

Classification

Physics Abstracts

42.30 — 61.16P — 07.80

A mathematical morphology approach to image formation and image restoration in scanning tunnelling and atomic force microscopies

Noël Bonnet, Samuel Dongmo, Philippe Vautrot and Michel Troyon

Laboratoire de Microscopie Electronique et Unité INSERM 314, 21 rue Clément Ader, 51100 Reims, France

(Received July 4; accepted December 16, 1994)

Abstract. — In the nanometer resolution range, image formation in STM and AFM cannot be described by a convolution process but is essentially governed by the geometrical interaction between the specimen surface and the tip surface. This non-linear process can be simply described by a dilation of the surface profile by a three-dimensional structuring element which has the shape of the tip. Accordingly, the restoration procedure using these concepts consists in performing the erosion of the image by the same structuring element. A preliminary investigation of a possible blind restoration procedure (i.e. restoration without a very precise knowledge of the tip shape and without the need of using a known test object) is performed.

1. Introduction.

Scanning Tunneling Microscopes (STM) and Atomic Force Microscopes (AFM) are able to provide images at very high resolution, in the subnanometer range. This is made possible by the direct interaction of a probe and of the specimen surface, without the need of any intermediates (lenses...). This direct interaction is also at the origin of the present time limitations of such microscopes: the shape, finite size and imperfections of the tip are responsible for imperfections of images.

Images can be considered as a function of the tip and of the specimen surface. Improving images necessitates a good knowledge of this imaging function and methods for performing the reverse transformation.

In the case of highly corrugated surfaces, it is clear that the tip influences the resolution and quality of images. Very often, this tip-specimen interaction which takes place during the imaging process is called a convolution. However, this term has a very precise meaning, which is described in many books concerning image formation. When an image-forming system is linear, the image which is produced can be described as a convolution between the object function and the impulse response of the system (the instrumental function).

In this paper, we would like to show that in STM and AFM, for resolution in the intermediate range (1 – 100 nm), the imaging process is not linear, and thus, speaking of convolution is not correct.

This is not a question of formal language: if the imaging process was linear and could be described by a convolution operation, it would be possible to restore a better image quality (i.e. with a higher resolution) by performing one of the diverse deconvolution methods available. If, on the contrary, the imaging process is not linear and cannot be described by a convolution product, these deconvolution methods are not useful and another approach has to be found for performing image restoration.

The fact that STM and AFM images of corrugated specimens in the 1 – 100 nm resolution range are not produced according to a linear process but to a non-linear process was previously described by a few authors [1-9]. Despite the fact that they use the term “convolution”, Reiss *et al.* [1] describe the imaging process as a non-linear process and show some implications of this, concerning the need to be very careful in the interpretation of images of corrugated specimens. These authors consider that “the tip and the surface show parallel tangential planes at the point of tunnelling contact”. Therefore, their algorithm consists of determining the point of contact (where the local tangent to the tip is parallel to the local tangent to the specimen surface) and in deducing the height of the tip when it is in contact with the specimen. They also suggest and test one method for image restoration [2].

In the works of Butt *et al.* [4] and of Putman *et al.* [7], a qualitative description of the same principle for image contrast is given in the case of rod-like structures, sine waves and spherical particles.

Keller *et al.* [5, 8] describe the image formation and restoration in terms of an envelope function. We have shown [6] that the mathematical background of this approach is in fact identical with the well known operations of mathematical morphology used in the field of digital image processing: dilation and erosion.

In this paper, the mathematics of these dual operations and their potentialities for restoring STM and AFM images are first recapitulated. Then, appropriate simulations and tests on a real image of a grating are presented. Finally, a potential development of this approach is discussed, i.e. a blind restoration procedure that does not need either the use of a known test object or a very precise knowledge of the tip shape.

2. Image formation and restoration.

For understanding the following, let us first recall how to express mathematically the image formation process in the case of linear interactions. The experimental image function $i(x, y)$ can then be described as a convolution between the object function $o(x, y)$, (which in case of corrugated surface is the local height of the specimen surface) and an instrumental function $h(x, y)$:

$$i(x, y) = o(x, y) * h(x, y) \quad (1)$$

The instrumental function (impulse response) in the case of near-field microscopies would be inversely proportional to the height of the tip surface point (x', y') relative to the front of the tip (at x, y); (h is generally a monotonically decreasing function of x and y).

According to this description, many points of the tip would participate in the formation of the image of a given specimen detail situated at coordinates (x', y') :

$$i(x, y) = \sum_{x'} \sum_{y'} o(x', y') \cdot h(x - x', y - y') \quad (2)$$

In reality, in the case of highly corrugated surfaces, the characteristic signal results from a localized contact point on the tip sides approaching adjacent regions of the specimen (see Fig. 1).

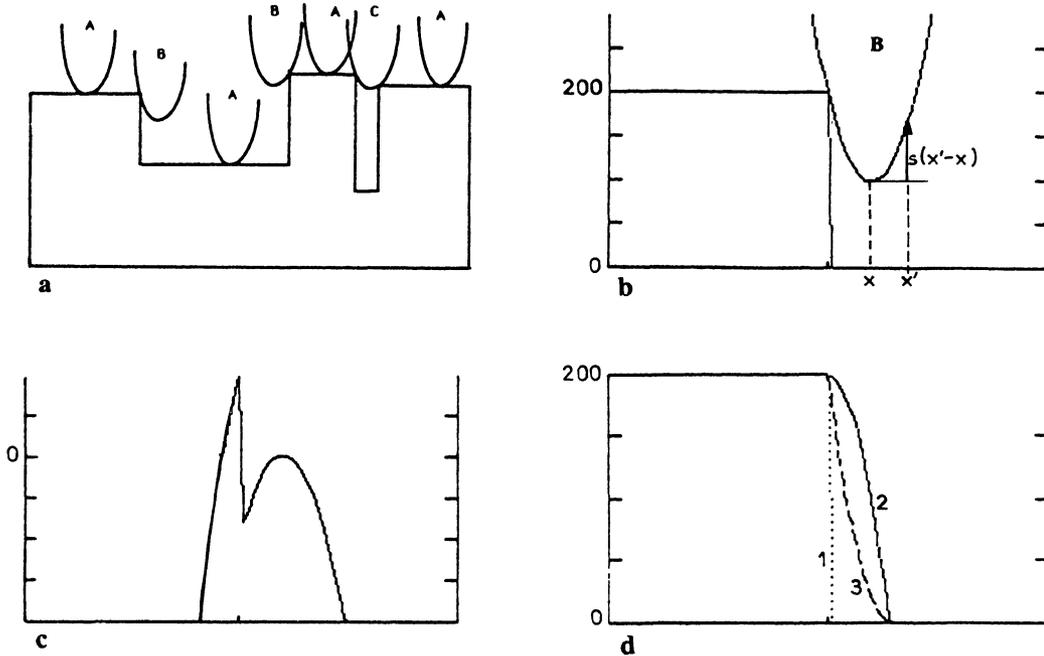


Fig. 1. — a) Schematization of the image formation process in near-field microscopy for resolution above one nanometer. The corrugated surface is represented by its height profile. The probing tip “tries” to follow this profile and the output image forming signal is assumed to be proportional to the height of a reference point on the tip. In situation A, the tip can stay in contact with the surface and the output signal gives precise information concerning the height of the surface regions. In situation B, the lowest part of the tip cannot be in contact with the corresponding point of the specimen, because another point of contact is reached. In these situations, the images of holes have a smaller width than the holes themselves. In situation C, the tip can never be in contact with the bottom of narrow holes, for the same reason. Therefore, not only the width, but also the height of these holes are not correctly imaged. b) A detailed part of figure 1a, illustrating the notations used in this paper. c) shape of $h(x, y, x', y') = o(x, y) - s(x' - x, y' - y)$, for the configuration displayed in figure 1b. d) the surface profile of figure 1b (dotted curve: 1), the computed image profile (full curve: 2) and the partially restored image profile (dashed curve: 3).

These considerations are not very new and are well-known by STM and AFM users who know that images have to be interpreted with care (dips can be deeper than apparent on images). In this case, the interactions cannot be considered as being linear and the convolution approach cannot be appropriate to describe the image formation process.

According to the qualitative description given above, the useful signal for a tip in position (x, y) is given by the height of the tip corresponding almost to a contact between tip and specimen (Fig. 1b). This can be described by:

$$i(x, y) = \max_R [h(x, y, x', y')] \tag{3}$$

where $h(x, y, x', y') = o(x', y') - s(x' - x, y' - y)$, s describes the three-dimensional shape of the tip, R is the region of the specimen, around (x, y) which can be in contact with the tip.

An illustration of the function h is given in figure 1c, for the one-dimensional geometry shown in figure 1b. The resulting image profile is given in figure 1d for a step edge.

When the shape of the tip is known, this algorithm can be applied very easily and an image can be simulated. In fact, this kind of non-linear operation is well known in the field of image processing where it is called "dilation". It belongs to the class of operations known as mathematical morphology (Coster and Chermant, [10]). It can be applied to grey level images as well as to binary images. It is a non-linear operation which consists of computing the maximum value of the original image in a given neighborhood around the pixel considered. This neighborhood is called a structuring element and can have different shapes (bipoint, segment, disc...) according to the aim of the morphological operation.

In the present situation, the structuring element is the tip itself. It is a three-dimensional structuring element and the dilation (Eq. (3)) is performed over the specimen surface represented by its height variations. The result is the image function $i(x, y)$.

The formalisation of image formation given above leads us now to the problem of image restoration. In mathematical morphology, the dual operation of dilation is erosion. It consists in computing the minimal value of the original (binary or grey value) image over a neighborhood defined by a structuring element. This results in the erosion of an object, the elimination of protuberances or bright peaks of size smaller than that of the structuring element. (Note that the erosion and dilation operations are not inverses: erosion · dilation is not equal to one).

In the present case, the restoration procedure according to an erosion by a three-dimensional structuring element can be formalized as:

$$\hat{o}(x, y) = \min_R [h'(x, y, x', y')] \quad (4)$$

where \hat{o} stands for the estimated (restored) object function, $h'(x, y, x', y') = i(x', y') + s(x - x', y - y')$.

This operation can be easily implemented when the tip shape $s(x, y)$ is known.

Of course, at the experimental level, the difficulty to restore an image is to know this exact shape of the tip. We will later propose a method of circumventing this problem. But let us first illustrate the previous considerations with some examples of simulation and restoration.

As an object model, we have chosen to take a two-dimensional grating whose periodicity is 100 nm. These kinds of objects are really produced as tests objects for near-field microscopes. They are made by electron microlithography. If the fabrication process was perfect, the grating obtained would consist of square studs ($50 \times 50 \text{ nm}^2$) separated by 50 nm and have a height of approximately 30 nm. In fact, a scanning electron micrograph of this object reveals that the studs are approximately hemispherical. Figure 2a shows the object model used for the simulation: the half-period is equal to the diameter of the hemispherical studs: 20 pixels. Figure 2b is the result of a simulation according to a convolution process. The impulse response was built with a value inversely proportional to the distance to the apex of the tip, with a value unity at the apex. Figure 2c displays the image obtained by AFM with a nanotip (Digital Instruments). The characteristics given by the manufacturer are such that the shape looks like a paraboloid with a radius of approximately 10 nm. The obtained AFM image presents elliptical studs revealing either an imperfect tip (elliptical instead of circular section) or a tip with its axis tilted with respect to the surface normal. Figure 2d is the result of applying the dilation process with the tip shape as the structuring element. The tip shape is parabolic:

$$s(x, y) = a \cdot i^2 + b \cdot j^2,$$

where (i, j) are the coordinates of a tip surface point (x, y) after rotation by an angle ϕ .

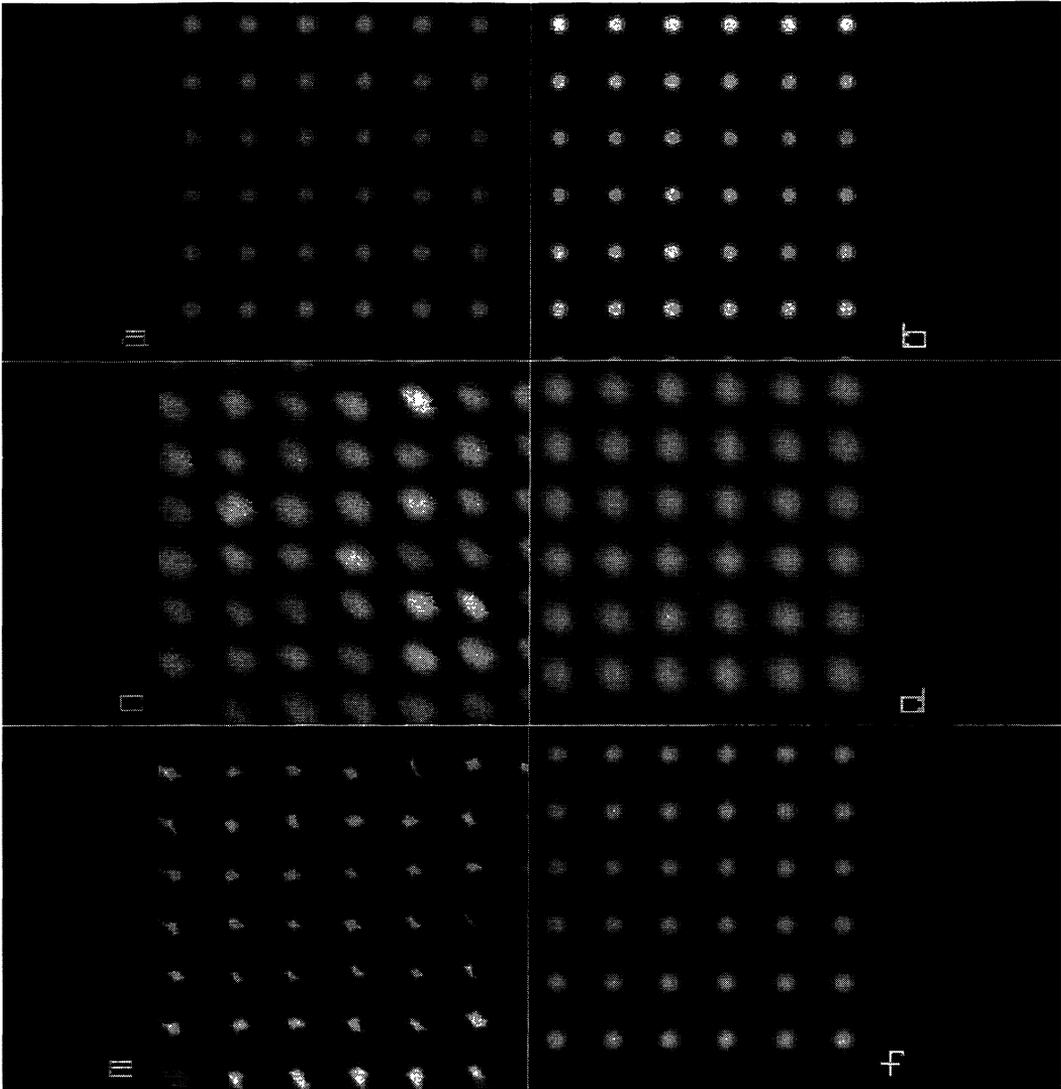


Fig. 2. — Simulation of image formation in near-field microscopy. a) Simulated object (height is coded as grey levels: white hemispheres represent studs); b) image simulated according to a convolution procedure. The instrumental function (impulse response) is the inverse of the height of a parabolic tip surface; c) an experimental image of such an object, recorded with an A.F.M. and a nanotip (theoretical radius $r = 10$ nm); d) image simulated by a dilation procedure (parabolic tip); e) restoration of the experimental image (c) by an erosion procedure; f) restoration of the simulated image (d) by an erosion procedure.

The parameters ($a = 0.08$; $b = 0.2$; $\phi = 50^\circ$) were chosen in order to get a simulated image close to the experimental one (Fig. 2c).

One can see directly that the effect of a dilation is much more drastic than the effect of a convolution, which does not strongly affect the size of the studs, but results mainly in a loss of fine details (smoothing). The results obtained by the dilation process (Fig. 2d) are much closer to the images obtained experimentally (Fig. 2c). This comparison of the two processing methods

with the experimental image shows without any doubt that the image formation is certainly not a linear process and thus not governed by a convolution.

Figure 2f shows the result of applying an erosion process (with the same structuring element as described above) to the dilated image (2d). One can see that the restored image is very close to the original object function (2a). This is an indication that useful restoration results can be obtained by this procedure.

Figure 2e is the result of applying the same erosion procedure (with the same structuring element) to the experimental image (2c). The results are not as good as with simulations, may be because some spurious effects (like the misorientation of the tip axis or saturation effects) are not taken into account. However, the improvement is still noticeable.

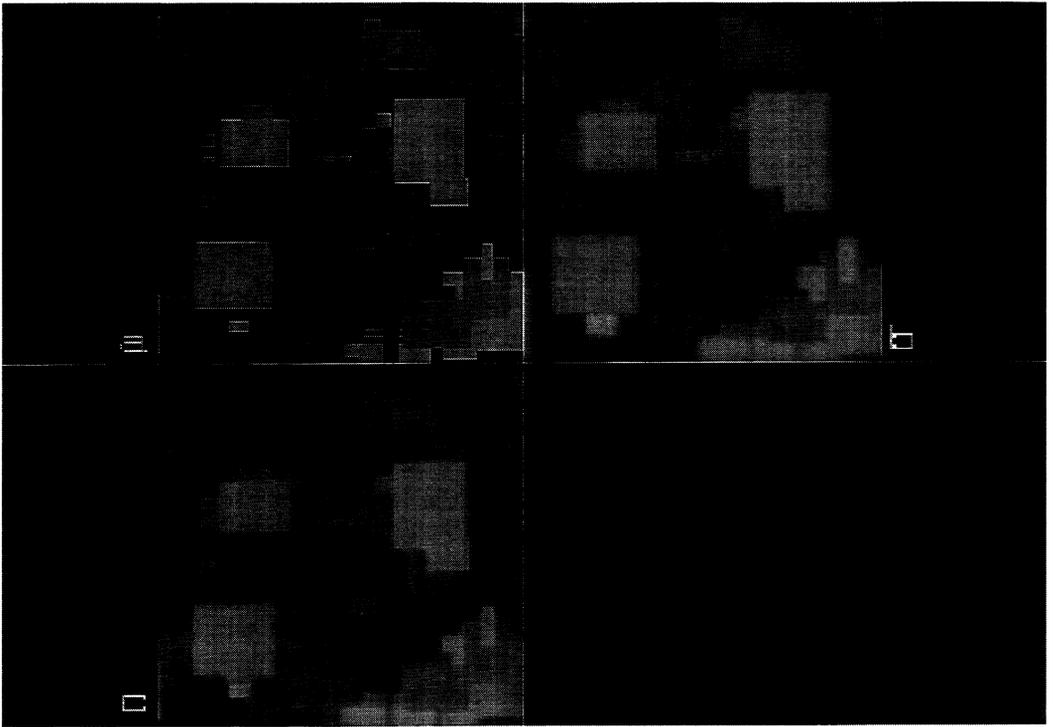


Fig. 3. — Simulation of image restoration in near-field microscopy of a highly corrugated surface. a) Simulated object (brighter structures represent higher parts of an object made of randomly dispersed rectangles of different sizes and heights); b) simulated image obtained by the dilation procedure with a tip of paraboloidal shape ($a = b = 1.5$, $\phi = 0$), c) restored object function obtained by an erosion procedure with the same structuring element.

Another example concerns the simulation of rectangular objects of different sizes and spacings (chosen at random). The “object” is displayed in figure 3a. The image simulated by the dilation procedure (parabolic shape with $a = b = 1.5$, $\phi = 0$) is shown in figure 3b while the restored image obtained by the erosion procedure (with the same structuring element) is shown in figure 3c. One can see that a strong improvement of the image characteristics and quality has been obtained. Edges are sharper. The size of the top-squares (i.e. which have a height greater than their

surroundings) is much closer to the true one than the one they had in the simulated image. However, one cannot expect miracles from this restoration method: the bottom of deep and narrow holes is not restored. This can be easily understood since the information concerning these parts of the objects was definitely lost during the non-linear imaging process (they were not “seen” by the tip which was not able to go down within these small holes; see Fig. 1).

3. Possible extension of the approach towards blind restoration.

Besides the fact that the above procedure is easy to implement, it can be hoped that these concepts will allow the limitations attained in the restoration procedure to be overcome. In this section, we would like to discuss such a possible extension, as a preliminary prospect.

One question that we have not discussed so far is: how can we estimate the tip shape function which is necessary for simulating an image from an object model as well as for restoring an object function from an experimental image?

Two answers to this question come immediately to mind:

- the first one is to observe the tip shape by electron microscopy, either scanning electron microscopy or better, transmission electron microscopy to get a high resolution image of the tip extremity. From this observation, the geometrical parameters of the tip (curvature radius, cone angle...) can be deduced. Let us point out that in order to get a complete knowledge of the three-dimensional tip shape, one needs to use a goniometer stage (keeping in mind the limitation that goniometer stages for electron microscopes offer, at best, the possibility of $\pm 60^\circ$ tilt).
- the second solution is to use a test object and to observe it with the tip that was used to record the image one wants to restore. Provided the test object is very accurately known, the near-field microscopy image of this test object can be compared to the “true” object function and the tip shape can be deduced from this comparison. This procedure has been performed by Reiss *et al.* [1] with an object whose profile has been precisely determined *a posteriori* by observing its cross-section in TEM. This very laborious procedure is still not satisfying because it only gives access to the tip profile in one direction. To get the 3D geometry, a 2D test object must be used and actually, the difficulty consists of fabricating 2D gratings presenting very sharp edges with heights known precisely.

Taking into account all the experimental difficulties related to the above solution, let us see whether there is a third possible solution that does not involve a well-known test object and needs only minimum knowledge on the tip shape.

It is clear that the experimental image contains not only information concerning the object but also information concerning the tip. The problem is to know how to split the information contained in the experimental image into information concerning the specimen and information related to the imaging instrument. This is a key point which has received full attention but which is only partially solved, even when a linear image formation process can be assumed (the terminology used is then: blind deconvolution [11, 12]).

This key point has received even fewer solutions when one is concerned with a non-linear imaging process, and we do not pretend to solve it completely within the framework of this paper. However, we believe that the dilation/erosion formalism (which is called a closing procedure in the terminology of mathematical morphology) described in the previous sections can help in finding ways to obtain some information concerning the tip from the image itself. These approaches rely on the concept of idempotence, which some operations of mathematical morphology process: for these, the second application of the operator does not change the result obtained after the first application (Coster and Chermant [10]).

Let us suppose, in order to simplify the explanation, that we have a "spherical" tip, the radius of which (r) is unknown. With this tip, we have obtained an image. Then we try to restore the object function by assuming that the tip radius is r_1 . The result of this restoration is an estimated object function \hat{o}_1 . Now, we perform a dilation of the restored object with the same structuring element of radius r_1 . This gives an estimated image \hat{i}_1 . Let us now compare the experimental image i with the computed image \hat{i}_1 . Two situations have to be considered: either $r_1 < r$ or $r_1 > r$ and the corresponding results are schematized in figure 4.

If $r_1 < r$, the erosion/dilation procedure (which is called an opening procedure in the terminology of mathematical morphology) restores the original experimental image ($\hat{i}_1 = i$) (Fig. 4a). If $r_1 > r$, the opening procedure does not restore the experimental image everywhere because too many structures are irretrievably destroyed at the erosion stage and cannot be recovered at the dilation stage (Fig. 4b).

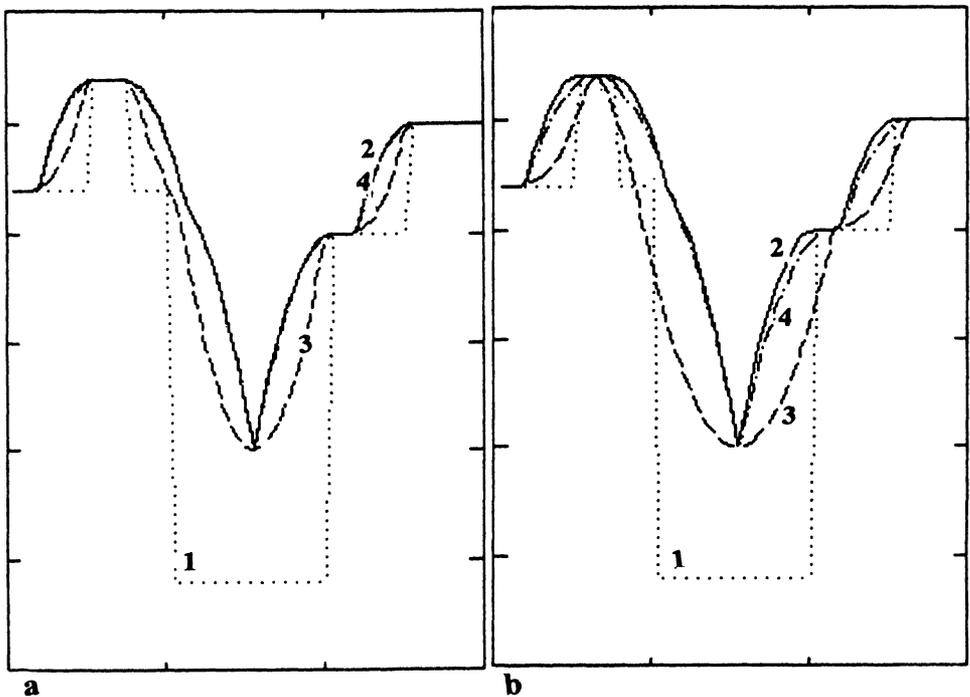


Fig. 4. — Schematization of the opening procedure with a structuring element of variable size. 1: object profile. 2: image profile. 3: eroded image profile. 4: open image profile. a) The structuring element has a size smaller than the "true" structuring element which produced the experimental image. In this case, the open image is identical to the original image; b) the structuring element has a size larger than the "true" structuring element. In this case, the narrowest structures of the image are destroyed during the erosion step and cannot be restored during the dilation step: the distance between the open image and the original one increases as the expected structuring element deviates from the true one.

The difference between the experimental image i and the open image \hat{i} can be quantified by a "distance" function:

$$d = \sum_x \sum_y |i(x, y) - \hat{i}(x, y)| \quad (5)$$

Thus, trying several restoration/degradation (i.e. erosion/dilation) procedures with several values of the tip radius, and evaluating the “distance” between images i and \hat{i} , allows an upper limit for the effective tip radius to be fixed.

A simulation of this procedure is given in figure 5. The simulated image of figure 5a was obtained from the “object” in figure 3a and from a 3D structuring element composed of a rod-like (cylindrical) tip terminated by a hemispherical extremity. The radius of the rod and of the extremity was set to 4 pixels. The simulated image is then eroded/dilated with a structuring element of the same shape, with a radius varying from 1 to 8 pixels. The computed images \hat{i} are displayed in figures 5b-d for radii of 3, 4 and 6 pixels respectively. The distances to the simulated experimental image are displayed in figure 6a as a function of the expected tip radius. One can see that an increase of the distance function occurs near the true tip radius ($r = 4$ pixels).

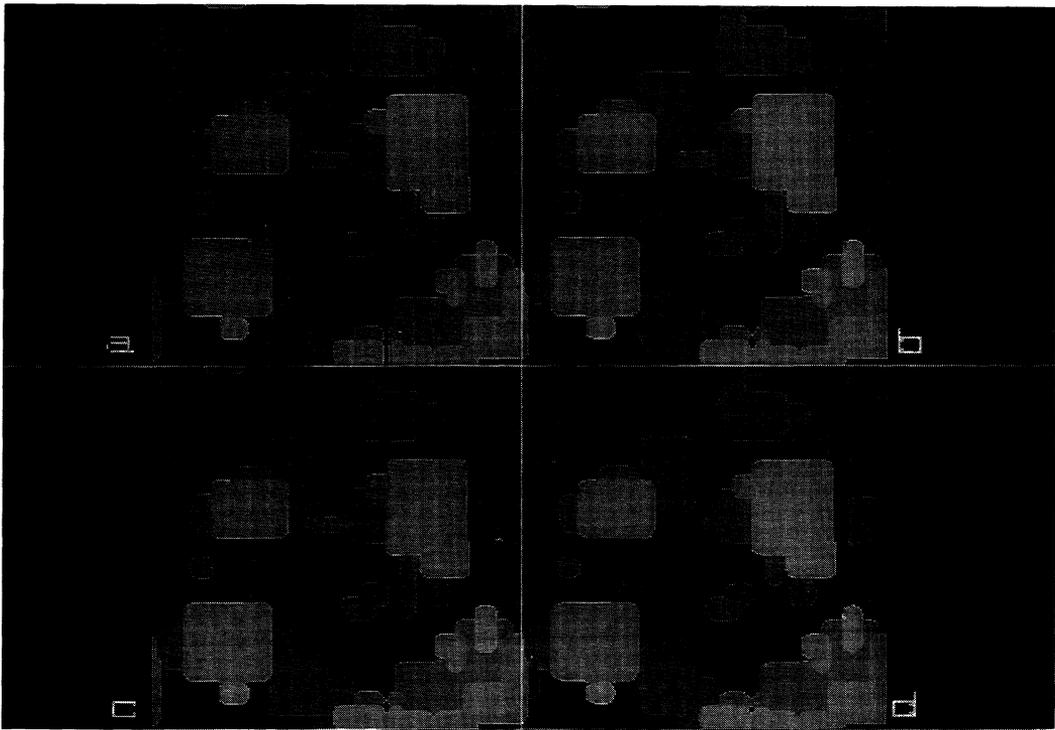


Fig. 5. — Illustration of the opening procedure schematized in figure 4. a) Simulated image computed from the object in figure 3a and a “spherical” tip ($r = 4$ pixels). Note that for such a highly corrugated surface, it is the lateral extension of the tip which governs image formation, rather than the shape of the extremity. Therefore, despite the fact that the radius of curvature at the tip extremity is greater than that of the parabolic tip, the effect of the “spherical” tip (which behaves like a rod-like tip) is less important -see figure 3- due to a smaller lateral extension. b, c, d) Open images obtained from a) and a spherical tip with $r = 3, 4$ and 6 pixels respectively. While figures b) and c) are close to the original image (a), figure d) is deteriorated, indicating that the assumed value of $r = 6$ pixels is higher than the true value ($r = 4$).

Of course, realistic models of the tip have to be more sophisticated and must include several parameters. The search procedure has to be implemented in a multidimensional space. An example is given in figure 6b where a paraboloid model was assumed. Figure 6b displays the distance

function in the (a, b) space ($\phi = 0$ was assumed in order to stay in a two-dimensional space). A decrease of the distance function is observed near the “true” value ($a = b = 1.5$).

Though the real experiments will lead to more complicated situations, we consider that the results obtained with the simple models described here are encouraging and that the investigation has to be pursued in this way.

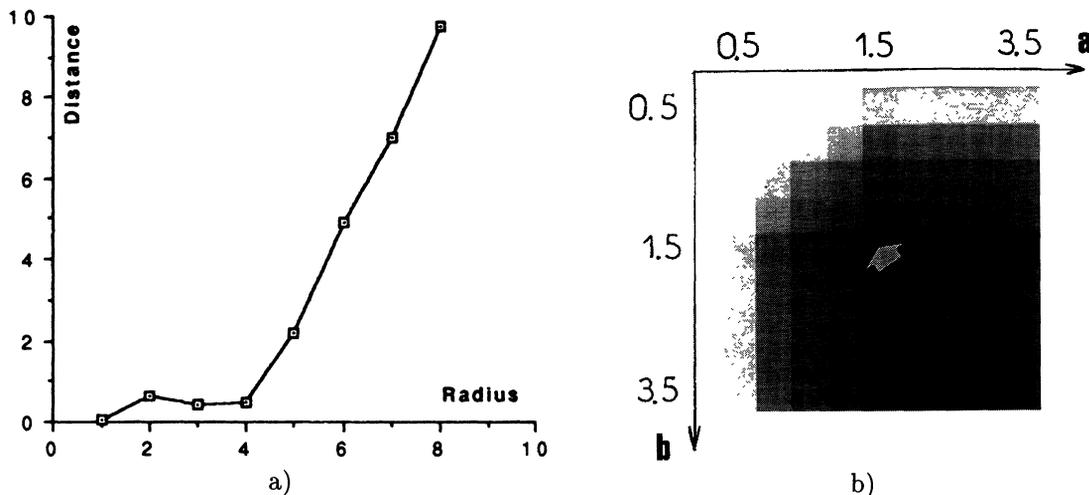


Fig. 6. — a) Plot of the “distance” (per pixel) between open images and the original image (Fig. 5a), as a function of the structuring element size (expected tip radius). b) 2D representation of the “distance” for a paraboloid tip, (with two parameters a and b ; ϕ is assumed to be zero here, but could also be estimated) as a function of parameters a (horizontal axis) and b (vertical axis). A step of the distance function is obtained for $a = b = 1.5$, which were the tip parameters used for simulating the “experimental” image (Fig. 3b).

4. Conclusion.

The tip effects in STM and AFM strongly affect the quality of the images. In the range of resolutions larger than one nanometre (corrugated surfaces), these effects are essentially non-linear and cannot be described by a convolution in the usual sense of this word. At first approximation, they can be described by a non-linear operation well-known in the field of image processing (mathematical morphology): the dilation by a structuring element.

The dual operation (also originating from the field of mathematical morphology) is the erosion. It provides a simple implementation for the process of image restoration. Of course, the restoration procedure is also non-linear. Therefore, there is no chance of restoring the object perfectly, even in the absence of noise, except if some *a priori* knowledge concerning the object can be incorporated.

Simulated images compared to an experimental image clearly demonstrate this non-linearity and that the erosion process allows to restore an image.

The dilation/erosion concepts open up the possibility of image restoration without *a priori* knowledge concerning the object one is looking at, and without the need to use an intermediate test object. The principle consists of using the property of idempotence for estimating parameters which characterize the tip size. An example of blind restoration on a simulated image is given.

This approach by non-linear processes may open up the way to efficient restoration procedures which anyway do not suppress, but attenuate some tip-related effects observed in STM and AFM.

Acknowledgements.

We would like to thank Mr P. Guerin from the CNET for providing the periodic test specimen.

References

- [1] Reiss G., Vancea J., Wittmann H., Zweck J. and Hoffmann H., *J. Appl. Phys.* **67** (1990a) 1156.
- [2] Reiss G., Schneider F., Vancea J. and Hoffmann H., *Appl. Phys. Lett.* **57** (1990b) 867.
- [3] Blackford B.L., Jericho M.H. and Mulhern P.J., *Scanning Microsc.* **5** (1991) 907.
- [4] Butt H.J., Guckenberger R. and Rabe J.P., *Ultramicroscopy* **25** (1992) 375.
- [5] Keller D.J., Franke F.S. and Chang L., Proc. 2nd International Reflection Workshop on "STM-AFM and Standard Biological Objects" Royaumont (1992) 70.
- [6] Bonnet N., Oral communication, Ultimatech (CNRS) meeting, Gif-sur-Yvette (France), 20 January (1993).
- [7] Putman C., de Grooth B.G., Hansma P.K., Van Hulst N.F. and Greve J., *Ultramicroscopy* **48** (1993) 177.
- [8] Keller D.J. and Franke F.S., *Surf. Sci.* **294** (1993) 409.
- [9] Pingali G.S., Jain R. and Kong L.C., Abstracts Int. Conf. STM, Beijing (1993) 11.
- [10] Coster M. and Chermant J.L., *Précis d'Analyse d'Images* (Editions du CNRS, 1985).
- [11] Stark H., *Image recovery: theory and applications* (Academic Press, 1987).
- [12] Katsaggelos A.K., *Digital image restoration* (Springer Verlag, 1991).