Scuola di Dottorato in Ingegneria Leonardo da Vinci – a.a. 2009/10

PROPRIETÀ MECCANICHE, OTTICHE, ELETTRONICHE DEI MATERIALI ALLE PICCOLE E PICCOLISSIME SCALE

Versione 1 – Ottobre 2010 – http://www.df.unipi.it/~fuso/dida

Parte 9
Analisi delle proprietà ottiche a scala piccola e piccolissima

Me 20.10 12.30-15.30 aula A DCCI

Outlook

- Optical properties at the small scale can be analyzed through (conventional) optical microscopy
- Origin and limitations of the optical microscopy are associated with fundamental aspects such as, numerical aperture and diffraction
- Analysis at the ultra-small scale requires to overcome diffraction effects, that can be attained by scanning near field optical microscopy (SNOM)
- Basics, a few details and examples of the powerful SNOM

Da Hecht Zajac
Optics
Addison-Wesley (1974)

Reminders of optical microscopy

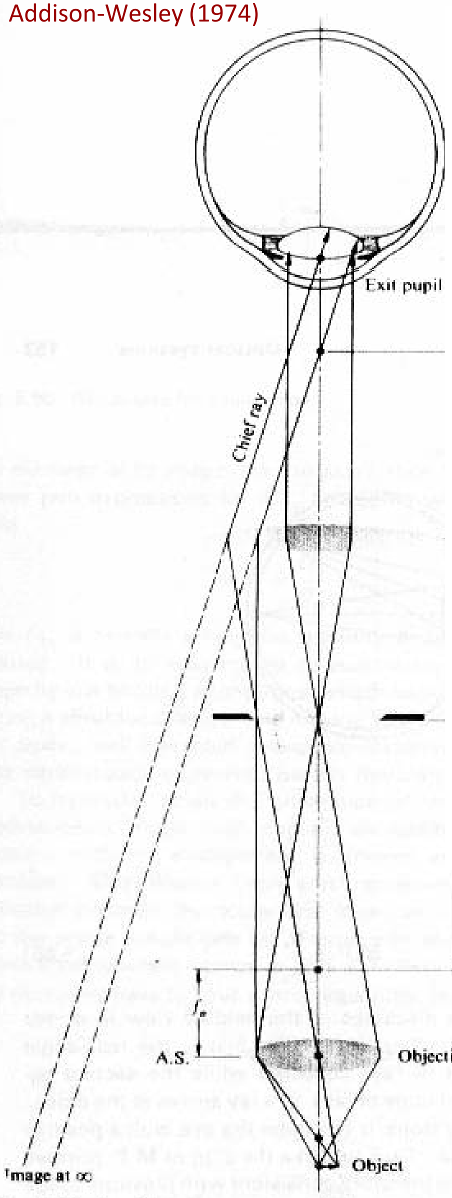


Fig. 5.86 A rudimentary compound microscope.

the magnifying power of the entire system is the product of the transverse linear magnification of the objective, M_{T_o}, and the angular magnification of the eyepiece, M_{A_e}, that is

$$\text{M.P.} = M_{T_o} M_{A_e}. \qquad (5.70)$$

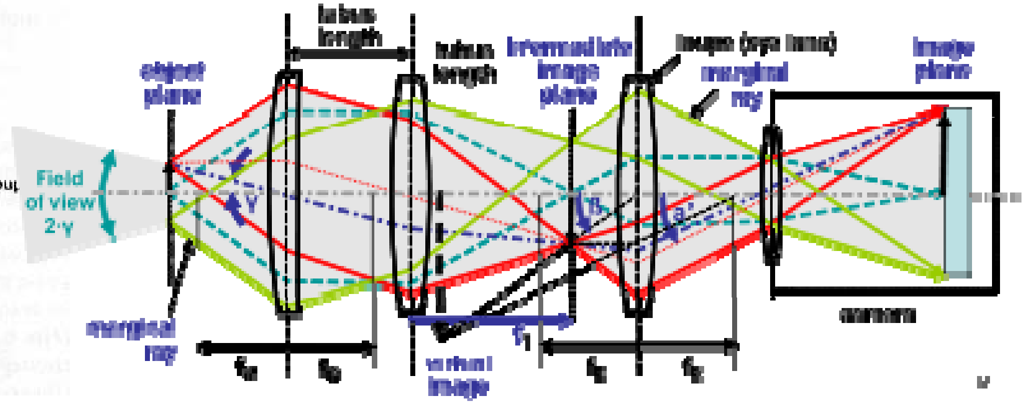
Recall that $M_T = -x_i/f$, (5.26), and with this in mind most, but not all, manufacturers design their microscopes such that the distance (corresponding to x_i) from the second focus of the objective to the first focus of the eyepiece is standardized at 160 mm. This distance, known as the *tube length*, is denoted by L in the figure. (Some authors define tube length as the image distance of the objective.) Hence, with the final image at infinity and the standard near point taken as 10 inches or 254 mm

$$\text{M.P.} = \left(-\frac{160}{f_o}\right)\left(\frac{254}{f_e}\right) \qquad (5.71)$$

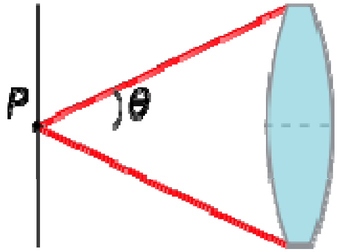
and the image is inverted (M.P. < 0). Accordingly, the barrel of an objective with a focal length f_o of say 32 mm will be engraved with the markings 5× (or ×5) indicating a power of 5. Combined with a 10× eyepiece (f_e = 1 inch) the microscope M.P. would then be 50×.

Magnification depends on lens features (focal length)

Arbitrary magnification *seems* feasible (by choosing arbitrarily short focal lengths for the optical components)



Numerical aperture



Great NA → large lens with a small focal length

In most areas of optics, and especially in microscopy, the numerical aperture of an optical system such as an objective lens is defined by

$$\mathrm{NA} = n \sin\theta$$

where n is the index of refraction of the medium in which the lens is working (1.0 for air, 1.33 for pure water, and up to 1.56 for oils), and θ is the half-angle of the maximum cone of light that can enter or exit the lens. In general, this is the angle of the real marginal ray in the system. The angular aperture of the lens is approximately twice this value (within the paraxial approximation). The NA is generally measured with respect to a particular object or image point and will vary as that point is moved.

In microscopy, NA is important because it indicates the resolving power of a lens. The size of the finest detail that can be resolved is proportional to λ/NA, where λ is the wavelength of the light. A lens with a larger numerical aperture will be able to visualize finer details than a lens with a smaller numerical aperture. Lenses with larger numerical apertures also collect more light and will generally provide a brighter image.

Numerical aperture is used to define the "pit size" in optical disc formats [1]

Numerical aperture versus f-number

[edit]

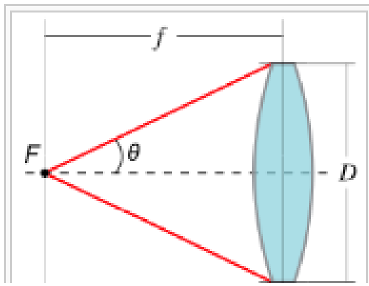
Numerical aperture is not typically used in photography. Instead, the angular acceptance of a lens (or an imaging mirror) is expressed by the f-number, written $f/\#$ or N, which is defined as the ratio of the focal length to the diameter of the entrance pupil:

$$N = f/D$$

This ratio is related to the numerical aperture with respect to the focal point of the lens. Based on the diagram at right, the numerical aperture of the lens in air is:

$$\mathrm{NA} = n\sin\theta = n\sin\arctan\frac{D}{2f} \approx \frac{D}{2f}$$

thus $N \approx \dfrac{1}{2\,\mathrm{NA}}$.



This approximation holds when the numerical aperture is small. The f-number describes the light-gathering ability of the lens in the case where the marginal ray

NA and DOF

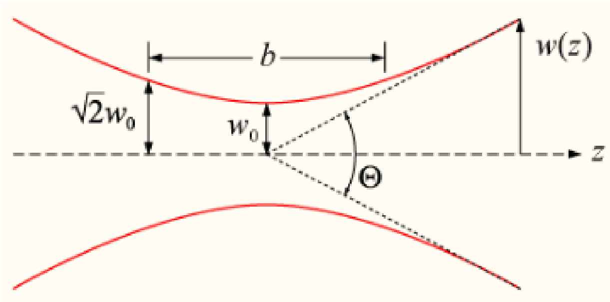
In optics and especially laser science, the **Rayleigh length** or **Rayleigh range** is the distance along the propagation direction of a beam from the waist to the place where the area of the cross section is doubled.[citation needed]

For a Gaussian beam, the Rayleigh length is given by

$$z_R = \frac{\pi w_0^2}{\lambda},$$

where w_0 is the radius of the beam at the waist. At distance z_R from the beam waist, the beam radius is increased by a factor $\sqrt{2}$.

A related parameter is the **confocal parameter** b, which is just two times the Rayleigh length:



Beam width or "spot size"

For a Gaussian beam propagating in free space, the spot size $w(z)$ will be at a minimum value w_0 at one place along the beam axis, known as the *beam waist*. For a beam of wavelength λ at a distance z along the beam from the beam waist, the variation of the spot size is given by

$$w(z) = w_0 \sqrt{1 + \left(\frac{z}{z_R}\right)^2}.$$

where the origin of the z-axis is defined, without loss of generality, to coincide with the beam waist, and where

$$z_R = \frac{\pi w_0^2}{\lambda}$$

is called the Rayleigh range.

The propagation of a Gaussian beam is fully specified by its beam waist and its divergence. For an ideal TEM_{00} beam, the product of the beam waist ω_0 times the divergence angle θ_0 can be expressed as

$$\omega_0 \theta_0 = \lambda/\pi$$

.3 Depth of Field and Depth of Focus

ce the resolution available for an object in focus in the image plane is limited by numerical aperture of the objective lens, it follows that the object need not be at exact object distance from the lens u, but may be displaced from this plane hout sacrificing any resolution (Fig. 3.12). The distance over which the object ains in focus is defined as the *depth of field*,

$$d = \delta/\tan\alpha \tag{3.2}$$

ure 3.12 Since the resolution is finite, the object need not be in the exact object-plane tion in order to remain in focus, and there is an allowed depth of field d. Similarly, the ge may be observed without loss of resolution if the image plane is slightly displaced, so there is an allowed depth of focus D

where α is half the angle subtended by the objective aperture at the focal point. Similarly, the image will remain in focus if it is displaced from its geometrically defined position at a distance v from the lens. The distance over which the image remains in focus is termed the *depth of focus*, as follows:

$$D = M^2 d \tag{3.3}$$

where M is the magnification. (Both of these expressions (equations (3.2) and (3.3)) are approximate and assume that the objective can be treated as a 'thin lens', which is never the case in a commercial instrument.) Since the resolution is given by $\delta = 0.61\lambda/\mu \sin\alpha = 0.61\lambda/NA$, it follows that the depth of field decreases as the numerical aperture increases. For the highest image resolution, the specimen should be positioned to an accuracy of better than 0.5 μm, which determines the required mechanical stability of the specimen stage.

The depth of focus is considerably less critical. Bearing in mind that a magnification of the order of 100 is necessary if all of the resolved detail is to be recorded, displacements of the order a millimetre are acceptable.

Da Brandon Kaplan
Microstruct. Charact.
of Materials
Wiley (1999)

Depth of field decreases as numerical aperture increases (related to *resolution*)

In lithography:
High magniication → short focal lenghts → increased *NA* → decreased *d* → thin resist needed!

Da Hecht Zajac, Optics
Addison-Wesley (1974)

Reminders on interference and diffraction

10.1.3 Several Coherent Oscillators

As a simple yet logical bridge between the studies of interference and diffraction, consider the arrangement of Fig. 10.6. The illustration depicts a linear array of N coherent point oscillators (or radiating antennas), which are each identical even to their polarization. For the moment, consider the oscillators to have no intrinsic phase difference, i.e. they each have the same epoch angle. The rays shown are all almost parallel, meeting at some very distant point P. If the spatial extent of the array is comparatively small, the separate wave amplitudes arriving at P will be essentially equal, having traveled nearly equal distances, that is

$$E_0(r_1) = E_0(r_2) = \cdots = E_0(r_N) = E_0(r).$$

The sum of the interfering spherical wavelets yields an electric field at P, given by the real part of

$$E = E_0(r)e^{i(kr_1-\omega t)} + E_0(r)e^{i(kr_2-\omega t)} + \cdots + E_0(r)e^{i(kr_N-\omega t)}. \quad (10.1)$$

It should be clear, from Section 9.1, that we need not be concerned with the vector nature of the electric field for this configuration. Now then

$$E = E_0(r)e^{-i\omega t}e^{ikr_1}[1 + e^{ik(r_2-r_1)} + e^{ik(r_3-r_1)} + \cdots + e^{ik(r_N-r_1)}].$$

The phase difference between adjacent sources is obtained from the expression $\delta = k_0\Lambda$ and since $\Lambda = nd\sin\theta$, in a medium of index n, $\delta = kd\sin\theta$. Making use of Fig. 10.6, it follows that $\delta = k(r_2 - r_1)$, $2\delta = k(r_3 - r_1)$ etc. Thus the field at P may be written as

$$E = E_0(r)e^{-i\omega t}e^{ikr_1}[1 + (e^{i\delta}) + (e^{i\delta})^2 + (e^{i\delta})^3 + \cdots (e^{i\delta})^{N-1}]. \quad (10.2)$$

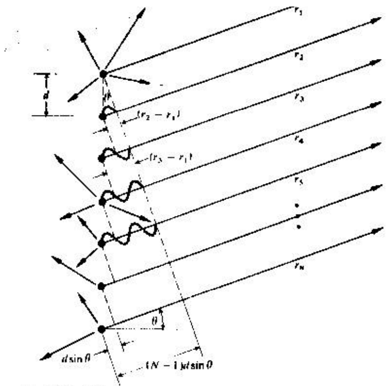


Fig. 10.6 A linear array of in-phase coherent oscillators. Note that at the angle shown $\delta = \pi$ while at $\theta = 0$ δ would be zero.

The bracketed geometric series has the value

$$(e^{i\delta N} - 1)/(e^{i\delta} - 1)$$

which can be rearranged into the form

$$\frac{e^{iN\delta/2}[e^{iN\delta/2} - e^{-iN\delta/2}]}{e^{i\delta/2}[e^{i\delta/2} - e^{-i\delta/2}]}$$

or equivalently

$$e^{i(N-1)\delta/2}\left[\frac{\sin N\delta/2}{\sin \delta/2}\right].$$

The field then becomes

$$E = E_0(r)e^{-i\omega t}e^{i[kr_1+(N-1)\delta/2]}\left(\frac{\sin N\delta/2}{\sin\delta/2}\right). \quad (10.3)$$

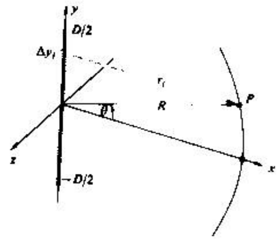


Fig. 10.8 A coherent line source.

Notice that if we define R to be the distance from the center of the line of oscillators to the point P, that is

$$R = \tfrac{1}{2}(N-1)d\sin\theta + r_1,$$

then Eq. (10.3) takes on the form

$$E = E_0(r)e^{i(kR-\omega t)}\left(\frac{\sin N\delta/2}{\sin\delta/2}\right). \quad (10.4)$$

Finally, then, the flux-density distribution within the diffraction pattern due to N coherent, identical, distant point sources in a linear array is proportional to $EE^*/2$ for complex E or

$$I = I_0\frac{\sin^2(N\delta/2)}{\sin^2(\delta/2)}, \quad (10.5)$$

where I_0 is the flux density from any single source arriving at P (see Problem 10.2 for a graphical derivation of the irradiance). For $N = 0$, $I = 0$, for $N = 1$, $I = I_0$, and for $N = 2$, $I = 4I_0\cos^2(\delta/2)$ in accord with Eq. (9.6). The functional dependence of I on θ is more apparent in the form

$$I = I_0\frac{\sin^2[N(kd/2)\sin\theta]}{\sin^2[(kd/2)\sin\theta]}. \quad (10.6)$$

The $\sin^2[N(kd/2)\sin\theta]$ term undergoes rapid fluctuations, while the function modulating it, $\{\sin[(kd/2)\sin\theta]\}^{-2}$, varies relatively slowly. The combined expression gives rise to a series of sharp principal peaks separated by small subsidiary maxima. The principal maxima occur in directions θ_m such that $\delta = 2m\pi$ where $m = 0, \pm1, \pm2, \ldots$. Because $\delta = kd\sin\theta$

$$d\sin\theta_m = m\lambda. \quad (10.7)$$

Since $[\sin^2 N\delta/2]/[\sin^2\delta/2] = N^2$ for $\delta = 2m\pi$ (from

Fraunhofer diffraction

L'Hôspital's rule) the principal maxima have values $N^2 I_0$. This is to be expected inasmuch as all of the oscillators are in phase at that orientation. The system will radiate a maximum in a direction perpendicular to the array ($m = 0$, $\theta_0 = 0$ and π). As θ increases, δ increases and I falls off to zero at $N\delta/2 = \pi$, its first minimum. Note that if $d < \lambda$ in Eq. (10.7), only the $m = 0$ or zero-order principal maximum exists. *If we were looking at an idealized line source of electron-oscillators separated by atomic distances, we could expect only that one principal maximum in the light field.*

The antenna array of Fig. 10.7 can then transmit radiation in the narrow beam or lobe corresponding to a principal maximum (the parabolic dishes shown reflect into the forward direction and the radiation pattern is no longer symmetrical around the common axis.) Suppose that we have a system in which we can introduce an intrinsic phase shift of ε between adjacent oscillators. In that case

$$\delta = kd \sin\theta + \varepsilon;$$

the various principal maxima will occur at new angles

$$d \sin\theta_m = m\lambda - \varepsilon/k.$$

Concentrating on the central maximum $m = 0$, its orientation θ_0 can be varied at will by merely adjusting the value of ε.

The principle of reversibility, which states that without absorption, wave motion is reversible, leads to the same field pattern for an antenna used as either a transmitter or receiver. The array, functioning as a radio telescope, can therefore be "pointed" by combining the output from the individual antennas with an appropriate phase shift, ε, introduced between each of them. For a given ε the output of the system corresponds to the signal impinging on the array from a specific direction in space.

Figure 10.7 is a photograph of the first multiple radio interferometer designed by W. N. Christiansen and built in Australia in 1951. It consists of 32 parabolic antennas, each 2 m in diameter, designed to function in phase at the wavelength of the 21 cm hydrogen emission line. The antennas are arranged along an east–west baseline with 7 m separating each one. This particular array utilizes the earth's rotation as the scanning mechanism.

Examine Fig. 10.8 which depicts an idealized line source of electron-oscillators (e.g., the secondary sources of the Huygens–Fresnel principle for a long slit whose width is much less than λ illuminated by plane waves). Each point emits a spherical wavelet which we write as

$$E = \left(\frac{\mathcal{E}_0}{r}\right) \sin(\omega t - kr)$$

explicitly indicating the inverse r-dependence of the amplitude. The quantity $\mathcal{E}_0$ is said to be the *source strength*. The present situation is distinct from that of Fig. 10.6 in that now the sources are very weak, their number, N, is tremendously large and the separation between them vanishingly small. A minute, but finite segment of the array Δy_i, will contain $\Delta y_i(N/D)$ sources where D is the entire length of the array.

Imagine then that the array is divided up into M such segments, i.e., i goes from 1 to M. The contribution to the electric field intensity at P from the ith segment is accordingly

$$E_i = \left(\frac{\mathcal{E}_0}{r_i}\right) \sin(\omega t - kr_i)\left(\frac{N\Delta y_i}{D}\right)$$

provided that Δy_i is so *small* that the oscillators within it have a negligible relative phase difference (r_i = constant) and their fields simply add constructively. We can cause the array to become a continuous (coherent) line source by letting N approach infinity. This description, besides being fairly realistic on a macroscopic scale, also allows the use of the calculus for more complicated geometries. Certainly as N approaches infinity, the source strengths of the individual oscillators must diminish to near zero if the total output is to be finite. We can therefore define a constant $\mathcal{E}_L$ as the *source strength per unit length* of the array, that is

$$\mathcal{E}_L \equiv \frac{1}{D} \lim_{N \to \infty} (\mathcal{E}_0 N). \tag{10.8}$$

The net field at P from all M segments is

$$E = \sum_{i=1}^{M} \frac{\mathcal{E}_L}{r_i} \sin(\omega t - kr_i)\, \Delta y_i.$$

For a continuous line source Δy_i can become infinitesimal ($M \to \infty$) and the summation is then transformed into a definite integral

$$E = \mathcal{E}_L \int_{-D/2}^{+D/2} \frac{\sin(\omega t - kr)}{r}\, dy, \tag{10.9}$$

where $r = r(y)$. The approximations used to evaluate Eq. (10.9) must depend on the position of P with respect to the array and will therefore make the distinction between Fraunhofer and Fresnel diffraction. The coherent *optical* line source does not now exist as a physical entity but we will make good use of it as a mathematical device.

10.2 FRAUNHOFER DIFFRACTION

10.2.1 The Single Slit

Return to Fig. 10.8 where now the point of observation is very distant from the coherent line source and $R \gg D$. Under these circumstances $r(y)$ never deviates appreciably from its midpoint value R so that the quantity $(\mathcal{E}_L/R)$ at P is essentially constant for *all* elements dy. It follows from Eq. (10.9) that the field at P due to the differential segment of the source dy is

$$dE = \frac{\mathcal{E}_L}{R} \sin(\omega t - kr)\, dy, \tag{10.10}$$

where $(\mathcal{E}_L/R)\, dy$ is the amplitude of the wave. Notice that the phase is very much more sensitive to variations in $r(y)$ than is the amplitude so that we will have to be more careful about introducing approximations into it. We can expand $r(y)$, in precisely the same manner as was done in Problem (9.4), to get it as an explicit function of y, thus

$$r = R - y\sin\theta + (y^2/2R)\cos^2\theta + \cdots, \tag{10.11}$$

where θ is measured from the xz-plane. The third term can be ignored so long as its contribution to the phase is insignificant even when $y = \pm D/2$, i.e. $(\pi D^2/4\lambda R)\cos^2\theta$ must be negligible. This will be true for all values of θ when R is adequately large and we again have the Fraunhofer condition. The distance r is then linear in y. Substituting into Eq. (10.10) and integrating leads to

$$E = \frac{\mathcal{E}_L}{R} \int_{-D/2}^{+D/2} \sin[\omega t - k(R - y\sin\theta)]\, dy, \tag{10.12}$$

and finally

$$E = \frac{\mathcal{E}_L D}{R} \frac{\sin[(kD/2)\sin\theta]}{(kD/2)\sin\theta} \sin(\omega t - kR). \tag{10.13}$$

To simplify the appearance of things let

$$\beta \equiv (kD/2)\sin\theta \tag{10.14}$$

so that

$$E = \frac{\mathcal{E}_L D}{R}\left(\frac{\sin\beta}{\beta}\right)\sin(\omega t - kR). \tag{10.15}$$

The quantity most readily measured is the irradiance (forgetting the constants) $I(\theta) = \langle E^2 \rangle$ or

$$I(\theta) = \frac{1}{2}\left(\frac{\mathcal{E}_L D}{R}\right)^2 \left(\frac{\sin\beta}{\beta}\right)^2, \tag{10.16}$$

where $\langle \sin^2(\omega t - kR) \rangle = \frac{1}{2}$. When $\theta = 0$, $\sin\beta/\beta = 1$ and $I(\theta) = I(0)$ which corresponds to the *principal maximum*. *The irradiance resulting from an idealized coherent line source in the Fraunhofer approximation* is then

$$I(\theta) = I(0)\left(\frac{\sin\beta}{\beta}\right)^2, \tag{10.17}$$

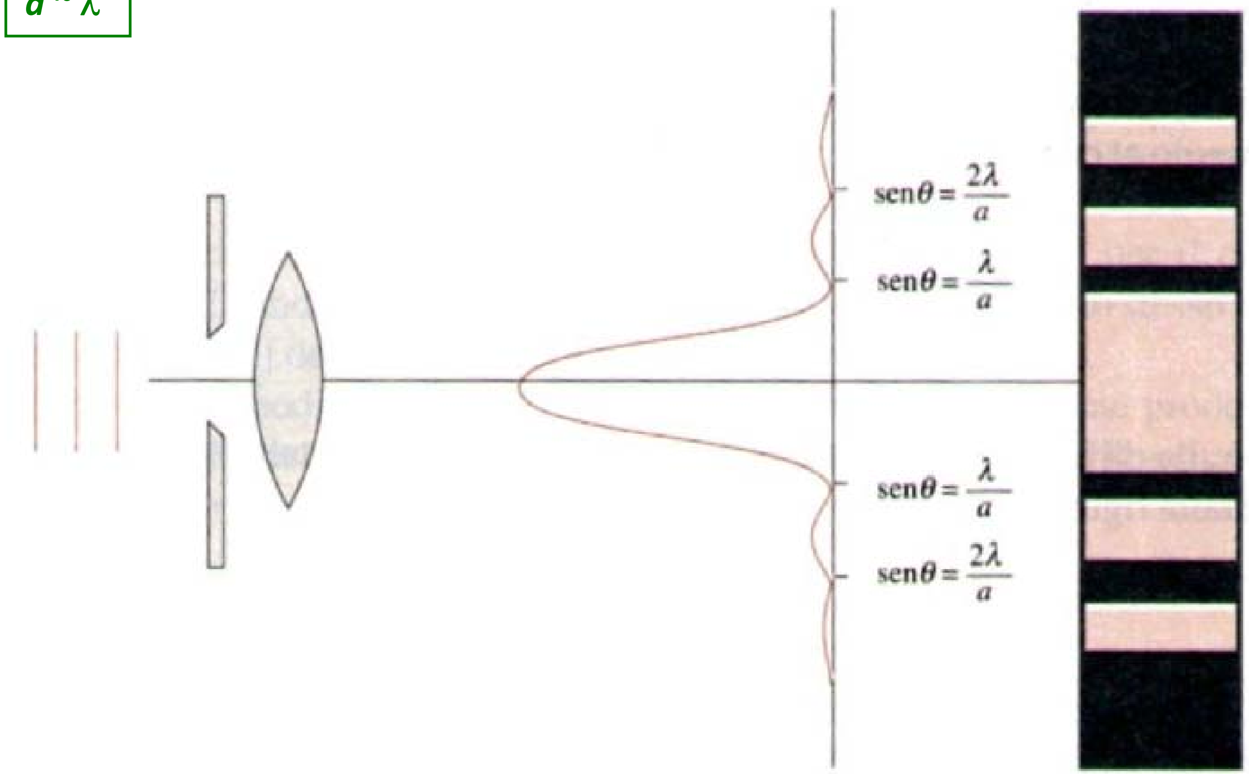
or using the *sinc function* (Section 7.9, and Table 1 of the Appendix)

$$I(\theta) = I(0)\, \mathrm{sinc}^2\, \beta.$$

There is symmetry about the y-axis and this expression holds for θ measured in any plane containing that axis.

Effects of diffraction

$a \sim \lambda$



"Diffusion cone": $\text{Sin}\theta \sim \lambda / a$

Diffraction ripples: $I \sim I_0 (\sin\alpha/\alpha)^2$ with: $\alpha = \pi\, a \sin\theta / \lambda$

Optical diffraction is for sure a *fundamental* limiting factor in optical microscopy

Criteria for space resolution (in optical microscopy)

3.1.2.1 POINT-SOURCE ABBE IMAGE

The calculated intensity distribution assumes a parallel beam of light travelling along the axis of a thin lens and brought to a focus at the focal distance (Fig. 3.8). For the *cylindrically symmetric* case, the ratio of the peak intensities for the primary and secondary peaks in the intensity distribution is *ca* 9 : 1, while the width of the primary peak is given by the *Abbe equation* as follows:

Abbe

$$\delta = 0.61 \frac{\lambda}{\mu \sin \alpha} \qquad (3.1)$$

where λ is the wavelength of the radiation, α is the aperture (half-angle) of the lens (determined by the ratio of the lens radius to its focal length), and μ is the refractive index of the medium between the lens and the focal point ($\mu \approx 1$ for air).

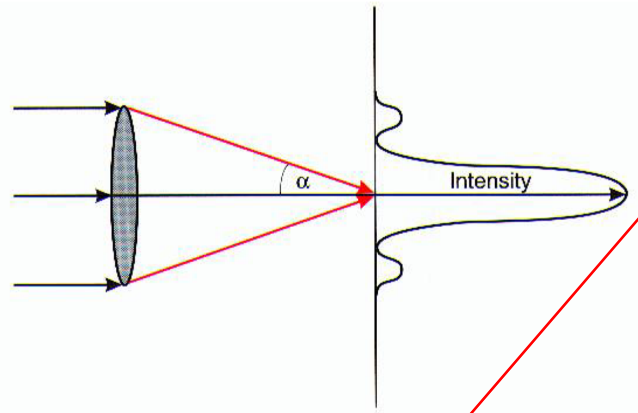
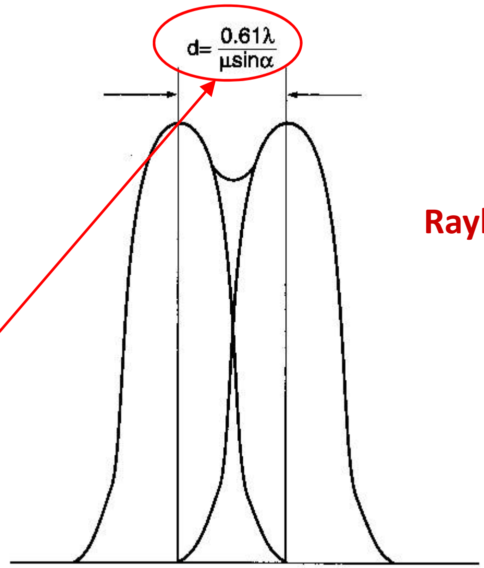


Figure 3.8 The Abbe equation gives the width of the first intensity peak for the image of point object at infinity in terms of the angular aperture of the lens α and the wavelength of tl radiation λ

Da Brandon Kaplan
Microstruct. Charact.
of Materials
Wiley (1999)



Rayleigh

Figure 3.10 The Raleigh resolution criterion requires that two point sources at infinity have an angular separation which is sufficient to place the maximum intensity of the primary image peak of one source at the position of the first minimum of the second

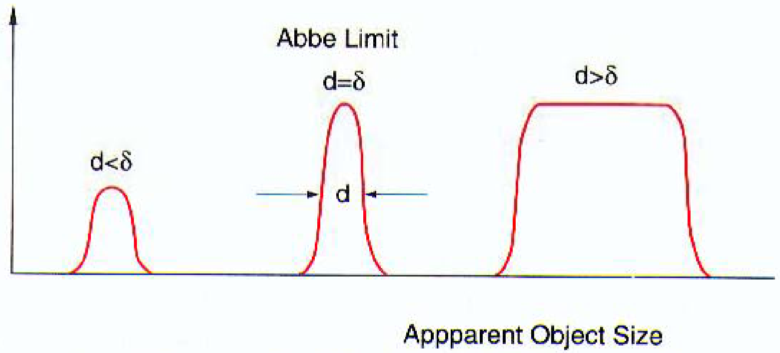


Figure 3.11 Large objects of diameter d are blurred by the diffraction limit δ derived from the Abbe relationship, but objects smaller than the Abbe width are still detectable in the microscope, although the intensity is reduced and they have an apparent width given by the Abbe equation

Maximum achievable space resolution *d* ~ 0.61 λ /(NA) > λ/2

(NA: numerical aperture of the optical system: NA = *n* sinα , with *n* refractive index)

A (qualitative) look at Fourier-transform optics

A surface pattern $f(x,y)$ (i.e., an object to be imaged!) can be always described in terms of a (2D) Fourier superposition

Fourier components are function of the reciprocal space $k_X\, k_Y$: $f(x,y) = \int f'(k_X, k_Y)\, exp(-ik_X x)\, exp(-ik_Y y)\, dk_X\, dk_Y$

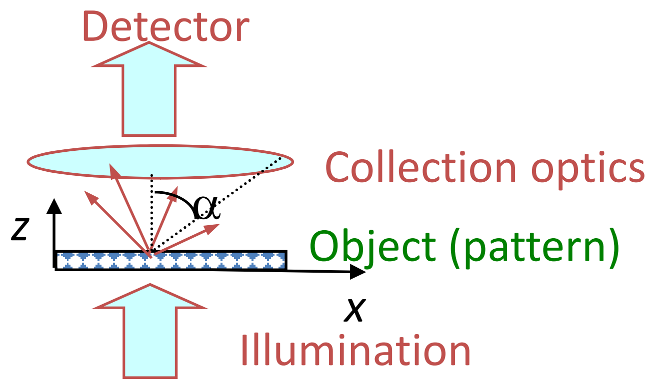
Large k_X (and k_Y) values imply large angular displacement

⇩

A faithful pattern reproduction (i.e., an accurate image) implies collecting the largest number of k_X and k_Y

⇩

Large numerical aperture ($NA=nSin\alpha$) needed for high resolution



but

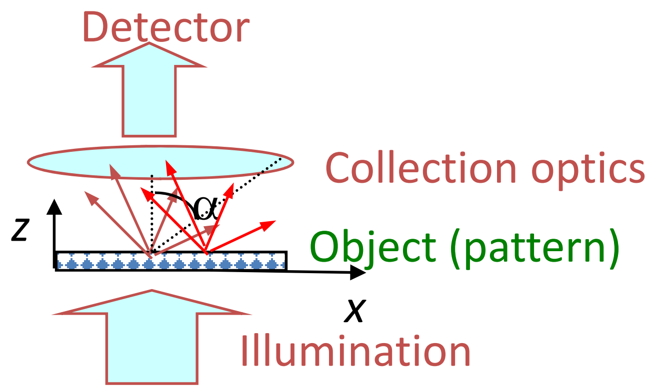
In non point-like illumination schemes (conventional) rays emitted from different points of the surface can be collected

⇩

Large numerical aperture leads to sensitivity to "stray light"

⇩

Contrast falls down and high space resolution is hampered

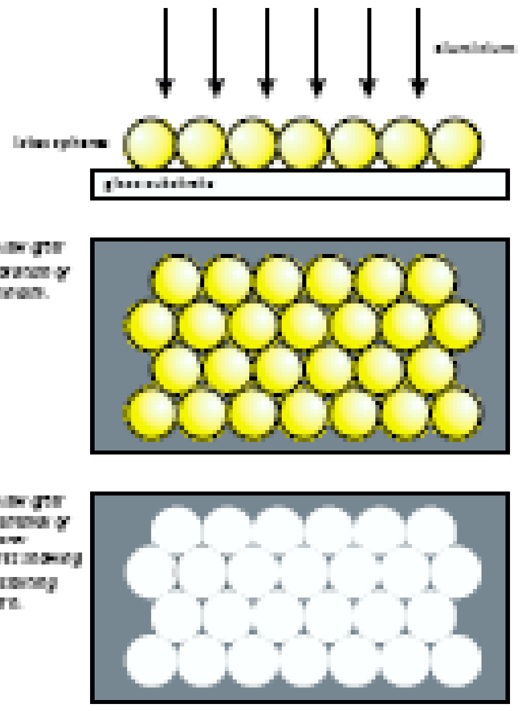


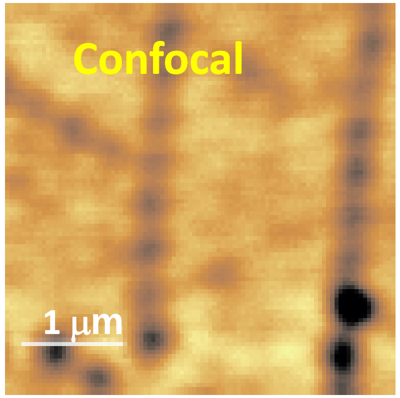
How to use extreme diffraction to get excellent resolution

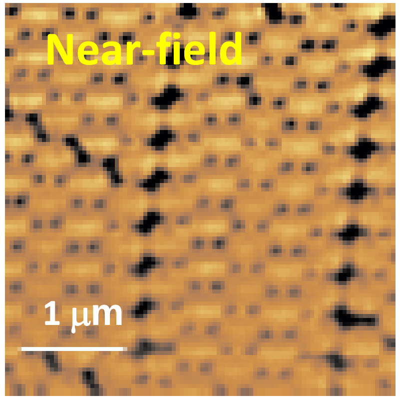
Comparison of optical resolution between Confocal Microscopy (CM, top image) and Scanning Near-field Optical Microscopy (SNOM, lower image). The scale bar is 1 micron.

The sample is a latex projection pattern. In the confocal image only the dislocation lines are visible, whereas the SNOM image also shows the Al islands clearly (same sample area).

Latex projection pattern are produced by evaporating aluminium onto a glass substrate covered with latex spheres. These latex spheres have a very uniform diameter. After the evaporation process, the latex spheres are removed.



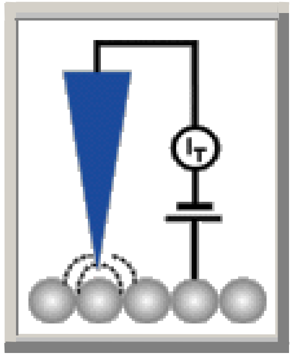
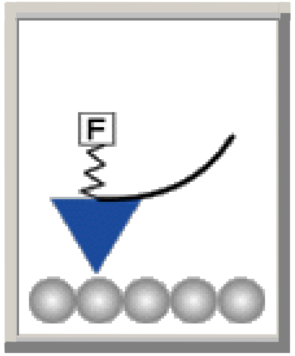
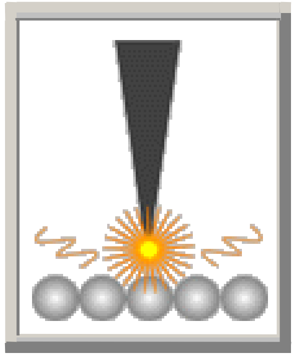




WITec

Near-field optics opens the way to overcome diffraction effects

Scanning Near Field Optical Microscopy (SNOM)

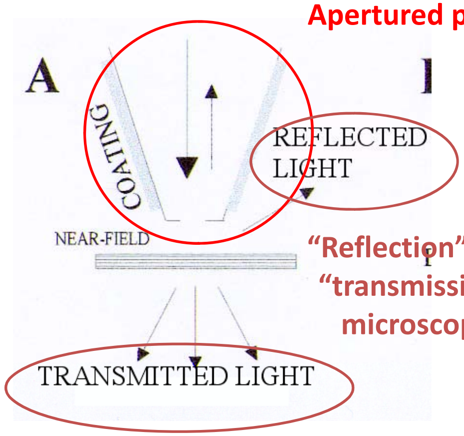
		
STM **Electron tunneling**	AFM **Force microscopy**	SNOM **Optical near-field** **Locally probed quantity**

SNOM holds the ***unique*** *ability to analyze optical properties with sub-diffraction space resolution*

Introducing near-field and SNOM

Main motivation: extending optical microscopy (and sepctroscopy) analyses into the nanoworld (which implies to overcome the diffraction limit!!)

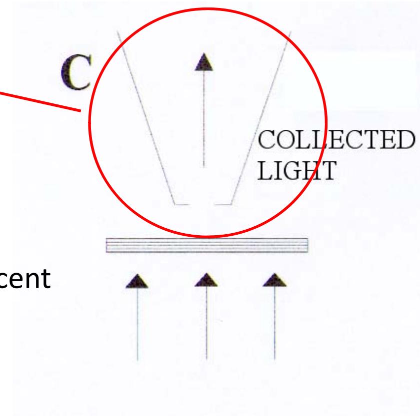
Optical near-field: an e.m. field with frequency in the optical range and a **non-propagating** nature
Near-field is the probed quantity in SNOM, and can be exploited in many different configurations



Apertured probe

"Reflection" and "transmission" microscopy

<u>Apertured tapered optical fibers</u>;
Hollow cantilevers;
Etc.



(Also for electroluminescent samples)

EMISSION MODE
(the near-field interacts with the sample and the result of the interaction is collected and analyzed in the far field)

COLLECTION MODE
(the near-field produced by conventional irradiation is collected by the near-field probe)

Other configurations can be involved, e.g., apertureless, "photon tunneling", etc.

SNOM probes (tapered fibers) I

Different configurations for SNOM exist

Here we will mention mostly **aperture-SNOM**, which often exploits tapered optical fibers as probes

Most common SNOM probe: tapered optical fiber, with metallization and apical aperture a<<λ (*aperture-SNOM*)

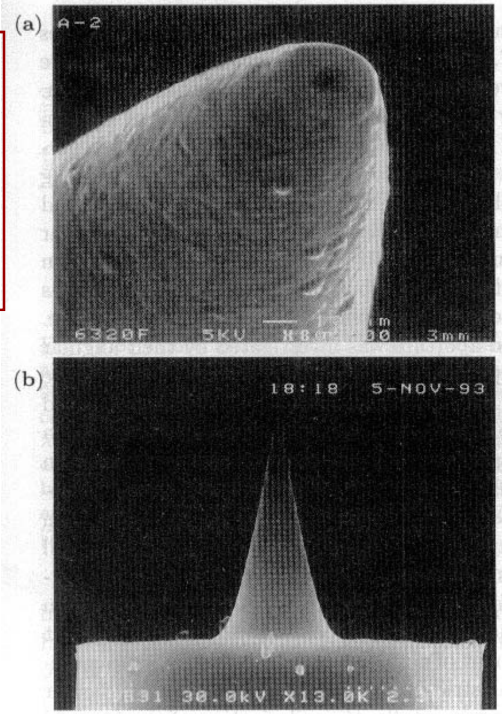
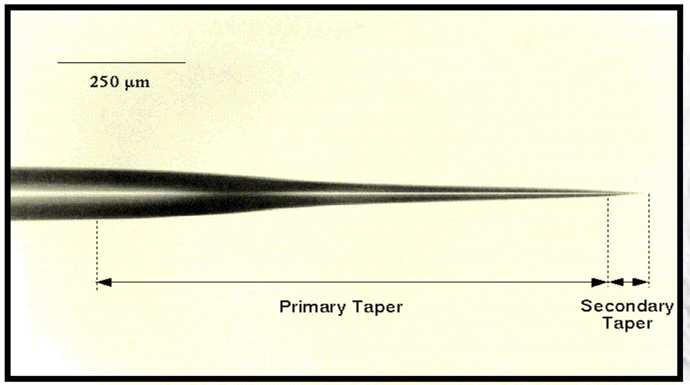
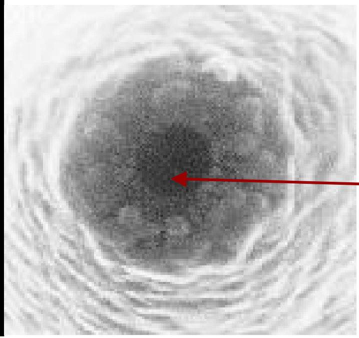


Fig. 7.7a–d. SEM micrographs of SNOM probes: (**a**) Tip of an aperture probe consisting of a thermally pulled tip of a quartz monomode fiber coated with aluminum. (By Courtesy of Sunney Xie). (**b**) Etched fiber tip according to Ohtsu [7.23]. The tip is fabricated by wet etching of a monomode quartz fiber. The thickness of the fiber coating is strongly reduced at the end of the fiber and a sharply pointed tip sticking out from the end is formed from the core. (**c**) Aperture probe fabricated on the basis of an etched tip, as shown in (**b**). The etched tip is coated with gold which is removed from the apex of the tip by a lithographic process such that a small aperture is formed with the tip sticking out [7.117]. (**d**) Tetrahedral tip. The tetrahedral tip consists of a glass fragment which is coated with metal. By courtesy of, R. Reichelt, Institute of medical Physics and Biophysics, University of Münster

(c) Core





Typical aperture diameter: 50-100 nm

SNOM probes (tapered fibers) II

Conventional probe fabrication

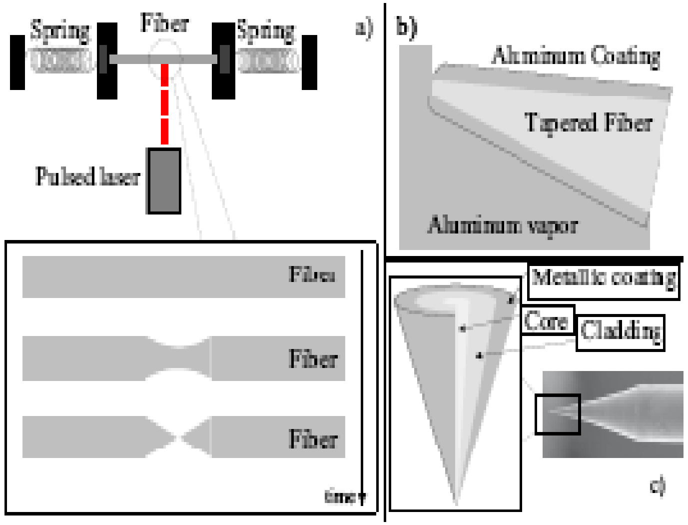


Figure 1.1: a) Heat and pull, one of the procedure to taper optical fibres for SNOM applications. b) Metallization of the SNOM fibre tip at steep angle to leave a sub-wavelength aperture at the end. c) SEM image of the resulting tip and sketch of its interior.

Many probes are available, including hollow cantilevers (similar to those for AFM, but with a pyramidal aperture)

Materiale tratto da Antonio Ambrosio
PhD Thesis Applied Physics, Pisa, 2005

An alternative technique

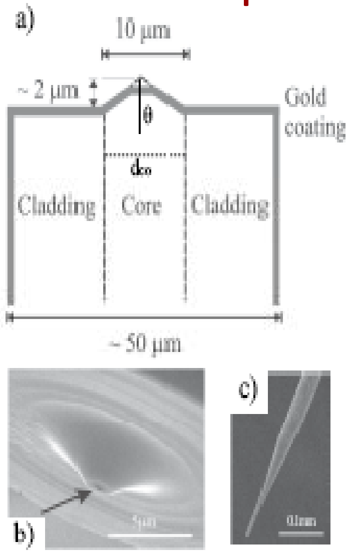


Figure 3.13: (a) Schematic of a SNOM fibre probe produced by selective chemical etching (SCE). The protruding cone is formed due to a slower etching rate of the core with respect to the cladding. After gold metallization (indicated by light gray lines), the tip is punched against a hard surface, producing a flattened apex with a sub-wavelength aperture at the center (black arrow), as evidenced by the SEM micrograph in (b).

An alternative probe

Figure 5.1: Back of an hollow cantilever. The square is the back aperture of the pyramidal hole produced by selective chemical etching. The inset shows a light spot coupled directly into the cantilever's hole by means of a microscope objective.

Example of non propagating fields: the evanescent wave

From Maxwell's equations, we know that the component of the electric field tangent to the interface of two dielectrics must be continuous across that interface. For a plane wave moving from one dielectric to another,

$$k_i \,\mathrm{Sin}[\theta_i] = k_t \,\mathrm{Sin}[\theta_t]$$

where k is the wave vector, theta is the angle between the wave propagation and the interface normal, and subscripts i and t stand for incident and transmitted wave fronts, respectively.

The frequency of the wave is identical on either side of the interface, so we have Snell's law:

$$n_i \,\mathrm{Sin}[\theta_i] = n_t \,\mathrm{Sin}[\theta_t]$$

with n the index of refraction of each of the media.

At incoming angles equal to and above the critical angle

$$\theta_c = \mathrm{ArcSin}\left[\frac{n_t}{n_i}\right]$$

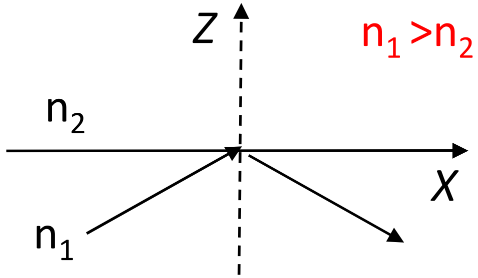
The reflectance is 1, the transmittance is 0, and all energy is reflected back to the incoming side of the interface. However, as electric fields impinge on our uncharged dielectric interface, the boundary conditions due to Maxwell's equations and the conservation of momentum demand that there be a matching field on the far side. The component of k_t parallel to the interface is still equal to the component of k_i to obey the boundary conditions.

We have an incoming plane wave

$$E_i[r, t] = \mathcal{E}_i \, e^{(i\,(k_{ix}\,\hat{x} + k_{iz}\,\hat{z} - \omega\, t))}$$

With amplitude *Ei*, wave vector k and frequency w. The z direction is normal to the plai of interface of two dielectrics, and the x direction is chosen so that the wave vector lies entirely in the xz plane.

Tom Hunt
http://www.physics.harvard.edu/~tomhunt/pubs/evanescent.pdf



$E_t[\mathbf{r},t] = E_{0t}\exp(i(\mathbf{k}_t\cdot\mathbf{r}-\omega t))$
$\mathbf{k}_t\cdot\mathbf{r}=k_{tx}x + k_{tz}z$
$k_{tx}=k_t\sin\theta_t$; $k_{tz}=k_t\cos\theta_t$

With Snell's law:
$k_t\sin\theta_t = n_1 k_t \sin\theta_i/n_2$
$k_t\cos\theta_t = \pm k_t\,(1-n_1^2\sin^2\theta_i/n_2^2)^{1/2}$
$= \pm i\, k_t\,(n_1^2\sin^2\theta_i/n_2^2-1)^{1/2} = \pm i\,\beta$

Hence:
$E_t[\mathbf{r},t] = E_{0t}\exp(-\beta z)\exp(i(k_x x-\omega t))$

Physical situations exist where e.m. field is not propagating (see, e.g., *evanescent waves*)

Near-field "conversion" to far-field

7.2.1 Ray Optics of a SNOM

SNOM techniques differ mainly in the types of probes which are used and also by their ray optical components, i.e., their optical scheme, which is useful for a classification of most types of SNOM concepts. As shown schematically in Fig. 7.4, three different regions where the rays propagate can be distinguished: I) the body of the probe, II) the outside and III) the substrate of the object. In general, regions I and III will have a higher refractive index than the outside region II. Different angular domains of rays propagating in regions I and III exist, which can be distinguished by the criterion of total reflection of a ray falling on its boundary or of it being partially refracted into the outside II. Thus, in the case of a transparent substrate III, we distinguish between the angular domain III_1 of angles ε with $-\varepsilon_t < \varepsilon < \varepsilon_t$, where ε_t is the critical angle of total reflection ($\varepsilon_t = 41.5°$ for glass of refractive index 1.5) and the angular domain III_2 with $90° > \varepsilon > \varepsilon_t$ or $-90° < \varepsilon < -\varepsilon_t$, which is sometimes called the range of forbidden light. Rays of the domain III_2 are totally reflected in the substrate, whereas rays of domain III_1 are partially refracted into the outside II. Also within the body of the tip I two different domains may be distinguished (Fig. 7.4). This figure only shows the case of a rectangular wedge, a two-dimensional analog of the three-dimensional body of the tip. For such a wedge, with a refractive index $n = 1.5$, rays entering at an angle within the angular domain ($-3.5° < \varepsilon < +3.5°$, region I_2) will be totally reflected back into a reflected ray of the same angle ε. Rays entering the wedge at different angles will also be reflected into the same angle and be partially refracted into the outside II of the wedge. This situation also applies, if the wedge is coated with a partially transparent metal film, as is typical for SNOM probes. Similar considerations also apply for a three-dimensional tip.

In summary, in many cases it is possible to distinguish in regions I and III between angular domains I_1 and III_1 where total reflection of the rays occurs into the same domain and the domains I_2 and III_2 from where light is partially refracted to the outside II.

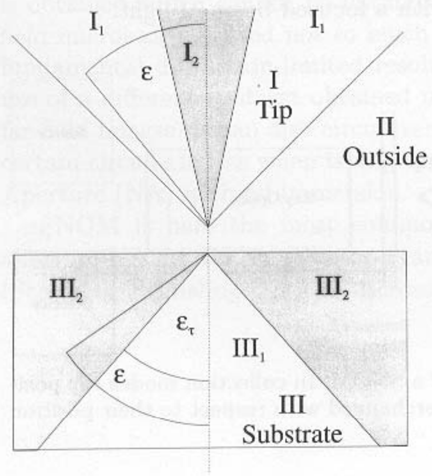


Fig. 7.4. Ray optics of a SNOM. One can distinguish between three different regions where the rays propagate; the body of the probe I, the outside II and the substrate III of the object. Different angular domains of rays propagating in regions III and I at an angle ε can be distinguished by the criterion of a ray falling on its boundary being totally reflected into the same domain (III_2, I_2) or being partially refracted into the outside II (III_1, I_1)

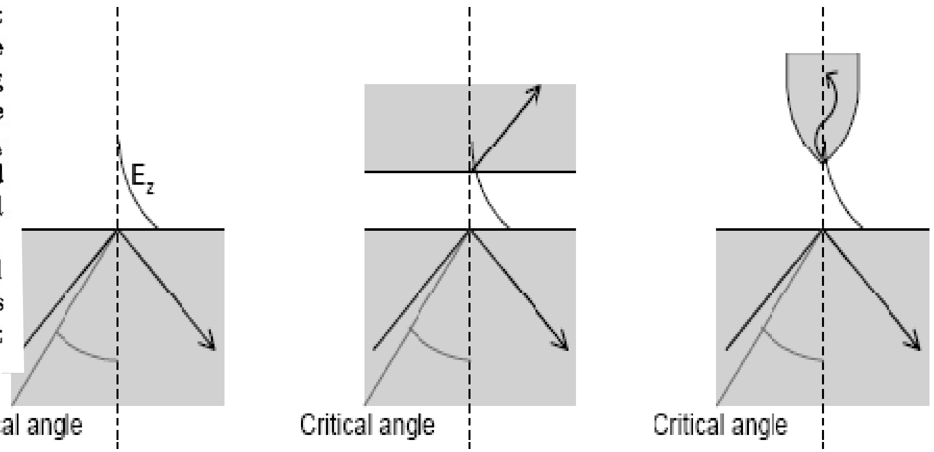


Figure 9: the left panel shows the penetration of an electromagnetic field in the less dense medium when total internal reflection occurs. The middle and the right panel schematically envision the energy flow between two dense media through an air gap (frustrated total internal reflection).

http://www.chem.vu.nl/~sneppen/literaturereport.pdf

Da Wiesendanger Ed., Scanning Probe Microscopies (Springer, 1998)

SNOM probe can "convert" near-field into far-field (especially relevant in collection mode)

Photon Tunnelling Microscopy

Non propagating fields and diffraction

Diffraction and the Heisenberg's microscope

The resolving power of a conventional microscope, i.e. the distance (Δx) between two object points which a microscope can just resolve, depends on the numerical aperture (NA) of the objective and the wavelength (λ) of the light used. The relation between these quantities is fixed by the *Rayleigh's criterion* that sets:

$$\Delta x = 0.61 \frac{\lambda}{NA} \tag{1.1}$$

The whole limits discussed above can be regarded in terms of quantum mechanics. In fact, applying the Heisenberg's uncertain principle to the components (x_i) of a photon's position and to those of the linear momentum (p_i) of the photon:

$$\Delta x_i \cdot |p_i| \geq \hbar \tag{1.2}$$

where i is an index indicating the projections along x, y or z axes. Each component of the linear momentum of the photon is related to the corresponding components (k_i) of the light wavevector ($\vec{k}$) by $p_i = \hbar k_i$. The relation 1.2 may then be written as:

$$\Delta x_i \geq \frac{1}{|k_i|} = \frac{1}{|n_i|}\frac{\lambda}{2\pi} \tag{1.3}$$

This formula fixes the physical limit for the linear dimensions of a focused beam as well as the achievable optical resolution. The possible values of k_i are limited by the mathematical condition being between each vector and its components:

$$|k| = \sqrt{k_x^2 + k_y^2 + k_z^2} \tag{1.4}$$

Classical optics and microscopy employ free propagating waves for which all the components k_i are real. In this case $k_i^2 = |k_i|^2$ and the relation 1.3 limits the best resolution achievable to values no much smaller than $\lambda/2$.

Materiale tratto da Antonio Ambrosio
PhD Thesis Applied Physics, Pisa, 2005

Propagating waves:
k_i are real and $|k_x| \leq k = 2\pi/\lambda$
$\Downarrow$
Heisenberg's principle: $k_x \geq 2\pi / \Delta x$

$$\Delta x \geq \lambda$$

(actual parameters give the Abbe's limit)

… but …

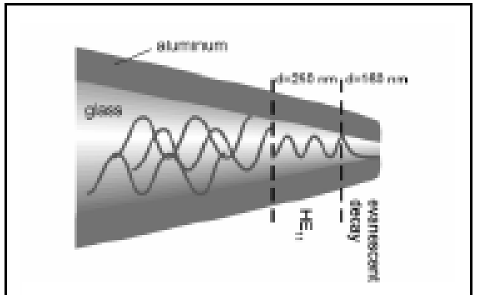
In **non-propagating (e.g., evanescent)** waves:
k_i can be imaginary, and, e.g.: $|k_x| \geq k = 2\pi/\lambda$

The Heisenberg's principle is no longer ruling the ultimate resolution!

Sub-diffraction space resolution associated with the non-propagating character of the field

A qualitative picture of tapered fibers

A whole slew of scanning near field optical microscopes (SNOMs) have been developed by researchers over the past 15 years. These microscopes smash the diffraction limit of far field microscopes, potentially achieving resolution an order of magnitude better than a standard confocal microscope [Hect 2000]. One such SNOM is the scanning tunneling optical microscope. This microscope uses a sharp glass tip to locally frustrate total internal reflection below a surface, indirectly imaging features on that surface at high spatial resolution. Aperture based SNOMs are more common and more practical. [Hect 2000] They produce an evanescent field by forcing light through a small aperture (see figure). The evanescent field locally illuminates the sample. Once free of the aperture, the field is no longer evanescent, and it expands in the far field to be picked up by a detector. To achieve high resolution, the aperture must be small, and close to the sample surface so the field is tightly confined when it interacts with the sample. Full analysis of the field-sample interaction of a SNOM is a difficult or impossible undertaking, but the data produced can yield important and detailed information about a sample surface.



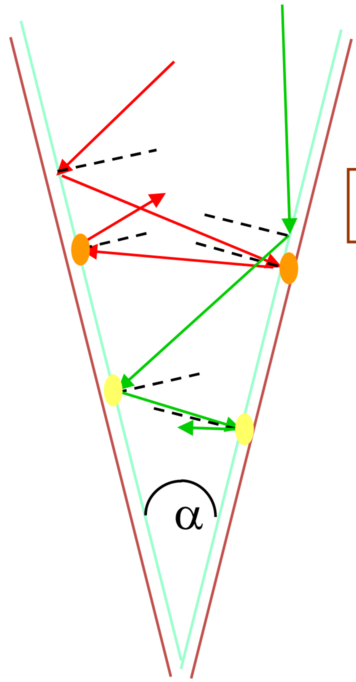
Evanescent fields produced by the tapered tip of a SNOM. [Hect, 2000]

Suggested reading: Hecht et al., J. Chem Phys. 112, 7761 (2000).

A simplified ray optics picture in the tapered region. Light can both:

- be back reflected

- (partially) absorbed by the metal layer



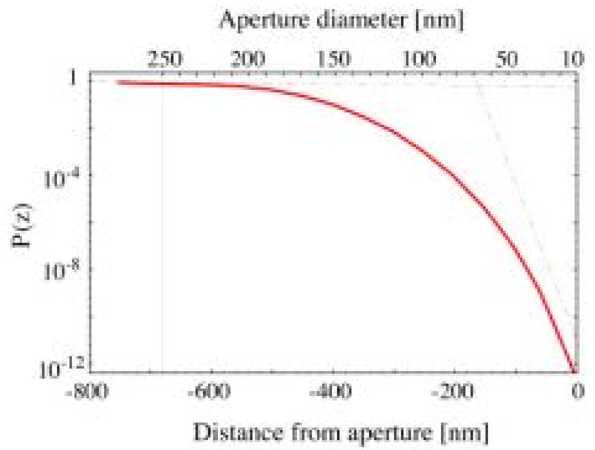
$$\theta_{i,n} = \theta_{i,n-1} - \alpha$$

Note: metal layer (typ Cr, Ni) can absorb radiation → power entering the fiber cannot exceed the mW range!

Note: probe “throughput” (i.e., ratio between output/input power) is quite low for fiber probes, ~ 1E-5 – 1E-8, but near field **intensity** can be large enough

Near-field throughput

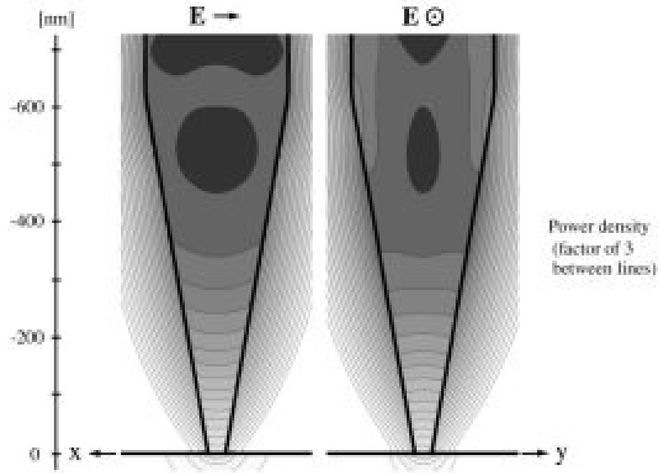
However, the amount of light that can be transmitted by a small aperture poses a limit on how small it can be made before nothing gets though. To a degree this can be lived with, as more optical power can be generated, but the cutoff is so severe that it cannot be made smaller. As the figures illustrate, this is not a subtle extinction.



L. Novotny and D. W. Pohl, in *Photons and Local Probes*, NATO ASI Series E, p.21-33, Kluwer Academic, 1995.

Little power is coupled into the near-field!

When the aperture is 100 nm, the cutoff is down four orders of magnitude, and when it reaches 50 nm, only one part in 10^8 makes it through. Furthermore, the input power cannot be increased arbitrarily because 1/3 of the power is absorbed in the coating. Increasing the input power above approximately 10mW will destroy the coating. This severly limits the signal-to-noise ratio of small apertures, and is the reason our group uses another approach.

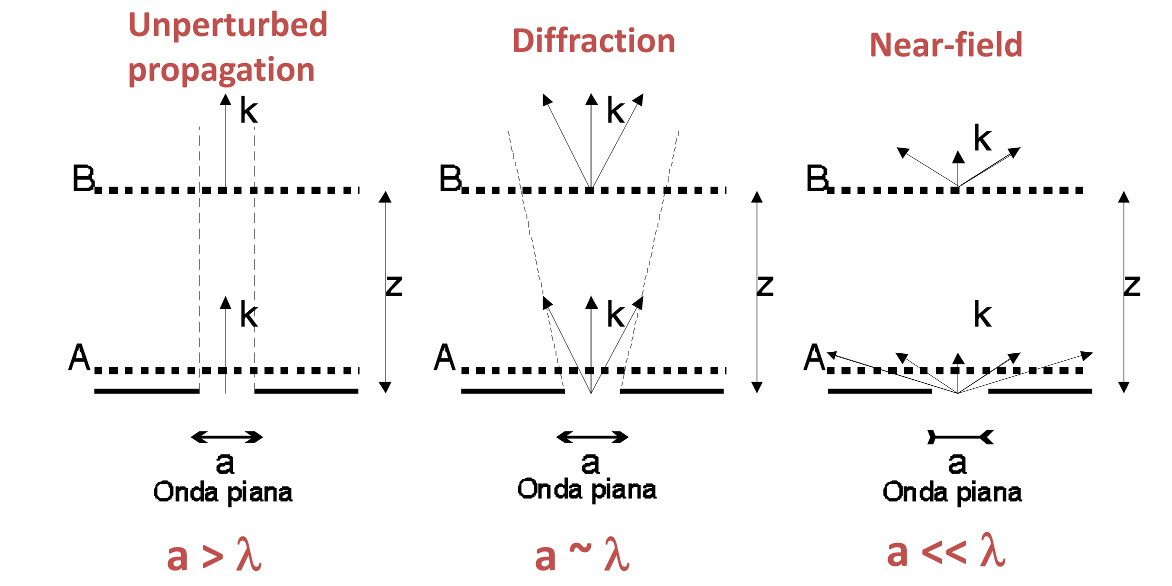


L. Novotny and D. W. Pohl, in *Photons and Local Probes*, NATO ASI Series E, p.21-33, Kluwer Academic, 1995.

http://xray.optics.rochester.edu/workgroups/novotny/snom.html

Optical Near Field and Fourier optics

Ideal case (e.g., studied in the 20's by Synge and reworked in the 40's by Bethe): radiation sent onto a conductive plane with a subwavelength circular or elliptic aperture



In terms of Fourier optics, the subwavelength aperture produces radiation with extremely high spatial frequencies (transverse wavevectors)
→ <u>space resolution no longer limited by diffraction</u>

When aperture diameter is (much) smaller than the wavelength, far-field (propagating) intensity gets negligible compared to near-field (**non-propagating**) intensity

The subwavelength aperture acts as a hi-pass filter for the spatial frequencies

Optical Near Field and Maxwell eqs. I

Typical approach to radiative problems is based on manipulations of Maxwell's eqs exploiting potentials (scalar and vector)

Vector potential

Possiamo infatti esprimere il campo $\vec{B}$ in termini del potenziale vettore $A(\vec{r})$, con:

$$\vec{\nabla} \times \vec{A} = \vec{B} \qquad (5)$$

dove $\vec{A} = \vec{A}(\vec{r}, t)$, con che l'equazione $\vec{\nabla} \cdot \vec{B} = 0$ é automaticamente soddisfatta.

Scalar potential (in the Lorentz gauge)

$$\vec{\nabla} \cdot \vec{A} + \epsilon\mu \frac{\partial V}{\partial t} = 0$$

$$\vec{\nabla} V = -(\vec{E} + \frac{\partial \vec{A}}{\partial t}) \qquad (6)$$

Potentials and charge, currents density

$$\vec{A}(\vec{r}, t) = \frac{\mu}{4\pi} \int_\tau \frac{\vec{J}(\vec{r'}, t - \Delta r/v)}{\Delta r} \, d\tau' \qquad (12)$$

$$V(\vec{r}, t) = \frac{1}{4\pi\epsilon} \int_\tau \frac{\rho(\vec{r'}, t - \Delta r/v)}{\Delta r} \, d\tau' \qquad (13)$$

dove $\vec{r}$ é il vettore tracciato a partire dall'origine al punto P dove si calcola i campi; $\vec{r'}$ é il vettore tracciato a partire dall'origine al generico punto dove é presente l'elemento di carica $\rho d\tau$ o l'elemento di corrente $\vec{J} d\tau$; $\Delta r = |\vec{r} - \vec{r'}|$ il vettore che va dal punto dove sono i generici elementi di carica/corrente al punto P; $v = 1/\sqrt{\epsilon\mu}$ é la velocitá di propagazione delle onde elettromagnetiche nel mezzo.

In the far-field (retardation effects!):

e possono essere trascurati. Quindi, a grandi distanze dal sistema di cariche, i potenziali diventano:

$$\vec{A}(\vec{r}, t) = \frac{\mu_0}{4\pi r} \int_\tau \vec{J}(\vec{r'}, t - r/c + \frac{\vec{r'} \cdot \vec{n}}{c}) \, d\tau' \qquad (19)$$

$$V(\vec{r}, t) = \frac{1}{4\pi\epsilon_0 r} \int_\tau \rho(\vec{r'}, t - r/c + \frac{\vec{r'} \cdot \vec{n}}{c}) \, d\tau' \qquad (20)$$

dove si é ipotizzato di essere nel vuoto ($\epsilon = \epsilon_0$, $\mu = \mu_0$) Il termine $\vec{r'} \cdot \vec{n}/c$ tra gli argomenti degli integrandi indica quanto l'onda elettromagnetica proveniente dalle parti piú distanti del sistema che irradia é ritardata rispetto all'onda proveniente dalle parti vicine. In altre parole, il termine $\vec{r'} \cdot \vec{n}/c$ determina il tempo che l'onda elettromagnetica impiega ad attraversare il sistema. Se la velocitá delle cariche é v, in tale tempo esse si saranno spostate di $v(r' \cdot \vec{n})/c$. Il ritardo interno al sistema sará trascurabile quando tale distanza sará piccola rispetto alle dimensioni del sistema, r'. Quindi la condizione é: $v(\vec{r'} \cdot \vec{n})/c \ll r'$ o, equivalentemente: $v \ll c$. In tal caso le cariche non avranno il tempo di cambiare apprezzabilmente le loro posizioni durante il tempo che l'onda impiega ad attraversare il sistema.

Da Chiara Roda
www.df.unipi.it/~roda/fisica2/postscript/potenziali.ps

Trascurando il ritardo interno al sistema, l'espressione del potenziale vettore diventa:

$$\vec{A}(\vec{r}, t) = \frac{\mu_0}{4\pi r} \int_\tau \vec{J}(\vec{r'}, t - r/c) \, d\tau' \qquad (21)$$

Sostituiamo in questa espressione a $\vec{J}$ il prodotto $\rho\vec{v}$, dove ora la densitá di carica ρ é calcolata al tempo $t - r/c$. Si ottiene quindi:

$$\vec{A}(\vec{r}, t) = \frac{\mu_0}{4\pi r} \int_\tau \rho(\vec{r'}, t - r/c) \vec{v} \, d\tau'$$

Notiamo ora che, se avessimo a che fare con una singola carica puntiforme e, l'integrale nell'ultima espressione scritta sarebbe semplicemente $e\vec{v}$. Per un sistema di cariche esso sará:

$$\sum_i e_i \vec{v}_i \,{}_{(t-r/c)}$$

dove la somma va calcolata al tempo ritardato $t - r/c$. Ora, il momento di dipolo di un sistema é definito come:

$$\vec{p} = \sum e_i \vec{r'}_i$$

per cui é:

$$\dot{\vec{p}} = \sum_i e_i \frac{d\vec{r'}_i}{dt}$$

ed infine, il potenziale vettore diviene:

$$\vec{A} = \frac{\mu_0}{4\pi r} \dot{\vec{p}}_{(t-r/c)}$$

Retardation (inside the emitter) is neglected, i.e., emitter is considered pointlike

Abbiamo in definitiva ottenuto delle espressioni per i potenziali vettore e scalare relativi ad un dipolo variabile nel tempo. Le espressioni ottenute sono:

$$V = \frac{1}{4\pi\epsilon_0 r^3} [\vec{p} + \frac{r}{c}\dot{\vec{p}}] \cdot \vec{r}_{\,(t-r/c)} \qquad (22)$$

$$\vec{A} = \frac{\mu_0}{4\pi r} \dot{\vec{p}}_{(t-r/c)} \qquad (23)$$

Retarted potential methods allow to derive potential wavefunctions

Optical Near Field and Maxwell eqs. II

Explicit field solution in the far-field

Le equazioni di Maxwell nel vuoto sono in genere risolte assumendo che la distanza del punto $\mathbf{r}$ in cui si vogliono calcolare i campi $\mathbf{E}(\mathbf{r})$ e $\mathbf{B}(\mathbf{r})$ dalla sorgente sia molto maggiore delle dimensioni della stessa. In realtà [21] le soluzioni *complete* delle equazioni per i campi generati da un dipolo elettrico $\mathbf{p}$ posto nell'origine sono:

$$\mathbf{E}(\mathbf{r}) = k^2(\mathbf{n}\times\mathbf{p})\times\mathbf{n}\frac{e^{ikr}}{r} + [3\mathbf{n}(\mathbf{n}\cdot\mathbf{p}) - \mathbf{p}]\left(\frac{1}{r^3} - \frac{ik}{r^2}\right)e^{ikr} \tag{27}$$

$$\mathbf{B}(\mathbf{r}) = k^2(\mathbf{n}\times\mathbf{p})\frac{e^{ikr}}{r}\left(1 - \frac{1}{ikr}\right) \tag{28}$$

Dove $\mathbf{n}$ è il versore di $\mathbf{r}$ e k è il modulo del vettore d'onda $\mathbf{k}$, con $k = \omega/c = 2\pi/\lambda$. Sia $\mathbf{E}$ che $\mathbf{B}$ sono dati dalla somma di due contributi con un diverso andamento in r. Nel limite di campo lontano, con $kr \gg 1$, si ritrovano le usuali soluzioni

$$\mathbf{E}_{far}(\mathbf{r}) \approx k^2(\mathbf{n}\times\mathbf{p})\times\mathbf{n}\frac{e^{ikr}}{r} \tag{29}$$

$$\mathbf{B}_{far}(\mathbf{r}) \approx k^2(\mathbf{n}\times\mathbf{p})\frac{e^{ikr}}{r} \tag{30}$$

per le quali vale $\mathbf{E} = \mathbf{B}\times\mathbf{n}$. Queste soluzioni rappresentano onde sferiche la cui ampiezza scala come $1/r$.

Explicit field solution in the near-field

Nel limite opposto in cui $kr \ll 1$ sviluppando e^{ikr} si ottengono i campi

$$\mathbf{E}_{near}(\mathbf{r}) \approx [3\mathbf{n}(\mathbf{n}\cdot\mathbf{p}) - \mathbf{p}]\frac{1}{r^3} \tag{31}$$

$$\mathbf{B}_{near}(\mathbf{r}) \approx ik(\mathbf{n}\times\mathbf{p})\frac{1}{r^2} \tag{32}$$

i quali sono nella stessa forma di quelli prodotti da un dipolo statico o lentamente variabile nel tempo, come ci si aspetta nel limite di $kr \to 0$.

Static dipole means non propagating (the oscillating temporal behavior is indeed preserved)!

Suggested reading: Jackson, Classical Electrodynamics

Per comodità consideriamo separatamente i campi nel caso $\mathbf{r}\perp\mathbf{p}$ e nel caso $\mathbf{r}\parallel\mathbf{p}$. Definendo il parametro adimensionale $x = kr$, le (27) e (28) si riscrivono

$$\mathbf{E}_\perp(x) = \frac{e^{ix}}{x}k^3\mathbf{p}\left(1 - \frac{1-ix}{x^2}\right) \tag{33}$$

$$\mathbf{B}_\perp(x) = i\frac{e^{ix}}{x}(\mathbf{n}\times\mathbf{p})\frac{1-ix}{x^2} \tag{34}$$

$$\mathbf{E}_\parallel(x) = 2\frac{e^{ix}}{x}k^3\mathbf{p}\frac{1-ix}{x^2} \tag{35}$$

$$\mathbf{B}_\parallel(x) = 0 \tag{36}$$

Nel primo caso $\mathbf{E}$ e $\mathbf{B}$ sono entrambi trasversi, mentre nel secondo caso il campo elettrico ha una componente longitudinale non nulla dovuta al contributo del campo prossimo. Tale componente tende a zero per $x \to \infty$. Le ampiezze dei campi $\mathbf{E}_\perp(x)$, $\mathbf{B}_\perp(x)$ e $\mathbf{E}_\parallel$ sono rispettivamente

$$|\mathbf{E}_\perp(x)| = k^3\frac{p}{x}\sqrt{1 - \frac{1}{x^2} + \frac{1}{x^4}} \tag{37}$$

$$|\mathbf{B}_\perp(x)| = k^3\frac{p}{x}\sqrt{1 + \frac{1}{x^2}} \tag{38}$$

$$|\mathbf{E}_\parallel(x)| = \frac{2k^3}{x^2}p\sqrt{1 + \frac{1}{x^2}} \tag{39}$$

Come si può vedere il campo $\mathbf{E}_\parallel$ diventa dominante per x piccoli. Questo significa che per valori di r molto più piccoli di λ gli effetti del campo prossimo diventano importanti. Questi risultati, benché ricavati nel caso particolare di un dipolo elettrico nell'origine, sono molto generali e almeno qualitativamente descrivono l'emissione in campo prossimo di ogni tipo di sorgente.

The field close to the emitter (the near field) holds unique features

Nicola Paradiso, Tesi di Laurea in Fisica, Pisa 2005

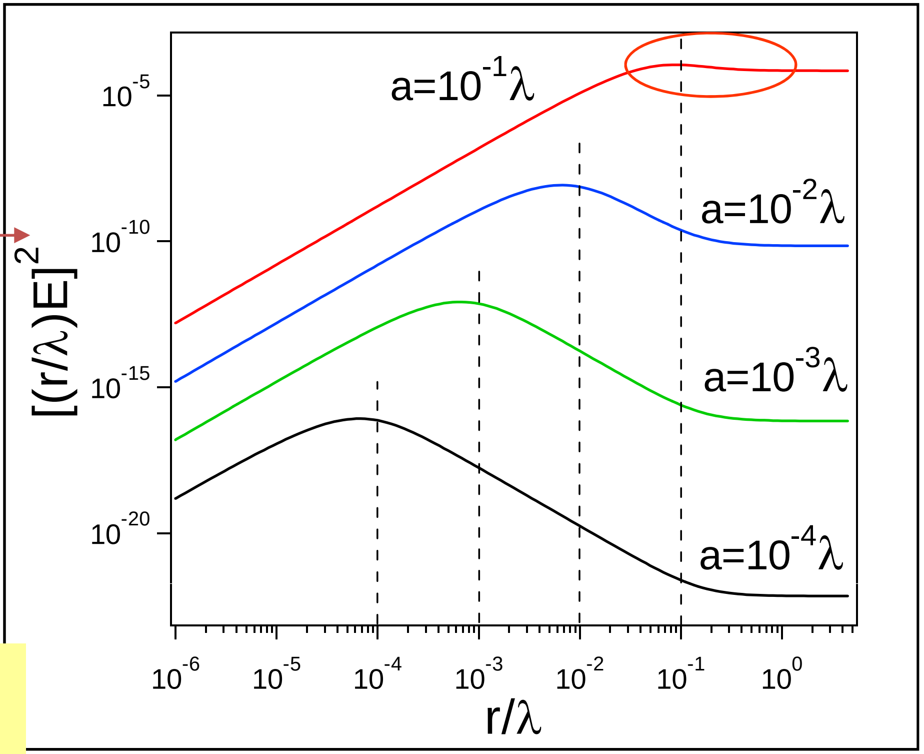
Near-field extinction

In the presence of a non-propagating (evanescent) wave, field undergoes a fast extinction as a function of the distance from the aperture (*z*)

$$I(0,0,z) \cong \left(\frac{8ka}{3\pi}\right)^2 e^{\frac{3\pi z}{4a}}$$

(ideal case, after Bethe)

"Normalized" field intensity as a function of the distance (***r***) from the aperture (estimation)



In practice:
a ~ 50 nm (~ $\lambda/10$) and the field drops rapidly to zero for *r* > *a* (i.e., for r/λ ~ 0.1)

The space distribution of the near-field (confined within a distance ~ *a*) gives SNOM its sub-diffraction space resolution

How to concern a sample with the near-field

A distinctive feature of near-field is that its amplitude rapidly drops to zero within a range ~*a*

⇩

During the scan, the probe tip must be kept “close” to the surface (typ at a distance < 10 nm)

“Constant gap” operation is strictly required for the SNOM images to be reliable

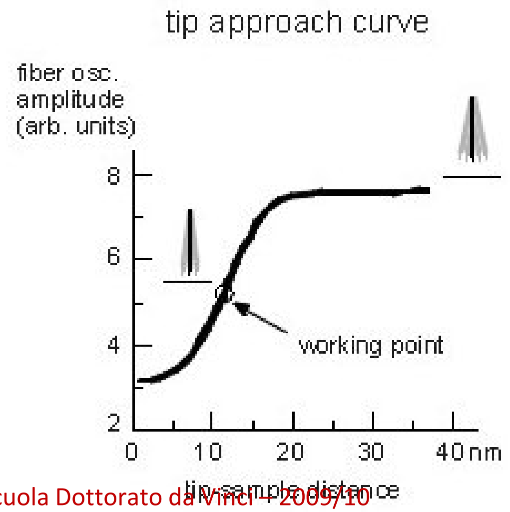
⬇

A method to continuously monitor tip/sample distance is needed

A feedback acting on the vertical piezo displacement is used (as, e.g., in non-contact AFM)

⬇

A topography image is simultaneously acquired during each scan, with a lateral resolution depending on the probe size, typ in the 100 nm range



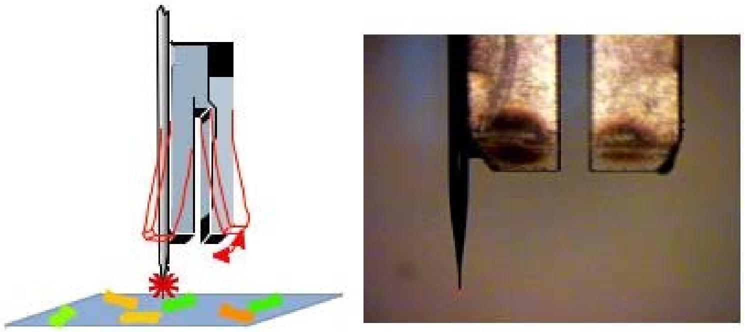
If the tip is kept in **longitudinal** oscillation, the oscillation amplitude depends on the distance due to shear-forces (mostly associated with friction of the air layers between tip and surface)

Notes:

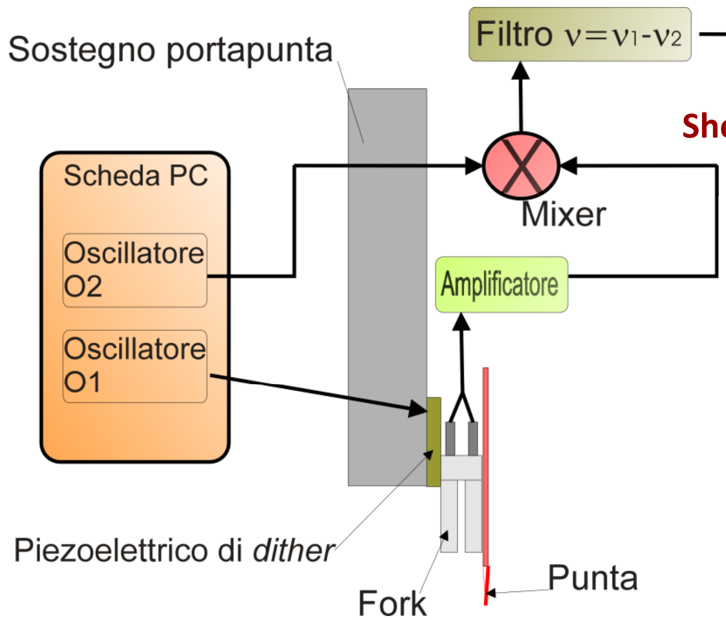
Oscillation amplitude must be small (typ ~ 1 nm) to prevent resolution loss

Oscillation at resonance frequency is required to get maximum sensitivity

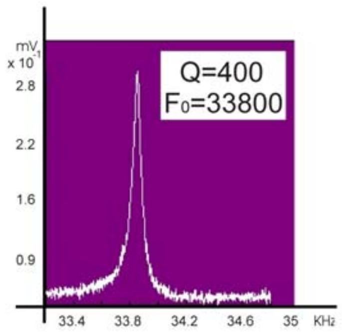
Shear-force method



Fiber tip is glued onto a *tuning fork*
(a quartz diapason acting as a mechanical oscillator and showing capabilities to measure the oscillation amplitude)



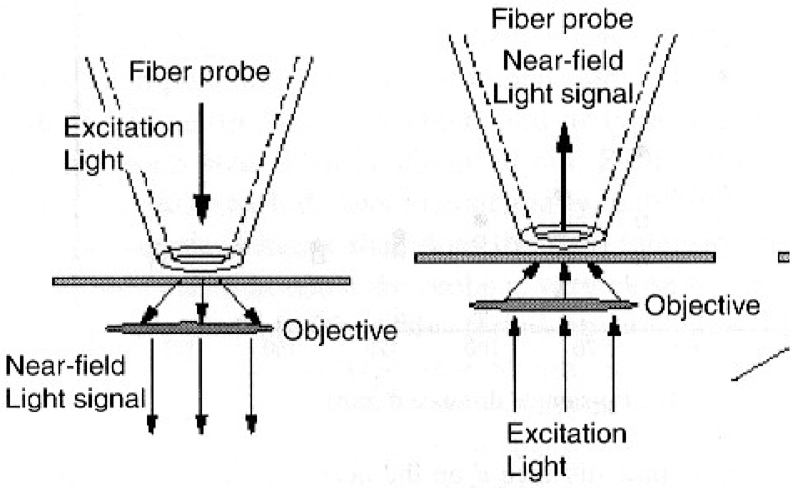
Shear-force method



Mech resonance spectrum
(undamped oscillation)

- A dithering piezoelectric transducer keeps the probe tip in oscillation along a direction **parallel** to the surface
- Oscillation amplitude is monitored by a *tuning fork*
- When the distance gets smaller (typ., below 10 nm), the oscillation is **damped** (and phase is changed) due to **shear-forces** involving many effects (e.g., viscous interaction of the air layer between tip and sample)
- Similar to AFM in tapping mode, but for the oscillation direction, the relevant distance and the involved mechanisms

Operation modes for SNOM



Illumination mode **Collection mode**

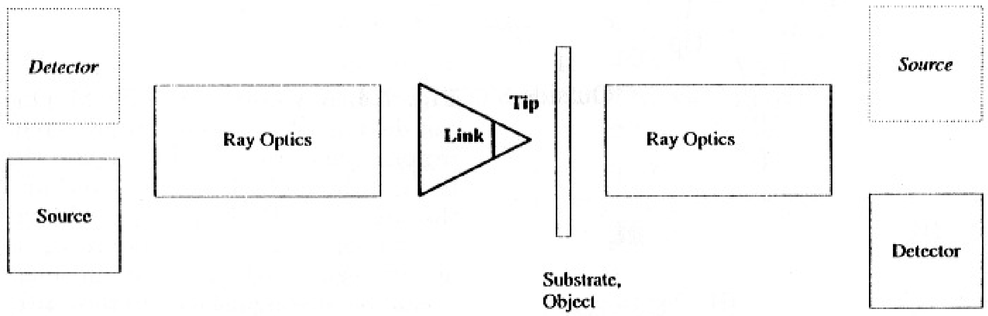


Fig. 7.3. The characteristic components of a SNOM. In collection modes the position of the detector and of the source is interchanged with respect to their position in the illumination modes

In **illumination mode** (the most common) the surface is concerned by the near field and the resulting scattered light is collected "in the far field" (either in transmission or reflection)

In **collection mode** the surface is illuminated by a propagating (conventional) field and the resulting scattered light is collected in the near field by the probe

Sub-diffraction space resolution is due to the non-propagating character of the near-field
(typ resolution comparable to the aperture size, i.e., tens of nanometers)

A very few words on apertureless SNOM

3.5 APERTURELESS NEAR-FIELD SPECTROSCOPY AND MICROSCOPY

As mentioned in Section 3.3, an emerging approach is the apertureless near-field spectroscopy and microscopy (Novotny et al., 1998; Sanchez et al., 1999; Bouhelier et al., 2003). The use of an aperture such as a tapered fiber opening poses a number of experimental limitations. Some of these are:

- Low light throughput due to the small fiber aperture and the finite skin depth (light penetration) into the aluminum metal coating around the tapered fiber.
- Absorption of light in the metal coating; this can produce significant heating that can create a problem in imaging, particularly of biological samples.
- Pulse broadening in the fiber, when using short pulses for nonlinear optical studies. Also, the fiber tip may be damaged by the high peak intensity as already discussed in Section 3.3.

The apertureless approach overcomes these limitations, at the same time providing a significantly improved resolution. It has been demonstrated by Novotny, Xie, and co-workers (Sanchez et al., 1999; Hartschuh et al., 2003; Bouhelier et al., 2003) that optical images and spectra of nanodomains ≤25 nm can be obtained using the apertureless … m.

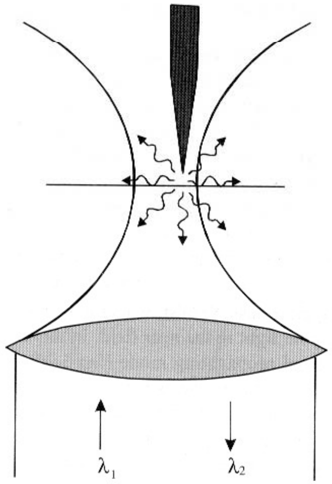


Figure 3.22. Metallic tip enhancing the local field by interacting with the focused beam at λ_1. The optical response at another wavelength λ_2 is collected by the same objective lens.

The two approaches used for apertureless NSOM are:

1. Scattering type, which involves nanoscopic localization and field enhancement of the electromagnetic radiation by scattering of the light from a metallic nanostructure. An example is provided by Figure 3.2 where the light is scattered by a sharp metallic tip. Scattering and field localization can also be produced by a metallic nanoparticle within nanometers of distance from the sample surface. The localization and enhancement of electromagnetic field by plasmon coupling to a metallic nanoparticle is discussed in Chapter 5 under "Plasmonics." This principle of obtaining nanoscopic resolution using scattering from a metallic nanoparticle also forms the basis of "plasmonic printing," discussed in Chapter 11 on "Nanolithography".
2. Field-enhancing apertureless NSOM, where a metallic tip is used to enhance the field of an incident light in the near field. In this case, the light is incident on the tip as a normal propagating mode (far-field). The strongly enhanced electric field at the metal tip produces nanoscopic localization of optical excitation. This approach offers simplicity and versatility of using light by just focusing on the metallic tip through a high-numerical-aperture lens. Hence it is described here in detail, with examples of some recent studies utilizing this approach.

A nanoparticle, or a nanosized tip, irradiated by a propagating field, acts as a quasi-pointlike source of the near field

What can be measured by SNOM

The effects of the **local interaction between the sample surface** (i.e., a layer with thickness comparable to the near field range) **and the near field** photons are recorded

They can be regarded as analogous (but for the sub-diffraction resolution and the surface origin) of conventional **optical transmission and/or reflection** measurements (depending whether the sample is transparent or opaque)

Non propagating behavior of the exciting near field can however play a role (for instance, specific polarization can give access to otherwise forbidden transitions, …)

Examples

- ✓ **Local variations of the "refractive index"** can be derived by analyzing the scattered radiation
- ✓ In case of emitting (photoluminescent) samples, fluorescence can be excited by the near field, and **photoluminescence** maps can be acquired
- ✓ By implementing a polarization control system *(see later on)*, **optical activity** of the sample (e.g., dichroism, birifringence) can be analyzed at the sub-diffraction level
- ✓ More advanced spectroscopy (e.g., Raman) can be carried out

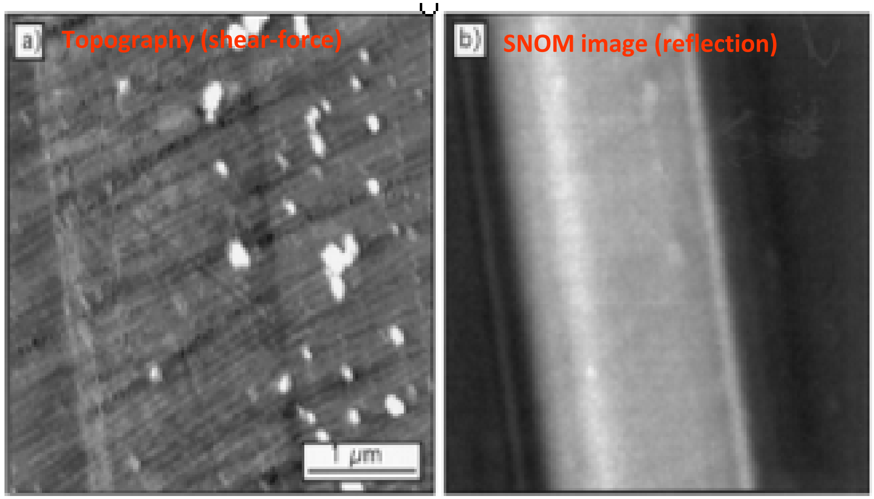
Examples

Collection mode can be used to map emission of, e.g., electroluminescent devices
Also, evanescent radiation, e.g., stemming from a waveguide surface, can be mapped

In addition, the tip/sample distance control, being based on a feedback system, allows acquisition of topography maps simultaneously with every SNOM scan (with a space resolution in the tens nm range)

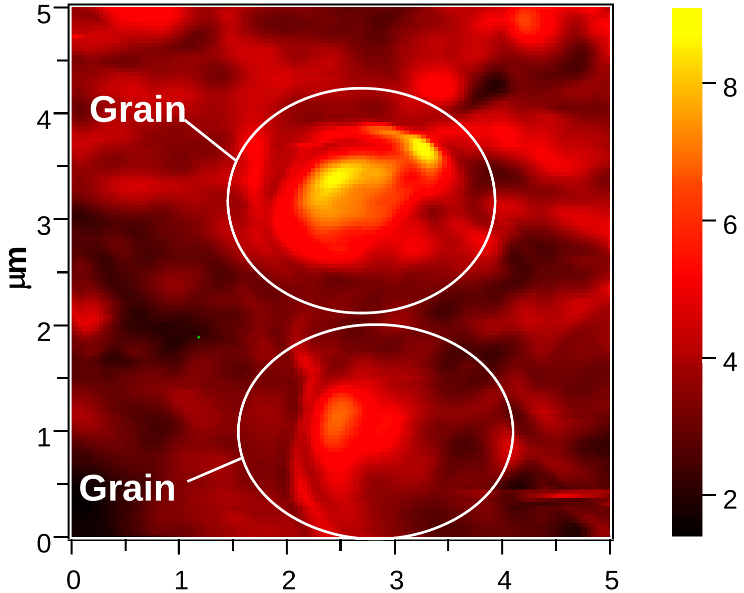
Morphological and optical information acquired and compared at a glance!

Examples of SNOM in reflection mode I



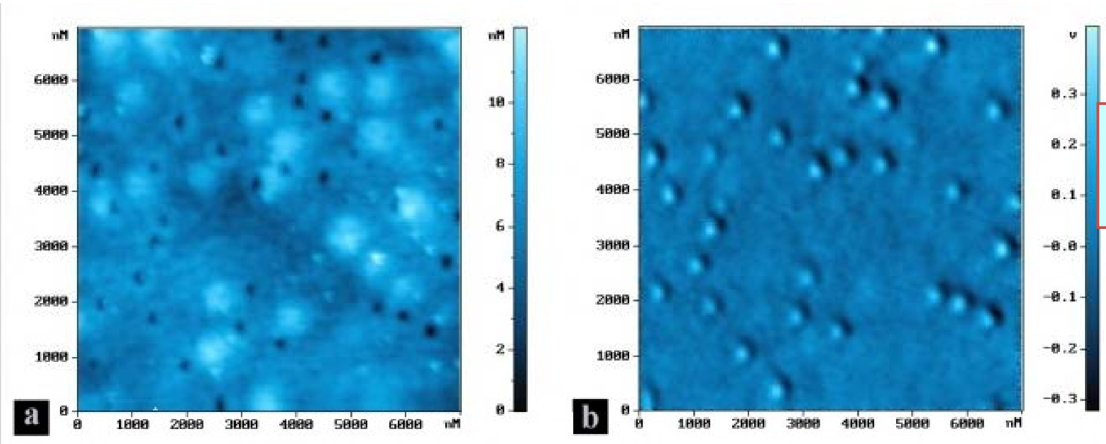
NiFe stripes embedded in an alumina matrix

YBCO/YSZ/NiFe multilayer deposited by PLD



Compositional differences (e.g., two-phase materials or structural fluctuations) associated with variations in the optical properties (i.e., refractive index) are evidenced in SNOM with a sub-diffraction space resolution

Examples of SNOM in reflection mode II



Contrast enhancement due to the use of optical signals

Quantum dots

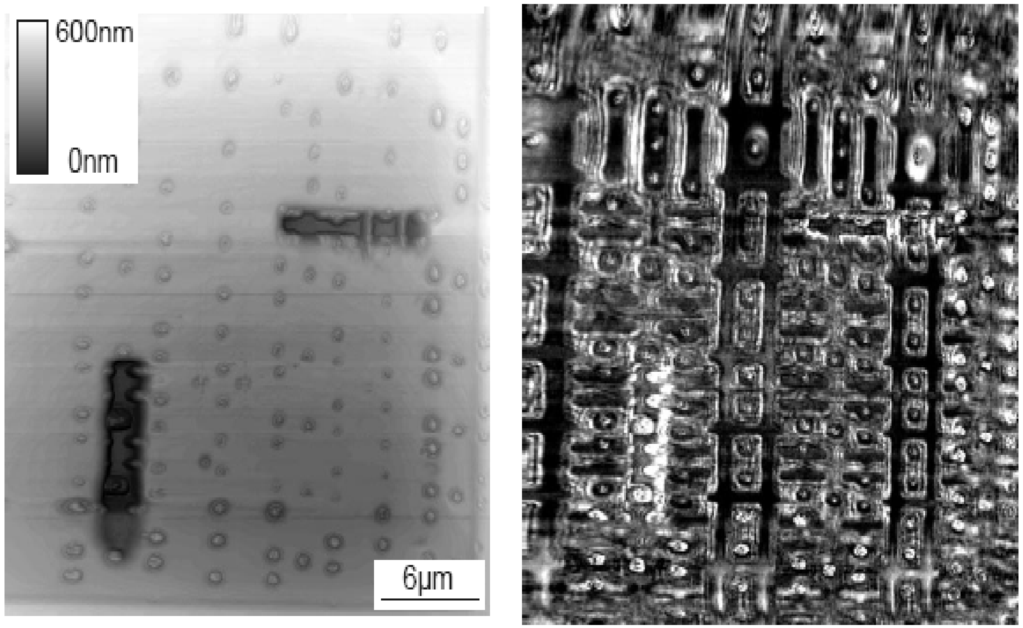
Shear Force (topography) (a) and reflection (b) images of In-Ga quantum dots made with the use of He-Cd 442 nm laser.

Images courtesy of Igor Dushkin, NT-MDT.

Scan size: 7x7 µm

Source MDT-file: download (515.06 Kb)

Buried structures can be detected (when using illumination light at a wavelength transmitted by the upper layers of the device)



g. 2. Topography (left) and oblique reflection (right) micrographs of aluminium structures buried underneath interlayer dielectrics.

Examples of SNOM photoluminescence

Saiki et al. (1998) and Matsuda et al. (2001) conducted room temperature photoluminescence study on a single quantum dot from InGaAs quantum dots grown on a GaAs substrate. Their result is shown in Figure 3.12. Because of the spectral resolution obtained by sampling only a single quantum dot (no inhomogeneous broadening), they were able to observe, at an appropriate excitation density, emission not only from the lowest level (subband) of the conduction band but also from higher levels. (See Chapter 4 for a description of these bands.) They were able to study the homogeneous line width, determined by the dephasing time of excitation (see Chapter 6 for a description of dephasing time), as a function of the interlevel spacing energy. They found that the line width was larger for a smaller-size quantum dot for which the interlevel spacing is larger. (This is predicted by a simple particle in a box model as the length of the box becomes smaller, see Chapters 2 and 4.)

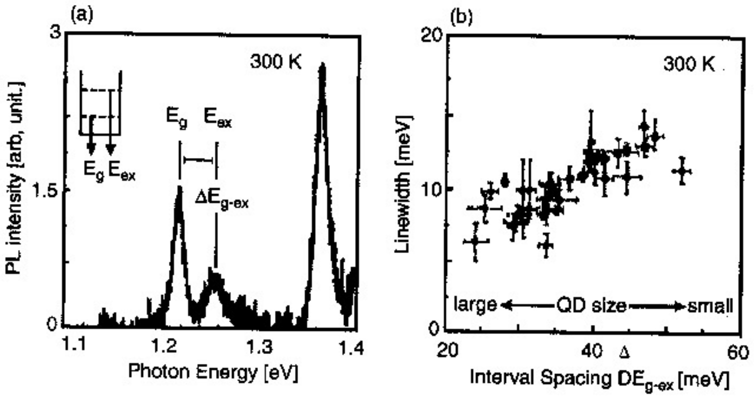


Figure 3.12. Photoluminescence spectrum of single QD at room temperature (a), and dependence of the homogeneous linewidth of ground-state emission on interval spacing, which is closely related to size of Qd's (b). From Saiki and Narita (2002), reproduced with permission.

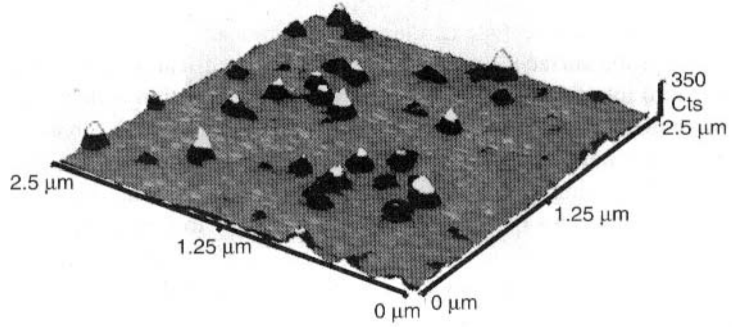


Figure 3.13. Fluorescence NSOM images of single molecules. From Professor D. Higgins and Professor P. Barbara, unpublished results.

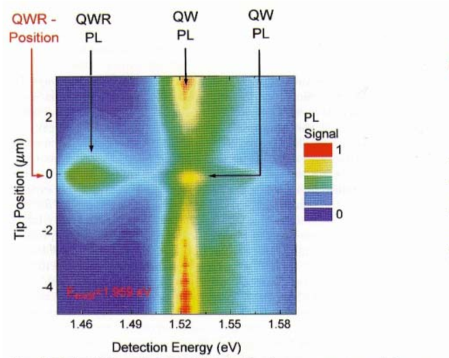


Fig. 7.23. Spatially resolved near-field photoluminescence spectrum of the quai tum wire heterostructure. The spectrum was recorded for spatially resolved exc tation of the sample at 1.959 eV. The tip was scanned along the lateral directic perpendicular to the wire. The PL intensity (in arbitrary units) is plotted as a fun tion of tip position and detection energy. The color red corresponds to high, whi purple corresponds to low PL intensity. The quantum wire emission is centere around 1.46 eV. Note that, in addition to the flat quantum well luminescence at 1.522 eV, a further, slightly blue shifted, sidewall quantum well emission is resolved

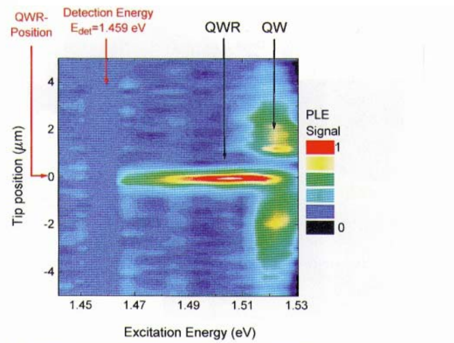


Fig. 7.24. Spatially resolved near-field photoluminescence excitation spectrum. The PLE signal is plotted as a function of tip position and excitation energy. The scan direction is perpendicular to the wire. The photoluminescence was detected at 1.46 eV. The color red corresponds to high and purple to low intensity

Da Wiesendanger Ed., Scanning Probe Microscopies (Springer, 1998)

Single, isolated nanostructures, nanoparticles, or emitting molecules can be analyzed in spectral terms

Conclusions

- ✓ Optical microscopy represents a well established tool to measure and analyze optical properties at the small scale
- ✓ Ultra-small features cannot be resolved due to optical diffraction, that limits the space resolution to approximately half the wavelength of the optical radiation used for the analysis (corresponds to roughly a few hundreds of nm)
- ✓ Near-field are non propagating (i.e., evanescent) fields which can be produced in several ways, e.g., by using very small apertures
- ✓ Near-fields are used in SNOMs, where extreme diffraction is used to overcome the diffraction itself, in a rather brilliant idea
- ✓ SNOMs share several aspects with other SPMs and can nowadays be routinely used to get optical information at the ultra-small scale