American Mineralogist, Volume 87, pages 5–16, 2002

Image distortion in scanning probe microscopy

K. HENRIKSEN AND S.L.S. STIPP*

Interface Geochemistry Laboratory, Geological Institute, Copenhagen University, Øster Voldgade 10, DK-1350 Copenhagen K, Denmark

ABSTRACT

Scanning probe microscopy (SPM) has become a common tool in mineralogy but distortion of images complicates interpretation and often limits the amount of information one can extract. Image distortion arises from a discrepancy between the intended and actual scan area caused by relative movement between tip and sample that is additional the intended scanning motion. We present a mathematical model to describe distortion in SPM images and provide a simple algebraic correction method. It uses Fourier periodicities for correcting high-resolution images; for micrometer-scale images, it can use any three non-colinear points that define a feature with known geometry. Observed distortion can be accounted for by two components: *drift*, the vector that quantifies the shape change of the intended scan area from a square into a parallelogram, and *scaling*, a constant that describes an isotropic change in dimension of the resulting scan area. The correction restores angular relationships and distances. The method was tested on the mineral graphite. In order to define the most important parameters affecting distortion, we made a sensitivity analysis by systematically varying temperature, scan speed, and time lapsed after the microscope was powered on. Neither drift nor scaling were found to be temperature dependent as such. However, both do depend on the time lapse after imaging begins. After the instrument is powered on, an initial 40 minute period of erratic drift is observed, whereafter drift velocity decreases with time while scaling increases slightly. Temperature variations in the range of 23 to 43 °C have negligible influence on distortion whereas scan speed affects scaling.

INTRODUCTION

Scanning probe microscopy (SPM) refers to a relatively new family of techniques that was born with the invention of scanning tunneling microscopy (STM; Binnig et al. 1982a, 1982b). It offers the possibility of observing surfaces under ambient conditions, essentially at local scale, where resolution is in the range of angstroms (10^{-10} m). This is several orders of magnitude smaller than with conventional microscopes. SPM offers tremendous possibilities for in situ studies of natural surfaces under experimental temperatures, pressures, and fluid compositions that approximate nature better than the vacuum conditions and sample preparations required by other high-resolution methods. From the time of publication of the first article using SPM on mineral surfaces (Hochella et al. 1989) there has been a steady increase in its application to improve fundamental understanding of surface structure and the molecular-level processes that control dissolution, precipitation, adsorption, and mineral transformation. As with all methods, data are more meaningful and interpretations more sound when the limitations of the technique and possible artifacts are known and considered in experimental planning. This has not always been the case with SPM research. Although Eggleston (1994) provides an introduction for the application of SPM to minerals, presents limitations of the technique, and describes many artifacts and how to deal with them, a simple but theoretically sound method for correcting distorted SPM images is lacking and would be beneficial to mineralogists.

Image distortion results from a discrepancy between the intended and the actual scan area caused by a relative movement between the tip and the sample that is additional to what is intended during sample scanning. The word "drift" is inextricably bound to definitions of image distortion in SPM because distortion seems to arise from the surface "drifting by" the tip when scanning. Distorted images still yield abundant information, so in studies where morphology or changes in morphology are the main focus, distortion causes no serious problems. One simply lives with it and sees around it. However, distortion destroys the angular relationships between features, invalidating angle measurements and introducing significant error into most distances. This is particularly unfortunate for mineralogists because ever since the first goniometer was applied to intersecting faces, we have relied on angular relationships to define symmetry and structural identity.

Distortion is a common feature of SPM imaging. Use of a fluid cell often aggravates it. In a few instrument designs, me-

E-mail: stipp@geo.geol.ku.dk

chanical modifications and software allow distortion to be corrected either during scanning or afterward. However, on the research instruments used in most mineralogy and geochemistry laboratories, there is often no easy way to add such modifications and the software package is a closed system from which it is impossible to extract numerical data to submit to polynomial correction. In the physics and materials science literature, there are reports of correction procedures for SPM images (Andersen and Møller 1994; Carrara et al. 1994; Garnæs et al. 1998; Jørgensen et al. 1998) but they are relatively involved mathematically, and most require that image data be in electronic form that can be operated on mathematically. Evidence that our community is not aware of the need for drift correction is a complete absence of citation of these methods in mineralogy articles using SPM while at the same time, there are many examples of published and unpublished (posters, talks, etc.) studies that interpret angular relationships from distorted images and make untenable conclusions.

Therefore, our aims in this work were (1) to develop a simple correction method that can be applied to SPM images, to allow extraction of valid angular relationships and distances, thus increasing the usefulness of SPM and (2) to present this method in an accessible and easy-to-apply form. In this article, we briefly describe how SPM works in order to provide a framework for the discussion of distortion, we present the theoretical mathematical basis for a correction model, then we apply the model to data from a mineral surface: some distorted images of graphite. In order to improve understanding of the nature of distortion, we report on a sensitivity analysis made by systematically varying some of the major parameters thought to cause it.

How SPM works

All scanning probe microscopes are based on the interaction of a tip with the sample surface. The techniques differ in the choice of probe and which physical parameter is measured as the tip is rastered over the surface. One member of the SPM family that is particularly useful to mineralogists is atomic force microscopy (AFM) because it can be used on all flat surfaces with no requirements for conductivity. An AFM image is constructed by collecting data from multiple parallel scan lines, either by having the tip move over the sample or the sample move beneath the tip. The latter is the design for the microscope used for this study, but in discussions here, the tip will be described as moving over the sample, because this is easiest to visualize. For the discussions about distortion in this article, microscope design is irrelevant. The scan line direction is denoted as the *x*-direction and each line is scanned twice, forth and back, whereafter the tip moves an increment perpendicular to *x* and a new line is scanned. Thus *x* is known as the fast-scan direction, and *y,* the slow-scan direction.

The high-resolution capability of SPM results from the possibility of making and accurately controlling very minute movements of the tip for which the key is the use of piezoelectric material, Pb-Zr-Ti ceramic. It is semiconducting and characterized by its ability to bend slightly in response to voltage. Movement of the tip (or the sample) can be controlled to only fractions of a nanometer by small applied voltages. Return to the original voltage results in relaxation and movement back to the original position. Unfortunately, relative movements between tip and sample are not always precisely as expected because of hysteresis in the piezo-ceramic's response to voltage and thermal expansion factors. Regardless of its cause, any difference between the intended and actual size and shape of the scanned surface area results in a distorted image.

All SPM images are susceptible to distortion; we chose to investigate the behavior and to develop the distortion correction method with AFM images for two reasons: (1) it works on insulating materials meaning it is more versatile for minerals, and (2) it uses a tip attached to a cantilever that, because of its length and often bi- or multi-metallic construction, makes it more susceptible to inhomogeneous expansion during temperature change, the most common explanation given in the literature for "drift." The method we present works equally well for STM and all other SPM methods.

Mineralogists have long been able to determine bulk mineral structure very precisely from X-ray diffraction (XRD). In spite of the high-resolution advantages, one would never use SPM to determine surface unit-cell dimensions because of the uncertainly associated with distortion. However, the resolving power of SPM is valuable for investigating local structural details and their change in dynamic systems. From XRD data, one can determine expected unit-cell dimensions on the termination of the bulk structure and then use this for distortion correction for investigating such details as structural differences in the mineral surface resulting from fracture or resulting from adsorption of other material.

When a mineral is fractured to create a surface, the disruption of symmetry results in a change in atomic bonding near it, requiring the near surface to rearrange. Surfaces usually relax so that the topmost layers of atoms have slightly shorter distances to the layers below than they would have had in the bulk. If the relaxation is uniform, the surface unit cell appears unchanged. Slight stretching or shrinking of the unit cell itself, in the directions parallel to the surface, is forbidden because this would set up tremendous strain in the surface region. Instead, some atoms relax deeper into the surface region than others. Such a reconfiguration may result in a surface unit cell that appears to have dimensions that are some factor larger than that expected from the termination of the bulk structure. This is expressed as a superstructure, which maintains the integrity of the distances and angles of the original surface unit cell by multiplication, i.e., 2×1, $\sqrt{3} \times \sqrt{3}$, 7×7, etc. Adsorption can result in a commensurate layer, where the attached material fits the surface sites and the unit cell has dimensions as expected for the underlying mineral. Alternatively, adsorption can result in an incommensurate layer, where size (or charge) constraints make a layer that does not fit exactly, which also produces a superstructure. SPM's high-resolution capabilities make it the only tool that can give information about such features at the local scale, in situ. Correction of image distortion makes these studies easier by restoring distances and angles.

The HOPG surface

One of the materials on which atomic resolution is relatively easy to achieve with AFM as well as STM is highly or-

dered pyrolytic graphite (HOPG). The surface of this material has therefore been linked to the ongoing debate about the imaging process and resolution capabilities of AFM. For consistency, we use it to develop the distortion correction model, so definition of its surface structure is important as a base. Graphite consists of layers of carbon atoms arranged in a honeycomb structure of hexagonal rings. Alternating layers are translated so that half the atoms of the upper layer are placed directly above the empty middle of a ring on the layer below; the other half are situated above atoms of the lower layer (Fig. 1).

The very first images ever made with an AFM were of HOPG (Binnig et al. 1987). They show a regular hexagonal pattern rather than the honeycomb structure described above, with the interatomic distance on the images larger than that expected from X-ray diffraction data. Much effort has been devoted to understanding why HOPG images show this deviation from the true surface structure, not least because the answer necessarily must shed light on the AFM imaging process in general, a topic vitally important to all AFM researchers. By developing models that simulate the tip-surface interaction in AFM, theoretical images can be produced and their correspondence to experimental images can be used to evaluate the model. One line of thought considers tip position probability densities on the surface, and has revealed that the tip moves in a "stick-slip" manner from one potential minimum to the next (Hölscher et al. 1999). Some researchers conclude that the image records the hollows in the atomic structure rather than any of the atoms (von Toussaint et al. 1997). Others report that varying the orientation of the model tip apex and the applied force gives simulated images showing either half or all of the atoms of the HOPG surface (Tang et al. 1993). The existence of experimental images showing the true honeycomb atomic arrangement (Gould et al. 1989) lends credence to this hypothesis. The research field of AFM imaging simulation is currently active and the source of the hexagonal atomic pattern commonly seen with HOPG is still a matter of debate. The pattern itself, however, is well known, making graphite a very useful mineral, both as an atomic scale grid for instrument calibration and as a substrate for adsorbing other materials such as colloidal particles where the atomic-scale pattern serves for internal calibration.

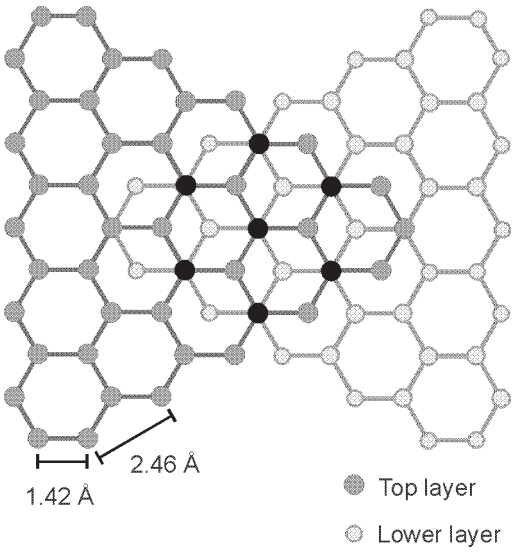

FIGURE 1. The structure of graphite as seen on the basal plane. The mineral consists of layers of C atoms arranged in a honeycomb structure of hexagonal rings. Beneath the empty middle of a ring is an atom of the layer below. SPM images of graphite most commonly show a regular hexagonal pattern (black) corresponding to resolution of only half the atoms of the surface layer. The inter-atomic distance of this pattern is 2.46 Å and the distance between parallel rows is 2.13 Å.

"DRIFT" IN SPM

"Drift" is a word often used in SPM articles and rarely defined. Sometimes referred to as "thermal drift," distortion is thought to be caused either by thermal expansion or contraction of the tip or sample or by non-linearity in the scanning process (Eggleston 1994). There is a tendency to use the word as a broad term encompassing image-distorting phenomena, the causes of which are often a matter of speculation. Yurov and Klimov (1994) define drift as a vector describing a change of the scan area from a square into a parallelogram. Assuming that the velocity of drift changes slowly compared to image acquisition time, they use two images taken sequentially—one scanning up and one scanning down—to find both drift and the slope of the sample on the sample holder. Andersen and Møller (1994) similarly describe drift with a vector that alters the scan area shape and they show how it is capable of changing the atomic structure displayed in the images, so a hexagonal surface pattern can appear as a tetragonal one. Surface distortion and calibration is also the topic of a study by Jørgensen et al. (1994). They consider the linear transformations between the image and the scanning area. They do not separate an actual drift vector, but use a transformation that takes into account change in length of the x axis as well as distortion of the y axis and that is, in essence, a different parameterization of the model presented here.

Aside from distortion corrections that are built directly into instrument hardware or software, at present there are three methods for dealing with image distortion: (1) The first is to do nothing. This is really quite an acceptable solution provided angular relationships are not used for interpretation, and one is willing to accept error in distance measurements. Distortion correction has not been necessary in many studies, such as examination of changes in morphology as a function of time or fluid conditions, or for relating an atomic-scale pattern to data from other techniques. (2) Another way is to use internal calibration. This is most commonly done by using a geometric feature with known proportions to determine a correction factor, which can then be applied to an unknown feature. This works fine for measuring distance along directions parallel to the direction used to derive the internal standard, but it does not work for angular relationships. Still, this method is sufficient for many studies where distance uncertainty better than 10% is necessary, but angles are not necessary. An example is the determination of modulation spacing from an atomic pattern (Henriksen et al., in review). (3) The third option is to use a rather complicated matrix correction method, such as those of Jørgensen et al. (1994) and Garnæs et al. (1998).

In light of "drift" being a much used concept that is relatively easy to visualize, our treatment of image distortion has the purpose of isolating and monitoring a drift vector, rather than imbedding it in a total distortion parameter. This makes it easier to conceptualize what happens during drift and clarifies its meaning and influence on the image. It is a consequence of this approach that a second scan area distorting phenomenon has to be introduced, denoted as scaling. Scaling has the form of an isotropic change of scan area axis lengths. Scaling in the absence of drift is a subtle cause of image distortion, expressing itself as a simple expansion or contraction of the structure shown on the images, without any telltale change in angles. Scaling effects in SPM imaging can be minimized by careful piezo-scanner calibration at the exact instrument conditions intended for imaging. Our experiments (described below) show that scaling effects vary with instrument conditions and time lapse after the instrument was turned on.

The conceptual distinction in this study between drift and scaling is in no way thought to correspond to two different causes of image distortion. Indeed, it is highly unlikely that such a separation is possible. The distinction is made primarily to achieve a clear definition of the word "drift," which can otherwise be a most confusing concept. It also makes it easier to use these separate parameters to un-distort data that is already in picture form. Drift is not a property of an image and arguments could be made to use a distortion parameter that is, for example the eccentricity of the ellipse representing the atomic structure of the image (the image plane ellipse) and the lengths of its axes. However, the status given to "drift" in the SPM community is quite entrenched and the term "thermal drift" is often used interchangeably with it. Our results show that this can be misleading because sample temperature in the range of 23 to 43 °C was observed to have *no* influence on drift whatsoever. Rather, an effect of time is seen, and it should be made clear that "warm-up time," for want of a better term, is used here to describe this effect, which has nothing to do with an actual temperature change of either the sample or the instrument, but rather with the acclimatization of the instrument to the electronics being powered on.

DISTORTION MODEL

Drift

The purpose of this model is to determine drift on the basis of AFM images. Mathematically, the model is based on linear transformations, which is the simplest possible approach that incorporates the image distortion observed. Clarifying what is meant by drift is the first step in recognizing it. The word refers to an illusory movement of the sample, which seems to be "drifting by" beneath the tip while scanning. Such motion is mathematically represented as a vector. Therefore, "drift" must be the vector that quantifies this apparent movement. The effect of drift is to cause a deformation of the area that is scanned on the surface, so that the atomic pattern of the sample, visualized by a circle, is distorted on the image where it can be represented by an ellipse. Finding drift from this ellipse is our purpose. The drift model is evolved on the basis of the following conceptualization of the drift problem.

Distortion means that there is a difference between what the image shows and what the sample truly contains. Therefore, it is useful to conceptualize two planes corresponding to image and sample (Fig. 2). Between the two, clear distinction must at all times be made, otherwise much confusion is possible. The first plane is denoted the real plane. It corresponds to the sample, and therefore contains the true atomic structure of the material studied. In the case of HOPG, we can describe the structure by a circle with radius equal to the distance between parallel rows of atoms that we expect to resolve on the (001) plane, i.e., 0.213 nm (Fig. 1). In addition to this circle, the real plane contains drift, the vector (**d'**) that distorts the scan area square into a parallelogram (Fig. 2). Drift never alters the scan area dimension along x, because this is the fast scan direction and drift velocity would have to be extremely large to have an effect on it. A large number of factors can affect the appearance of the real plane, which in turn defines the appearance of the images. Some of these factors have to do with the instrumental setup that is controlled by the operator, for example, the assigned scan area. Among the real plane defining factors are possible causes of drift, such as non-linearity in the scanner's response to voltage and factors that may influence drift, such as temperature, scan speed, and time, which are investigated in this paper. Although often referred to as the *cause* of image distortion, drift, in itself, is an *effect* of these factors. The real plane is inaccessible to direct scrutiny because the objects and areas of interest are too small, but it is represented on the images, and these form the basis of our knowledge.

Therefore, the second conceptual plane is denoted as the image plane. The image plane always has the dimensions of a square of the intended scan size. Because the actual shape of the area scanned is a parallelogram, the data collected, which corresponds to the true atomic structure of the sample, is deformed when shown on the image. Thereby, the circle representing the atomic structure is turned into an ellipse. This ellipse is a property of the image plane and therefore directly accessible to us. Clearly, the shape of the ellipse must in some way be proportional to the drift velocity on the real plane. Drift in

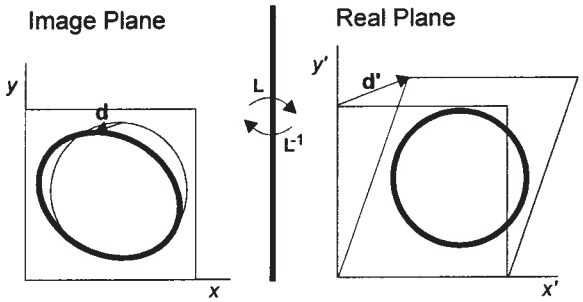

FIGURE 2. Illustration of the drift model. Drift is the vector **d'**. The mapping, **L,** brings the image plane over into the real plane and is drift plus the unit matrix. The inverse mapping, $\mathbf{L}^{-1}$, is the unit matrix plus the vector **d**, (d_x,d_y) defined by Equation 5. If **d** is obtained, **d'** can be determined. By placing a circle, which has a diameter equal to the width of the image plane ellipse, where it is broadest along x so that its lowest point touches the lowest point on the ellipse, **d** can be constructed as the vector from circle top point to ellipse top point.

the *y* direction means that the area scanned is too large along *y*, and when shown on the image, the atomic structure is compressed in this direction. The corresponding ellipse has its shortest axis along *y*, and the larger the drift is, the shorter this axis is. In order to find drift, understanding the actual transformations between real and image plane is necessary. Because we have defined drift as a vector, these transformations are linear.

The simplest transformation needed to map the image plane over into the real plane is represented by a 2×2 matrix, the two columns containing the vectors that are the transforms of the unit vectors, **i** and **j**. If we label coordinates of the image plane *(x,y)* and the corresponding real plane coordinates *(x′,y′)*, the transformation, **L**, that brings the ellipse of the image plane over into the circle of the real plane, must have the form of the unit matrix plus the drift vector **d'**, which is (d'_x, d'_y).

$$\mathbf{L} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & d'_x \\ 0 & d'_y \end{pmatrix} = \begin{pmatrix} 1 & d'_x \\ 0 & 1 + d'_y \end{pmatrix} \tag{1}$$

Note that the *x* axis is undisturbed by **L**. We can now make the important definition: Drift is the vector, **d′**, which quantifies the mapping, **L**, from image plane to real plane. The inverse transformation, which brings the real plane over into the image plane is found by inverting the matrix **L**:

$$\mathbf{L}^{-1} = \frac{1}{1 + d'_y} \begin{pmatrix} 1 + d'_y & -d'_x \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -d'_x / (1 + d'_y) \\ 0 & 1 / (1 + d'_y) \end{pmatrix} \tag{2}$$

Substituting

$$\frac{1}{1 + d'_y} = \frac{1 + d'_y - d'_y}{1 + d'_y} = 1 - \frac{d'_y}{1 + d'_y} \tag{3}$$

into Equation 2, yields

$$\mathbf{L}^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & -d'_x / (1 + d'_y) \\ 0 & -d'_y / (1 + d'_y) \end{pmatrix} = \begin{pmatrix} 1 & d_x \\ 0 & 1 + d_y \end{pmatrix} \tag{4}$$

where

$$d = (d_x, d_y) = \left(\frac{-d'_x}{(1 + d'_y)}, \frac{-d'_y}{(1 + d'_y)} \right) \tag{5}$$

is the vector that quantifies the transformation of the circle of the real plane onto the ellipse of the image plane. This vector is a property of the image plane and is associated with drift, being of opposite direction but different length. It is a measure of the image deformation, which shows greatest contraction in the direction of maximum real plane stretch. When **d** is known, **d'** can be calculated from:

$$(d'_x, d'_y) = \left(\frac{-d_x}{(1 + d_y)}, \frac{-d_y}{(1 + d_y)} \right) \tag{6}$$

by solving the two dependent equations.

We can find **d** graphically from the image plane (see Fig. 2). The image plane ellipse is derived by Fourier transforming the image, which yields the dominant atomic row directions as well as the distance between parallel rows. By this transformation, an atomic scale image of graphite is represented by six points situated symmetrically about the origin. An image taken under the influence of drift yields six points that define an ellipse. To obtain **d** for an up-scan image, the ellipse is first drawn through the Fourier points. Because drift does not affect the *x* axis, the width of the ellipse where it is broadest along *x* must be the diameter of the real plane circle. Next this circle is constructed and placed so that its lowermost point sits over the lowest point on the ellipse. This point is identical in the real plane and the image plane. The vector going from the top point of the circle to the top point of the ellipse is **d**. To obtain **d** for a down-scan image, the circle must be placed so that its top point touches the highest point on the ellipse, and **d** is the vector from the lowermost point on the circle to the lowermost point on the ellipse.

Expansion and contraction

When using the model described above, it became apparent that there was a discrepancy between the diameter of the circle that fitted the *x*-axis width of the ellipse and the diameter expected from the known distances on the basal plane of HOPG (0.426 nm). In other words, the images show a disturbance along the *x*-axis, which is a phenomenon that cannot be accounted for by drift. An additional, isotropic alteration of the unit vector lengths, called "scaling," *s*, is required. The corresponding linear transformation from image plane to real plane is **S**:

$$\mathbf{S} = \begin{pmatrix} s & 0 \\ 0 & s \end{pmatrix} \tag{7}$$

We define *s* as the expected circle diameter divided by observed circle diameter, so that a scaling factor greater that one means that the observed circle is contracted and the area scanned on the real plane was larger than intended. Drift can account for any change in shape of the real plane circle into an ellipse, and scaling can account for any change in size. In the absence of drift, scaling deforms the intended scan area on the real plane into a square of a different side length, which causes an isotropic contraction or expansion of the atomic structure seen on the image plane. This is an important and subtle phenomenon that must be kept in mind when interpreting images. Scaling could be caused by an actual surface contraction or expansion, just as an apparent drift could be caused in the event that the surface was somehow strained. But the results of this study (see experimental section) must lead us to be wary, and conclude that only if the results can be reproduced at different times and scan speeds is interpretation involving surface structure distortion called for. Figure 3 shows how drift and scaling together dictate the deformation of the scan area, which control the appearance of the image.

Calculating drift and scaling

To develop the correction model, we have explained how the drift vector and scaling can be determined graphically. This is particularly useful for correcting drift from micrometer scale

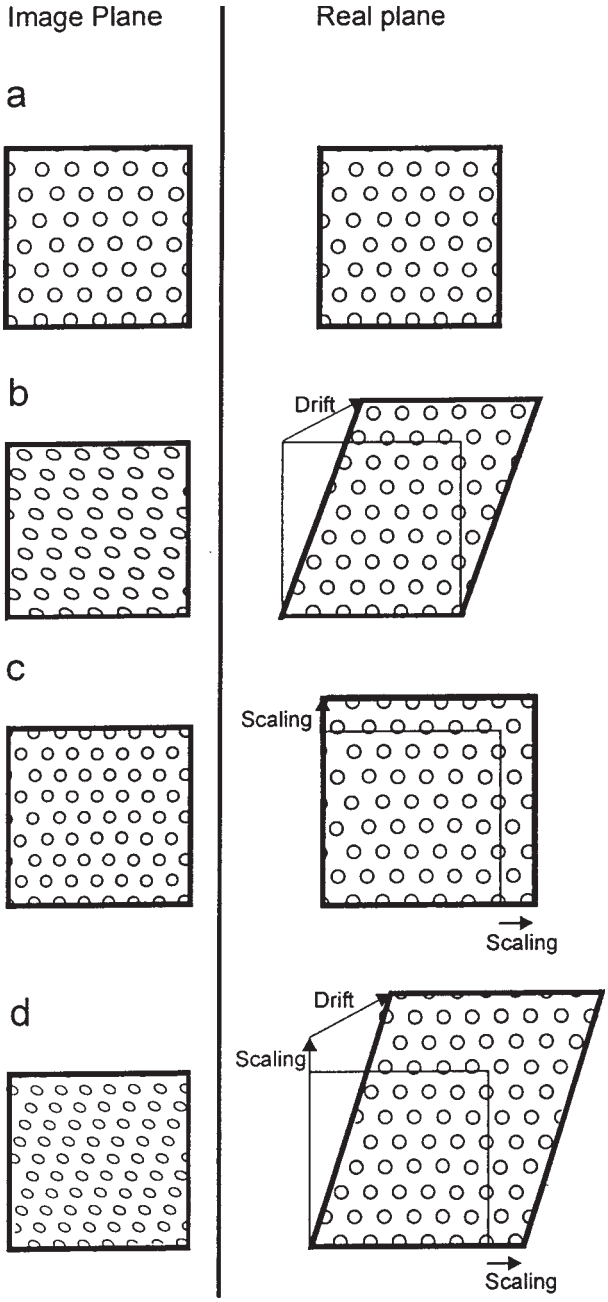

FIGURE 3. Four possible distortion scenarios. (**a**) No distortion. The image plane shows the true atomic structure of the sample. (**b**) Drift only. The image plane atomic structure shows distortion of lengths and angles, but there is no length alteration along *x*. (**c**) Scaling only. The atomic structure on such an image is either compressed or expanded, but there is no change in angles. (**d**) Drift and scaling. The image shows an atomic structure that is distorted in angles and lengths, including lengths parallel to *x*.

images where one or more features on the image have an independently known geometry. Choice of any three non-colinear points can be used to determine distortion. For high-resolution images, a quick and more precise approach is to derive average row spacings directly from Fourier periodicity data and apply an algebraic formulation. The method considers the mapping, **L**, from image to real plane and uses the scaling factor, *s*. The complete transformation from image plane to real plane can then be written as:

$$\mathbf{M} = s\begin{pmatrix} 1 & d'_x \\ 0 & 1+d'_y \end{pmatrix} = s\begin{pmatrix} 1 & q_x \\ 0 & q_y \end{pmatrix} \tag{8}$$

where (q_x,q_y) is simply $(d'_x, 1 + d'_y)$. We have three points on the image plane $a = (a_x,a_y)$, $b = (b_x,b_y)$ and $c = (c_x,c_y)$ found by Fourier transformation and known to sit on the image plane ellipse. The transformation, **M**, maps all points on this ellipse over into the real plane circle (Fig. 4). The presence of the scaling factor, *s,* ensures that this circle has the radius 0.213 dictated by the atomic structure of HOPG. But if we define the radius of this circle, $r = 1$, we can make our calculations for a general case rather than HOPG specifically, and we obtain:

$$k^2[(a_x + q_x a_y)^2 + (q_y a_y)^2] = 1 \tag{9}$$
$$k^2[(b_x + q_x b_y)^2 + (q_y b_y)^2] = 1 \tag{10}$$
$$k^2[(c_x + q_x c_y)^2 + (q_y c_y)^2] = 1 \tag{11}$$

where *k* is a constant given by

$$s = kr \leftrightarrow k = \frac{s}{r} \quad . \tag{12}$$

Using these equations we can isolate q_y^2 as a function of q_x, *a*, *b*, and *c*:

$$q_y^2 = \frac{a_x^2 + 2a_x a_y q_x + a_y^2 q_x^2 - b_x^2 - 2b_x b_y q_x - b_y^2 q_x^2}{b_y^2 - a_y^2} \tag{13}$$

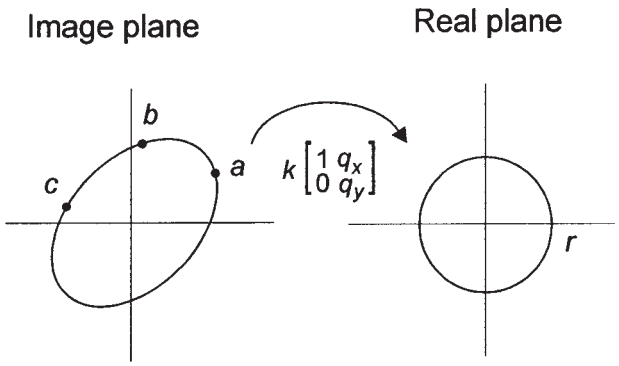

FIGURE 4. A linear transformation relates the three Fourier peak points *a*, *b*, and *c* to the real plane circle. The value of the constant, *k*, depends on the size of *r*. If *r* is 0.213 Å corresponding to HOPG, *k* is *s* (and the mapping is **M**, Eq. 8). Our calculations are made for the general case $r = 1$, giving $k = s/0.213$.

$$q_y^2 = \frac{a_x^2 + 2a_x a_y q_x + a_y^2 q_x^2 - c_x^2 - 2c_x c_y q_x - c_y^2 q_x^2}{c_y^2 - a_y^2} \quad (14)$$

$$q_y^2 = \frac{b_x^2 + 2b_x b_y q_x + b_y^2 q_x^2 - c_x^2 - 2c_x c_y q_x - c_y^2 q_x^2}{c_y^2 - b_y^2} \quad (15)$$

and we can write k^2 as a function of a, b, and c:

$$k^2 = \frac{a_y^2(b_x b_y - c_x c_y) - b_y^2\left(a_x a_y - c_x c_y\right) + c_y^2\left(a_x a_y - b_x b_y\right)}{a_x a_y(b_x^2 c_y^2 - c_x^2 b_y^2) - b_x b_y\left(a_x^2 c_y^2 - c_x^2 a_y^2\right) + c_x c_y\left(a_x^2 b_y^2 - b_x^2 a_y^2\right)} \quad (16)$$

Finally, we can solve for q_x as a function of a, b, and c:

$$q_x = \frac{1}{2}\frac{a_y^2 b_x^2 - a_x^2 b_y^2 - a_y^2 c_x^2 + b_y^2 c_x^2 + a_x^2 c_y^2 - b_x^2 c_y^2}{a_y^2(c_x c_y - b_x b_y) + b_y^2(a_x a_y - c_x c_y) + c_y^2(b_x b_y - a_x a_y)} \quad (17)$$

Method summary

We now have the necessary equations to find drift, $\mathbf{d}' = (d'_x, d'_y)$, and scaling, s, from the points a, b, and c, derived either from Fourier periodicities or three non-colinear points from a feature with known geometry measured graphically from the images.

The steps are:

(1) Determine q_x from Equation 17.

(2) Insert this value into Equation 13, 14, or 15, and solve for q_y using $q_y > 0$.

(3) Determine k from Equation 16, using $k > 0$.

(4) Find scaling, s, from Equation 12. In the case of HOPG, $s = 0.213\ k$

(5) Obtain drift, $\mathbf{d}' = (d'_x, d'_y)$ from $d'_x = q_x$ and $d'_y = q_y - 1$.

Once the drift vector and scaling are known, an image can be precisely reformed using a graphics software package to reestablish the original shape of the surface, so the ellipse on the image plane is transformed back to the expected circle, and true surface structure is shown. This can be done in a graphics software package such as Photoshop or CorelDraw, by transforming the square image into a parallelogram of the size and shape representing the surface that was actually scanned.

(6) Using the software function for homogeneous stretching or shrinking figures, transform the image into a square of side-length equal to the image side-length times the scaling factor, s.

(7) Using the software function for distorting figures (nonhomogeneous stretching), apply both magnitude and direction of the drift vector ($\mathbf{d}'$) to transform the image square into the parallelogram of data that the image truly represents.

Experimental study

The purpose of the experimental investigation was to test the distortion correction method, and to examine the relationship between the two distortion parameters, drift and scaling, to variations in temperature, time, and scan speed. Drift and scaling data were obtained by Fourier transforming atomic resolution images of the basal plane of HOPG and inserting the Fourier peaks into the equations developed above.

Materials and techniques

We used highly ordered pyrolitic graphite (HOPG). Samples were cleaved before each experiment. All experiments were carried out on a Digital Instruments Multimode SPM running in contact AFM mode. In order to minimize building vibration and noise, the instrument was hung on a vibration-dampening platform suspended on rubber cords from the ceiling and covered with hoods of aluminum and foam. The microscope uses tubular piezoelectic elements for scanning the sample. We used the shortest scanner, with maximum x and y offset of about 1 micrometer. Probes were standard commercially available integrated pyramidal tips of Si_3N_4 on a gold-covered cantilever with a spring constant of 0.6 nN. The experiments were conducted in air; we monitored but did not control humidity. Because capillary forces resulting from adsorbed water are nearly impossible to quantify, net force of the tip on the sample was not determined. Forces were minimized by frequently adjusting the set point to values just below that resulting in disengagement. The scan angle was kept at 0° throughout, to avoid its possible influence on drift.

The scanner was calibrated against HOPG before any of the experiments began by the following method: Two hours after the instrument was powered on, images were taken using the instrument conditions (tip spring constant, force, scan angle, image size, scan rate, gains, etc.) intended for standard experiments. After reaching stable running conditions, images collected with minimal distortion could be used for the calibration routine available in the instrument control software package.

Distortion in images is most frequently explained as "thermal drift." Thus we attempted to investigate the dependence of drift and scaling on temperature fluctuations from 23 to 43 °C. The temperature of the instrument changes after the power is turned on. It is important to conceptually recognize the difference between the temperature change and the stabilization of the electronics that occurs during the same time interval. Thus the term "warm-up time" should be used with caution to avoid confusion with actual temperature change. We heated and cooled the microscope by installing and removing the hoods. During scanning with the DI Multimode AFM, heat is introduced to the system mainly from the electronics controlling the piezo-scanner in the base, and from the laser and detector in the head. As a midpoint compromise for temperature sensor position, we attached a thermocouple to the sample holder with a very flexible wire that passed through a mouse-hole on the sound-shielding metal cover so that we could monitor temperature while scanning. Separate sensors outside detected room temperature and humidity. All were connected through the software Lab-View to record continuously.

Temperature vs. time. The thermocouple allowed temperature to be monitored as a function of time during operation of the microscope, but noise transmitted along the wire made the quality of atomic scale images taken on the thermocouple-sample holder too poor to be useful for determining distortion as a function of time directly. However, the experiments gave highly reproducible data for the relationship between temperature and the time after the instrument was turned on, while the microscope was operating with parameters typical of image taking. Thus, experiments to record the relationships between

distortion and the other parameters could be linked indirectly to temperature through time. Heating/cooling experiments were carried out over periods of 12 to 14 hours while the instrument was continually scanning. The microscope was covered with the hoods from the beginning of the experiment until an hour or two before the experiment ended, at which point they were removed to induce rapid cooling.

Distortion vs. time. A series of experiments was designed to monitor distortion as a function of imaging time. A sample of freshly cleaved HOPG was used, and atomic scale resolution was achieved as quickly as possible. Images were taken in the same scan direction, except nine images of reverse direction that were captured during one experiment for comparison. Image size was kept constant at 16 nm, so scan rate could be held constant. Once it was possible to record atomic scale images of reasonable quality, images were captured periodically. After 1–2 hours of imaging, the hoods were removed, while imaging continued. This was done to make an important separation of the time and temperature parameters, necessary because temperature is dependent on time under normal scanning conditions until temperature reaches its maximum, equilibrium value. This makes it difficult to discriminate the influence of time or temperature individually, on drift and scaling. If temperature has the controlling influence on distortion, its rapid decrease when the hoods are removed should be reflected in the drift and scaling parameters.

Distortion vs. scan speed. In order to evaluate the influence of scan speed on image distortion, these experiments were carried out after about 1.5 h of scanning to allow for a period of instrument warm-up (electronics stabilization), and as quickly as possible to minimize the influence of operation time. The scan speed parameter was controlled by varying the scan size and the number of lines scanned per second (frequency), but sampling density was always selected to be 512 pixels on each side. Scanning speed in the slow scan direction was determined as image side length divided by the imaging time. Imaging time was found by dividing the number of sample lines (512) by the scanning frequency in lines per second. All images were taken in the same scan direction.

Results

Temperature vs. time. Figure 5 shows a typical plot for temperature as a function of time. The sample and microscope were initially at room temperature, but the sample heated rapidly under the hoods as the instrument electronics and laser warmed the system. An equilibrium temperature was reached about five hours after scanning began. When the hoods were removed, temperature decreased rapidly at first, and then equilibrated at a value about 8 °C higher than room temperature. Three experiments proved excellent reproducibility.

Drift and scaling vs. time. Drift and scaling results are shown on Figure 6. Removal of the hoods (indicated by arrows), thus a rapid decrease in temperature, had absolutely no influence on either drift velocity or scaling. For Experiment 1, atomic scale imaging was achieved immediately. There was a period of about 40 minutes where drift and scaling varied erratically, followed by about 20 minutes with less scatter and finally, fairly constant distortion parameters. Experiment 2,

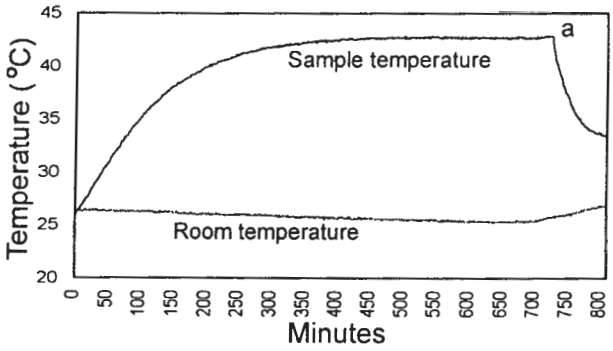

FIGURE 5. Sample temperature as a function of scan time. Sample temperature initially rises rapidly and comes to equilibrium. At point **a**, the hoods were removed and a rapid fall in temperature resulted with a simultaneous slight rise in room temperature recorded by the probe located at the base of the microscope.

where atomic scale images were only obtained after 36 minutes, had a period of about 20 minutes of moderate scatter, but an hour after the instrument was turned on, drift and scaling factors were fairly constant.

The scaling factor reached an equilibrium value of about 1.1, meaning that the tip covered an area of the sample that was 10% larger than intended. Note that at the beginning of Experiment 1, the scaling factor is 1. This means that the scanner calibration is fine for imaging during the first 10 minutes (about long enough to calibrate a scanner!) but that at no time afterward is the calibration correct. A scanner calibrated after about an hour or two of scanning gives more stable results, but scaling is a function of all scan parameters (as demonstrated further below), so complete elimination of scaling is impossible. Obviously, even a perfectly calibrated scanner cannot eliminate drift.

During experiment 2, some images of reverse (up) scan direction were captured (shown as crosses on Fig. 6). It is interesting that these images have drift values that are significantly different from those of down scan direction, taken alternately with them in time. The drift of the up-scan images follows a trend similar to the one observed for down-scan images, so drift decreases with time, but has a different value. This evidence that drift may differ considerably between successive images of opposite scan direction indicates the limit of the usefulness of visualizing the cause of drift as motion of the sample beneath the scanning tip. If this was what actually took place, it would be extremely difficult to explain alterations in drift when scan direction is reversed. We know drift to be a composite of a number of factors, among which is the response of the piezo to voltage; its behavior probably alternates when the sign of the voltage is changed.

Drift and scaling vs. scan speed. Scan speed had no significant influence on drift (Fig. 7). In both experiments, drift velocities were very low and constant. Scaling was slightly influenced by scan speed; it decreased as scan speed decreased, i.e., when either scan size or scan frequency was reduced. As expected, values of 1.1 were found for 16 nm size images scanned at 24.5 Hz (where pixel density was

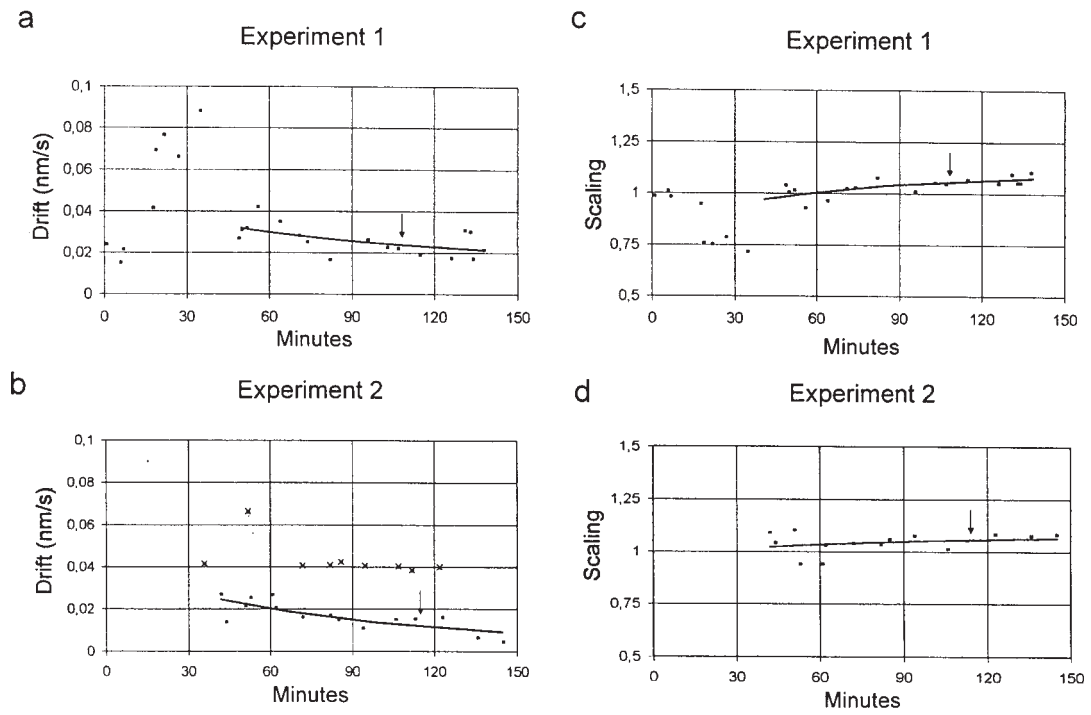

FIGURE 6. Drift velocity and scaling as functions of time for the two experiments. Experiment 1 (**a** and **c**), where atomic scale images were achieved almost immediately had an initial 40 minute period of wide scatter in both drift and scaling, whereafter drift decreased with time, while scaling values rose slightly. Experiment 2 (**b** and **d**), where useable images were not achieved until after 36 minutes of scanning shows an approach to more stable conditions after the first hour. Drift values for images of reversed (up) scan direction are shown as crosses on **b**. Note that these values are significantly different than their down scan counterparts, but seem to follow a similar trend that decreases with time. The arrows mark the times when the hood was removed and a sharp drop in temperature occurred.

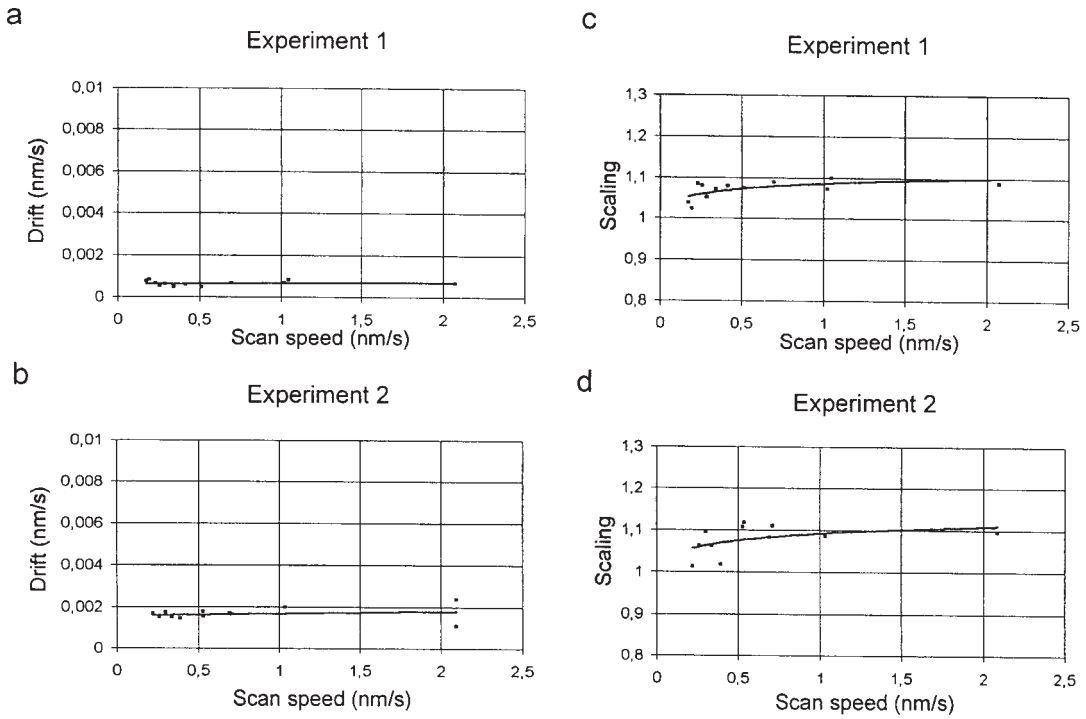

FIGURE 7. The variation of drift (**a** and **b**) and scaling (**c** and **d**) with scan speed. The experiments were conducted after 1.5 hours of scanning. For both experiments, drift is negligible, while a slight rise in scaling values with scan speed is indicated.

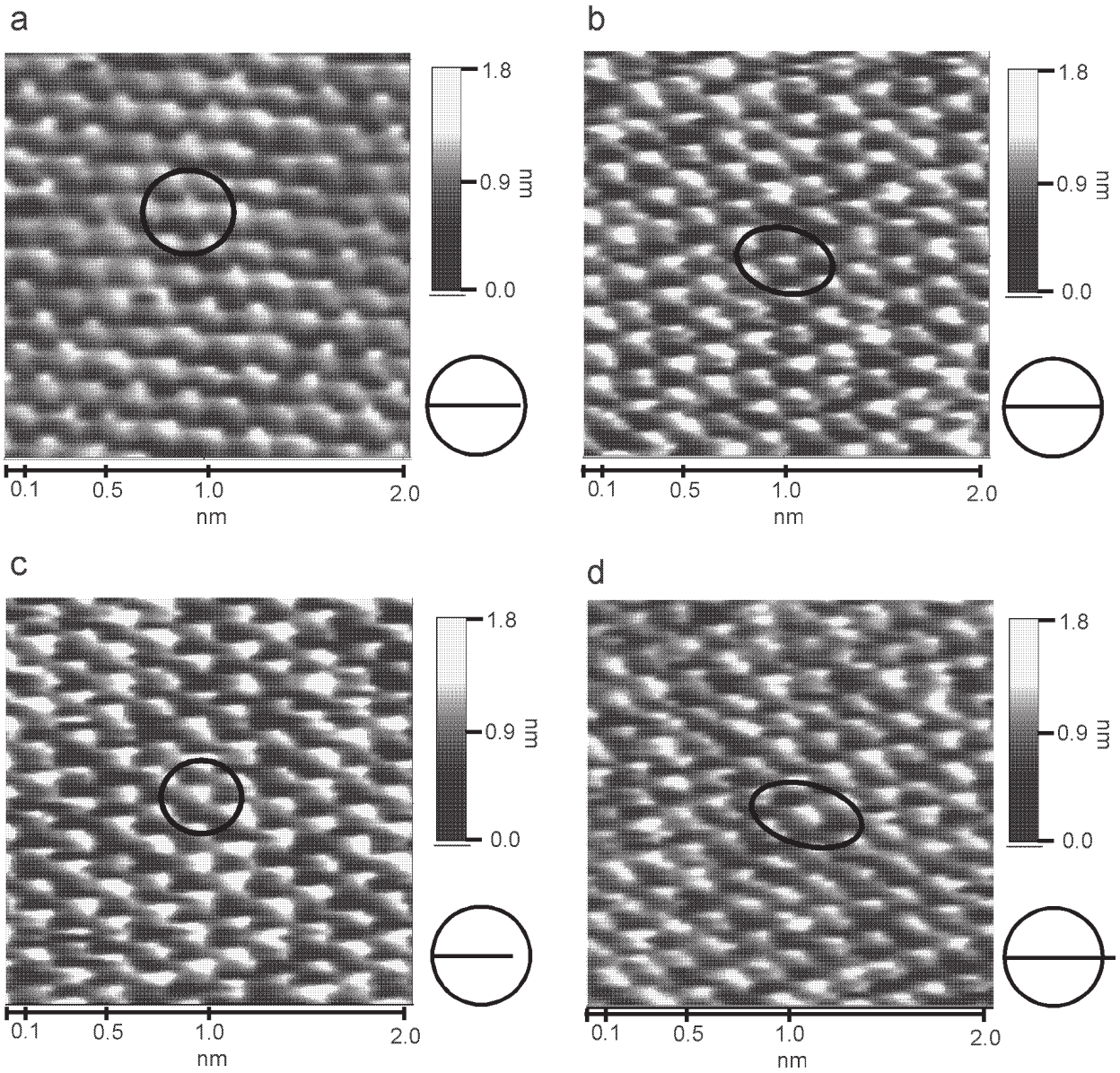

FIGURE 8. AFM images of HOPG at the atomic scale comparing possible distortion combinations. Image plane ellipses are drawn through the points of a hexagon. A real plane circle, representing the real diameter for graphite (drawn to scale) is shown beside each image. The line bisecting the circle has length equal to the width of the ellipse where it is broadest along *x*, so that it represents the expansion or contraction on the image. The ratio of the circle diameter to line length is the scaling factor, *s*. (**a**) Image taken under influence of very slight drift and scaling (nearly corresponding to Fig. 3a). (**b**) Image influenced by drift but not scaling (corresponds to Fig. 3b). (**c**) Image influenced by scaling but not drift (corresponds to Fig. 3c). (**d**) Image under influence of drift and scaling (corresponds to Fig. 3d). Note that the atomic lattice of **c** appears contracted, whereas on **d**, it is expanded. Images are unfiltered.

512), corresponding to 0.77 nm/s, the conditions of the scaling vs. time experiments.

DISCUSSION AND CONCLUSIONS

Although temperature of the sample can increase significantly with time of scanning, the experiments prove that temperature, as such, is not the main parameter controlling drift or scaling within the range of 23 to 43 °C. The major controlling parameter is time lapse after the electronics of the microscope are powered on. Large and unpredictable values for both drift and scaling were observed during the first 40 minutes after operation began so obviously, high-precision atomic scale work cannot be done during this time. Waiting 60 to 90 minutes to allow the electronics to stabilize before starting to image minimizes drift. Effects of scaling are diminished by calibrating the piezo only after the stable period has begun, and by using the same instrument conditions and parameters (scan size, range, etc.) intended for use during image collection. Hysteresis is more pronounced when the scanner is operating over large proportions of its range. Response of the piezoelectric ceramic to voltage is essentially linear for small voltage sweeps (van de Leemput et al. 1991; Libioulle et al. 1991).

Four AFM images of HOPG (Fig. 8) illustrate the range of apparent structure that can result from distortion. Figure 8a, where drift velocity and scaling factor are very small, can be compared with images affected by drift alone (Fig. 8b), scaling alone (Fig. 8c), and both drift and scaling (Fig. 8d). If measurement of distances and angles is an important aspect of a study, evidence of *distortion* should be the first check of an operator's routine tests for imaging artifacts. The most reliable warning signs for drift and/or scaling are: (1) distances and/or angles are not what is expected for the mineral examined; (2) the surface seems to move, so that features drift by during scanning; (3) morphology or the atomic pattern on up-scan images looks different than on down-scan images (see Fig. 9); or (4) image morphology changes when scan angle or scan frequency is changed, or when zooming in or out, or as time passes (for non-dynamic surfaces).

Each distorted image has its own drift vector and scaling factor. These parameters cannot be applied to any other image, though as the experiments presented here show, once into the stable period of instrument function, values on subsequent images are similar as long as instrument and imaging conditions (including scan direction) are the same. Each ceramic element behaves differently under operation, so even for exactly the same conditions, distortion parameters for images taken from two different scanners are not the same. However, the model presented here provides an easy, step-by-step way to quantify and adjust for drift and scaling on any images where there is a feature that can be represented as a circle on the real plane and identifiable as an ellipse on the image plane. In addition to atomic periodicity, this can also include symmetric features on micrometer scale images, for example etch pits or spiral growth hillocks, etc.

The model described here requires that distortion is homogeneous, which in our experience covers almost all images taken with reasonable imaging parameters after instrument electronics have had time to stabilize. When distortion is non-linear,

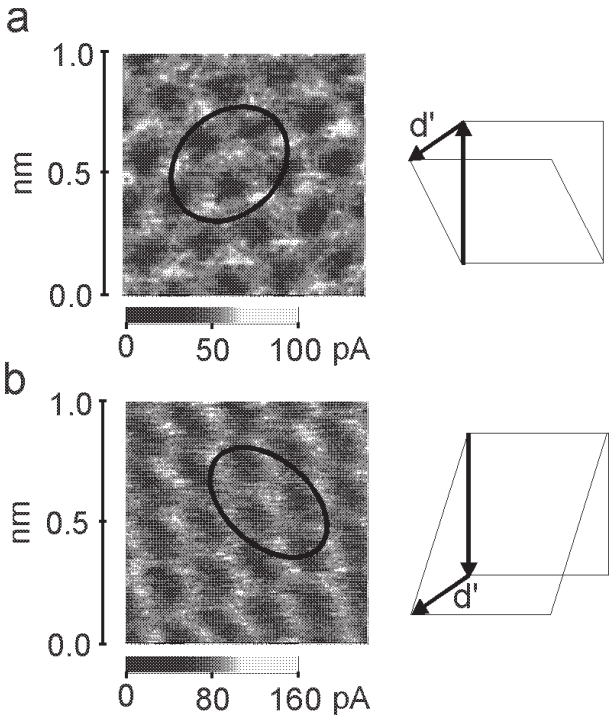

FIGURE 9. Two AFM images of HOPG illustrating how reversing scan direction results in dramatically different image morphologies, even if drift is constant. Sketches of the real plane situation are shown. For both images the scan area parallelogram is determined by drift and the slow scan vector. When this vector changes direction, the parallelogram changes shape. Our results indicate that the drift vector often changes as well (Fig. 7 and corresponding text), so morphology alterations when scan direction is reversed is a common and reliable sign of distortion. The images are unfiltered.

which can be recognized most easily in features that are known to be straight appearing curved, this method cannot be used. The more general distortion correction method presented by Garnæs et al. (1998) is recommended. Once drift and scaling parameters have been found, the image can be treated with a standard computer graphics program such as Illustrator or Corel Draw to reverse distortion. The drift vector is used to stretch the image into the parallelogram representing the real data imaged on the sample surface and the scaling factor is used to correct spatial representation so that the true surface proportions are restored. By correcting images for distortion using known structural relationships, local differences in structure or their change with time are more apparent. Distortion correction allows much more information to be extracted from SPM images, specifically, angular relationships that are valid and distances that are less uncertain.

ACKNOWLEDGMENTS

We thank C. Henriksen for profitable discussions of the model, P. Jørgensen for building the thermal sensor, O. Johnsen for supplying a natural graphite sample used in the initial experiments, T. Balic Zunic for help with X-ray diffraction, and our patient librarians, K. Knudsen, I. Hansen, and D. Korinth Jeppesen. The manuscript was improved by comments from M. Hochella, J. Rakovan, B. Bickmore, and three anonymous reviewers. Funding was provided by the Danish Natural Sciences Research Council.

REFERENCES CITED

Andersen, J.E.T. and Møller, P. (1994) Analysis and calibration of in situ scanning tunnelling microscopy images with atomic resolution influenced by surface drift phenomena. Surface and Coatings Technology, 67, 213–220.

Binnig, G., Rohrer, H., Gerber, C., and Weibel, E. (1982a) Tunneling through a controllable vacuum gap. Applied Physics Letters, 40 (2), 178–180.

———(1982b) Surface studies by scanning tunneling microscopy. Physical Review Letters, 49 (1), 57–61.

Binnig, G., Gerber, C., Stoll, E., Albrecht, T.R., and Quate, C.F. (1987) Atomic resolution with atomic force microscope. Europhysics Letters, 3, 1281.

Carrara, S., Facci, P., and Nicolini, C. (1994) More information on the calibration of scanning stylus microscopes by two-dimensional fast Fourier-transform analysis. Reviews of Scientific Instruments, 65, 2860–2863.

Eggleston, C. (1994) High resolution scanning probe microscopy: Tip-surface interaction, artifacts, and applications in mineralogy and geochemistry. In Nagy, K.L. and Blum, A.E., Eds., CMS Workshop Lectures Volume 7: Scanning probe microscopy of clay minerals, 1–90. Clay Mineral Society, Aurora, CO, U.S.A.

Garnæs, J., Nielsen, L., Dirscherl, K., Jørgensen, J.F., Rasmussen, J.B., Lindelof, P.E., and Sørensen, C.B. (1998) Two-dimensional nanometer-scale calibration based on one-dimensional gratings. Applied Physics A, 66, 831–835.

Gould, S.A.C., Burke, K., and Hansma, P.K. (1989) Simple theory for the atomic-force microscope with a comparison of theoretical and experimental images of graphite. Physical Review B, 40 (8), 5363–5366.

Hochella, M.F. Jr., Eggleston, C.M., Elings, V.B., Parks, G.A., Brown, G.E., Jr., Wu, C.M., and Kjoller, K. (1989) Mineralogy in two dimensions: Scanning tunneling microscopy of semiconducting minerals with implications for geochemical reactivity. American Mineralogist, 74, 1235–1248.

Hölscher, H., Raberg, W., Schwarz, U.D., Hasbach, A., Wandelt, K., and Wiesendanger, R. (1999) Imaging of sub-unit-cell structures in the contact mode of the scanning force microscope. Physical Review B, 59 (3), 1661–1664.

Jørgensen, J.F., Madsen, L.L., Garnæs, J., Carneiro, K., and Schaumburg, K. (1994) Calibration, drift elimination and molecular structure analysis. Journal of Vacuum Science and Technology B, 12 (3), 1698–1701.

Jørgensen J.F., Jensen C.P., and Garnæs, J. (1998) Lateral metrology using scanning probe microscopes, 2D pitch standards and image processing. Applied Physics A 66, S847–S852.

Libioulle, L., Ronda, A., Taborelli, M., and Gilles, J.M. (1991) Deformations and nonlinearity in scanning tunneling microscope images. Journal of Vacuum Science and Technology B, 9 (2), 655–658.

Tang, H., Joachim, C., and Devillers, J. (1993) Interpretation of AFM images: The graphite surface with a diamond tip. Surface Science, 291, 439–450.

van de Leemput, L.E.C., Rongen, P.H.H., Timmerman, B.H., and van Kempen, H. (1991) Calibration and characterization of piezoelectric elements as used in scanning tunneling microscopy. Reviews of Scientific Instruments, 62 (4), 989–992.

von Toussaint, U., Schimmel, Th., and Küppers, J. (1997) Computer simulation of the AFM/LFM imaging process: Hexagonal versus honeycomb structure on graphite. Surface and Interface Analysis, 25, 620–625.

Yurov, V.Y. and Klimov, A.N. (1994) Scanning tunneling microscope calibration and reconstruction of real image: Drift and slope elimination. Reviews of Scientific Instruments, 65 (5), 1551–1557.

MANUSCRIPT RECEIVED DECEMBER 4, 2000
MANUSCRIPT ACCEPTED AUGUST 1, 2001
MANUSCRIPT HANDLED BY JOHN RAKOVAN