsilon < 0$). The directions of the respective **K**-vectors of the magnetic fields are also shown in Fig. 31 for the case of NIM with propagating **K**. From the figure, one can see that as the normal component of the electric field in medium 2 has to change sign from that in medium 1 due to Eq. (48), negative refraction results. The same arguments goes for the case of a Transverse Magnetic (TM or s polarized) wave impinging on a material interface with NIM or NMM. The NIM will give negative refraction for both TE and TM polarized waves but NDM and NMM will do so only for TM and TE (or p and s) polarized wave, respectively. As Pendry pointed out, an ideal NIM slab has no reflection at the air-NIM interface due to impedance matching with air ($Z = \sqrt{\mu/\varepsilon}$), which is another interest property of superlens based on NIM, NDM, or DMM [231].

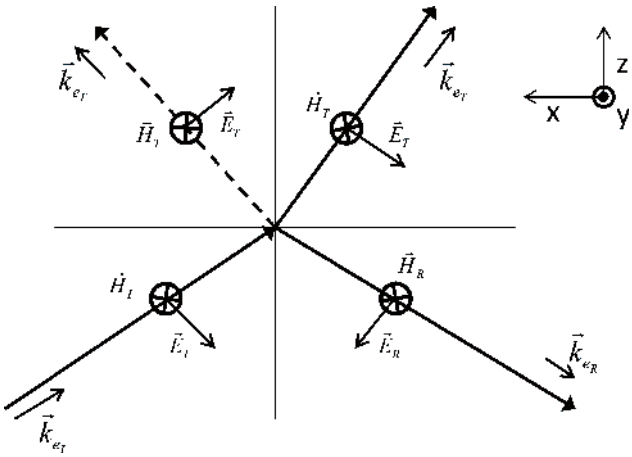

Figure 31 Representative figure of a monochromatic plane wave impinging on a material interface from medium 1 to medium 2 for the case of a transverse magnetic (TM or p polarized) wave, where subscripts "I", "R", and "T", label the incident, reflected, and transmitted fields of their **K**-vectors, respectively. The solid line in the upper half of the figure denotes the transmitted fields for the case of a normal ($\varepsilon > 0$) medium and the dotted line denotes the transmitted fields for the case of a NIM ($\varepsilon < 0$).

Super resolution via superlens

The optical imaging problem can be thought of as propagating an image on a 2D plane through a space of distance L to be reconstructed on another 2D plane as illustrated in Fig. 32. This image propagation problem can be largely understood by assuming scalar waves and ignoring the vectorial property of electromagnetic fields. In the figure, we denote the two dimensional spatial coordinates of the input plane by $\overline{q_1} = (x_1, y_1)$ and that of the output plane by $\overline{q_2} = (x_2, y_2)$. The monochromatic complex scalar field is denoted by $\varphi(\bar{r})$ so the field at the input plane is given by $\varphi_I(\overline{q_1}) = \varphi(x_1, y_1, z = 0)$ and at the output plane is given by $\varphi_O(\overline{q_2}) = \varphi(x_2, y_2, z = L)$.

The field in between $z = 0$ and $z = L$ obeys the wave equation

$$\nabla^2\varphi(\bar{r}) + K^2\varphi(\bar{r}) = 0\,, \tag{51}$$

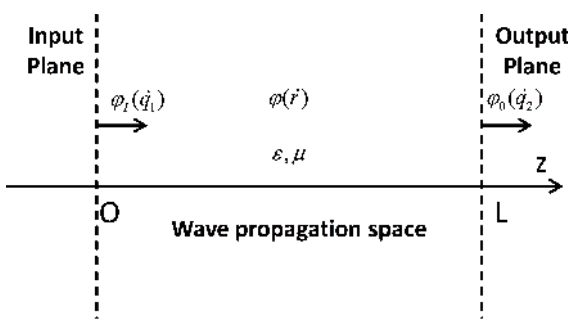

Figure 32 The general geometry of imaging problem showing the propagation of an image on a 2D plane at $z = 0$ through a space of distance L to be reconstructed on another 2D plane at $z = L$.

where

$$K^2 = \mu\varepsilon\left(\frac{\omega}{c}\right)^2. \tag{52}$$

Note the dependence of K^2 on the medium's μ and ε. The field transformation acts like a linear system due to the linear superposition property of Maxwell's equations. Thus, we can employ the powerful Fourier decomposition technique by deposing the propagating wave in terms of plane waves such that:

$$\begin{aligned}\varphi_I(\overline{q_1}) &= \int d\bar{f}\tilde{\varphi}_I\left(\bar{f}\right)e^{+i2\pi\bar{f}.\overline{q_1}}\\ &= \iint df_x df_y\tilde{\varphi}_I\left(\bar{f}\right)e^{i2\pi\bar{f}.\overline{q_1}},\end{aligned} \tag{53}$$

where the tilde "$\sim$ " denotes the field in the spatial frequency domain, and $\bar{f} = (f_x, f_y)$ denotes the two-dimensional spatial frequency in the transverse plane perpendicular to the direction of propagation. The spatial frequencies in this transverse plane carry the spatial features on the object involved. When each of the plane-wave components $e^{i2\pi(f_x.x_1+f_y.y_1)}$ in Eq. (54) is propagated by a distance L, it becomes $e^{i2\pi(f_x.x_1+f_y.y_1)+i2K_ZL}$, where

$$\begin{aligned}K_z &= \sqrt{K^2 - K_x^2 - K_y^2}\\ &= \sqrt{K^2 - (2\pi.f_x)^2 - (2\pi.f_y)^2}\\ &= K\sqrt{1 - \left|\lambda\bar{f}\right|^2},\end{aligned} \tag{54}$$

for which $\lambda = 1/K$ is the wavelength in the medium. By linear superposition of solutions, the output field at $z = L$ can be obtained, giving:

$$\varphi_O(\bar{q}_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} df_x df_y\tilde{\varphi}_I(\bar{f})e^{i2\pi(f_xx_1+f_yy_1)+k_zL}. \tag{55}$$

At high spatial frequency $|\lambda\bar{f}| > 1$, k_z becomes complex so that $k_z = i|k_z|$ and these high spatial frequency wave components will vary in z as e^{-K_zz} and will decay rapidly. Thus, they will hardly propagate to the output plane at $z = L$ when the propagation distance L is large. This means that the image at the output plane will not be able to reproduce spatial features smaller than approximately half the wavelength. While all these high spatial frequency components exist in the near field, they decay away in the far field when $|K_z|L \gg 1$. Only low spatial frequencies with $\left|\lambda\bar{f}\right| < 1$ are propagating waves that can propagate to the far field, representing the propagation of image from the object with spatial feature size larger than about half a wavelength. The near field is the region where $|K_z|L < 1$, at which the non-propagating waves have not decayed. Now $|K_z|L > KL$ according to Eq. (54) for those non-propagating high frequency waves where $\left|\lambda\bar{f}\right| >> 1$. Thus in near field $KL < |K_z|L < 1$ or $L \ll \lambda/(2\pi)$. Thus, the near field is effectively a region within a fraction of a wavelength from the object, beyond which the high spatial frequency waves will decay away and cannot be recovered under normal imaging situations. This is the basic reason of why a regular microscope can only image with a resolution no smaller than $\lambda/2$, and the resolution is said to be diffraction limited or rather "high-spatial-frequency-decaying limited". The high-spatial-frequency components that have decayed away simply cannot be recovered at the image plane. Pendry showed that these decaying waves can be amplified back to their original amplitudes at the output plane when propagated through a NIM, thereby resulting in perfect image recovery at the output plane. For an ideal lossless NIM, the object (and image) can in principle be far away from the surface of the NIM as long as the NIM has a lateral extent that is much larger than the object distance so that all the propagating or decaying **K**-vectors propagating at small or large angles can reach the NIM. In practice, the loss in a realistic NIM makes the NIM unable to give too large an amplification, so the object and image still have to be near the NIM but still further than the usual near-field distance. For NDM or NMM, only half of the energy with the right polarization will achieve the super focusing effect [231].

For NIM, Pendry pointed out that the electric fields at the air-NIM boundary have the following field transmission coefficient for a TE (s polarized) field [231]:

$$E_2 = tE_1, \tag{56}$$

$$t = \frac{2\mu K_{1z}}{\mu K_{1z} + K_{2z}}, \tag{57}$$

$$r = \frac{\mu K_{1z} - K_{2z}}{\mu K_{1z} + K_{2z}}, \tag{58}$$

where E_1 is the field in air and E_2 is in the NIM, t is the field transmission coefficient at the interface, and K_{1z} is the z-component of the **K**-vector in air and K_{2z} is the z-component of the **K**-vector in the medium. For the high spatial frequency components these **K**-vectors are given by:

$$K_{1z} = i\sqrt{(2\pi f_x)^2 + (2\pi f_y)^2 - (\omega/c)^2}, \tag{59}$$

$$K_{2z} = i\sqrt{(2\pi f_x)^2 + (2\pi f_y)^2 - \mu\varepsilon(\omega/c)^2}. \tag{60}$$

Since $\mu = -1$ and $\varepsilon = -1$, we have $\mu\varepsilon = 1$ and $K_{1z} = K_{2z}$, which means the waves are all decaying waves varying as e^{-K_1z} both in the air and in the NIM medium. Where then is the origin of the field amplification? If we look at Eq. (57) and (59), we will find that the field transmission coefficient "t " becomes infinitely large when $\mu = -1$. The field reflection coefficient "r " will also become infinitely large (see [231]). The same thing occurs at the NIM-to-air interface when the waves exit the NIM medium. It is these unusual behaviors that cause amplification of the decaying waves. The total transmission through the two interfaces can be calculated by adding up the repeatedly reflected fields between the two interfaces. It turns out that the total transmission is finite and results in a net amplification for the decaying waves that enables the decaying waves to exactly recover their amplitudes at the original object plane when

they are propagated to the image plane. These infinite transmission and reflection coefficients at the air-NIM or NIM-air interfaces are obviously an idealization due to the lossless assumption of the NIM. When dispersion and loss are included, these transmission coefficients can still be large though no longer infinite. More detailed investigation link these high fields at the boundary to resonant surface waves. For the case of NDM realized with a metal thin film, these resonant surface waves correspond to the familiar resonant plasmon excitations. These resonances help to build up the high field strengths. In the case of an ideal lossless NIM, the two interfaces can support an infinite number of resonances all degenerate at the frequency ω, which can give an infinite response when stimulated with a small field. More details of these processes can be found in [318, 321–323].

Other interests

Other interests in superlenses include molding light using a muti-layer superlens structure. This enables transformation of the decaying field to a propagating field essentially by magnifying the image via a multi-layer structure so that it is no longer diffraction limited, resulting in what is referred to as hyperlens [262, 269, 324–327]. Another related work is the interest in realizing "flat lenses", designed via transformation optics, that are also capable of image magnification of a sub-diffraction-limited object [262, 269, 324–327]. There is also interest in using the time-reversal property of a right-handed material or NIM to correct interference and achieve sub-diffraction limited imaging [314, 328, 329]. The employment of optical gain in NIM to potentially overcome the high loss of NIM with realistic materials has also been modeled [330]. A number of useful review articles on superlenses and NIM can be found in [225, 331–333].

A super-lens achieves superresolution imaging through negative index materials. Another way to achieve superresolution imaging is to generate nano-sized light spot using optical antennas.

3.2.3. Optical antennas

One effective way to obtain superresolution light spot is to use optical antennas [334–336], which are capable of coupling, enhancing and localizing optical waves, where surface plasmons, light excited plasma waves in the free electrons of a metal surface, play an important role [337, 338]. Surface plasmons have a higher spatial frequency than the excitation light frequency, and are directly related to the plasma frequency.

The plasma frequency can be obtained by following the Drude model [337, 338]

$$\varepsilon(\omega) = 1 - \frac{\omega_p^2}{\omega^2 + i\omega\gamma}, \tag{61}$$

where ε is the permittivity of the metal at frequency ω, ω_p is the plasma frequency, ω is the light frequency and γ the damping frequency. By expressing the permittivity of the metal as

$$\varepsilon = \varepsilon_r + j\varepsilon_i, \tag{62}$$

the plasma frequency can be obtained by solving Eq. (61) and (62)

$$\omega_p = \omega\left[1 - \varepsilon_r + \frac{\varepsilon_i^2}{1-\varepsilon_r}\right]^{1/2}, \tag{63}$$

$$\gamma = \frac{\omega\varepsilon_i}{1-\varepsilon_r}. \tag{64}$$

To excite a surface plasmon, the real part of the permittivity for the metal must have negative value at the excitation light frequency, and the imaginary part is usually a positive value, and, therefore, the plasma frequency is always higher than the incident light frequency. At light frequencies, $\omega >> \gamma$, thus the real part of Eq. (61) can be approximated as [337, 338],

$$\varepsilon_r = 1 - \frac{\omega_p^2}{\omega_{\rm sp}^2}, \tag{65}$$

and the surface plasmon resonance condition requires that $\varepsilon_r = -\varepsilon_d$, where ε_d is the permittivity of the immediate dielectric material. For a large planar-structure thick metal film (usually over 50 nm), its surface plasmon frequency is calculated as [337]

$$\omega_{\rm sp} = \omega_p/\sqrt{1+\varepsilon_d}. \tag{66}$$

The surface plasmon frequency also depends on the shape of the metal. For a nanosphere, its resonance condition requires that $\varepsilon_r = -2\varepsilon_d$, and thus its frequency is [338]

$$\omega_{\rm sp} = \omega_p/\sqrt{1+2\varepsilon_d}. \tag{67}$$

The resonance frequency of a general spherical particle is more complicated [338]. For a nanorod, its resonance frequency is related to its length and radius [334, 339], with

$$\lambda_{\rm sp} = \lambda_0 + \lambda_1\omega_p/\omega, \tag{68}$$

$$\omega_{\rm sp} = 2\pi c/\lambda_{\rm sp}, \tag{69}$$

where λ_0 and λ_1 are constants that depend on the geometry and dielectric parameters of the antenna [334, 339].

Optical antennas can self-excite surface plasmons. This is because an antenna has features much smaller than the wavelength of the excitation light, which can excite evanescent waves that cover a wide frequency range, including the surface plasmon frequency. The evanescent waves that match the surface plasmon frequency are coupled to the surface plasmon. These can be enhanced to achieve high intensity and nano-sized field localization [340–345].

Generally, there are two types of optical antennas, the resonance type [339–342, 345–360] and the transmission type [201, 204, 206, 208, 212, 343, 344, 361–380]. Metal nanorod, metal tip or nanoparticle [346–353], paired nanorod or nanoparticle [354–357], and bowtie antennas [343, 344] are all resonant type optical antennas, where strong resonance occurs between the boundaries of these antennas and with field localization on the boundaries of the antennas. Circular aperture [201, 204, 206, 208, 212, 361–372], bowtie aperture [376–380], C-shaped aperture [381–384],

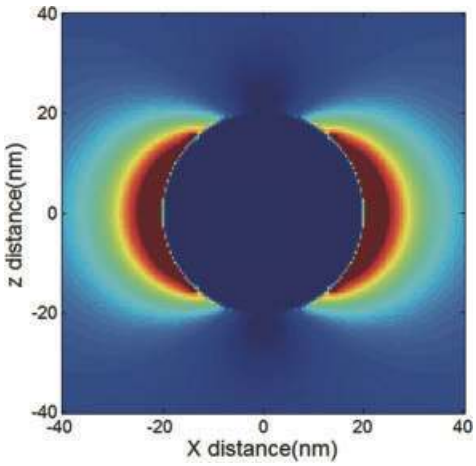

Figure 33 (online color at: www.lpr-journal.org) Field localization by a gold nanoparticle with diameter of 40 nm when the incident light is of transverse polarization; the field is localized on the sides ofthe particle.

and H-shaped aperture [378, 379], inside a metal film, are all transmission type optical antennas, where strong field localization occurs at the center of their transmission apertures.

A metal particle is the simplest form of optical antenna, which can enhance and localize light at its boundary along the polarization direction of the incident light as shown in Fig. 33. Its surface plasmon frequency is described by Eq. (67), and the size of the localized light spot immediately next to the particle is comparable to the radius of the particle. The field can be further enhanced if two or more particles are put close to each other [357].

A bowtie antenna uses the resonance between the two arms [343, 344, 359, 360]. The surface plasmon wavelength propagating along its surface is defined by Eq. (66) and Eq. (69), and strong field localization occurs between the gap between the nearest tips. To obtain best performance, the size of the bowtie is usually chosen as a half wavelength of the excitation light, similar to the half-wavelength dipole antenna used with radio waves [358]. For a paired nanorod, the half-wavelength is replaced with half of the surface plasmon wavelength shown in Eqs. (68) and (69) [334, 339].

All transmission type optical antennas have a hole surrounded with metal film, the shape and size of the holes depending on the surface plasmon wavelength and its applications. The simplest form of aperture is a circular shape, which has been extensively explored [201,204,206,208,361–372]. Its transmission was thought to be proportional to $(a/\lambda)^4$ when $a \ll \lambda$, where a is the hole radius and λ is the light wavelength. However, it was later found that the transmission could be orders higher than the theoretical prediction [204]. This is because of the excitation of high momentum surface waves [212, 364]. The transmission of a hole is affected by its surrounding structures. A concentric corrugation structure can enhance the transmission, because of the excitation of extra surface waves [206, 208, 212, 361–364]. However, to obtain best performance, the period of the corrugation structure should match the wavelength of the excitation light and the phase of the electromagnetic field should also match [212, 362–364]. When multiple holes exist in a metal plate, the individual hole size, shape and the intervals between the holes will affect the transmission and transmission spectrum [364–372]. The transmission of the holes can be enhanced or suppressed, depending on the relative phase of the surface plasmon introduced by the intervals between them [373]. It can also affect the results of Young's slit experiment [373–375].

For an aperture inside a metal film with an infinite boundary, its transmission is more affected by the length of the aperture in the direction normal to the polarization of light [212, 385]. Usually, for a high efficiency transmission aperture, it should be larger than half of the surface plasmon wavelength to allow a single mode surface plasmon to go through. However, along the polarization direction, no matter how small the size is, a transmission mode always exists [212]. Therefore, for a high efficiency rectangular aperture-type superresolution optical antenna [385], the achievable beam size along the direction normal to the light polarization is always larger. The transmission of such a rectangular aperture can be further enhanced by introducing resonances from boundaries [380, 385]. This is because the surface plasmon excited by the aperture propagates outward and is reflected back to the aperture, and also the boundaries can generate surface plasmons and propagate toward the aperture to contribute to the transmission [385]. To further reduce the size of the light spot normal to the polarization direction, one needs to introduce resonant arms inside such an aperture, like an H shaped aperture [378, 379] or bowtie shaped aperture [376–380]. For a bowtie aperture, it is possible to achieve a beam spot below 10 nm with high optical efficiency [380]. Now we see that transmission type optical antennas also involve resonances, which can occur between the boundaries and the holes or between the resonant arms inside the holes.

The polarization state of the excitation light also affects the behavior and design of optical antennas. For longitudinally polarized light [55], or tightly focused radially polarized light [386, 387] where strong longitudinal light exists, a nanotip or nanosphere can directly localize a nanosized light spot below it [385–389] as shown in Fig. 34, which makes it convenient for many tip based applications [388–391].

In summary, optical antennas can excite surface plasmons [334–399], enhance them and condense them to nanometer scale through either transmission holes structures or resonant arms structures. The transmission of a single nanometer hole can be enhanced by using periodic concentric corrugation structures, with peak transmission spectrum determined by the period of the corrugations [208, 212, 361–364, 397]. High optical efficiency can be achieved if the size of an aperture normal to the polarization direction of excitation light is larger than half of the surface plasmon wavelength [212, 385]. However, to compensate for the size increase, resonant tips along the polarization direction inside the aperture are necessary, like the C-shaped aperture, the H-shaped aperture and the bowtie-

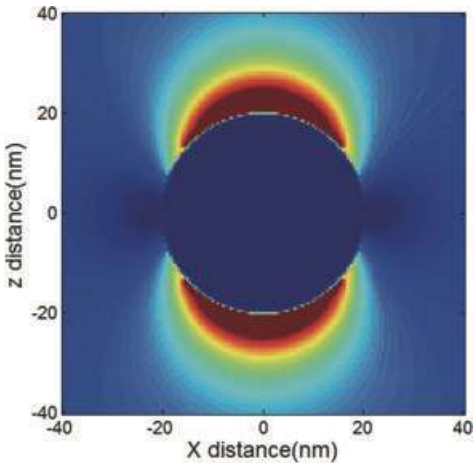

Figure 34 (online color at: www.lpr-journal.org) Field localization by a gold nanoparticle with diameter of 40 nm when the incident light is longitudinally polarized; the field is localized below and above the particle.

shaped aperture [376–384]. The tips can further localize light inside the aperture to nanometer scale. Surface plasmon resonance between the boundaries of a single nanoparticle or nanorod, or paired nanoparticles or nanorods, or bowties can also localize light to nanometer scale [345–357, 398]. The size of the localized light spot is determined by size of the particles, the rod diameter or the width of the tips of bowties. A nano-sized light spot can be directly generated below a nano-sphere or a nano-rod when longitudinally polarized light or light with a strong longitudinal field component obtained through tightly focusing of radially polarized light [386–390].

4. Conclusion and outlook

The unfavorable aspects of diffraction, namely beam divergence and limited resolution, can be overcome through both far field apodization techniques and near field diffraction structures. Apodization techniques can be applied to the pupil of a lens in the form of a pupil mask, which can be a pure phase type or an amplitude type, or a combination of the two. These pupil masks can be used to reduce or eliminate the divergence of light beam in the focal region of a lens [17–33,36–40,45,55], or generate nondiffracting beams with limited propagation distance [45,55,59–62]. The principle of reducing the divergence for phase type and amplitude type apodizers is different. For concentric-annular phase type apodizers [37–39, 45, 55], the divergence of the beam in the focal region is eliminated through generating multiple foci along the optical axis, and the defocusing spherical aberration between the neighboring focal points is totally offset, resulting in zero divergence [37, 38, 45, 55]. In the radial direction, the interference of light field from different belts of the phase apodizer can result in a superresolution light spot [37–39,45,55]. For amplitude type apodizers, like the annular aperture [17–22,32–34,59,60], the mask extends the angular spectrum of light, so that light rays from different angular spectrum components are focused at different positions on the optical axis, and therefore, the divergence of the beam is reduced. Besides this, the use of an annular aperture also results in superresolution, because the aperture blocks the lower spatial frequency light, which increases the relative ratio of high spatial frequency light [33].

A generalized apodization technique also covers approaches that purposely change the light distribution on the object plane [131–133, 400], like the structured illumination microscopy, which projects grating or fringe patterns on the object to increase the band width of the imaging system up to 2 times, resulting in superresolution. Stimulated emission depletion microcopy is a more complicated apodization technique, involving the change of light distribution on the object plane and non-linear effects of fluorescence materials [148–165], which can result in a resolution of $\lambda/45$ [149].

The divergence of the light beam from subwavelength apertures can be reduced through converting high wave number light rays to lower wave number light rays using near field grating-like structures [211–220]. Superresolution can be achieved through turning high wave number evanescent wave into low wave number propagating waves [262–304], or generating a subwavelength localized light spot using optical antennas [334–389]. A superlens can realize ideal imaging through reconstructing the object field using all the wave numbers of light from the object, including evanescent waves. Therefore the image contains all the details of the object [221, 231], but, in practice, when such a high resolution image is captured by a detector, the pixel size of the detector determines the final resolution that can be achieved, and to see more details of the object, the ideal image of the object has to be magnified through turning evanescent waves into propagating waves [262–282].

As an outlook, for the apodization techniques, we are expecting to see more of its applications in some fields, like free space communication, micro-lithography, laser material processing, etc. Its combination with nonlinear effects can generate nano-sized resolution, as in STED microscopy. For near field diffraction structures, like the super lens, new designs of super-lens materials need to be found, because the current materials, like silver, are very dissipative, limiting the thickness of the super-lens, which only allows the thickness to be around 35 nm, determining the working distance to be in 20 nm range. The combination of a super-lens with an optical magnification system through turning evanescent waves into propagating waves is very promising, but, so far, all the designed magnification systems, like the hyperlens and grating structures are also band limited, and the demonstrated resolution is about within the limit of far field resolution. For example, the far field confocal scanning system has a resolution between $\lambda/(3\,\mathrm{NA})$ and $\lambda/(4\,\mathrm{NA})$, which is in the same range as that experimentally demonstrated using a super-lens working in the near field. If the scanning system is combined with a SIL lens with effective refractive index of 2.0 working in the near field, effective resolutions between $\lambda/(6\,\mathrm{NA})$ and $\lambda/(8\,\mathrm{NA})$ can be achieved. Therefore, a practical super-lens needs to have a working distance

of 100 μm or above, and with resolution beyond at least $\lambda/4$ without scanning. If working in the near field, it should be beyond $\lambda/8$ without scanning. At present, even the far field super lens and hyper lens are still working in the near field region of the sample under investigation. Another property a super-lens or hyperlens should have is the capability for two dimensional planar imaging. Some scientists are working on this issue, and have shown promising general designs for a super lens or hyperlens, but there is still a long way to go, and more effort on study of the fundamental principles of nano-imaging is expected. Optical antennas can enhance a light field and localize it to the nanometer scale. This has already been used in many fields, but the requirement for different applications are different: for example, in nano-imaging, we care more about the field enhancement and the size of the localized light spot, but in the data storage field, we also need to take care of the absorptions of the recording materials. Different needs results in differences in the design of the antennas, but optical antennas will certainly be one of the key elements of future data storage systems.

Received: 7 February 2011, **Revised:** 21 June 2011, **Accepted:** 27 June 2011

Published online: 23 August 2011

Key words: Diffraction, nondiffracting, diffraction limit, super-resolution, evanescent wave, time reversal, STED microscopy, super lens, optical antenna, STORM microscopy, high focal depth, apodization, longitudinal polarization, nano-optics, nano-photonics, nanoscopy, microscopy, nano-focusing, nano-lithography, beam shaping.

Haifeng Wang was born in 1971 and received his Ph. D. in 2001 from Shanghai Institute of Optics and fine Mechanics (SIOM), Chinese Academy of Sciences (CAS). After his PhD, he first worked as a post doctoral research fellow in Delft University of Technology (TUDelft), the Netherlands, from 2001 to 2003, where he studied surface plasmon, optical wave guide, vector diffraction theory and radially polarized light. This was followed by a stint at the Free University of Amsterdam (VUA) between 2003 and 2004, where his research was concerned with electromagnetic theory. He joined Data Storage Institute, Singapore in January 2005 where he now works as a Research Scientist. Haifeng discovered the special effect of longitudinal polarization enhancement in binary phase elements, conceived and demonstrated a needle of longitudinally polarized light. His work has been published in journals such as Nature Photonics, Applied Physics Letters, Optical Society of America(OSA) journals, etc., and his research topics cover a wide range, including Nanoscopy, super-resolution bio-imaging, non-diffracting beams, photon spin, abnormal optical polarization engineering, optical antennas, microwave photonics and optical/magnetic data storage. He is a member of the OSA and also holds the position of Guest Professor at the SIOM.

Colin Sheppard, born in 1947, received the MA and PhD degrees in Engineering from Cambridge University, UK. From 1974 to 1989 he was at Oxford University, as a University Lecturer in Engineering Science and Fellow of St. John's and Pembroke College. In 1986 he was awarded a DSc from Oxford in Physical Sciences. From 1989 to 2003 he was Professor of Physics and Head of the Physical Optics Department at the University of Sydney, Australia, and also the Director of Research at the Australian Key Centre for Microscopy & Microanalysis. Since 2003 he has been a Professor at National University of Singapore, where he has been Head of the Division of Bioengineering while simultaneously holding appointments with the Departments of Diagnostic Radiology and Biological Sciences. He has held appointments at University College London, University of Tokyo, Technical University of Delft, University of Queensland, University of Western Australia, EPFL Switzerland, and Massachusetts Institute of Technology. He was elected FIEE (1983) and FInstP (2004), and received an Alexander von Humboldt Research Award (2003) and the Optics & Photonics Division Prize of UK Institute of Physics (2006). He was the Vice-President of the International Commission for Optics (1999–2002) and the Editor-in-Chief of the Journal of Optics A: Pure & Applied Optics (2002–08).

Koustuban Ravi, born in 1987, received the B. Eng degree (First Class Honors) in Electrical and Electronic Engineering from Nanyang Technological University, Singapore, in 2009, sponsored by the Singapore International Airlines –Neptune Orient Lines (SIA-NOL) undergraduate scholarship. Since January 2008, he has worked on the development of computationally efficient and accurate Finite Difference Time Domain (FDTD) models to govern light-matter interaction for the simulation of active optoelectronic devices, first as a student and later as Research Engineer at the Nanophotonics and Electronic-Photonic Integration (NEI) group, Data Storage Institute (DSI), Singapore. His research interests span the nature of light-matter interaction and their implications on applications as diverse as Photonic Integrated Circuits, communication, imaging and data storage.

Seng-Tiong Ho, born in 1958, received B. S. degrees in physics and electrical engineering, and M. S. in electrical engineering and computer science from Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts, in 1984 and Ph. D. in Electrical Engineering from MIT in 1989. From 1989 to 1991, he was a Member of the Technical Staff at AT&T Bell Laboratories, Murray Hill, New Jersey. From 1982–1984, he was also a research member at Microring Laser Gyro Group at Northrop Corporation, Norwood,

Massachusetts. In 1991 he joined the faculty of Northwestern University, Evanston, Illinois and is currently a Professor in the Department of Electrical and Computer Engineering. His contributions have spanned experimental and theoretical aspects in diverse fields ranging from Fundamental Quantum Optics to Microcavity lasers to Organic Photonics and Integrated photonics. He is a recipient of the National Science Foundation (NSF)-Early CAREER Award and was nominated a Fellow of the Optical Society of America (OSA) for his outstanding contributions to the advancement of Microcavity Devices. Prof. S. T. Ho also serves as a visiting professor and advisor to the Nanophotonics and Electronic-Photonic Integration (NEI) group at Data Storage Institute (DSI), Singapore.

Guillaume Vienne, born in 1966, was educated in Africa, France, Germany, and England. In 1997, he received a PhD from the Optoelectronics Research Center at the University of Southampton for his thesis on Erbium-Ytterbium co-doped optical fibers. He has since worked for both industry and academia in Japan, France, Denmark, China (Hong Kong and Hangzhou) and Singapore. Prior to joining the Data Storage Institute (DSI) in July 2009, he was an Associate professor at the Nanophotonics group at Zhejiang University, China. He is currently a research scientist in the Advanced Concept Group (ACG) at DSI.

References

[1] H. Gernsheim, The History of Photography (Oxford University Press, New York, 1955), p. 1.
[2] J. Petzval, Wien Ber. **25**, 33 (1857); Philos. Mag. **17**, 1 (1859).
[3] Lord Rayleigh, Philos. Mag. **31**, 87 (1891); Scientific Papers, Vol. I (Dover Publications, New York, 1965), p. 513.
[4] E. W. H. Selwyn, Photogr. Z. **90B**, 47 (1950).
[5] K. Sayanagi, J. Opt. Soc. Am. **57**, 1091–1099 (1967).
[6] S. R. Wilk, Opt. Photon. News **17**, 12–13 (2006).
[7] L. Kipp, M. Skibowski, R. L. Johnson, R. Berndt, R. Adelung, S. Harm, and R. Seemann, Nature **414**, 184–188 (2001).
[8] Q. Cao and J. Jahns, J. Opt. Soc. Am. A **19**, 2387–2393 (2002).
[9] J. Jahns, Q. Cao, and S. Sinzinger, Laser Photonics Rev. **2**, 249–263 (2008).
[10] F. M. Huang, T. S. Kao, V. A. Fedotov, Y. Chen, and N. I. Zheludev, Nano Lett. **8**, 2469–2472 (2008).
[11] M. Born and E. Wolf, Principles of Optics, 7th ed. (Cambridge University Press, Cambridge, 1999).
[12] G. Farnell, Can. J. Phys. **35**, 777–783 (1957).
[13] C. J. R. Sheppard and K. G. Larkin, J. Opt. Soc. Am. A **17**, 772–779 (2000).
[14] C. J. R. Sheppard and P. Török, J. Opt. Soc. Am. A **15**, 3016–3019 (1998).
[15] B. E. A. Saleh and M. C. Teich, Fundamentals of Photonics, (John Wiley & Sons, New York, 1991), Chap. 3, pp. 80–107.
[16] H. F. Wang, G. Q. Yuan, W. L. Tan, L. P. Shi, and T. C. Chong, Opt. Eng. **46**, 065201(2007).
[17] G. B. Airy, Philos. Mag. **18**, 1–10 (1841).
[18] L. Rayleigh, Mon. Notes R. Astron. Soc. **33**, 59–63 (1872).
[19] E. H. Linfoot and E. Wolf, Proc. Phys. Soc. B **66**, 145–149 (1953).
[20] G. Lansraux, Rev. Opt. **32**, 475 (1953).
[21] W. H. Steel, Rev. Opt. **32**, 4–26 (1953).
[22] W. T. Welford, J. Opt. Soc. Am. **50**, 749–753 (1960).
[23] G. Toraldo di Francia, Atti. Fond. Giorgio Ronchi **7**, 366–372 (1952).
[24] G. Toraldo di Francia, Nuovo Cimento Suppl. **9**, 426–435(1952).
[25] B. J. Thompson, J. Opt. Soc. Am. **55**, 145–149 (1965).
[26] J. Ojeda-Castañeda, L. R. Berriel-Valdos, and E. Montes, Opt. Lett. **10**, 520–522 (1985).
[27] C. J. R. Sheppard and Z. S. Hegedus, J. Opt. Soc. Am. A **5**, 643–647 (1988).
[28] S. Mezouari and A. R. Harvey, Opt. Lett. **28**, 771–773 (2003).
[29] C. J. R. Sheppard, J. Mod. Opt. **43**, 525–536 (1996).
[30] J. Durnin, J. Opt. Soc. Am. A **4**, 651–654 (1987).
[31] C. J. R. Sheppard and T. Wilson, IEE J. Microw. Opt. Acoust. **2**, 105–112 (1978).
[32] J. Durnin and J. J. Miceli, Jr., and J. H. Eberly, Phys. Rev. Lett. **58**, 1499–1501 (1987).
[33] H. F. Wang and F. X. Gan, Opt. Eng. **40**, 851–855 (2001).
[34] J. H. McLeod, J. Opt. Soc. Am. **44**, 592–597 (1954).
[35] N. Davidson, A. A. Friesem, and E. Hasman, Opt. Lett. **16**, 523–525 (1991).
[36] Z. Jaroszewicz, J. Sochacki, and A. Kolodzjejczyk, and L. R. Staroński, Opt. Lett. **18**, 1893–1895 (1993).
[37] H. F. Wang and F. X. Gan, Appl. Opt. **41**, 5263–5266 (2002).
[38] H. F. Wang and F. X. Gan, Appl. Opt. **40**, 5658–5662 (2001).
[39] C. J. R. Sheppard, J. Campos, J. C. Escalera, and S. Ledesma, Opt. Commun. **281**, 3623–3630 (2008).
[40] C. J. R. Sheppard, IEE J. Microw. Opt. Acoust. **2**, 163–166(1978).
[41] C. J. R. Sheppard and M. Martinez-Corral, Opt. Lett. **33**, 476–478 (2008).
[42] B. Richards and E. Wolf, Proc. Roy. Soc. A **253**, 358–379 (1959).
[43] A. Boivin and E. Wolf, Phys. Rev. B **138**, 1561–1565 (1965).
[44] P. Török, P. D. Higdon, and T. Wilson, Opt. Commun. **148**, 300–315 (1998).
[45] H. F. Wang, L. P. Shi, G. Q. Yuan, X. S. Miao, W. L. Tan, and T. C. Chong, Appl. Phys. Lett. **89**, 171102 (2006).
[46] D. Pohl, Appl. Phys. Lett. **20**, 266–267 (1972).
[47] S. C. Tidwell, D. H. Ford, and W. D. Kimura, Appl. Opt. **29**, 2234–2239 (1990).
[48] M. Stalder and M. Schadt, Opt. Lett. **21**, 1948–1950(1996).
[49] R. Dorn, S. Quabis, and G. Leuchs, Phys. Rev. Lett. **91**, 233901 (2003).
[50] A. S. van de Nes, L. Billy, S. F. Pereiraand, and J. J. M. Braat, Opt. Express **12**, 1281–1293(2004).
[51] H. P. Urbach and S. F. Pereira, Phys. Rev. Lett. **100**, 123904 (2008).
[52] C. J. R. Sheppard and P. Török, Optik **104**, 175–177(1997).
[53] C. J. R. Sheppard and A. Choudhury, Appl. Opt. **43**, 4322–4327 (2004).
[54] S. F. Pereira and A. S. van de Nes, Opt. Commun. **234**, 119–124(2004).
[55] H. F. Wang, L. Shi, B. Luk'yanchuk, C. J. R. Sheppard, and T. C. Chong, Nature Photon. **2**, 501–505(2008).

[56] Q. Zhan and J. R. Leger, Opt. Express **10**, 324–331(2002).
[57] T. D. Visser and J. T. Foley, J. Opt. Soc. Am. A **22**, 2527–2531 (2005).
[58] Q. Zhan, Adv. Opt. Photon. **1**, 1–57 (2009).
[59] M. Mazilu, D. J. Stevenson, F. Gunn-Moore, and K. Dholakia, Laser Photonics Rev. **4**, 529–548 (2010).
[60] D. McGloin and K. Dholakia, Cont. Phys. **46**, 15–28 (2005).
[61] J. Durnin, J. J. Jr Miceli, and J. H. Eberly, Phys. Rev. Lett. **58**, 1499–1501 (1987).
[62] J. Durnin, J. Opt. Soc. Am. A **4**, 651–654 (1987).
[63] J. Stratton, Electromagnetic Theory (McGraw-Hill, New York, 1941).
[64] C. J. R. Sheppard and T. Wilson, IEE J. Microw. Opt. Acoust. **2**, 105–112 (1978).
[65] C. J. R. Sheppard, IEE J. Microw. Opt. Acoust. **2**, 163–166 (1978).
[66] S. Franke-Arnold, L. Allen, and M. Padgett, Laser Photonics Rev. **2**, 299–313 (2008).
[67] S. Chávez-Cerda, M. Padgett, I. Allison, G. New, J. Gutiérrez-Vega, A. O'Neil, I. Macvicar, and J. Courtial, J. Opt. B **4**, S52-S57 (2002).
[68] J. Gutiérrez-Vega, M. Iturbe-Castillo, and S. Chávez-Cerda, Opt. Lett. **25**, 1493–1495 (2000).
[69] M. V. Berry and N. L. Balazs, Am. J. Phys. **47**, 264–267 (1979).
[70] G. A. Silviloglou, J. Borky, A. Dogariu, and D. N. Christodoulides, Phys. Rev. Lett. **99**, 213901 (2007).
[71] E. R. Dowsky and W. T. Cathey, Appl. Opt. **34**, 1859–1866 (1995).
[72] A. Chong, W. H. Renninger, D. N. Christodoulides, and F. W. Wise, Nature Photon. **4**, 103–106 (2010).
[73] A. Chong, W. H. Renninger, D. N. Christodoulides, and F. W. Wise, Opt. Photon. News **21**, 52–52 (2010).
[74] Y. Kartashov, V. Vysloukh, and L. Torner, Prog. Opt. **52**, 63–148 (2009).
[75] A. Couairon and A. Mysyrowicz, Phys. Rep. **441**, 47–189 (2007).
[76] P. Rohwetter, J. Kasparian, K. Stelmaszczyk, Z. Hao, S. Henin, N. Lascoux, W. M. Nakaema, Y. Petit, M. Queißer, R. Salamé, E. Salmon, L. Wöste, and J. Wolf, Nature Photon. **4**, 451–456 (2010).
[77] W. T. Welford, J. Opt. Soc. Am. **50**, 749–753 (1960).
[78] J. W. S. Rayleigh, Mon. Notes Astron. Soc. **33**, 59–63 (1872).
[79] C. J. R. Sheppard, Optik **48**, 329–334 (1977).
[80] G. Inbedetouw, J. Opt. Soc. Am. A **6**, 150–152 (1989).
[81] A. J. Cox and D. C. Dibble, J. Opt. Soc. Am. A **9**, 282–286 (1992).
[82] J. H. McLeod, J. Opt. Soc. Am. **50**, 166–169 (1960).
[83] J. Fujiwara, J. Opt. Soc. Am. **52**, 287–292 (1962).
[84] J. Dyson, Proc. Roy. Soc. A **248**, 93–106 (1958).
[85] J. Amako, D. Sawaki, and E. Fujii, J. Opt. Soc. Am. B **20**, 2562–2568 (2003).
[86] W. H. Steel, in: Optics in Metrology, edited by P. Mollet (Pergamon, 1960), pp. 181–193.
[87] B. Salik, J. Rosen, and A. Yariv, J. Opt. Soc. Am. **12**, 1702–1706 (1995).
[88] T. Grosjean, S. S. Saleh, M. A. Suarez, I. K. Ibrahim, V. Piquerey, D. Charraut, and P. Sandoz, Appl. Opt. **46**, 8061–8067 (2007)
[89] S. Ramachandran and S. Ghalmi, in: Proceedings of the Conference on Lasers and Electro-Optics/Quantum Electronics and Laser Science Conference and Photonic Applications Systems Technologies, OSA Technical Digest (CD), (Optical Society of America, 2008), paper CPDB5.
[90] J. K. Kim, J. Kim, Y. Jung, W. Ha, Y. S. Jeong, S. Lee, A. Tünnermann, and K. Oh, Opt. Lett. **34**, 2973–2975 (2009).
[91] S. R. Lee, J. Kim, S. Lee, Y. Jung, J. K. Kim, and K. Oh, Opt. Express **18**, 25299–25305 (2010).
[92] Q. Zhan, Opt. Lett. **31**, 1726–1728 (2006).
[93] W. Chen and Q. Zhan, Opt. Lett. **34**, 722–724 (2009).
[94] M. K. Bhuyan, F. Courvoisier, P. A. Lacourt, M. Jacquot, R. Salut, L. Furfaro, and J. M. Dudley, Appl. Phys. Lett. **97**, 081102 (2010).
[95] X. Tsampoula, V. Garcés-Chávez, M. Comrie, D. J. Stevenson, B. Agate, C. T. A. Brown, F. Gunn-Moore, and K. Dholakia, Appl. Phys. Lett. **91**, 053902 (2007).
[96] Z. H. Ding, H. W. Ren, Y. H. Zhao, J. S. Nelson, and Z. P. Chen, Opt. Lett. **27**, 243–245 (2002).
[97] K. S. Lee and J. P. Rolland, Opt. Lett. **33**, 1696–1698 (2008).
[98] F. O. Fahrbach, P. H. Simon, and A. Rohrbach, Nature Photon. **4**, 780–785 (2010).
[99] T. A. Planchon, L. Gao, D. E. Milkie, M. W. Davidson, J. A. Galbraith, C. G. Galbraith, and E. Betzig, Nature Methods, DOI:10.1038/NMETH.1586 (2011).
[100] P. Dufour, M. Piché, and Y. D. Korninck, and N. McCarthy Appl. Opt. **45**, 9246–9252 (2006).
[101] J. Sharpe, U. Ahlgren, P. Perry, B. Hill, and A. Ross, J. Hecksher-Sørensen, R. Baldock, and D. Davidson, Science **296**, 541–545 (2002).
[102] C. A. McQueen, J. Artl, and K. Dholakia, Am. J. Phys. **67**, 912–915 (1999).
[103] A. Boivin, J. Opt. Soc. Am. **42**, 60 (1952).
[104] C. J. R. Sheppard, G. Calvert, and M. Wheatland, J. Opt. Soc. Am. A **15**, 849–856 (1998)
[105] R. Luneberg, Mathematical Theory of Optics, 2nd ed. (University of California Press, Berkeley and Los Angeles, California, 1966).
[106] P. Jacquinot and M. B. Roizen-Dossier, Prog. Opt. **3**, 67–108(1964).
[107] C. J. R. Sheppard, Micron **38**, 165–169 (2007).
[108] M. Martinez-Corral and G. Saavedra, Prog. Opt. **53**, 1–67 (2009).
[109] C. W. McCutchen, J. Opt. Soc. Am. **59**, 1163–1171 (1969).
[110] A. Papoulis, J. Opt. Soc. Am. **62**, 1423–1429 (1972).
[111] B. J. Thompson, J. Opt. Soc. Am. **55**, 145–149 (1965).
[112] M. P. Cagigal, J. E. Oti, V. F. Canales, and P. J. Valle, Opt. Commun. **241**, 249–253 (2004).
[113] C. J. R. Sheppard, J. Campos, J. C. Escalera, and S. Ledesma, Opt. Commun. **281**, 913–922 (2008).
[114] V. F. Canales, J. E. Oti, and M. P. Cagigal, Opt. Commun. **247**, 11–18 (2005).
[115] R. Boivin and A. Boivin, Opt. Acta **27**, 857–610 (1980).
[116] C. K. Sieracki and E. W. Hansen, Proc. SPIE **2184**, 120–126 (1994).
[117] C. J. R. Sheppard, Optik **99, 1**, 32–34 (1995).
[118] M. Martinez-Corral, P. Andrés, C. J. Zapata-Rodriguez, and M. Kowalczyk, Opt. Commun. **165**, 267–278 (1999).
[119] M. Martinez-Corral, C. Ibanez-Lopez, and G. Saavedra, Opt. Express **11**, 1740–1745 (2003).
[120] D. M. de Juana, V. F. Canales, P. J. Valle, and M. P. Cagigal, Opt. Commun. **229**, 71–77 (2004).

[121] C. J. R. Sheppard, J. Campos, J. C. Escalera, and S. Ledesma, Opt. Commun. **281**, 3623–3630 (2008).
[122] C. J. R. Sheppard, M. A. Alonso, and N. J. Moore, J. Opt. A, Pure Appl.Opt. **10**, 0333001 (2008).
[123] C. J. R. Sheppard and M. Martinez-Corral, Opt. Lett. **33**, 476–478 (2008).
[124] C. J. R. Sheppard and E. Y. S. Yew, Opt. Lett. **33**, 497–499 (2008).
[125] C. J. R. Sheppard, N. K. Balla, and S. Rehman, Opt. Commun. **282**, 727–734 (2009).
[126] T. G. Jabbour and S. M. Kuebler, Opt. Express **14**, 1033–1043 (2005).
[127] Z. S. Hegedus, Optica Acta **32**, 815–826 (1985).
[128] Z. Hegedus and V. Sarafis, J. Opt. Soc. Am. A **3**, 1892–1896 (1986).
[129] M. V. Berry and S. Popescu, J. Phys. A, Math. Gen. **39**, 6965–6977 (2006).
[130] N. Zheludev, Nature Mater. **7**, 420–422 (2008).
[131] W. Lukosz and M. Marchand, Optica Acta, **10**, 241–255 (1963).
[132] C. W. McCutchen, J. Opt. Soc. Am. **57**, 1190–1192 (1967).
[133] I. J. Cox, C. J. R. Sheppard, and T. Wilson, Optik **60**, 391–396 (1982).
[134] S. B. Ippolito B. B. Goldberg, and M. S. Ünlü, Appl. Phys. Lett. **78**, 4071–4073 (2001).
[135] S. H. Goh, C. J. R. Sheppard, A. C. T. Quah, C. M. Chua, L. S. Koh, and J. C. H. Phang, Rev. Sci. Instrum. **80**, 013703 (2009).
[136] M. G. L. Gustafsson, L. Shao, P. M. Carlton, C. J. R. Wang, I. N. Golubovskaya, W. Z. Cande, D. A. Agard, and J. W. Sedat, Biophys. J. **94**, 4957–4970 (2008).
[137] C. J. R. Sheppard and R. Kompfner, Appl. Opt. **17**, 2879–2882 (1978).
[138] C. J. R. Sheppard and M. Gu, Optik **86**, 104–106 (1990).
[139] G. Bouwhuis and J. H. M. Spruit, Appl. Opt. **29**, 3766–3768 (1990).
[140] S. W. Hell, Nature Biotechnol. **21**, 1347–1355 (2002).
[141] M. G. L. Gustafsson, Proc. Natl. Acad. Sci. **102**, 13081–13086 (2005).
[142] C. J. R. Sheppard and H. J. Matthews, J. Opt. Soc. Am. A **4**, 1354–1360 (1987).
[143] C. J. R. Sheppard and C. J. Cogswell, Reflection and transmission confocal microscopy, presented at the Optics in Medicine, Biology and Environmental Research: Proceedings of the International Conference on Optics Within Life Sciences, Garmisch-Partenkirchen, Germany, 1993.
[144] S. Hell and E. H. K. Stelzer, Opt. Commun. **93**, 277–282 (1992).
[145] M. G. L. Gustafsson, D. A. Agard, and J. W. Sedat, J. Microsc. **195**, 10–16 (1999).
[146] L. Shao, B. Isaac, S. Uzawa, D. A. Agard, J. W. Sedat, and M. G. L. Gustafsson, Biophys. J. **94**, 4971–4983 (2008).
[147] R. Schmidt, C. A. Wurm, S. Jacobs, J. Englehardt, A. Egner, and S. W. Hell, Nature Methods **5**, 539–544 (2008).
[148] S. W. Hell and J. Wichmann, Opt. Lett. **19**,780–782 (1994).
[149] S. W. Hell, Science **316**, 1153–1154(2007).
[150] Y. Garini, B. J. Vermolen, and I. T. Young, Curr. Opin. Biotechnol. **16**, 3–12 (2005).
[151] S. Inoue, in: Handbook of Biological Confocal Microscopy, 3rd ed., edited by J. B. Pawley (Springer, New York, 2006).
[152] T. A. Klar, S. Jakobs, M. Dyba, A. Egner, and S. W. Hell, Proc, Natl. Acad. Sci. **97**, 8206–8210 (2000).
[153] T. A. Klar, E. Engel, and S. W. Hell, Phys. Rev. E **64**, 66613-1–9 (2001).
[154] R. Menon, P. Rogge, and H. Y. Tsai, J. Opt. Soc. Am. A **26**,297–304 (2009).
[155] S. W. Hell, Nature Methods, **6**, 24–32 (2009).
[156] S. W. Hell, K. I. Willig, M. Dyba, S. Jakobs, L. Kastrup, and V. Westphal, in: Handbook of Biological Confocal Microscopy, 3rd ed., edited by J. B. Pawley (Springer, New York, 2006).
[157] V. Westphal, C. M. Banca, M. Dyba, L. Kastrup, and S. W. Hell, Appl. Phys. Lett. **82**, 3125–3127(2003).
[158] T. A. Klar and S. W. Hell, Opt. Lett. **24**, 954–956 (1999).
[159] S. W. Hell, R. Schmidt, and A. Egner, Nature Photon. **3**, 381–387(2009).
[160] M. Dyba and S. W. Hell, Phys. Rev. Lett. **88**, 1–4 (2002).
[161] V. Westphal and S. W. Hell, Phys. Rev. Lett. **94**, 1–4 (2005).
[162] C. V. Middendorff, A. Egner, C. Geisler, S. W. Hell, and A. Schonle, Opt. Express **16**, 20774–20788 (2008).
[163] G. Moneron and S. W. Hell, Opt. Express, **17**, 14567–14573 (2009).
[164] E. Riitweger, K. Y. Han, S. E. Irvine, C. Eggeling, and S. W. Hell, Nature Photon. **3**,144–147(2009).
[165] J. Vogelsang, C. Steinhauer, C. Forthmann, I. H. Stein, B. Person-Segro, T. Cordes, and P. Tinnefeld, Chem. Phys. Chem. **11**, 2475–2490 (2010).
[166] S. Bretschneider, C. Eggeling, and S. W. Hell, Phys. Rev. Lett. **98**, 1–4 (2007).
[167] E. Rittweger, D. Wildanger, and S. W. Hell, Europhys. Lett. **86** 14001(2009).
[168] P. C. Maurer, J. R. Maze, P. L. Stanwix, L. Jiang, A. V. Gorshkov, A. A. Zibrov, B. Harke, J. S. Hodges, A. S. Zibrov, A. Yacoby, D. Twitchen, S. W. Hell, R. L. Walsworth, and M. D. Lukin, Nature Phys. **6**, 912–918 (2010).
[169] M. Hofmann, C. Eggeling, S. Jakobs, and S. W. Hell, Proc. Natl. Acad. Sci. **102**, 17565–17569 (2005).
[170] G. Donnert, J. Keller, R. Medda, M. A. Andrei, S. O. Rizzolim, R. Luhrmann, R. Jahn, C. Eggeling, and S. W. Hell, Proc. Natl. Acad. Sci. **103**, 11440–11445(2006).
[171] J. Vogelsang, R. Kasper, C. Steinhauer, B. Person, M. Heilemann, M. Sauer, and P. Tinnefeld, Angew. Chem. **47**, 5465–5469(2008).
[172] J. Vogelsang, T. Cordes, C. Forthmann, C. Steinhauer, and P. Tinnefeld, Nano Lett. **10**, 672–679(2010).
[173] M. Heilemann, P. Dedecker, J. Hofkens, and M. Sauer, Laser Photonics Rev. **3**, 180–202 (2009).
[174] I. Willig, S. O. Rizzoli, V. Westphal, R. Jahn, and S. W. Hell, Nature **440**, 935–939 (2006).
[175] U. V. Nagerl, K. I. Willig, B. Hein, S. W. Hell, and T. Bonhoeffer, Proc. Natl. Acad. Sci. **105**, 18982–18987(2008).
[176] R. R. Kellner, C. J. Baier, K. I. Willig, S. W. Hell, and J. Barrantes, Neurosci. **144**, 135–143(2007).
[177] D. J. Stephens and V. J. Allan, Sci. **300**, 82–86 (2003).
[178] R. Yueste, Nature Methods **2**, 902–904(2005).
[179] K. I. Willig, R. R. Kellner, R. Medda, B. Hein, S. Jakobs, and S. W. Hell, Nature Methods **3**, 721–723 (2006).
[180] M. Dyba, S. Jakobs, and S. W. Hell, Nature Biotechnol. **21**, 1303–1304 (2003).
[181] J. B. Ding, K. T. Takasaki, and B. L. Sabatini, Neuron **63**,429–437(2009).
[182] W. E. Moerner, Proc. Natl. Acad. Sci. **104**, 12596–12602(2007).
[183] B. Huang, Curr. Opin. Chem. Biol. **14**, 10–14 (2010).

[184] J. L. Schwartz and S. Manley, Nature Methods **6**, 21–23(2009).
[185] G. Donnert, J. Keller, C. A. Wurm, S. O. Rizzoli, V. Westphal, A. Schonle, R. Jahn, S. Jakobs, C. Eggeling, and S. W. Hell, Biophys. J., Biophys. Lett. **92**, 67–69 (2007).
[186] D. Neumann, J. Buckers, L. Kastrup, S. W. Hell, and S. Jakobs, PMC Biophys. **3**, 4 (2010).
[187] S. Manley, J. Gillette, and J. L. Schwartz, Methods Enzymol. **475**, 109–120 (2010).
[188] E. Betzig, G. H. Patterson, R. Sourgat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess, Science **313**, 1642–1645 (2006).
[189] T. J. Gouldm, V. V. Verkhusha, and S. T. Hess, Nature Protocols **4**, 291–308 (2009).
[190] M. J. Rust, M. Bates, and X. Zhuang, Nature Methods **3**, 793–795 (2006).
[191] B. Huang, W. Wang, M. Bates, and X. Zhuang, Science **319**, 810–813 (2008).
[192] B. Huang, S. A. Jones, B. Brandenburg, and X. Zhuang, Nature Methods **5**, 1047–1052 (2008).
[193] R. Henriques, C. Griffiths, E. H. Rego, and M. M. Mhlanga, Biopolymers **95**, 322–331 (2011).
[194] M. Bates, B. Huang, G. T. Dempsey, and X. Zhuang, Science **317**, 1749–1752 (2007).
[195] F. V. Subach, G. H. Patterson, S. Manley, J. M. Gillette, J. Lippincott-Schwartz, and V. V. Verkhusha, Nature Methods **6**, 153–159 (2009).
[196] M. Heilemann, S. van de Linde, M. Schüttpelz, R. Kasper, B. Seefeldt, A. Mukherjee, P. Tinnefeld, and M. Sauer, Angewandte Chemie **47**, 6172–6176(2008).
[197] J. Folling, M. Bossi, H. Bock, R. Medda, C. A. Wurm, B. Hein, S. Jakobs, and C. Eggeling, and S. W. Hell, Nature Methods **5**, 943–945(2008).
[198] P. Hoyer, T. Staudt, J. Engelhardt, and S. W. Hell, Nano Lett. **11**, 245–250 (2011).
[199] H. Shroff, C. G. Galbraith, J. A. Galbraith, and E. Betzig, Nature Methods **5**, 417–423 (2008).
[200] S. Manley, J. M. Gillette, and J. Lippincott-Schwartz, Nature Methods **5**, 155–157 (2008).
[201] H. A. Bethe, Phys. Rev. **66**, 163–182(1944).
[202] H. F. Wang, F. H. Groen, S. F. Pereira, and J. J. M. Braat, Appl. Phys. Lett. **83**, 4486–4487(2003).
[203] H. F. Wang, F. H. Groen, S. F. Pereira, and J. J. M. Braat, Opt. Rev. **14**, 29–32(2007).
[204] T. W. Ebbesen, H. Z. Lezec, H. F. Ghaemi, T. Thio, and P. A. Wolf, Nature **391**, 667–669(1998).
[205] D. E. Grupp, H. J. Lezec, T. Thio, and T. W. Ebbesen, Adv. Mater. **11**, 860–862 (1999).
[206] T. Thio, K. M. Pellerin, R. A. Linke, H. J. Lezec, and T. W. Ebbesen, Opt. Lett. **26**, 1972–1974 (2001).
[207] A. Nahata, R. A. Linke, T. Ishi, and K. Ohashi, Opt. Lett. **28**, 423–425 (2003).
[208] T. Thio, H. J. Lezec, T. W. Ebbesen, K. M. Pellerin, G. D. Lewen, A. Nahata, and R. A. Linke, Nanotechnology **13**, 429–432 (2002).
[209] M. J. Lockyear, A. P. Hibbins, J. R. Sambles, and C. R. Lawrence, Appl. Phys. Lett. **84**, 2040–2042 (2004).
[210] H. Cao, A. Agrawal, and A. Nahata, Opt. Express **13**, 763–769 (2005).
[211] A. Degiron and T. W. Ebbesen, Opt. Express **12**, 3694–3700 (2004).
[212] C. Genet and T. W. Ebbesen, Nature **445**, 39–46 (2007).
[213] H. Z. Lezec, A. Degiron, E. Deveaux, R. A. Linke, L. Martin-Moreno, F. J. Garcia-Vidal, and T. W. Ebbesen, Science **297**, 820–822(2002).
[214] L. Martín-Moreno, F. J. García-Vidal, H. J. Lezec, A. Degiron, and T. W. Ebbesen, Phys. Rev. Lett. **90**, 167401 (2003).
[215] A. Kock, W. Beinstingl, K. Berthold, and E. Gornik, Appl. Phys. Lett. **52**, 1164–1166(1988).
[216] A. Kock, E. Gornik, M. Hauser, and W. Beinstingl, Appl. Phys. Lett. **57**, 2327–2329 (1990).
[217] N. Yu, J. Fan, Q. Wang, C. Pflugl, L. Diehl, T. Edamura, M. Yamanishi, H. Kun, and F. Capasso, Nature Photon. **2**, 564–570 (2008).
[218] N. Yu, R. Blanchard, J. Fan, F. Capasso, T. Edamura, M. Yaminishi, and H. Kan, Appl. Phys. Lett. **93**, 181101–181103 (2008).
[219] P. Kramper, M. Agio, C. M. Soukoulis, A. Birner, F. Muller, R. B. Wehrspohn, U. Gosele, and V. Sandoghdar, Phys. Rev. Lett. **92**, 113903 (2004).
[220] I. Bulu, H. Caglayan, and E. Ozbay, Opt. Lett. **30**, 3078–3080 (2005).
[221] V. G. Veselago, Sov. Phys. Uspekhi-USSR **10**, 509–514 (1968).
[222] J. B. Pendry, A. J. Holden, W. J. Stewart, and I. Youngs, Phys. Rev. Lett. **76**, 4773–4776 (1996).
[223] J. B. Pendry, A. J. Holden, D. J. Robbins, and W. J. Stewart, IEEE Trans. Microwave Theory Tech. **47**, 2075–2084 (1999).
[224] C. M. Soukoulis, S. Linden, and M. Wegener, Science **315**, 47–49 (2007).
[225] V. M. Shalaev, Nature Photon. **1**, 41–48 (2007).
[226] D. R. Smith, W. J. Padilla, D. C. Vier, S. C. Nemat-Nasser, and S. Schultz, Phys. Rev. Lett. **84**, 4184–4187 (2003).
[227] R. A. Shelby, D. R. Smith, and S. Schultz, Science **292**, 77–79 (2001).
[228] M. Bayindir, K. Aydin, E. Ozbay, P. Markos, and C. M. Soukoulis, Appl. Phys. Lett. **81**, 120–122 (2002).
[229] C. G. Parazzoli, R. B. Greegor, K. Li, B. E. C. Koltenbah, and M. A. Tanielian, Phys. Rev. Lett. **90**, 107401 (2003).
[230] R. B. Greegor, C. G. Parazzoli, and M. H. Tanielian, Appl. Phys. Lett. **82**, 2356–2358 (2003).
[231] J. B. Pendry, Phys. Rev. Lett. **85**, 3966–3969 (2000).
[232] D. R. Smith, D. Schurig, R. Rosenbluth, S. Schultz, S. A. Ramakrishna, and J. B. Pendry, Appl. Phys. Lett. **82**, 1506–1508 (2003).
[233] N. Fang and X. Zhang, Appl. Phys. Lett. **82**, 161–163 (2003).
[234] V. A. Podolskiy, Appl. Phys. Lett. **87**, 231113 (2005).
[235] I. A. Larkin and M. I. Stockman, Nano Lett. **5**, 339–343 (2005).
[236] Z. Liu, N. Fang, T. Yen, and X. Zhang, Appl. Phys. Lett. **83**, 5184–5186 (2003).
[237] N. Fang, Z. Liu, T. J. Yen, and X. Zhang, Opt. Express **11**, 682–687 (2003).
[238] A. Grbic and G. V. Eleftheriades, Phys. Rev. Lett. **92**, 117403 (2004).
[239] B. I. Popa and S. A. Cummer, Phys. Rev. E **73**, 016617 (2006).
[240] A. A. Houck, J. B. Brock, and I. L. Chuang, Phys. Rev. Lett. **90**, 137401 (2003).
[241] A. N. Lagarkov and V. N. Kissel, Phys. Rev. Lett. **92**, 077401 (2004).
[242] N. Fang, H. Lee, C. Sun, and X. Zhang, Science **308**, 534–537 (2005).

[243] H. Lee, Y. Xiong, N. Fang, W. Srituravanich, S. Durant, M. Ambati, C. Sun, and X. Zhang, New J. Phys. **7**, 255 (2005).
[244] D. Melville and R. Blaikie, Opt. Express **13**, 2127–2134 (2005).
[245] E. Shamonina, V. A. Kalinin, K. H. Ringhofer, and L. Solymar, Electron. Lett. **37**, 1243–1244 (2001).
[246] S. A. Ramakrishna, J. B. Pendry, M. C. K. Wiltshire, and W. J. Stewart, J. Mod. Opt. **50**, 1419–1430 (2003).
[247] P. A. Belov and Y. Hao, Phys. Rev. B **73**, 113110 (2006).
[248] B. Wood, J. B. Pendry, and D. P. Tsai, Phys. Rev. B **74**, 115116 (2006).
[249] T. Taubner, D. Korobkin, Y. Urzhumov, G. Shvets, and R. Hillenbrand, Science **313**, 1595 (2006).
[250] M. Notomi, Phys. Rev. B **62**, 10696–10705 (2000).
[251] C. Luo, S. G. Johnson, J. D. Joannopoulos, and J. B. Pendry, Phys. Rev. B **65**, 201104 (2002).
[252] C. Luo, S. G. Johnson, J. D. Joannopoulos, and J. B. Pendry, Phys. Rev. B **68**, 045115 (2003).
[253] A. L. Efros and A. L. Pokrovsky, Solid State Commun. **129**, 643–649 (2004).
[254] S. Foteinopoulou, and C. M. Soukoulis, Phys. Rev. B **67**, 235107 (2003).
[255] E. Cubukcu, K. Aydin, E. Ozbay, S. Foteinopoulou, and C. M. Soukoulis, Nature **423**, 604–605 (2003).
[256] P. V. Parimi, W. T. Lu, P. Vodo, and S. Sridhar, Nature **426**, 404 (2003).
[257] E. Cubukcu, K. Aydin, E. Ozbay, S. Foteinopolou, and C. M. Soukoulis, Phys. Rev. Lett. **91**, 207401 (2003).
[258] Z. Y. Li and L. L. Lin, Phys. Rev. B. **68**, 245110 (2003).
[259] G. Shvets and Y. A. Urzhumov, Phys. Rev. Lett. **93**, 243902 (2004).
[260] E. Schonbrun, Q. Wu, W. Park, T. Yamashita, C. J. Summers, M. Abashin, and Y. Fainman, Appl. Phys. Lett. **90**, 041113 (2007).
[261] V. A. Podolski and E. E. Narimanov, Opt. Lett. **30**, 75–77 (2005).
[262] S. Durant, Z. Liu, J. M. Steele, and X. Zhang, J. Opt. Soc. Am. B **23**, 2383–2392 (2006).
[263] Z. Liu, S. Durant, H. Lee, Y. Pikus, N. Fang, Y. Xiong, C. Sun, and X. Zhang, Nano Lett. **7**, 403–408 (2007).
[264] Z. Liu, S. Durant, H. Lee, Y. Pikus, Y. Xiong, C. Sun, and X. Zhang, Opt. Express **15**, 6947–6954 (2007).
[265] Y. Xiong, Z. Liu, S. Durant, H. Lee, C. Sun, and X. Zhang, Opt. Express **15**, 7095–7102 (2007).
[266] Y. Xiong, Z. Liu, C. Sun, and X. Zhang, Nano Lett. **7**, 3360–3365 (2007).
[267] Z. Jacob, L. V. Alekseyev, and E. Narimanov, Opt. Express **14**, 8247–8256 (2006).
[268] A. Salandrino and N. Engheta, Phys. Rev. B **74**, 075103 (2006).
[269] H. Lee, Z. Liu, Y. Xiong, C. Sun, and X. Zhang, Opt. Express. **15**, 15886–15891 (2007).
[270] J. B. Pendry and S. A. Ramakrishna, J. Phys. Condens. Matter **14**, 8463–8479 (2002).
[271] J. B. Pendry, Opt. Express **11**, 755–760 (2003).
[272] Z. Liu, H. Lee, Y. Xiong, C. Sun, and X. Zhang, Science **315**, 1686–1686 (2007).
[273] I. I. Smolyaninov, Y. J. Huang, and C. C. Davis, Science **315**, 1699–1701 (2007).
[274] Z. Jacob, L. V. Alekseyev, and E. Narimanov, J. Opt. Soc. Am. A **24**, A52-A59 (2007).
[275] H. Lee, Z. Liu, Y. Xiong, C. Sun, and X. Zhang, Opt. Express **15**, 15886–15891 (2007).
[276] L. Chen, X. Y. Zhou, and G. P. Wang, Appl. Phys. B **92**,127 (2008).
[277] J. G. Hu, P. Wang, Y. H. Lu, H. Ming, C. C. Chen, and J. X. Chen, Chin. Phys. Lett. **25**, 4439 (2008).
[278] X. Zhang and Z. Liu, Nature Mater. **7**, 435–441(2008).
[279] Z. Liu, Z. Liang, X. Jiang, X. Hu, X. Li, and J. Zi, Appl. Phys. Lett **96**, 113507 (2010).
[280] C. Ma, R. Aguinaldo, and Z. Liu, Chin. Sci. Bull. **55**, 2618–2624(2010).
[281] J. Li, L. Fok, X. Yin, G. Bartal, and X. Zhang, Nature Mater. **8**, 931–934 (2009).
[282] A. V. Kildishev, U. K. Chettiar, Z. Jacob, V. M. Shalaev, and E. E. Narimanov, Appl. Phys. Lett. **94**, 071102 (2009).
[283] Y. Xiong, Z. Liu, and X. Zhang, Appl. Phys. Lett. **94**, 203108 (2009).
[284] L. Chen and G. Wang, Opt. Express **17**, 3903–3912 (2009).
[285] W. Wang, H. Xing, L. Fang, Y. Liu, J. Ma, L. Lin, C. Wang, and X. Luo, Opt. Express **16**, 21142–21148(2008).
[286] E. H. Synge, Philos. Mag. **6**, 356–362(1928).
[287] E. H. Synge, Philos. Mag. **13**, 297–300(1932).
[288] B. Hecht, B. Sick, U. P. Wild, V. Deckert, R. Zenobi, O. J. F. Martin, and D. W. Dieter, J. Chem. Phys. **12**,7761–7774(2000).
[289] S. Kawata, M. Ohtsu, and M. Irie, Nano-optics (Springer-Verlag, Berlin, Heidelberg, 2002).
[290] E. A. Ash and G. Nicholls, Nature **237**, 510–512(1972).
[291] A. Lewis, M. Isaacson, A. Harootunian, and A. Murray, Ultramicroscopy **13**, 227–231(1984).
[292] Y. Oshikane, T. Kataoka, N. Okuda, S. Hara, H. Inoue, and M. Nakano, Sci. Technol. Adv. Mater. **8**, 181–185(2007).
[293] D. W. Pohl, W. Denk, and M. Lanz, Appl. Phys. Lett. **44**, 651–653(1984).
[294] M. Specht, J. D. Pedarnig, W. M. Heckl, and T. W. Hänsch, Phys. Rev. Lett. **68**, 476–479 (1992).
[295] R. Toledo-Crow, P. C. Yang, Y. Chen, and M. Vaez-Iravani, Appl. Phys. Lett. **60**, 2957–2959(1992).
[296] L. Novotny, D. W. Pohl, and B. Hecht, Opt. Lett. **20**, 970–972(1995).
[297] S. M. Mansfield and G. S. Kino, Appl. Phys. Lett **57**, 2615–2616(1990).
[298] J. K. Trautman, E. Betzig, J. S. Weiner, D. J. DiGiovanni, T. D. Harris, F. Hellman, and E. M. Gyorgy, J. Appl. Phys. **71**, 4659–4663(1992).
[299] B. D. Terris, H. J. Mamin, D. Rugar, W. R. Studenmund, and G. S. Kino, Appl. Phys. Lett. **65**, 388–390(1994).
[300] L. P. Ghislain and V. B. Elings, Appl. Phys. Lett. **72**, 2779–2781(1998).
[301] T. D. Milster, J. S. Jo, and K. Hirota, Appl. Opt. **38**, 5046–5057 (1999).
[302] W. Kim, Y. Yoon, H. Choi, N. Park, and Y. Park, Opt. Express **16**, 13933–13948 (2008).
[303] F. Zijp, M. B. van der Mark, C. A. Verschuren, J. I. Lee, J. van den Eerenbeemd, H. P. Urbach, and M. A. H. van der Aa,, IEEE Trans. Magn. **41**, 1042–1046 (2005).
[304] A. S. van de Nes, J. J. M. Braat, and S. F. Pereira, Rep. Prog. Phys. **69**, 2323–2363(2006).
[305] N. C. Pauoiu and R. M. Osgood, Opt. Commun. **223**, 331–337 (2003).

[306] N. Katsarakis, T. Koschny, M. Kafesaki, E. N. Economou, and C. M. Soukoulis, Appl. Phys. Lett. **84**, 2943–2945 (2004).
[307] T. J. Yen, W. J. Padilla, N. Fang, D. C. Vier, D. R. Smith, J. B. Pendry, D. N. Basov, and X. Zhang, Science **303**, 1494–1496 (2004).
[308] S. Linden, C. Enkrich, M. Wegener, J. Zhou, T. Koschny, and C. M. Soukoulis, Science **306**, 1351–1353 (2004).
[309] N. Katsarakis, G. Konstantinidis, A. Kostopoulos, R. S. Penciu, T. F. Gundogdu, M. Kafesaki, E. N. Economou, T. Koschny, and C. M. Soukoulis, Opt. Lett. **30**, 1348–1350(2005).
[310] V. M. Shalaev, W. Cai, U. K. Chettiar, H. K. Yuan, A. K. Sarychev, V. P. Drachev, and A. V. Kildishev, Opt. Lett. **30**, 3356–3358 (2005).
[311] V. A. Podolskiy, A. K. Sarychev, V. M. Shalaev, and J. Nonlinear, Opt. Phys. Mater. **11**, 65–74 (2002).
[312] V. A. Podolskiy, A. K. Sarychev, and V. M. Shalaev, Opt. Express. **11**, 735–745 (2003).
[313] A. K. Sarychev and V. M. Shalaev, Proc. SPIE, **5508**, 128–137 (2004).
[314] J. B. Pendry, Science **322**, 71–73 (2008).
[315] D. O. S. Melville and R. J. Blaikie, J. Vac. Sci. Technol. B **22**, 3470–3474 (2004).
[316] D. O. S. Melville and R. J. Blaikie, J. Opt. A **7**, S176–S183 (2005).
[317] S. A. Ramakrishna, D. Schurig, D. R. Smith, S. Schultz, and J. B. Pendry, J. Mod. Opt. **49**, 1747–1762 (2002).
[318] J. B. Pendry, Contemp. Physics **45**, 191–202 (2004).
[319] P. M. Valanju, R. M. Walser, and A. P. Valanju, Phys. Rev. Lett. **88**, 187401 (2002).
[320] X. Huang and W. L. Schaich, Am. J. Phys. **72**, 1232–1240 (2004).
[321] R. Ruppin, J. Phys., Condens. Matter **13**, 1811–1819 (2001).
[322] H. Wallen, H. Kettunen, and A. Sihvola, Metamaterials **2**, 113–121 (2008).
[323] W. C. Chow, Progr. Electromagn. Res. **51**, 1–26 (2005).
[324] W. Zhang, H. Chen, and H. O. Moser, Appl. Phys. Lett. **98**, 073501 (2011).
[325] A. Kildshev and V. M. Shalaev, Opt. Lett. **33**, 43–45 (2008).
[326] S. Han, Y. Xiong, D. Genov, Z. Liu, G. Bartal, and X. Zhang, Nano Lett. **8**, 4243–4247 (2008).
[327] J. Rho, Z. Ye, Y. Xiong, X. Yin, Z. Liu, H. Choi, G. Bartal, and X. Zhang, Nature Comms. **1**:143 (2010).
[328] L. Y. Mario, X. Yuan, M. K. Chin, and J. Nonlinear, Opt. Phys. Mater. **14**, 245–257 (2005).
[329] G. Lerosey, J. Rosny, A. Tourin, and M. Fink, Science **315**, 1120–1122 (2007).
[330] Y. Sivan, S. Xiao, U. K. Chettiar, A. V. Kildshev, and V. M. Shahaev, Opt. Express **17**, 24060–24074 (2009).
[331] S. A. Ramakrishna, Rep. Prog. Phys. **68**, 449–521 (2005).
[332] X. Zhang and Z. Liu, Nature Mater. **7**, 435–441 (2008).
[333] R. E. Collin, Prog. Electromagn. Res. B **19**, 233–261 (2010).
[334] P. Bharadwaj, B. Deutsch, and L. Novotny, Adv. Opt. Photon. **1**, 438–483 (2009).
[335] J. Alda, J. M. Rico-García, J. M. López-Alonso, and G. Boreman, Nanotechnology **16** S230-S234(2005)
[336] L. Novotny and B. Hecht, Principles of Nano-Optics (Cambridge University Press, Cambridge, 2006).
[337] H. Raether, Surface Plasmons on Smooth and Rough Surfaces and on Gratings (Springer-Verlag, Berlin, 1988)
[338] Craig F. Bohren, and D. R. Huffman, Absorption and Scattering of Light by Small Particles (John Wiley & Sons, New York, 1998)
[339] L. Novotny, Phys. Rev. Lett. **98**, 266802 (2007).
[340] B. C. Stipe, T. C. Strand, C. C. Poon, H. Balamane, T. D. Boone, J. A. Katine, J. Li, V. Rawat, H. Nemoto, A. Hirotsune, O. Hellwig, R. Ruiz, E. Dobisz, D. S. Kercher, N. Robertson, T. R. Albrecht, and B. D. Terris, Nature Photon. **4**, 484–488 (2010).
[341] W. A. Challener, C. Peng, A. V. Itagi, D. Karns, W. Peng, Y. Peng, X. Yang, X. Zhu, N. J. Gokemeijer, Y.-T. Hsia, G. Ju, R. E. Rottmayer, M. A. Seigler, and E. C. Gage, Nature Photon. **3**, 220–224 (2009).
[342] E. Yablonovitch and M. Staffaroni, in: Proceedings of the INSIC EHDR Program Meeting, Berkeley, CA, 2010.
[343] K. Şendur, W. Challener, and C. Peng, J. Appl. Phys. **96**, 2743–2752 (2004).
[344] A. V. Itagi, D. D. Stancil, J. A. Bain, and T. E. Schlesinger, Appl. Phys. Lett. **83**, 4474–4476 (2003).
[345] W. A. Challener, E. Gage, A. Itagi, and C. Peng, Jpn. J. Appl. Phys. **45**, 6632–6642(2006).
[346] T. H. Taminiau, F. D. Stefani, F. B. Segerink, and N. F. Vanhulst, Nature Photon. **2**, 234–237(2008).
[347] K. Kempa, J. Rybczynski, Z. Huang, K. Gregorczyk, A. Vidan, B. Kimball, J. Carlson, G. Benham, Y. Wang, A. Herczynski, and Z. F. Ren, Adv. Mater. **19**, 421–426, (2007).
[348] K. B. Crozier, A. Sundaramurthy, G. S. Kino, and C. F. Quate, J. Appl. Phys. **94**, 4632–4642(2003).
[349] L. Novotny, E. J. Sanchez, and X. S. Xie, Ultramicroscopy **71**, 21–29 (1998).
[350] L. Novotny and S. J. Stranick, Annu. Rev. Phys. Chem. **57**, 303–331 (2006).
[351] A. Hartschuh, M. R. Beversluis, A. Bouhelier, and L. Novotny, Philos.Trans. R. Soc. Lond. A **362**, 807–819 (2003).
[352] A. Hartschuh, Angewandte Chemie, **47**, 8178–8191(2008).
[353] E. Bailo and V. Deckert, Chem. Soc. Rev. **37**, 921–930 (2008).
[354] P. Muühlschlegel, H.-J. Eisler, O. J. F. Martin, B. Hecht, and D. W. Pohl, Science **308**, 1607–1609(2005).
[355] E. Cubukcu, E. A. Kort, K. B. Crozier, and F. Capasso, Appl. Phys. Lett. **89**, 093120(2006).
[356] A. N. Grigorenko, N. W. Roberts, M. R. Dickinson, and Y. Zhang, Nature Photon. **2**, 365–370 (2008).
[357] K. Li, M. I. Stockman, and D. J. Bergman, Phys. Rev. Lett. **91**, 227402 (2003)
[358] T. Milligan, Modern Antenna Design (McGraw-Hill, New York, 1985).
[359] R. D. Grober, R. J. Schoelkopf, and D. E. Prober, Appl. Phys. Lett. **70**, 1354–1356 (1997).
[360] P. J. Schuck, D. P. Fromm, A. Sundaramurthy, G. S. Kino, and W. E. Moerner, Phys. Rev. Lett. **94**, 017402 (2005).
[361] H. F. Ghaemi, T. Thio, D. E. Grupp, T. W. Ebbesen, and H. J. Lezec, Phys. Rev. B **58**, 6779–6782 (1998).
[362] A. Degiron, H. J. Lezec, W. L. Barnes, and T. W. Ebbesen, Appl. Phys. Lett. **81**, 4327–4329 (2002).
[363] O. T. A. Janssen, H. P. Urbach, and G. W. t' Hooft, Phys. Rev. Lett. **99**, 043902 (2007).
[364] H. J. Lezec and T. Thio, Opt. Express **12**, 3629–3651(2004).
[365] W. L. Barnes, W. A. Murray, J. Dintinger, E. Devaux, and T. W. Ebbesen, Phys. Rev. Lett. **92**, 107401 (2004).
[366] E. Popov, M. Nevière, S. Enoch, and R. Reinisch, Phys. Rev. B **62**, 16100 (2000).

[367] L. Martín-Moreno, F. J. García-Vidal, H. J. Lezec, K. M. Pellerin, T. Thio, J. B. Pendry, and T. W. Ebbesen, Phys. Rev. Lett. **86**, 1114–1117 (2001).
[368] A. Krishnan, T. Thio, T. J. Kim, H. J. Lezec, T. W. Ebbesen, P. A. Wolff, and J. Pendry, L. Martín-Moreno, and F. J. García-Vidal, Opt. Commun. **200**, 1–7 (2001).
[369] J.-M. Vigoureux, Opt. Commun. **198**, 257–263 (2001).
[370] S. Collin, F. Pardo, R. Teissier, and J.-L. Pelouard, Phys. Rev. B **63**, 33107 (2001).
[371] H. X. Yuan, B. X. Xu, H. F. Wang, and T. C. Chong, Jpn. J. Appl. Phys. **45**, 787–791(2006).
[372] H. X. Yuan, B. X. Xu, H. F. Wang, and T. C. Chong, Jpn. J. Appl. Phys. **45**, 6974–6980(2006).
[373] C. H. Gan, G. Gbur, and T. D. Visser, Phys. Rev. Lett. **98**, 043908 (2007).
[374] H. F. Schouten, N. Kuzmin, G. Dubois, T. D. Visser, G. Gbur, P. F. A. Alkemade, H. Blok, G. W. t' Hooft, D. Lenstra, and E. R. Eliel, Phys. Rev. Lett. **94**, 053901 (2005).
[375] N. Kuzmin, G. W. t' Hooft, E. R. Eliel, G. Gbur, H. F. Schouten, and T. D. Visser, Opt. Lett. **32**, 445–447(2007).
[376] L. Wang, S. M. Uppuluri, E. X. Jin, and X. Xu, Nano Lett. **6**, 361–364(2006).
[377] E. X. Jin and X. Xu, Appl. Phys. Lett. **88**, 153110(2006).
[378] E. X. Jin and X. Xu, Appl. Phys. Lett. **86**, 111106(2005).
[379] E. X. Jin and X. Xu, J. Quant. Spectrosc. Radiat. Transf. **93**, 163–173(2005)
[380] H. F. Wang, T. C. Chong, and L. Shi, in: Proceedings of the Optical Data Storage Topical Meeting 2009, Lake Buena Vista, FL, USA (2009).
[381] X. Shi, L. Hesselink, and R. L. Thornton, Opt. Lett. **28**, 1320–1322 (2003).
[382] X. Shi and L. Hesselink, J. Opt. Soc. Am. B **21**, 1305–1317 (2004).
[383] F. Chen, A. Itagi, J. A. Bain, D. D. Stancil, T. E. Schlesinger, L. Stebounova, G. C. Walker, and B. B. Akhremitchev, Appl. Phys. Lett. **83**, 3245 (2003);
[384] J. S. Shumaker-Parry, H. Rochholz, and M. Kreiter, Adv. Mater. **17**, 2131–2134(2005)
[385] H. F. Wang, B. X. Xu, H. X. Yuan, and T. C. Chong, in: Proceedings of the International Symposium on Optical Memory 2007, Technical Digest, Pan Pacific Hotel Singapore 2007, pp. 174–175.
[386] W. Chen and Q. Zhan, Opt. Express **15**, 4106–4111 (2007).
[387] T. J. Antosiewicz, P. Wróbel, and T. Szoplik, Opt. Express **17**, 9191–9196 (2009).
[388] A. Bouhelier, J. Renger, M. R. Beversluis, and L. Novotny, J. Microscopy **210**, 220–224 (2003).
[389] W. Chen and Q. Zhan, Chin. Opt. Lett. **5**, 709–711 (2007).
[390] P. Bharadwaj, P. Anger, and L. Novotny, Nanotechnology **18**, 044017(2007).
[391] P. Verma, T. Ichimura, T. Yano, Y. Saito, and S. Kawata, Laser Photonics Rev. **4**, 548–561 (2010).
[392] P. Vasa, C. Ropers, R. Pomraenke, and C. Lienau, Laser Photonics Rev. **3**, 483–507 (2009).
[393] M. Pelton, J. Aizpurua, and G. Bryant, Laser Photonics Rev. **2**, 136–159 (2008).
[394] K. F. MacDonald and N. I. Zheludev, Laser Photonics Rev. **4**, 562–567 (2010).
[395] J. A. Polo, Jr. and A. Lakhtakia, Laser Photonics Rev. **5**, 234–246 (2011).
[396] I. P. Radko, V. S. Volkov, J. Beermann, and A. B. Evlyukhin, T. Søndergaard, A. Boltasseva, and S. I. Bozhevolnyi, Laser Photonics Rev. **3**, 575–590 (2009).
[397] R. Gordon, A. G. Brolo, D. Sinton, and K. L. Kavanagh, Laser Photonics Rev. **4**, 311–335 (2010).
[398] A. Wax and K. Sokolov, Laser Photonics Rev. **3**, 146–158 (2009).
[399] P. R. West, S. Ishii, G. V. Naik, N. K. Emani, V. M. Shalaev, and A. Boltasseva, Laser Photonics Rev. **4**, 795–808 (2010).
[400] C. Maurer, A. Jesacher, S. Bernet, and M. Ritsch-Marte, Laser Photonics Rev. **5**, 81–101 (2011).