On resolution with far field optics

Zheng Xi and Paul Urbach • Optics Research Group, Delft University of Technology, 2628CJ, Delft, The Netherlands

Paul Urbach did his PhD at Groningen University. Then he joined Philips Research in Eindhoven. In 2000 he became part-time professor diffraction optics at Delft University of Technology. In 2008 he became leader of the Optics Research Group of this university. He has been president of the European Optical Society from 2012-2014 and again in 2017-2018. He is member of the Advisory Board on metrology of the minister of Economic Affairs and Climate.

Dr. Zheng Xi obtained his PhD degree from the University of Science and Technology of China in 2014 with thesis on the interaction of fluorescent molecules with optical antennas. After obtaining his PhD, he went to the Optics Research Group at the Delft University of Technology, for post doctoral research. His current interests include the design of diffractive optical element for LED illumination shaping as well as the study of scattering of optical antennas by cylindrical polarized beams for superresolution metrology applications.

Abstract

It is common knowledge in optics that, when only far fields are measured, there is a fundamental limit to the resolution given by the wavelength of the light. Yet research continues to beat this fundamental limit. Even a Nobel Prize was awarded for super-resolution some years ago. In this article we consider the apparent contradiction and explain under what circumstances the classical limit of resolution can be beaten. As will be seen, these circumstances are quite special, but if they can be realised super-resolution is possible.

1. The classical resolution limit

Light of a single colour (i.e. single wavelength) has in general a state of polarisation that varies from point to point. This happens for example in the focal region of a strongly focusing lens. Although in high performance optical systems such lenses occur frequently, we shall here mostly discard polarisation. The topic of resolution can already be well explained by using the so-called scalar theory, where only one component of the electric field is considered.

In optical design, geometrical optics is used to find suitable lenses and refractive indices to realise a certain image quality for a desired field of view. In demanding applications, it is necessary that aberrations are very small so that the design achieves the diffraction limit. This means that the ultimately realised resolution is limited by diffraction. In such demanding cases Fourier optics is a suitable tool to study the limit of resolution.

In Fourier optics the interaction of light with a thin object is described by a complex reflection or transmission function. The amplitudes of these functions are smaller than unity when part of the light is absorbed. The argument of the complex function describes the change of phase upon reflection or transmission. We consider reflection and assume that the object is described by the reflection function $r(x,y)$. If $U_i(x,y)$ is the field that illuminates the object, the reflected field is:

$$U_s(x,y) = r(x,y)U_i(x,y), \tag{1}$$

The function r will be referred to briefly as "the object" and the aim of imaging is to retrieve as much information as possible about this function from measured intensities.

Imaging in the broad sense is transformation of information about an object to a detector by

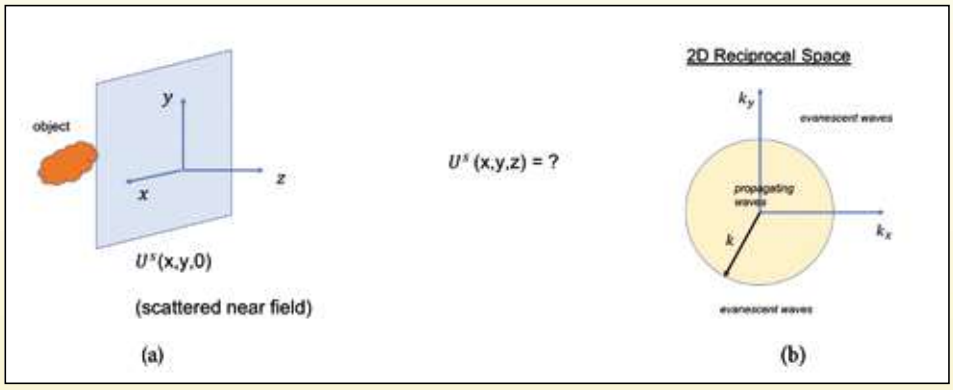

Fig. 1. (a) Propagation of a field in the positive z-direction. b) Two sets of spatial frequencies in two-dimensional reciprocal space.

optical fields. In classical imaging one tries to realise an intensity distribution in the detector plane that resembles as much as possible the intensity distribution at the object. In contrast, in computational imaging the detector is often not in the image plane but instead in another plane. The information about the object then has to be retrieved by processing the measured intensities. In scatterometry intensities are measured in the far field or in a plane conjugate to it. Also in this case the measured data have to be processed to retrieve information about the object.

Whatever method is used, details of the field smaller than the wavelength are lost already during propagation through homogenous space. A field $U(x,y,0)$ in the plane $z=0$ can be propagated to a plane $z>0$ through homogeneous space by the angular spectrum method (Fig. 1a). First $U(x,y,0)$ is Fourier transformed with respect to x and y. This gives the complex amplitudes of the plane waves with wave vectors whose x- and y-components k_x, k_y are the spatial frequencies of the Fourier transform. By propagating these plane waves to a plane z= constant >0 and adding the amplitudes of the plane waves in that plane (which amounts to performing a back Fourier transform), the field $U(x,y,z)$ is obtained.

It turns out that there are two types of plane waves (see Fig. 1b). Waves with spatial frequencies satisfying: $k_x^2 + k_y^2 < k^2$, where $k=2\pi/\lambda$ with λ the wavelength, are propagating waves. The amplitudes of the propagating waves remain constant during propagation and only their phase changes. The waves with high spatial frequencies $k_x^2 + k_y^2 > k^2$ have amplitude which decreases exponentially with distance z and are called evanescent. In contrast to the propagating waves, the evanescent waves do not propagate energy in the direction z. They can not contribute to the field after propagation over macroscopic distances, therefore only the amplitudes of spatial frequencies $k_x^2 + k_y^2 < k^2$ are preserved. By the uncertainty principle it follows that details of the field of order smaller than $\frac{2\pi}{k} = \lambda$ are lost. Hence, irrespective of the quality of the components in an optical system, propagation through homogeneous space already causes the loss of all information of details smaller than the wavelength. On top of that, because of their finite size, optical components can not capture all the propagating waves. Only the propagating waves with $k_x^2 + k_y^2 < k^2\mathrm{NA}^2$, where NA<1 is the numerical aperture of the optical system, can be measured.[1] Hence the actual resolution that can be obtained with a system of numerical aperture NA is λ/NA where λ is the wavelength in the medium between the object and the detector.

We now ask the question what information of the reflection function r can be transferred? By the convolution theorem, the Fourier transform of Eq. (1) becomes

$$\mathcal{F}(U_s)(k_x,k_y) = \iint \mathcal{F}(r)(k_x - k'_x,\, k_y - k'_y)\mathcal{F}(U_i)(k'_x,k'_y)\mathrm{d}k'_x\mathrm{d}k'_y\,, \quad (2)$$

where $\mathcal{F}$ is the two-dimensional Fourier transform with respect to x,y. If both detection and illumination are done with an optical system of numerical aperture NA, both $\mathcal{F}(U_i)(k_x,k_y)$ and $\mathcal{F}(U_s)(k_x,k_y)$ can be nonzero only when the spatial frequencies satisfy $k_x^2 + k_y^2 < k^2\mathrm{NA}^2$. If the illumination is perpendicular to the object, i.e. U_i is a plane wave with wave vector parallel to the optical z-axis, the Fourier transform of the incident field is nonzero only for $k_x = k_y = 0$. It is seen from Eq. (2) that then the part of the spectrum of the reflection function that can influence the detected field is the same as that of the scattered field, i.e. it is limited by the circle with radius $k\mathrm{NA}$ and hence details of the reflection function that can be resolved are larger than λ/NA. However, if the incident field contains an oblique plane wave, e.g. $k'_x = -k\mathrm{NA}$, and If the spatial frequency of the detected field is maximum: $k_x = k\mathrm{NA}$, then $k_x - k'_x = 2k\mathrm{NA}$. Hence Eq. (2) implies that in this case the largest spatial frequency of the reflection function r that influences U_s is $k_x - k'_x = 2k\mathrm{NA}$. Hence, if the illuminating field contains plane waves with spatial frequencies that are maximum for the given numerical aperture of the system, the spatial frequencies of the reflection function of the object that influence the measurement are up to twice that for the case of perpendicular illumination, and hence the resolution is up to twice as large: $\lambda/(2\mathrm{NA})$. One says that by oblique illumination high spatial frequencies of the object which otherwise can not enter the optical system are "folded into the numerical aperture". This folding effect is the explanation for the resolution obtained in optical recording and in confocal microscopy (Fig. 2). In both cases the sample (i.e. a pit structure on an optical disc in the case of optical recording) is illuminated with a strongly focused spot which is a superposition of plane waves with maximum spatial frequency $k\mathrm{NA}$ and the reflected field is imaged on a detector by the same objective as used in the illumination. The lateral resolution therefore is roughly $\lambda/(2\mathrm{NA})$. In the case of the microscope, the higher resolution comes at the cost of slower imaging because the object has to be scanned, but the confocal detection has the additional benefit of better axial resolution.

In total internal reflection microscopy, the sample is positioned on a substrate with high refractive index and is illuminated

[1] It is customary to explain the higher resolution of an immersion microscope by saying that the numerical aperture is larger than 1, but actually the higher resolution is because in a medium with refractive index $n>1$ the wavelength is shorter.

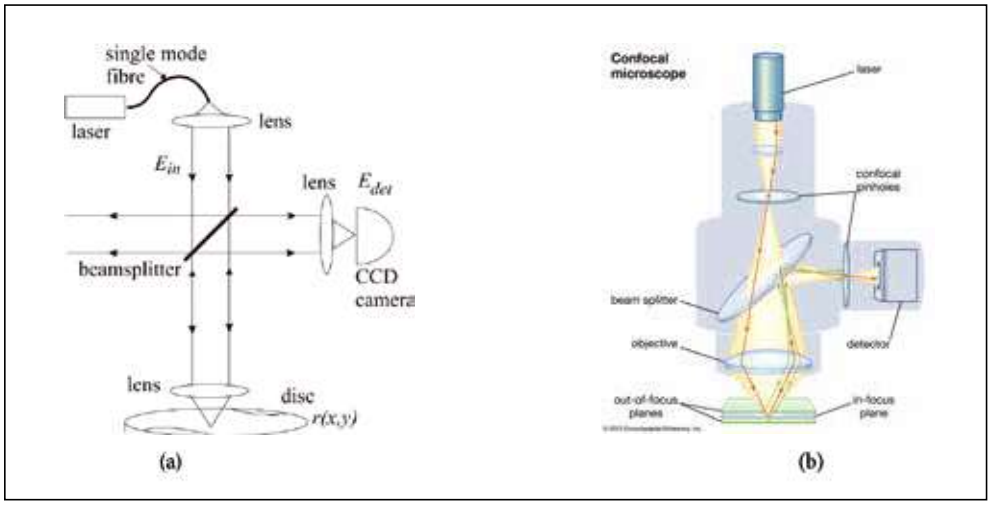

Fig. 2.
Two cases where a scanning focused spot with high numerical aperture is used:
a) optical recording,
b) confocal microscope
(From Ecyclopedia Brittanica).

through the substrate by a beam at an angle of incidence above the critical angle of total internal reflection. As illustrated in Fig. 3, the field is then evanescent in air and the period of the wave at the surface is shorter than the wavelength in air. The sample is thus illuminated by an evanescent wave and the folding effect causes that relatively high spatial frequencies of the object can be detected by the objective of the microscope.

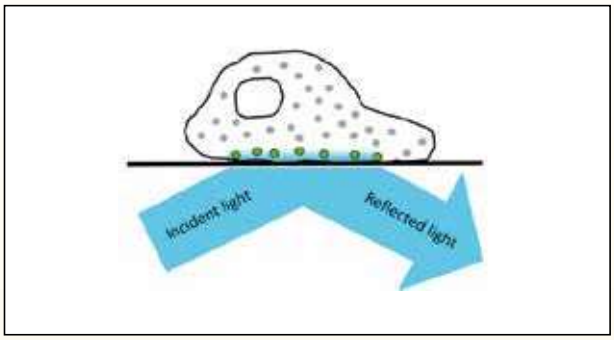

Fig. 3. Illumination using a beam incident above the angle of total internal reflection (from Cell Imaging Centre, Drexel University).

2. Higher resolution

We mention now some methods to achieve high resolution. A solid immersion lens (SIL) enhances the resolution to the extent that in the relation $\lambda/(2NA)$ the wavelength should be replaced by $\frac{\lambda}{n}$ or even by λ/n^2, depending on whether the SIL is a half sphere or a Weierstrass optic (Fig. 4a). Here $n>1$ is the refractive index of the material of the SIL. The focused spot is formed at the bottom of the SIL. To achieve higher resolution, the SIL should either be positioned on the sample or else the air gap should be very small [1].

Very high resolution is obtained with a scanning near field optical microscope (SNOM). In the type of SNOM shown in Fig. 4b, the light is focused through a small aperture coated with a metal on an AFM cantilever which is kept very close to the sample. The transmitted light is collected by a microscope objective [2]. Because the distance between the aperture and the sample is of the order of nanometers, a small spot is formed. The resolution that can obtained is limited by the

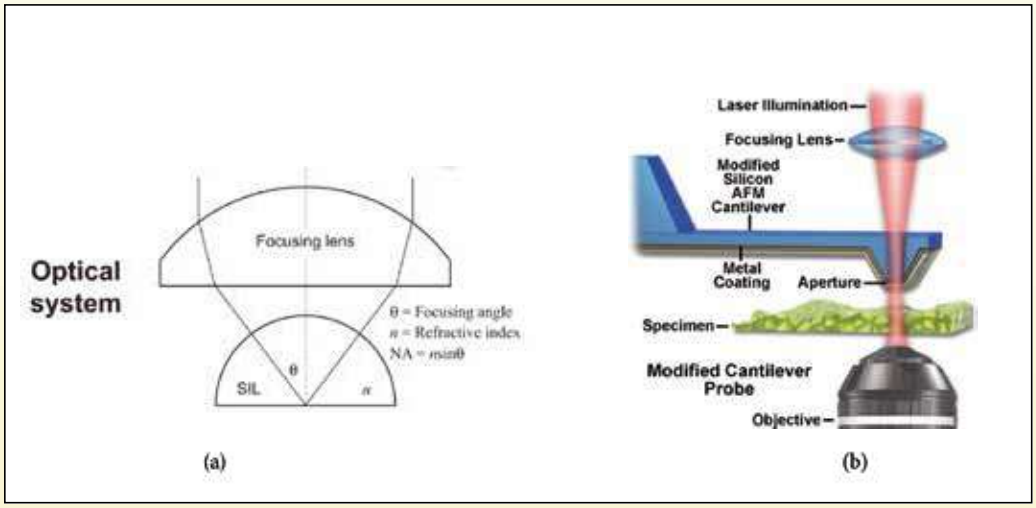

Fig. 4.
a) Solid immersion lens.
b) Scanning near field optical microscope
(courtesy Olympus).

size of the aperture, which is of the order of 50 nm. Since the aperture must be brought almost in contact with the sample, the SNOM falls outside of the realm of far field optics and therefore is out of the scope of this paper. This holds also for the SIL mentioned above because it should also be brought very close to the sample.

High resolution can also be achieved by nonlinear effects. The stimulated emission depletion (STED) microscope, invented by Stefan Hell, is a type of fluorescence microscope where two focused beams are used: one to excite the molecules and the other to deplete them by producing stimulated emission from the excited to the ground state [3]. The depletion beam has a small region with zero intensity in the centre to prevent depletion there. Because of saturation, the region of non-depletion can be made much smaller than a normal diffraction limited spot. The spatial extent of the fluorescent region is therefore substantially reduced.

Another fluorescence microscope that yields nanometer resolution is the Photoactivated Localisation Microscope (PALM) invented by Betzig [4]. By illuminating a sample of inactive fluorescence molecules by a brief pulse of ultraviolet or violet light, a small percentage (±1%) of the molecules is activated. These are then imaged and subsequently eliminated by bleaching. Then a next uv pulse excites another 1% of the molecules and so on. Because most fluorescent molecules are at relatively large distance to each other, they can be distinguished in the image. Furthermore, by determining the centre of mass of the images of the molecules their positions can be determined with precision of a few nm. The 2014 Nobel Prize was shared between the inventors of the two new fluorescence microscopes.
In 1952 Toraldo di Francia showed that an annular pupil with proper transmission values can lead to a diffraction pattern with a central core that can be made arbitrary narrow at the cost of very little intensity and extremely large side lobes [5], [6], [7]. When for example a central spot is made of which the first circle of zeros is at radius half that of the Airy pattern, there occur side lobes with amplitudes several orders of magnitude larger than the central maximum and these occur at a radius of roughly 4 times that of the first ring of zeros (Fig. 5a). These large side lobes make the method unpractical.

In 2000 Pendry published his paper on the perfect lens [8]. The work was based on an earlier paper of the Sovjet scientist Veselago [9]. The lens consists of a slab made of material with negative permittivity and negative permeability, both opposite to the values of the ambient medium. It was derived that such a slab should work as a perfect lens and that an ideal image would occur inside and on the other side of the lens (Fig. 5b). The decrease of the evanescent amplitudes in the ambient medium is restored inside the slab. This may however easily lead to instabilities. Furthermore, at optical frequencies materials with the required properties do not seem to exist. The paper of Pendry nevertheless was the starting point of a lot of research on metamaterials, i.e. artificial materials with special optical properties that do not exist in nature.
A metamaterial that is interesting for super-resolution is the so-called hyperbolic material. It is an uniaxial anisotropic material with permittivity tensor with eigenvalues of different sign. In a uniaxial medium two types of plane waves exist called the ordinary and extra-ordinary waves, each having a particular linear state of polarisation. The ordinary waves behave in the normal manner in that there are propagating and evanescent ordinary waves. The extraordinary polarised waves, however, are special in that there are no evanescent extraordinary waves. No matter how large the spatial

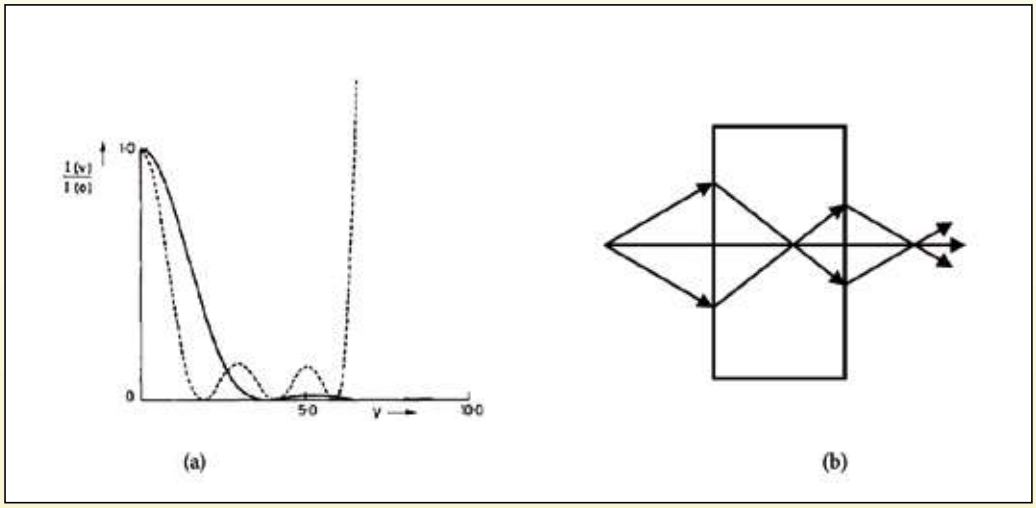

Fig. 5 a) Focused spot (dashed) generated by a Toraldo filter in the lens pupil. The central lobe is smaller than the Airy spot (continuous line) but large side lobes are caused (from [7]). b) Pendry's perfect lens made of negative permittivity and permeability (from [8]).b) Pendry's perfect lens made of negative permittivity and permeability (from [8]).

frequencies are, the wave vector always has three real components. The reason is that the wave vector components are for given frequency of the light on a hyperbolic surface in reciprocal space, instead of on a sphere as in the case of ordinary waves. Unfortunately, hyperbolic materials exist only in very limited frequency ranges in the optical domain and are absorbing [10].

3. New route to super-resolution

The discussion so far suggests that the classical resolution limit $\lambda/(2NA)$ is the maximum that can be achieved with far field optics without special nonlinear effects and without metamaterials. This is basically correct. However, under very special circumstances, better resolution can nevertheless be achieved. To explain why this is in principle possible we return to the approximate model in Eq. (1) which we have used for the interaction of the illumination with the object. In the study of super-resolution, this model is not accurate enough. From Maxwell's equations it follows that, for a thin object, Eq. (2) has to be replaced by

$$\mathcal{F}(U_s)(k_x,k_y) \propto \iint \mathcal{F}(\Delta\epsilon)(k_x - k'_x,\, k_y - k'_y)\mathcal{F}(U_{tot})(k'_x,k'_y)\mathrm{d}k'_x\mathrm{d}k'_y, \quad (3)$$

where $\mathcal{F}(\Delta\epsilon)$ is the Fourier transform of the difference $\Delta\epsilon(x,y)$ between the relative permittivity of the object, averaged over the thickness of the object, and the background. Furthermore, U_{tot} is the total field inside the object. Here the total electric field is again written as a scalar field. By comparing Eq. (3) with Eq. (2) we see that in the accurate theory the role of the reflection function is replaced by the permittivity contrast and that the incident field is replaced by the total field. The total field differs from the incident field by the scattered field, which is the field generated by multiple scattering by the dipoles in the sample. if only single scattering is considered, the total field becomes the incident field.

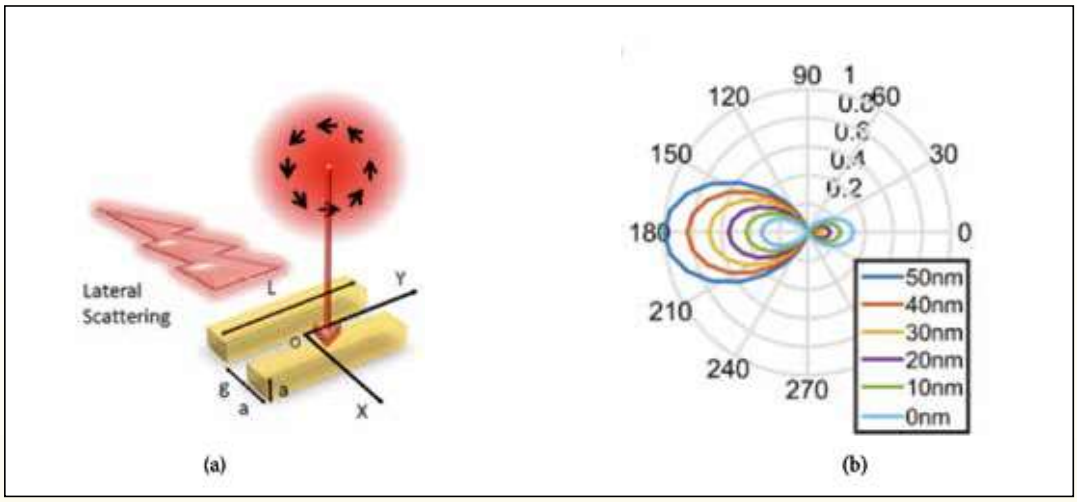

Fig. 6.
a) Two gold bars illuminated by an azimuthally polarised spot.
b) Asymmetric lateral scattering for different displacements.

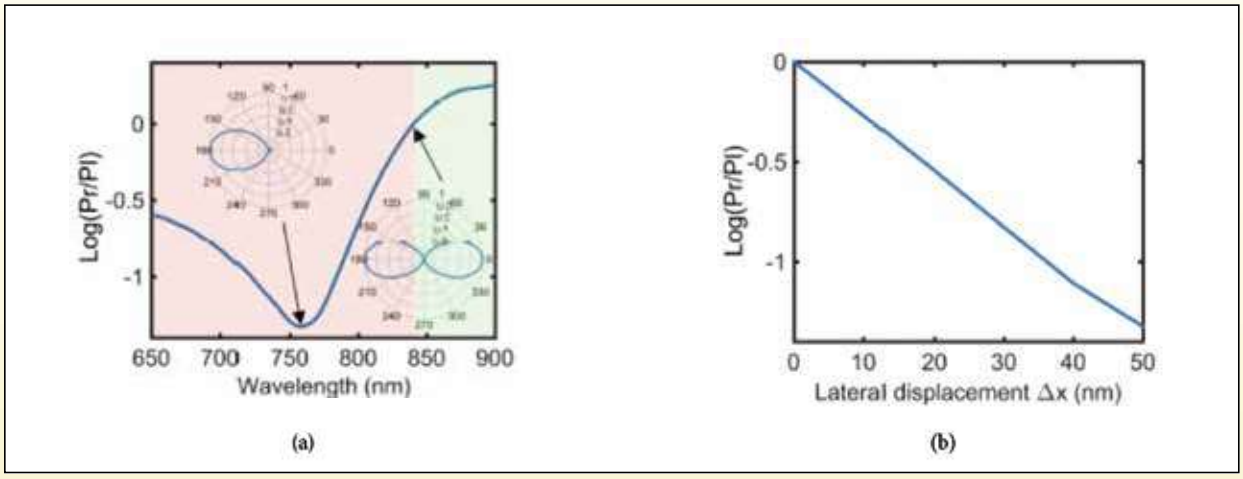

Fig.7. a) Asymmetric scattering as function of the wavelength. At 750 nm, where the asymmetry is largest, the plasmon resonance occurs. b) Log of the power ratio of left and right scattered power as function of the displacement.

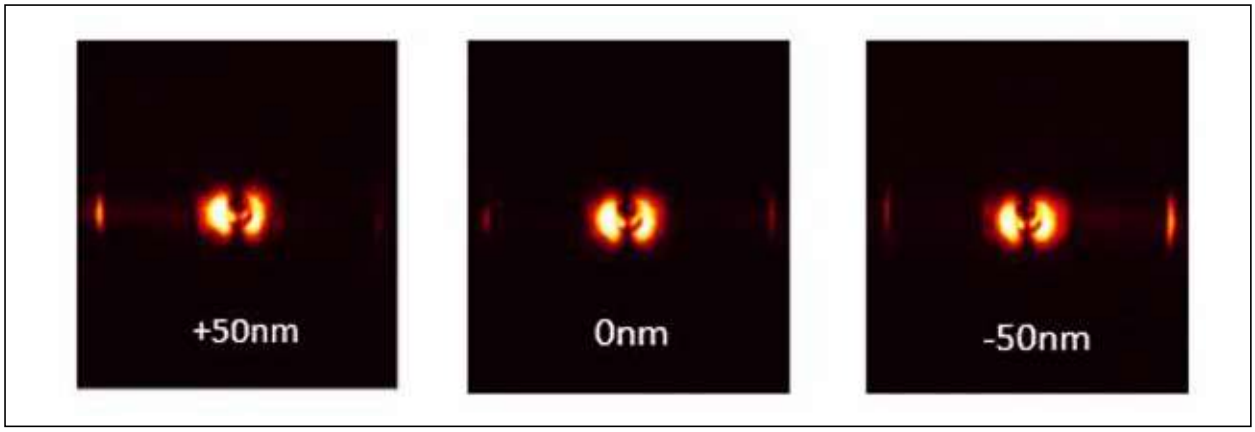

Fig. 8. Experimental demonstration of the asymmetric scattering. A small scatter is put to the left and the right of the nanobars to scatter the laterally scattered light into the camera.

Now, under special circumstances, in particular in the case of a resonance, multiple scattering is large. Then the total field can act as a nearby illumination with high spatial frequencies. Then high spatial frequencies of $\Delta\epsilon$ can influence the measured intensities in the far field by the folding principle explained after Eq. (2). A structure with metals can exhibit plasmon resonances having high local fields and high spatial frequencies. The incident field should be such that the resonance is excited. A priori knowledge about the sample should be available and then only a relatively small number of parameters of the sample can be retrieved with deep sub-wavelength accuracy. To illustrate this, we give two examples where both the structure and the illumination are optimised such that a deep subwavelength change of one or two parameters leads to drastic changes in the far field.

The first example uses a localised plasmon resonance. The structure consists of two golden nanobars each 50×50×100 nm with a gap of 50 nm between them [11]. The resonance wavelength is 750 nm. The structure is illuminated by a focused azimuthally polarised spot (Fig. 6a). This means that the electric field is linearly polarised parallel to the azimuthal direction and that the field at the centre of the spot is zero. When the centre of the focused spot is exactly halfway between the golden bars, the configuration is symmetric and just as much light is scattered laterally to the left as to the right. But at the resonance wavelength, the ratio between the powers scattered to the left and the right increases drastically when the spot is displaced over small distances (see Figs. 6b). Due to the displacement, one of the bars is in the dark centre of the spot and hence its local field is mainly excited by the scattered field of the other bar. The interference between the scattered fields of the two bars at resonance causes unilateral scattering (Fig. 7a). As shown in Fig. 7b, the logarithm of the ratio of the laterally scattered powers depends linearly on the displacement. Fig. 8 shows the experimental confirmation where two additional scattering structures are put on the substrate to the left and right of the nanobars to couple the laterally scattered light into the camera. For zero displacement the scattering is symmetric whereas for a displacement over 50 nm the scattering is strongly asymmetric. If the bars are rotated over a small angle, the scattering lobe rotate as well. This offers a highly sensitive method of determining rotation angles (Fig. 9). So both displacements and rotations can be determined very accurately from the scattered far field. Because the nanobars are small, the amount of light scattered is low. The scattered light can be enhanced by using an array of dimers. Such a configuration starts to work as a periodic grating which diffracts the light in narrower angular ranges with higher intensity.

As second example of structures of which small perturbations induce large changes in the far field, we discuss a silicon

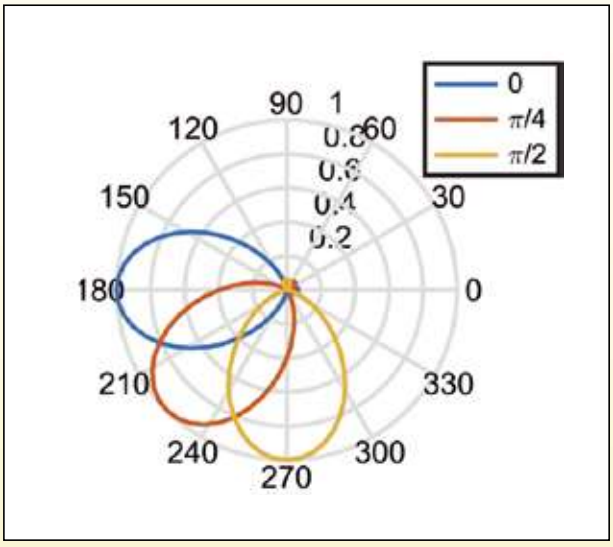

Fig. 9. Sensitivity of the far field to rotations.

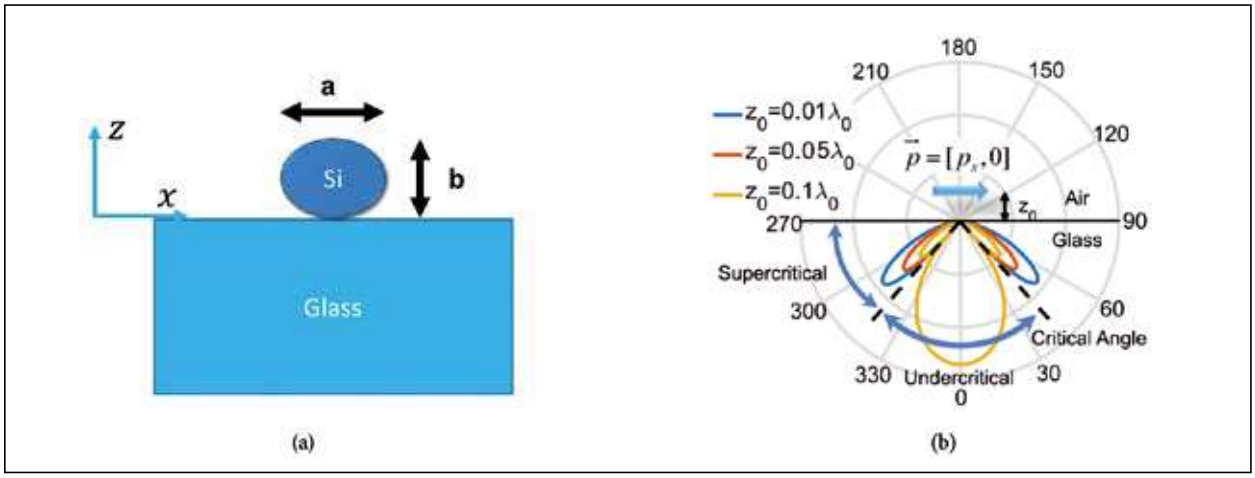

Fig. 10. a) Cylinder with elliptical cross-section on a glass substrate. b) Radiation pattern in the glass substrate when the incident field is a perpendicular x-polarised plane wave.

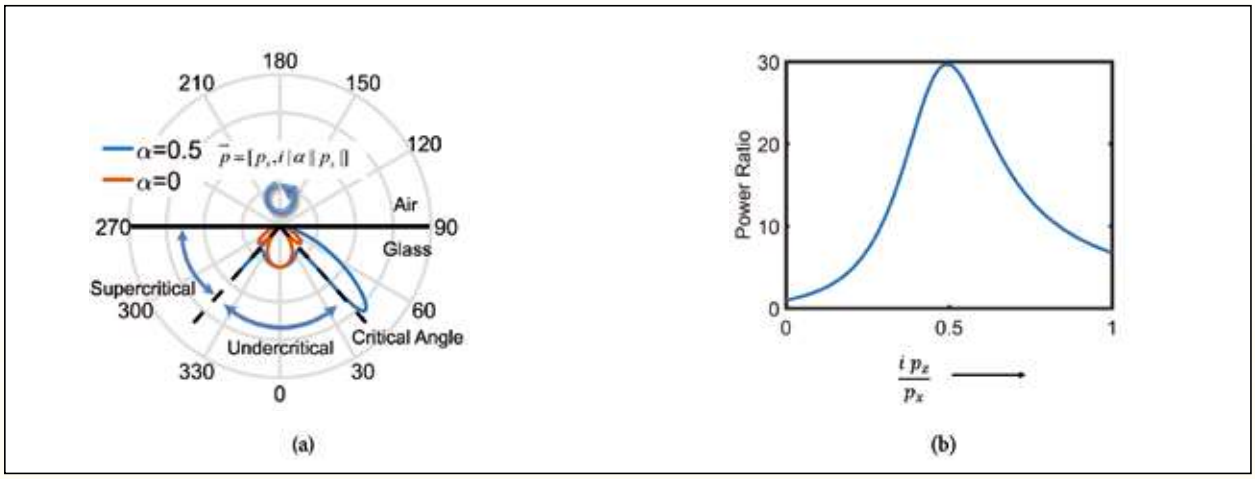

Fig. 11. a) Radiation pattern in the glass substrate when the incident field is a perpendicular x-polarised plane wave.
b) The asymmetry is maximum when $\frac{p_z}{p_x} = 0.5i$.

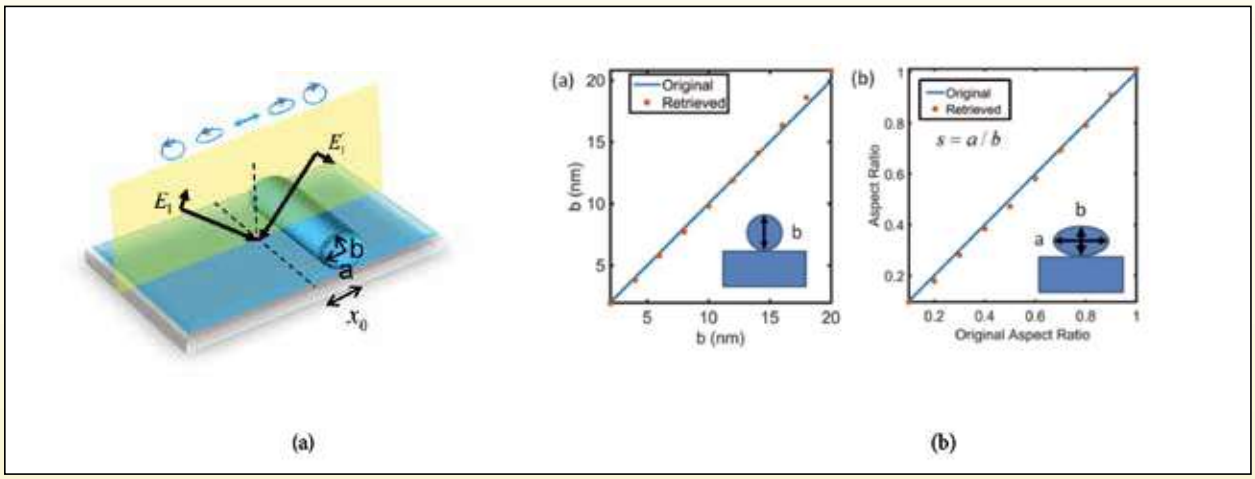

Fig. 12. a) Illumination using two p-polarised plane waves. In the standing wave pattern all kind of states of polarisation occur which can be used to probe the particle. b) Comparison between reconstructed b and $\frac{a}{b}$ (with in the latter case $b=10$ nm) and nominal values.

cylinder with ellipsoidal cross-section on top of a glass substrate as shown in Fig. 10a. The aim is to determine a and b from far field intensities measured through the glass substrate [12]. The wavelength used is 1000 nm and the refractive indices of silicon and glass are $n_{si} = 3.5$ and $n_g = 1.5$. As the cross-section is very small, the scattering can be explained by an electric dipole model. The retrieval of a and b is done in two steps. First the cylinder is illuminated from the air by a perpendicular incident plane wave polarised parallel to the x-direction. This implies that the dipole is polarised horizontally and that the scattering pattern as function of angle is symmetric. As is seen in Fig. 10b we distinguish two angular ranges: the under critical range: $-\theta_c < \theta < \theta_c$ and the super critical ranges: $\theta_c < |\theta| < \frac{\pi}{2}$, where θ_c is the angle with the normal of total internal reflection of a wave incident from the glass on the surface with air. The intensity radiated into the super critical angular ranges are excited by evanescent waves emitted by the dipole whereas the field radiated in the under critical range is excited by waves that are propagating in air. Therefore, the intensities of the former quickly decrease when the distance between the dipole and the surface is increased, while the intensities of the second hardly depend on this distance. By measuring the ratio of the power radiated in the super critical and under critical ranges, we can very sensitively determine the distance of the dipole to the interface and hence the value of b.

For an elliptically polarised dipole, the scattering is asymmetric. The amount of light scattered in the two super critical angular ranges: $-\frac{\pi}{2} < \theta < -\theta_c$ and $\theta_c < \theta < \frac{\pi}{2}$ differs and depends on the ellipticity (Fig.11). It turns out that maximum asymmetry occurs when the ratio of the components of the dipole is $\frac{p_z}{p_x} = 0.5i$. The ratio p_z/p_x can be expressed in terms of the product of the ratio of the polarisability in the z- and x-directions and the ratio of the z- and x-components E_z^i *and* E_x^i of the incident electric field that excites the dipole. It is an interesting observation that by using two p-polarised incident waves as shown in Fig. 12a, all kinds of states of polarisation occur in the standing wave. By scanning the standing wave over the particle and determining the position x where maximum asymmetry occurs in the super critical angle ranges, one can retrieve the ratio of the polarisations. This ratio can be expressed in terms of the permittivity of silicon and the ratio $\frac{a}{b}$. Because b has already been determined, a can now be found as well. The retrieved values for b and a/b are compared to the nominal values in Fig. 12b. It is seen that the accuracy is better than 5 nm, i.e. better than $\frac{\lambda}{200}$. ◆

References

[1] M-S Kim et al., "Submicron hollow spot generation by solid immersion lens and structured illumination", New Journal of Physics, **14** (2012).

[2] Lucas Novotny and Bert Hecht *Principles of Nano-Optics*, Cambridge University Press, Cambridge, UK. 2006.

[3] Stefan Hell and Jan Wichmann, "Breaking the diffraction resolution limit by stimulated emission: stimulated emission-depletion fluorescence microscpy", *Opt. Lett.*, **19**, No. 11 (1994).

[4] E. Betzig et al., "Imaging intracellular fluorescence proteins at nanometer resolution", *Science*, **313** (2006).

[5] G. Toraldo di Francia, "Super-gain antennas and optical resolving power", *Nuovo Ciomento Suppl.*, **9**, 426-438 (1952).

[6] R. Boivin and A. Boivin, "Optimized amplitude filtering for superresolution over a restricted field I. Achieving of maximum central irradiance under an energy constraint", *Opt. Acta*, **27**, 1641-1670 (1980).

[7] I.J. Cox, C.J.R. Sheppard and T. Wilson, "Repraissal of arrays of concentric annuli as superresolving filters", *JOSA Letters*, **27**, No. 9 (1982).

[8] J.B. Pendry, "Negative refraction makes a perfect lens", *Phys. Rev. Lett.*, **85**, No. 18 (2000). [9] V.G. Veselago, *Sov. Phys. Usp.*, 10, 509 (1968).

[10] Karolina Korzeb, Marcin Gajc, and Dorota Anna Pawlak, "Compendium of natural hyperbolic materials", *Opt. Express* **23**, 25406-25424 (2015).

[11] Zheng Xi, Lei Wei, A.J.L. Adam and H.P. Urbach, "Accurate feeding of nanoantenna by singular optiics for nanoscale translationala nd rotational displacement sensing", *Phys. Rev. Lett.*, **117**, 113903 (2016).

[12] Xheng Xi and H.P. Urbach, "Retrieving the size of deep-subwavelength objects via tunable optical spin-orbit coupling", *Phys. Rev. Lett.*, **120**, 253901 (2018).