Chapter 1

History of Near-field Optics

This first chapter deals with the notions of imaging systems, vision, object and image in a historical context.

1.1 Notion of imaging system

The optical system has never been invented: it is one of the five senses of most living beings and for a very long time man has tried to improve his ability to see.

It is admitted that spectacles were in use in China well before they were known and used in Western countries. These spectacles turned out to be planar without focusing properties. The purpose was mainly cosmetic.

The first chronicle relating the use of an optical component for improving human eyesight is due to Pliny the Elder† who wrote in 23–79 A.D. that : "*emeralds are usually concave so that they may concentrate the visual rays. The Emperor Nero used to watch in an emerald the gladatorial combats*". That is the first description of the use of a piece of a transparent material as a monocle for correcting short-sightedness.

There is very little evidence of the use of lenses in Antiquity. The *Lanyard lens* discovered in Nimroud (Neoassyrian city) dated to 721–705 B.C. and often cited as the first case of convex lens is probably a mere piece of jewellery.

However, a well-known example of light ray focusing, not directly connected with imaging, are the famous burning-mirrors invented by Archimedes† for the defence of Syracuse.

The physical mechanisms describing the capacity of glass to bend light rays have been known for a long time. In the 2nd century B.C. C. Ptolemy† described a method to determine the refractive index of water by measuring the apparent bending of a stick partially immersed in a pool of water. In the 1st century A.D. Seneca† proposed the use of a sphere filled with water as a converging system. The technique is still in use in jewellery.

In the 10th century A.D. the Arabian scholar Alhazen† published the first major optical work –*Opticae Thesaurus*– in which the basic principles of optics were discussed.

Finally the Englishman R. Bacon† wrote in his *Perspectiva* (1267): "*Great things can be performed by refracted vision. If the letters of a book, or any minute object, be viewed through a lesser segment of a sphere of glass or crystal, whose plane is laid upon them, they will appear far better and larger*".

For his pertinent approach of light refraction R. Bacon is generally considered as the inventor of the glass lens.

In the 17th century, A. Leeuwenhoeck† is credited with the invention of the microscope and the report of the first magnified images of biological samples. Contemporary with A. Leeuwenhoeck, R. Hooke† probably participated in the improvement of the microscope. He is often considered as the founder of microbiology. The very first microscopes were very simple and mainly composed of a single lens which was held up to the eye (like a modern magnifying lens). The sample was placed on a small pin just on the other side of the lens. Despite the simplicity of the tool, Leeuwenhoeck's microscopes were excellent and drew attention to that emerging field. During the same period appeared the compound microscope probably due to Hans and Zacharias Jannsen† (father and son) who constructed a simple instrument made of a pair of lenses mounted in a sliding tube. J. Kepler† described a similar system around 1611 (despite his preference for simple lens microscopes). The name "microscope" has first been proposed by the naturalist G. Faber† in 1625. Two years before, Galileo Galilei† used, in a communication, the generic term of "telescope".

These microscopes were the ancestors of our modern tools. Figure 1.1 shows a British microscope built in 1730. Its overall aspect is not far from our modern tools.

The first attempt to increase the resolution in a non-classical way is due to G.B. Amici† who introduced the notion of immersion† in the 19th century. At the end of the 19th century the scientist E. Abbe† and later

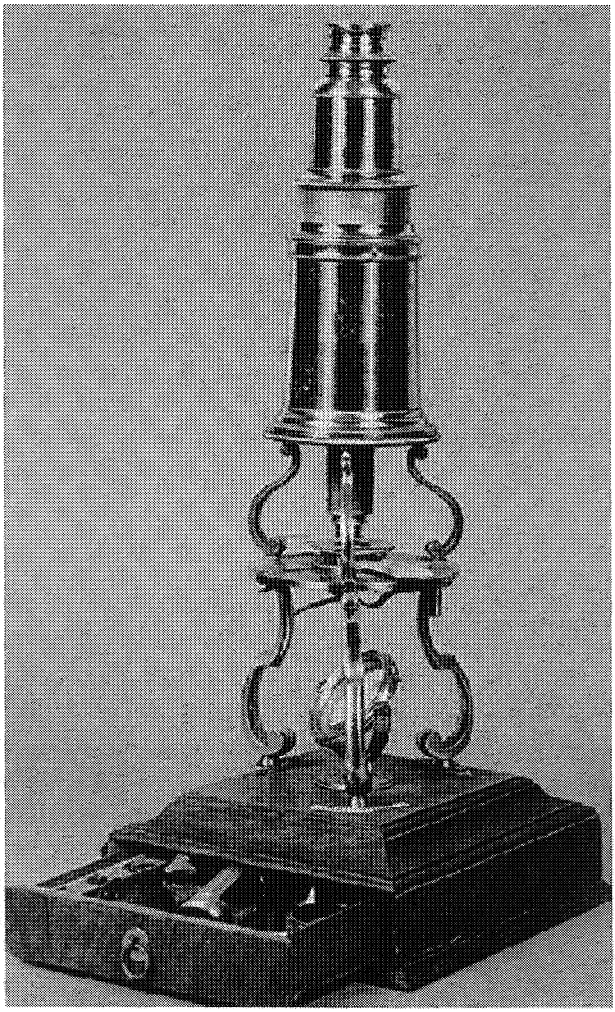


Fig. 1.1 Early microscope (Culpepper-Scarlett, 1730).

Lord Rayleigh† demonstrated the existence of a limit of resolution ruining the hopes of scientists to observe the ultimate structure of matter by means of optical tools [Abbe, 1873, Rayleigh, 1879, Rayleigh, 1880].

Let us also mention the invention by A. Köhler† of a particular illumination combination. The arrangement enables both a uniform illumination of the sample and an optimal light transmission through the various optical components of the microscope.

Finally in the 1950's, F. Zernike† invented the phase contrast microscope. He was awarded the Nobel Prize for that discovery. It is the very first case of image improvement by encoding (the image is no longer a mere magnified copy of the sample as seen by a human eye).

1.2 Bases of imaging

1.2.1 *Vision*

Notions which seems obvious today such as light propagation were not so clear some centuries ago. Our perception of light and the natural feeling that light propagates from the source to the detector have been discussed and controversed for several centuries in Antiquity. The feeling that there is no other way to interpret light propagation is mainly due to an accumulation of knowledge which sometimes has had a negative effect on progress. The conviction of scientists that the smallest possible periodicity to be imaged by optical procedure could not be smaller than half a wavelength is due to misinterpretation of Abbe's work by the scientific world.

In the 5th and 4th centuries B.C. Greece knew its highest degree of civilization. Some schools of thoughts working and debating around the problem of vision already existed [Ronchi, 1970]. Three hypotheses coexisted to explain the mechanism of perception. The first hypothesis was that "something" emanated or was emitted by the eye and travelled towards the object. The idea was accepted and defended by the Pythagoricians†. The second hypothesis was that "something" was emitted by the object and then caught by the eye. The Atomists of the School of Democritus† (470–370 B.C.) were in favour of such a theory. The third proposition defended by Empedocles† (5th century B.C.) was a combination of the other two.

A fourth possibility was proposed by Aristotle† (384–322 B.C.) who emitted some strong reserves concerning such hypotheses. His feeling was that something was propagating through the medium from the object to the eye. The main difference with Pythagoricians' thought was that the emission by the object was not an emanation of the latter. It was something else we can call *light*. Exegesis on Aristotle's texts presents him as a precursor of the modern wave theory though he has not been able to formulate his ideas with clarity.

(a) Pythagorician's perception of vision.

(b) Democritus point of view.

(c) Empedocles' concept.

(d) Aristotle approach.

Fig. 1.2 Perception of sight in Antiquity.

These different points of view concerning the mechanisms of vision coexisted for a long time demonstrating the difficulty for philosophers and scientists to have a clear notion of light propagation.

If it is definitely admitted that the eye cannot be the source of the emission of light, Empedocles' hypothesis cannot be so easily rejected. In the case of near-field detection –as it will be shown and explained in this book– the emitter and the detector often play symmetrical roles because the light emitted by the sample is re-emitted by the collector, which means that Empedocles' concepts were close to the actual mechanisms taking place in near-field microscopy. Moreover, in the case of evanescent waves detection, Empedocles' hypotheses are remarkably verified since evanescent waves are bound to the object and contain information about the latter. Although they do not carry energy they contain something which can be transmitted by non-radiative transfer. Maybe Empedocles' ethereal link between the object and the eye is simply what we call information today.

1.2.2 *Image*

The notion of *image* is also very old in the history of science. Assuming a physical being called *object*, the *image* of this object is defined as the result of a given transform of the object. The similarity between object and image is not compulsory. The image of a star is weakly related to the star itself. The Fourier spectrum of a given field distribution is also mathematically speaking an image. We are obviously more accustomed to the usual notion of image in which the relationship between object and image tends towards a simple affine transform or a similarity.

$$Object \xrightarrow{T} Image$$

The relation between an object and an image is never simple because the object itself is generally a piece of matter (a star, a bacteria, an atom...) that is a three-dimensional material substance. On the contrary, the image is often a two-dimensional distribution of light intensity.

1.2.3 *Far-field imaging systems*

The first way of describing an imaging system consists in considering the light beam as a bundle of elementary light rays obeying the conservation laws (*Maupertuis's principle, Fermat's principle, Clausius's law*, etc.) For most usual applications such a description is precise enough and reflects the physical reality rather correctly. In such a description each light ray is infinitely thin and can be analysed independently from the others. Since its behaviour is fully described by energy conservation laws, the possible diffraction of light is not taken into account. Therefore, in an aberration free system (a constraint which can be approached thanks to the new materials and the computing capacities of modern computers) the image can be a quasi-perfect copy of the object itself.

In the case of microscopy, such a rough modelling is not sufficient and the wave theory is a necessity. One of the precursors of the wave theory applied to imaging system is the scientist E. Abbe who demonstrated that every object behaves as a grating which diffracts more or less the light going through it. When the grating period is small, the diffraction angle is large and only a small amount of light can be collected by the imaging lens. On the contrary when the period (*i.e.* the detail) is large, the light beam is slightly diffracted and the maximum of light is caught by the lens. By this simple approach one can understand that high frequencies will be less easily transmitted than low frequencies and also that there must be a cutoff frequency necessarily connected to the geometry of the collecting system. Figures 1.3 (a) and (b) sketch this behaviour.

Let us note that this approach supposes coherent illumination. It is quite easy to extend it to partially coherent light, see for example the textbooks of M. Beran & G. B. Parrent and of J. Peřina [Beran & Parrent, 1964, Peřina, 1971].

Following Abbe's theory, a light beam impinging on a flat (bi-dimensional object) is diffracted by the object features. According to the diffraction properties, the smaller the details, the higher the diffraction orders. The resulting diffracted light cone is then totally or partially collected by the imaging system as shown in figure 1.3.

After World War II, imaging devices improved dramatically with the transposition in optics of the well-known transfer laws used and developed for electronics. It is out of the scope of this book to treat the general problem of imaging. The reader could consult several reference books such as "*Principle of Optics*" by M. Born and E. Wolf [Born & Wolf, 1999] or "*Introduction to Fourier Optics*" by Goodman [Goodman, 1968], etc.

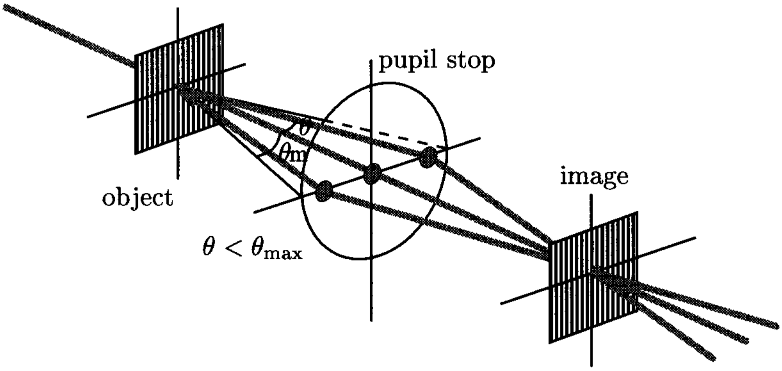
1.2.4 *Notion of superresolution*

If evanescent fields and more generally non-propagating fields belong to the physical background of near-field microscopy, the search for superresolution is the main motivation of the activity in this new field of investigation.

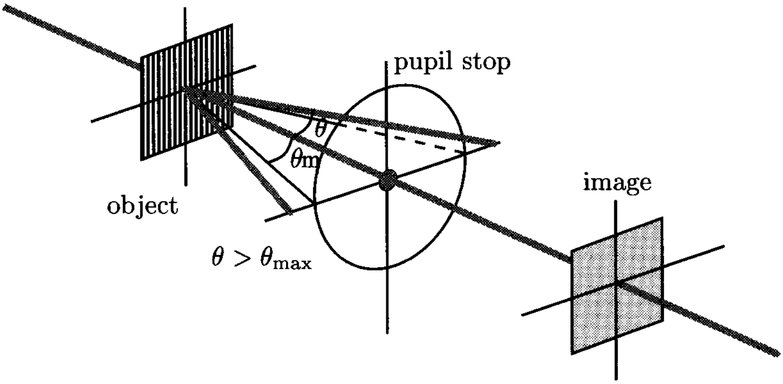
The concept of superresolution is certainly not new, some scientists such as E. Synge† introduced the notion of "ultra-microscopy" in the twenties but the true takeoff of the concept of superresolution began after World War II with the works of scientists such as H. Hopkins†, P.M. Duffieux†, A. Lohmann† and several others (see author index).

Note that the development of the concept of superresolution is directly related to the development of *Fourier Optics*† which is nothing but the transposition of frequency analysis used in electronics in the field of optics. Starting from the notion of frequency bandwidth, the temptation to break the classical cutoff frequency of a system was a natural feeling in search of high resolution in optics.

Scientists such as W. Lukosz [Lukosz, 1967] demonstrated that the bandwidth was not the definitive limit in imaging systems as it was supposed from Fourier analysis. The degrees of freedom were the invariant quantity which was impossible to pass. A great deal of works were published in the period 1960–1980 proposing new ways to go beyond the classical cutoff


pupil stop
object
image
θ
θm
θ < θmax

(a) Imaging of a coarse grating: the diffraction orders pass through the lens pupil.


pupil stop
object
image
θ
θm
θ > θmax

(b) Imaging of a fine grating: the diffraction orders are stopped by the lens pupil.

Fig. 1.3 Abbe theory illustration.

frequency. The *theta-modulation* of J. Armitage [Armitage & Lohmann, 1965] & al. is one of the best examples of the possibility of reshuffling the parameters defining the degrees of freedom.

1.2.5 *Near-field imaging systems*

As it will be shown in chapter 2 and throughout the whole book, the notion of near-field –and its corollaries– is recent because the observation of non-propagating fields can never be as straightforward as the observation of stars in astronomy and microbes in far-field microscopy.

Near-fields are usually very confined in the vicinity of matter so that their detection is not easy and needs nanometre accuracy actuators capable of moving a suitable sensor a few nanometres from the sample.

The first evidence of the existence of near-fields is due to I. Newton[†] who discovered an anomaly in light behaviour with his famous "*Two-prism-experiment*" carried out at the end of the 17th century [Newton, 1952, Newton, 1955].

In his experiment, I. Newton studied the total reflection of a light beam inside a glass prism. The ability of a simple bare piece of glass to play the role of a perfect mirror intrigued Newton. To understand the light mechanism he tried to frustrate the reflection by fastening a second identical prism against the first one. The expected result was the transmission of the light beam from the first to the second prism without major perturbation. If I. Newton really observed such a behaviour of the light beam, he also discovered that *it travelled from the first to the second prism* ***before*** *the two flat prism surfaces were in contact that is* ***before*** *the material continuity was ensured between the two glass pieces.* The gap estimated by Newton was exceedingly small (about 10 nanometres) but it **was**. At that time, with the classical approach implying tiny light rays or even light particles, Newton was unable to explain the real mechanism taking place in his experiment. He suggested that because of their speed the light particles were unable to stop at the interface between the glass and the air and leave the glass before being attracted by the glass according to the gravitation principle Newton was elaborating. Surprisingly, the strange property of the evanescent wave which is not detectable *stricto sensu* escaped I. Newton who saw in the total reflection a problem of attraction of particles by a body.

Newton's experiment is probably the first reported manifestation of the existence of the evanescent light wave and also probably the first non-classical behaviour of a physical phenomenon observed in a laboratory room.

If near-field optics is mainly based on the existence of evanescent waves, a few other fundamental notions are crucial to understand the superresolution capacity of near-field microscopes. One of them is the notion of non-propagating field. We will see in chapter 2 that such a notion must be handled carefully because it can cover different concepts which have no strong connection with one another.

The emission of the dipole as described in Born and Wolf's book [Born & Wolf, 1999] and many other physics treatises, points out the dependence of the field emitted by the dipole versus the distance r from its centre as varying in inverse power of r, r^2 and r^3. It is easy to show that the first term (varying in r^{-1}) is associated with a propagating field obeying the energy conservation law. The other two terms (in r^{-2} and in r^{-3}) cannot be explained simply: they carry no energy and therefore do not propagate. They are often called non-radiating terms. However, they cannot be neglected in near-field optics for they are one of the bases of sub-wavelength resolution.

Finally, the emergence of a new technology is often the result of the convergence of non related scientific progress. It is the case of near-field optics which has been reactivated thanks to the invention of the *Scanning Tunnelling Microscope* usually called STM[‡]. The microscope invented in Switzerland by G. Binnig, H. Rohrer[††] with the participation of C. Gerber has certainly been the most significant progress in metrology for the last two decades. That tool is not only a technological *tour de force* but also a conceptual revolution in the way of observing an object. *Instead of putting the detector far away from the sample and using the free or guided propagation of the physical quantity (photons, electrons, ions, waves...) to transfer the information from the sample to the detector, the detector is set very close to the sample in a zone known as near-zone characterized by particular physical properties.* In such situations the notion of propagation becomes meaningless and it becomes impossible to differentiate the respective role of the sample and of the detector. As often asserted in local probe microscopy[†], "*the tip gives an image of the sample and the sample gives an image of the tip in a similar way*".

This fantastic tool created a great emulation in the world of metrology. Therefore, it was immediately followed by a great deal of microscopes based on the same basic principle *i.e.* the detection of a field such as a force, an electromagnetic or even an acoustic field in their near-zone.

This new approach completely upset our perception of the physical world because until then, the knowledge of the world had been exclusively based on the detection of *propagating fields* that is electromagnetic fields propagating from the sample to the detector.

In electron scanning tunnelling microscopy and more generally in every other local probe microscope, the signal collector is so close to the sample that *it is no longer dissociable from it.* The non-separability is not a simple technological problem due to a physical proximity between two pieces of matter, it is a fundamental principle due to the nature of electromagnetic fields. The non-separability of the detector and the sample is certainly one of the most disconcerting and frustrating property of near-field microscopy. This point will be discussed in chapter 4.

Another embarrassing question is the description of the field in the near-zone. It is often read in scientific literature that the '*field in the vicinity of an object is composed of both non-propagating (evanescent) components and propagating ones*'. This physical duality of the field has the merit of being simple but it conceals the fact that these components cannot be dissociated: the propagating field exists *because* the non-propagating terms exist. Only the propagating components can be defined without ambiguity in the far-zone (the distance playing the role of an efficient filter for evanescent components). It seems more suitable to see the near-field as a whole even if that curious field cannot be sketched as simply as a propagating one. That complexity is a consequence of the analycity of the field amplitude. E. Wolf & al. demonstrated that inseparability in a article often cited in near-field literature [Wolf & Nieto-Vesperinas, 1985].

Another key element in near-field detection process is the detection of evanescent field components known as *tunnelling effect*. As it will be shown in chapter 3, the paradox of tunnelling effect is not only in the peculiar capacity of a particle to pass through a potential barrier higher than its own energy, but also in the fact that it can be interpreted both as a non-classical "specifically quantum effect" as asserted by quantum physicists and as a simple classical phenomenon which can be explained in terms of boundary continuity conditions –as shown in reference [Born & Wolf, 1999]– or even in terms of simple energy and momentum conservation laws.

Near-field optics and more specifically near-field microscopy open a new era in electromagnetism *i.e.* in the knowledge of electromagnetic fields in the vicinity of matter. Near-fields and evanescent fields are certainly more interesting than propagating waves as 'laboratory object'. Propagating components far away from the source or the object which emits them are strongly filtered. The evanescent field is a more or less faithful imprint of the sample. It has probably not yet revealed all its mysteries.

1.3 History of near-field microscopy

It appears from the previous comments that near-field microscopy was born from the association of a few concepts such as *evanescent wave detection*, *superresolution*, *local detection*, *scanning detection*, *tunnelling detection*, etc. These concepts have appeared in the course of the last 100 years independently. Therefore, there is no unique vital lead to retrace the rather erratic story of near-field microscopy.

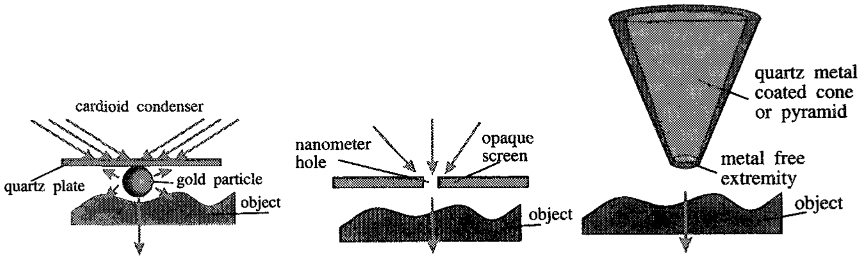
1.3.1 *Synge's speculation*

The first mention of near-field microscopy appeared in the article published by the Irishman E. Synge [Synge, 1928] who suggested the use of a new concept to break the diffraction limit. The article entitled '*Suggested method for extending microscopic resolution into the ultra-microscopic region*' pointed out clearly the ambition of his author. That speculative work contains the bases of modern near-field microscopy, introducing the notions of superresolution, local detection and above all, evanescent detection and even a differential screw system for moving the detector. As emphasized by the English historian D. Macmullan who recently exhumed Synge's article [McMullan, 1990], E. Synge exchanged correspondence with A. Einstein about the possibility of using other devices (a gold particle, a metallized quartz cone, etc.) as a local collector for increasing the spatial resolution. A few years later, E. Synge proposed a way to move with great precision the small hole in the last few nanometres from the surface [Synge, 1932]. The solution was based on the use of piezo-electric actuators in place of the usual differential screw systems. Synge's ideas for moving the detector are the two main ways still in use in local microscopy.

Note that work has been ignored by scientists until the work of D. Macmullan on microscopy history gave Synge's work all the credit it deserved.

1.3.2 *J. O'Keefe's letter*

In a short letter J. O'Keefe proposed a simple way to increase the spatial resolution in microscopy [O'Keefe, 1956]. Based on an experimental approach, he suggested to use a small pinhole as a detector instead of the usual objective. By scanning the hole along the illuminated sample, it was


cardioid condenser
quartz plate
gold particle
object
nanometer hole
opaque screen
object
quartz metal coated cone or pyramid
metal free extremity
object

(a) Basic's Synge idea.

(b) Einstein's answer.

(c) Final Synge's proposition.

Fig. 1.4 Three speculative Synge-Einstein's ideas to increase the spatial resolution in microscopy.

possible to detect a superresolved image whose resolution would be limited by the hole diameter only.

1.3.3 *E. Ash and G. Nicholls realization*

At the beginning of the seventies an article was published by E. Ash and G. Nicholls [Ash & Nicholls, 1972] describing with details a micro-wave realization allowing a sub-wavelength resolution. Again, the basic component was a small pinhole scanning the object surface. Despite the use of microwaves instead of visible light, it was the very first realization of a near-field microscope.

1.3.4 *Superresolution in imaging systems*

Independently of Synge's work, superresolution is one of the motivation for the development of near-field microscopy. After the last war, 'Fourier physicists' demonstrated that it was possible to go beyond the diffraction limit by ingeniously exploiting the degrees of freedom of the message to be transmitted through the optical system. A. Lohmann, J. Armitage, and a great number of others used the two-dimensional Fourier spectrum of the object in order to optimize the transfer. The basic idea was to match the object spectrum to the optical transfer function (or the contrary). The so-called '*theta modulation*' [Armitage & Lohmann, 1965] is probably one of the most accomplished technique for transmitting information with the minimal loss of resolution. Because of the need for pre-encoding the object, it has not been developed as a superresolution alternative in modern microscopy.

Despite the relative failure of superresolution techniques based on object spectrum-transfer function matching, the possibility of passing the diffraction limit was admitted in the scientific community and a search into a more efficient technique of superresolution went on on a completely different basis.

1.3.5 *Scanning tunnelling microscopy*

As mentioned previously, G. Binnig and H. Rohrer [Binnig & Rohrer, 1982] published an article describing a new instrument based on the measurement of the tunnelling current between a small metallic tip[†] and a conducting sample. The feat stressed in scientific magazines was the exceedingly small size of the apex of the tip often ending with a single atom: that single atom transmitting the tunnelling current to the amplifier. Since the tunneling current can be detected over a distance of about one nanometre, the tip was necessarily very close to the sample. The most significant improvement in the experiment has been the introduction of a feedback loop maintaining the tunnelling current constant and therefore allowing the tip to move a few angströms from the sample surface. The xyz tip motion reproduced very faithfully the exact electronic behaviour of the sample under test.

That fantastic tool was followed in the mid-eighties by a series of transposed technologies involving ions, photons, acoustic waves, radio-frequency waves, capacity measurements, etc.

The *Near-Field Optical Microscope* NFOM[‡] belongs to that family. However, the works reporting the first experiments in near-field optical microscopy do not mention the physical connection between Electron Tunnelling Microscopy and Scanning near-field Optical Microscopy.

1.3.6 *Early optical near-field microscopes*

The story of near-field microscopy is still too recent to be dealt with in this historical chapter (see chapter 9 instead). However, we can consider that near-field optics was born in the eighties both in Switzerland and in the United States. In the mid-eighties, D. Pohl & al. proposed a characterization tool called ***optical stethoscope*** for its capacity of imaging details far smaller than the wavelength (such as an acoustical stethoscope) [Pohl *et al.*, 1984]. The same year, G. Massey using a Fourier approach showed that a resolution of about 30 nanometres was probably achievable [Massey, 1984] and A. Lewis & al. proposed a superresolving instrument of 50 nanometres resolution [Lewis *et al.*, 1984]. Such instruments have been called *Scanning Near-Field Optical Microscope* (SNOM)‡ or *Near-Field Scanning Optical Microscope* (NSOM)‡. A turning point in this short history is the invention of the *Scanning Tunnelling Optical Microscope* (STOM)‡ also called PSTM‡ for *Photon Scanning Tunnelling Microscope*, which was born simultaneously in France and in the United States [Courjon *et al.*, 1989b,Reddick *et al.*, 1989,de Fornel *et al.*, 1989]. That microscope established the link between near-field microscopy and tunnelling effect. Since then the notion of non-radiating detection has experienced a rapid development and theories and new modellings involving evanescent waves have begun to flourish.