CONFERENCES AND SYMPOSIA PACS numbers: 01.10 Fv

Scientific session of the Division of General Physics and Astronomy of the Russian Academy of Sciences (26 November 1997)

A scientific session of the Division of General Physics and Astronomy of the Russian Academy of Sciences was held on 26 November 1997 at the P L Kapitza Institute for Physical Problems, RAS. The following reports were presented at the session:

(1) **Zhdanov G S, Libenson M N, Martsinovskiĭ G A** (S I Vavilov State Optical Institute, St Petersburg) "Optics in the diffraction limit: principles, results, and problems";

(2) **Zabrodskiĭ A G** (A F Ioffe Physicotechnical Institute, Russian Academy of Sciences, St Petersburg) "Coulomb gap and metal – insulator transitions in doped semiconductors".

Brief presentations of both reports are given below.

PACS number: **07.79.–v**

Optics in the diffraction limit: principles, results, and problems

G S Zhdanov, M N Libenson, G A Martsinovskiĭ

Lately there has been an upsurge of interest in the possibility of studying and forming nanometer structures by optical methods. This possibility, which not so long ago seemed highly hypothetical, emerged in connection with the development of near-field optics (NFO), a new scientific and technological line of research which, from the physical viewpoint, is based on the presence in the far zone (the Fraunhofer region) of traces of interaction of the light and the object in the near field. Technically, NFO is a combination of elements of ordinary optics and scanning probe microscopy (SPM). A distinctive element of near-field devices is an optical probe, which is usually a fiber with a sharpened end, *1* whose surface is covered, except for the tip of the cone, with an opaque metal layer, *2* (Fig. 1).

A fraction of the light flux traveling along the fiber passes through a diaphragm in a metal screen and reaches the sample, which is in the 'near field' (NF) of the source. If the distance z to the sample surface and the radius a of the aperture are much smaller than the wavelength λ of the light, $a, z \ll \lambda$, the size of the light spot on the sample is close to the size of the aperture. When the probe is moved along the sample, the resolution may not be restricted by diffraction (such resolution is known as 'super-resolution').

Uspekhi Fizicheskikh Nauk **168** (7) 801 – 808 (1998)
Translated by E Yankovsky; edited by A Yaremchuk

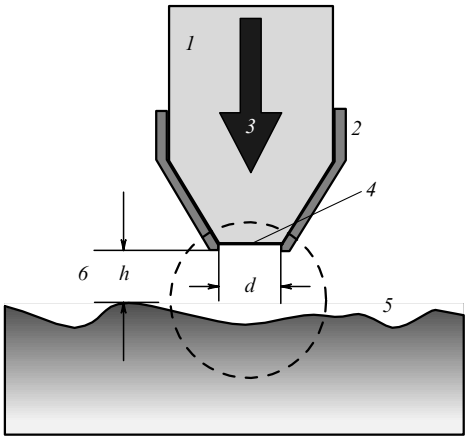

Figure 1. Design of an aperture NSOM: *1*, sharpened optical fiber; *2*, metallic coating; *3*, light transmitted by the probe; *4*, output probe aperture, $d \ll \lambda$; *5*, sample surface; and *6*, the distance h between the surface under investigation and the probe aperture, $h \ll \lambda$. The dashed circumference indicates the region of near-field contact.

Although such an idea was proposed as long ago as 1928 by Syngh [1], it was far ahead of its time and was practically left unnoticed. Its first confirmation was obtained in 1972 in microwave experiments carried out by Ash and Nichols [2]. At the beginning of the 1980s Pohl, Denk, and Lanz of the IBM Zurich Research Laboratory 'penetrated' the diffraction limit and demonstrated a resolution of $\lambda/20$ on a device operating in the visible optical range, which became known as a 'near-field scanning optical microscope' (NSOM) [3]. Somewhat earlier the first scanning tunneling microscope (STM) was developed in this same laboratory, which made it world-famous.

In contrast to the tunneling microscope and the atomic-force microscope (AFM), which immediately won wide popularity, the NSOM was left in the shade for some time. The unique possibilities of NSOMs were fully realized only at the beginning of the 1990s, after two very important technical problems had been solved: the energy efficiency of the probes was raised considerably, and a reliable scheme for monitoring the distance between the probe tip and the sample was developed. Today NSOMs are successfully used in dozens of laboratories for solving a broad range of problems of the physics of surfaces, biology, various techniques for recording and retrieving information, and the like. In 1993 the commercial production of NSOMs began.

There are about 20 types of NSOM, differing in optical systems and functional design. Depending on the presence or absence of an aperture at the tip of the probe, NSOMs can be

classified into two types: 'aperture,' and 'non-aperture.' The design of an aperture NSOM, which is the most common type currently in operation, is illustrated in Fig. 1 by the block diagram of such a microscope [4].

The laser beam passes through a matching device, enters the sharpened metallized fiber, and at the exit narrows to the aperture size. The mutual motion of tip and sample in three dimensions, x, y, and z, is performed by piezoelectric propulsive devices. The photons that have passed through the sample or have been reflected and scattered are detected by one of the two microscope objectives (*2* or *1*, respectively, in Fig. 2) and are directed to the photomultiplier (PM). Usually such a microscope objective is a component of an ordinary optical microscope, which makes it possible to select the section for observation and to fix this section in relation to a broader field. This scheme usually applies to devices operating in the illumination mode. Devices that operate in the photon collection mode are also widespread. In this mode the probe carries the photons from the sample illuminated, say, through a microscope objective to the detector. In the combined (illumination/collection) mode the probe performs both functions simultaneously.

To place the tip at a required height above the sample, all scanning probe microscopes use the dependence of the intensity I of the signal being registered on z. In the majority of NSOM types the function $I(z)$ is multivalued since, in addition to a near-field signal I_1 there is also a signal I_2 caused by the interference of the incident and reflected waves, and this signal is a periodic function of z. This hinders reliable monitoring of z by the value of $I = I_1 + I_2$ as the tip moves closer to the sample (or even makes such monitoring impossible). The best way to solve this problem is to add to NSOM auxiliary devices, which makes it possible for the microscope to combine the functions of an STM and an AFM, and in this case determining z is easy. In such combined devices the imaging is done through two channels simultaneously, one of which reproduces the surface pattern and the other, the distribution of the refractive index. The possibility of distinguishing between optical and topographic contrasts considerably simplifies the problem of interpreting the image. The most popular method of monitoring z is by varying the shear force with which the tip acts on the sample [5].

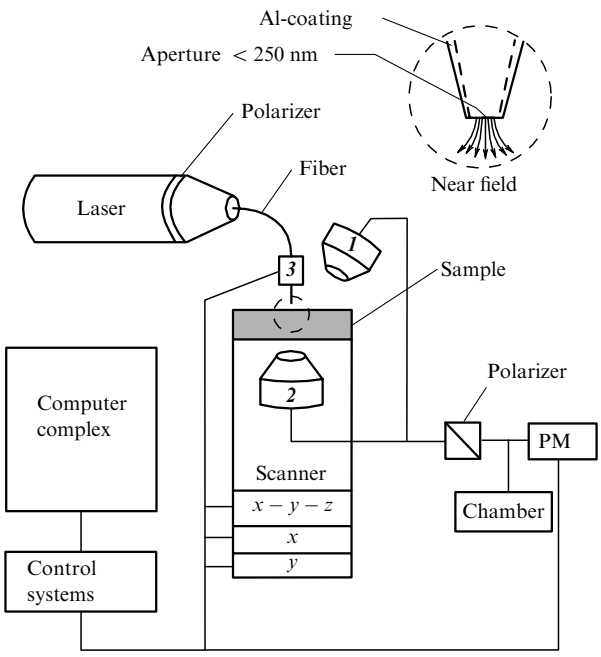

Figure 2. Block diagram of a near-field microscope: *1*, microscope objective operating in reflected light; *2*, microscope objective operating in transmitted light; and *3*, piezoelectric propulsive device for moving the probe. Other notation is indicated in the figure. The dashed circumference indicates the region of near-field contact.

The basic characteristic of an NSOM is its spatial resolution, which largely depends on the conditions of illumination or, in the more general case, observation of the sample, the structure of the sample surface, and the microgeometry of the probe. As is known, the pulse response function of a diffraction-limited optical system is described by the Airy distribution. The halfwidth of the principal maximum of the distribution corresponds to Rayleigh's resolution, $\Delta x = 0.61\lambda/\sin\varphi$, where φ is the aperture angle. In the $\varphi \to \pi/2$ limit, $\Delta x \to \Delta x_{\min} = 0.61\lambda$. As the light travels through the small aperture, the spatial frequency spectrum is distorted due to scattering, and this distortion can be described by the Airy distribution (just as the smearing of a point can be described by the Airy distribution). The physically natural restriction on a perturbation provided by the aperture size is accompanied by a broadening Δf of the spatial frequency spectrum, with $\Delta f = 0.61/a$. As $a \to 0$, the wave field directly after the aperture contains arbitrarily large spatial frequencies, i.e. $\Delta x_{\min} \to 0$. In a real situation, due to the finite conductivity of the metal screen (coating), the minimum effective radius of the aperture is determined by the thickness δ of the skin layer. Allowing for all this, the expected highest resolution for a probe with an aluminum coating in the visible spectral range is $\Delta x_{\min} \approx 2\delta \approx 13$ nm, which agrees with the best experimental data of Betzig et al. [6]. The absence of physical restrictions on the size of the probe's tip in non-aperture NSOMs allows for a resolution better than 1 nm [7].

The Rayleigh criterion is an illustration of the Heisenberg uncertainty principle, according to which any attempt to raise the degree of localization or the accuracy Δx in determining the position of the light sources leads to an increase in the indeterminacy Δp_x in the conjugate photon momentum. When the photons are scattered within the maximum range of angles $-\pi/2 \leqslant \varphi \leqslant \pi/2$, we have $\Delta p_x = \hbar\Delta k_x = 4\pi\hbar/\lambda$ (here $\hbar$ is Planck's constant, and k_x is the x-component of the wave vector k) and $\Delta x \geqslant \lambda/2$. It would seem that the possibility of a resolution $\Delta x \ll \lambda/2$ contradicts a basic physical principle. However, one must bear in mind that the uncertainty relation in its most general form refers to the position of a particle in momentum-coordinate space. Hence, while limiting one of the components of the wave vector, it still allows variation of the other components. For instance, we can put $k_y = 0$ and $k_z = -\mathrm{i}\gamma$, where γ is a positive real number. Then $k_x = (k^2 - k_z^2)^{1/2} = (k^2 + \gamma^2)^{1/2} > k$. As $\gamma \to \infty$, the range of admissible values of k_x broadens without limit and Δx may be made as small as desired.

Imaginary values of k_z correspond to damped waves. Hence, for super-resolution to manifest itself, the antenna–probe must be within the damped field near the sample surface, i.e. certainly at $z \leqslant \lambda$.

Now we can refine the concept of the near field, associating it with the region where there are damped (and, hence, non-radiative) waves whose amplitude varies with the distance z from the interface between the media, or a small

scatterer providing the law $E(z) = E(0)\exp(-\gamma z)$, where $\gamma > 0$. The quantity γ^{-1} is the penetration depth for the damped wave and is equal in order of magnitude to the subwave scatterer. In particular, for an aperture of radius a in a thin conducting screen, $\gamma^{-1} \approx 2a$. For a surface with a complex relief the value of γ^{-1} is determined by the total contribution of the components of the spatial frequency spectrum, with the mth component of period $d_m \ll \lambda$ detectable at a distance $z \leqslant \gamma_m^{-1} \approx d_m/2\pi$. (In the photon collection mode the accuracy with which the surface's profile is reproduced increases with the number m of components of the damped field participating in image formation, which means it increases as z gets smaller.) In the Fraunhofer region, where $z \geqslant \lambda$, there are only ordinary propagating waves, to which the laws and restrictions of classical optics can be applied.

Hans Bethe [8] was the first to solve the problem of light passing through an aperture of radius $a \ll \lambda$ in an infinitely thin conducting screen. He found that an aperture of such dimensions transmits much less light than expected from extrapolating the results of calculations for $a > \lambda$. In particular, the scattering cross section σ of unpolarized light is linked to a and the wave number $k = 2\pi/\lambda$ through the following relationship:

$$\sigma = \frac{64}{27\pi} k^4 a^6 \left(1 - \frac{3}{8}\sin^2\theta\right), \qquad (1)$$

where θ is the angle of incidence of the light.

In addition to having a different numerical factor, the expression on the right-hand side of Eqn (1) differs from that used in calculations by the Kirchhoff method by having an additional factor $(ka)^2 \ll 1$. The reason is that a considerable fraction of the electromagnetic energy is converted into a nonradiative form, which cannot be recorded by a distant observer. From the viewpoint of such an observer, the aperture scatters the light as if it were a pair of mutually perpendicular dipoles: an electric dipole directed along the aperture axis, and a magnetic dipole, whose moments are, respectively,

$$\boldsymbol{\mu}_e = \frac{a^3}{3\rho} E_0\,, \quad \boldsymbol{\mu}_m = -\frac{2a^3}{3\rho} H_0\,, \qquad (2)$$

where E_0 and H_0 are the electric and magnetic fields in front of the screen. A sphere of radius a with a dielectric constant $\varepsilon \approx 2$ may also act as an aperture.

The rapid decay (as $I \sim a^6$) of the detected signal I as the radius a of the aperture or the probe tip gets smaller is the main reason why achieving the highest resolution in NSOMs is so difficult. The most commonly used apertures have $a \approx 50$ nm, a value stemming from a compromise between resolution and the admissible signal-to-noise ratio. Here the transmission coefficient or efficiency of the probe amounts to $\sim 10^{-6}-10^{-4}$, if one allows for the losses in the probe's conical part in front of the aperture.

The somewhat more exact formulas obtained by Bethe and later by other researchers, formulas that describe the field's distribution directly in front of the aperture ($z \to +0$) are still the only analytic representation of the near field. Lately the focus has been on developing numerical methods in which the space near the tip of the probe is partitioned into cells and a solution of the electrodynamic problem of propagation of a perturbation from one cell to the others is sought [9]. The resolution of the mathematical models is determined by the cell size, i.e., in the final analysis, by a reasonable time for the calculations to be completed, and amounts to roughly 1 nm.

Computer simulation has made it possible to graphically represent the field's structure not only in the gap between the probe and sample but in the part of the probe adjacent to the tip and close to the cut-off section of the fundamental wave mode. It has been found, in particular, that the distribution of the wave in the probe is characterized by a complex combination of standing waves that emerge because of multiple reflections of light off the cone's surface.

The possibility of increasing by a factor of ten or more the degree to which the optical methods used in surface studies are local is important in solving a broad range of scientific and applied problems. When analyzing the interaction of light and an inhomogeneous surface by methods of classical optics one is forced to average the effects of many defects within the limits of the illuminated section. Using NSOMs facilitates the study of individual inhomogeneities of nanometer dimensions. The discovery of 'one-particle plasmons' generated by light in metallized latex spheres was the first confirmation of this feature.

Among the objects for which the problem of localization of optical analysis is most important are heterostructures with quantum-size properties. An NSOM can be used not only to localize individual luminescence centers, which is a problem in its own right, but also to separate the spectra of such centers [10]. Such studies provide valuable information about the structural features of the system, including the roughness (at the atomic level) of the interface and the mechanism of diffusion and decay of excitons. Studies of the induced-photocurrent effect in NSOMs make it possible to expose surface defects in semiconductor samples with a resolution almost ten times higher than the resolution of the widely used OBIC (optical beam induced current) and EBIC (electron beam induced current) methods.

The possibility of discovering, with NSOMs, luminescent markers capable of attaching themselves selectively to various elements of intercellular structure or sections of DNA is of great interest to biologists. The objective of reducing the size of the markers to the molecular-size limit requires greater precision from the device.

The ability of near-field optical microscopy to register individual fluorine molecules was first demonstrated by Betzig and Chichester [11] and was later corroborated by other researchers. Multiple scanning of a surface makes it possible to monitor the dynamics of processes associated with changes in the position of molecules, the orientation of the molecules in space, the strength of the bonds of the molecules with the ambient matrix, etc., including the case of pulsed irradiation with nano- and picosecond time resolution.

NFO methods make it possible not only to study surfaces with a high degree of localization but also modify the surface structure. The characteristic size of the patterns that an NSOM can imprint on various samples amounts to 50–70 nm, which is much smaller than patterns imprinted by ordinary optical devices. The possibility of increasing the resolution in photolithography severalfold and raising the information storing density by a factor of ten or more, e.g. in magneto-optical media, is extremely promising and is stimulating extensive research into these problems. However, the transition from laboratory tests to the development of industrial technologies is hampered by the slow rate at

which a pattern is imprinted as the surface is scanned by the probe. The required scanning rate is related to the illuminating power, which is limited by the probe's thermal stability. As noted earlier, in typical conditions only one-millionth of the light flux reaches the sample, while the main fraction is absorbed by the metallic coating of the probe and heats it up. Kurpas et al. [12] found that the temperature distribution in the probe strongly depends on the probe's microgeometry and the structure of the field near the probe's tip. Conditions may be such that the most heated region is far from the tip. The calculated ratio of the maximum rise in the probe's temperature, ΔT, to the absorbed power P for ordinary aperture angles of the cone is $\Delta T/P \approx 10^5$ K W^{-1}, which agrees with the results of measurements. At $P \sim 10$ mW or at $P_0 \sim 10$ nW for the radiation that has reached the sample, the probe may disintegrate because of melting of the aluminum coating.

A considerable increase in P_0 can be achieved if the standard quartz probes with an aluminum coating are replaced by all-metal probes. Gurevich and Libenson [13] suggested using metal for the rod, the light being fed to the tip of the rod by exciting a cylindrical surface electromagnetic wave (SEW). This eliminates the difficulties caused by the field cut-off in a sharpened quartz fiber probe and the associated large energy losses. Analysis shows that an SEW field at the probe's tip is concentrated within a region comparable in size to the diameter of the tip.

When examining the interaction of light and matter in the near-field contact region (see Fig. 1), one must bear in mind that the mean free path of the non-equilibrium carrier generated in the process of light absorption and the size of the effective interaction zone may be much larger than the size of the light spot. The zone is formed as a result of a cascade of processes in which the electron and phonon subsystems of the sample participate [14, 15]. In the steady state, the maximum rise in temperature in this zone is a function of the parameter $\zeta = R/\sqrt{D\tau}$, where R is the radius of the light spot, and D and τ are the diffusion coefficient and the lifetime of the non-equilibrium carriers.

The range of applications of NFO is rapidly broadening. A number of new tracks of research are in the stage of idea formulation or in the experimental stage. One of these ideas is related to the possibility of using the NFO method to control the elements of high-power optics [16]. Usually, optical breakdown of the materials and elements of optics is initiated by defects whose nature is not always known. The most natural approach to the detection of such defects consists in analyzing the surface and thin layers by radiation with the same frequency as that of the high-power light. The possibility of visualizing small optical inhomogeneities and doing a spectral analysis of these inhomogeneities in an NSOM suggests that using this device constitutes an effective way of solving the problem.

References

1. Synge E H *Philos. Mag.* **6** 356 (1928)
2. Ash E A, Nichols G *Nature* (London) **238** 510 (1972)
3. Pohl D W, Denk W, Lanz M *Appl. Phys. Lett.* **44** 651 (1984)
4. Cline J A, Isaacson M *Appl. Opt.* **34** 4869 (1995)
5. Betzig E, Finn P L, Weiner J S *Appl. Phys. Lett.* **60** 2484 (1992)
6. Betzig E et al. *Science* **251** 1468 (1991)
7. Zenhausern F, Martin Y, Wickramasinghe H K *Science* **269** 1083 (1995)
8. Bethe H A *Phys. Rev.* **66** 163 (1944)
9. Novotny L, Pohl D W, Regli R *J. Opt. Soc. Am. A* **11** 1768 (1994)
10. Hess H F et al. *Science* **264** 1740 (1994)
11. Betzig E, Chichester R J *Science* **262** 1422 (1993)
12. Kurpas V, Libenson M, Martsinovsky G *Ultramicroscopy* **61** 187 (1995)
13. Gurevich V, Libenson M *Ultramicroscopy* **57** 277 (1995)
14. Libenson M N, Martsinovsky G A *Proc. SPIE* **2714** 305 (1995)
15. Guzovskiĭ Yu G, Libenson M N, Martsinovskiĭ G A *Izv. Ross. Akad. Nauk, Ser. Fiz.* **61** 1301 (1997) [*Bull. Russ. Acad. Sci., Fiz.* **61** 1014 (1997)]
16. Gruzdev V E, Libenson M N, Martsinovskiĭ G A *Poverkhnost'* (2) 37 (1998)

PACS number: 71.30.+h

Coulomb gap and metal – insulator transitions in doped semiconductors

A G Zabrodskiĭ

1. Gapless models of localization and hopping transport

Doped semiconductors belong to disordered systems that are widely used in studying problems associated with the metal – insulator transition and low-temperature electron transport. At a certain critical value N_c of the concentration of the primary impurities, 'metallization' of the impurity levels near the Fermi level occurs in such systems. But when $N > N_c$, the reverse metal – insulator transition may be induced by introducing a compensating impurity, which captures the primary charge carrier. The compensation lowers the Fermi level and introduces what is known as 'vertical' disorder into the system in addition to the 'horizontal' disorder due to the random distribution of the impurities.

Since the 1970s, the common approach to describing such a metal – insulator transition is to use the one-electron Anderson model, in which a random field leads to localization of the states near the Fermi level [1] when the level coincides with the mobility edge. This model makes it possible, among other things, to explain why a compensated semiconductor with a partially filled ground-state impurity band does not have metallic conduction at an arbitrarily small impurity concentration N. What is important to the discussion below is that in the Anderson model the density of states $g(E)$ has no singularity and remains finite near the Fermi level. In this sense we call this model 'gapless,' because the insulator state in it corresponds not to a gap near the Fermi level but to what became known as the 'mobility gap' between the Fermi level and the mobility edge.

According to Mott [2], in the insulator state with a finite density of states near the Fermi level E_F, low-temperature electron transport occurs via tunnel hopping near the Fermi level with variable activation energy and range (variable range hopping, or VRH) and is accompanied by emission or absorption of phonons, which in the three-dimensional case yields the well-known $T^{-1/4}$-law for the electrical conductivity,

$$\rho \propto \exp\left(\frac{T_0}{T}\right)^x, \qquad (1)$$

where $x = 1/4$, the parameter $T_0 \propto \left[a^3 g(E_F)\right]^{-1}$, and a is the localization length.