Spectroscopy with the scanning tunnelling microscope: a critical review

1989 J. Phys.: Condens. Matter 1 10211

(http://iopscience.iop.org/0953-8984/1/51/001)

J. Phys.: Condens. Matter 1 (1989) 10211–10228. Printed in the UK

REVIEW ARTICLE

Spectroscopy with the scanning tunnelling microscope: a critical review

Ruud M Tromp
IBM Research Division, Thomas J Watson Research Center, PO Box 218, Yorktown Heights, NY 10598, USA

Received 10 July 1989, in final form 24 August 1989

Abstract. In this paper we discuss the use of the scanning tunnelling microscope as a spectroscopic tool. Several methods of obtaining spectroscopic information are reviewed. The strengths and weaknesses of scanning tunnelling microscopy are discussed in comparison with more conventional surface spectroscopy techniques.

1. Introduction

In the scanning tunnelling microscope (STM) a metallic tip is advanced towards the sample to the point where tip and sample wavefunctions start to overlap. Upon application of a small bias voltage between tip and sample, an electron current will flow through the gap separating sample and tip due to tunnelling. The tunnelling current depends critically on the wavefunction overlap. If this overlap is changed—for instance, by changing the applied voltage difference between tip and sample—the tunnelling current will change with it. This allows the use of the STM as a spectroscopic tool [1–3].

The simplest demonstration of the sensitivity of the STM to the wavefunctions of the sample is obtained by measuring topographs on one sample at different applied voltages. Such topographs are obtained by scanning the tip in a raster-like fashion across the sample. As the tunnelling current tends to change from one place to the next, a feedback circuit dynamically adjusts the distance between tip and sample to keep the tunnelling current constant. One measures the correction voltage applied by the feedback circuit to the z-piezo (controlling sample–tip distance) along the raster scan, which can then be displayed as an atomically resolved height map of the surface. Such an image contains geometric information about the sample (such as the presence of atomic steps), but it also contains information on the surface electronic structure (such as the presence of localised surface states, band bending, or work-function variations).

Two images of the Si(1 1 1)-(7 × 7) surface obtained at −2 and +2 V applied bias (tip grounded) are shown in figures 1(a) and (b), respectively [4]. Both images show the familiar diamond-shaped unit cell with six maxima in each half of the cell, and a pronounced depression at the corners (the so-called 'corner hole'). Since the geometric structure of the surface is identical in both cases, the similarity of the two images is not surprising. What is of interest for someone interested in spectroscopy is the difference between the two images. Figure 1(b) shows mirror symmetry of the unit cell across the short diagonal (all bumps look the same), whereas figure 1(a) shows a pronounced

0953-8984/89/5110211 + 20 $02.50

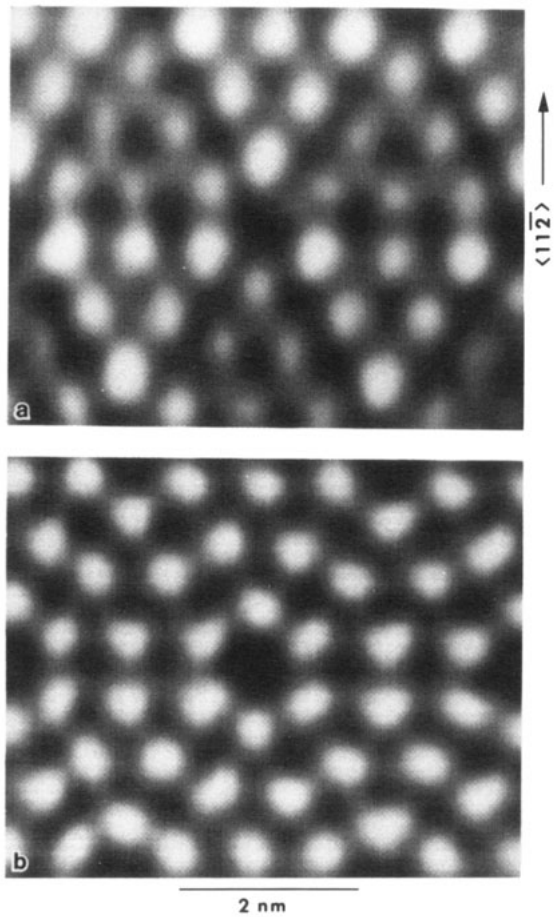

Figure 1. Constant-current STM images obtained on the Si(1 1 1)-(7 × 7) surface at (*a*) −2 V and (*b*) +2 V sample bias.

asymmetry with one half of the unit cell apparently higher than the other half, and the bumps near the corner holes higher than the bumps between them. In figure 1(*b*) the empty states of the sample are probed, in figure 1(*a*) the filled states. Evidently, there is a pronounced difference in electronic structure between these two tunnelling conditions, which gives rise to the observed difference in images.

The questions addressed in this paper are the following. How can one obtain detailed spectroscopic information on a surface using the scanning tunnelling microscope? Figures 1(*a*) and (*b*) contain both *structural* and *electronic* information. To what extent can these be identified and separated? Is there an 'ideal and correct' way to perform an STM experiment? And, maybe even more importantly, can one understand the result one obtains?

This paper is organised as follows. First we will briefly review the theory of current flow in the STM. Next we will take a look at some experimental results obtained at a fixed location, not attempting to image the electronic structure. Then we will discuss various efforts to obtain real-space images of the surface electronic structure. In our discussion

we will take a critical look at the progress that has been made and the problems that remain.

2. Theory

The first theory of the STM was put forward by Tersoff and Hamann [5] and considered the case of vanishingly small bias voltage. In this limit it was shown that the images measured at constant tunnelling current correspond to surfaces of constant charge density at the Fermi level, measured at the centre of curvature of the tip. The theory did not consider finite bias voltages and made, therefore, no statement about spectroscopic measurements with the STM.

If we have a tip t and a sample s with a bias voltage V applied between tip and sample, then the tunnelling current flowing between these two electrodes is given by [6]

$$I = \int_0^{eV} \rho_s(r, E)\rho_t(r, \pm eV \mp E)T(E, eV, r)\,\mathrm{d}E. \tag{1}$$

Here ρ_s and ρ_t are the densities of states of sample and tip at energy E (relative to the Fermi level) and at position r. The upper signs are for positive sample bias, and the lower signs for negative sample bias. The tunnelling transmission probability density T is given by

$$T(E, eV) = \exp\left[-\frac{2z(2m)^{1/2}}{\hbar}\left(\frac{\varphi_s + \varphi_t}{2} + \frac{eV}{2} - E\right)^{1/2}\right] \tag{2}$$

and T depends exponentially on the sample–tip distance z and the square root of the sum of the work functions (for small bias voltages).

Lang [7] considered the following situation. Take two semi-infinite jellium electrodes at distance s. On one or on both of the electrodes an atom may be adsorbed. The wavefunctions for these two electrodes may be solved from first principles. When a bias voltage V is applied, a tunnelling current I will flow in accordance with equation (1). Since the electrodes are semi-infinite, this current is infinite, but by subtracting the tunnelling current between the two electrodes without the adsorbed atom(s) one obtains the relevant part, the tunnelling current flowing through the adsorbed atoms, which one may arbitrarily label 'tip' and 'sample'.

The theory allows the simulation of an STM experiment, i.e. it can calculate the distance s at which the tunnelling current remains constant as the tip atom on one electrode scans across the sample atom on the second electrode. The result of such a study is shown in figure 2, where a Na tip scans over various sample atoms (Na, S and He), at small bias voltage. Na has a higher density of states at the Fermi level than S and gives rise to a larger displacement of the tip at constant tunnelling current. He is different in that it screens the density of states of the jellium substrate and forces the tip to approach the sample more closely in order to maintain a constant tunnelling current.

Also, the theory can calculate the tunnelling current as a function of applied bias voltage and at constant tip separation. Lang [7] has shown that the quantity $\mathrm{d}\ln I/\mathrm{d}\ln V$ closely resembles the combined density of states of sample and tip in the tunnelling region. Figure 3(a) shows the density of states induced on the jellium electrode by the adsorption of Na and Ca. The energy scale for the Na is reversed, so that it can be interpreted as the tip. Figure 3(b) (full curve) is a calculation of $\mathrm{d}\ln I/\mathrm{d}\ln V$ versus V

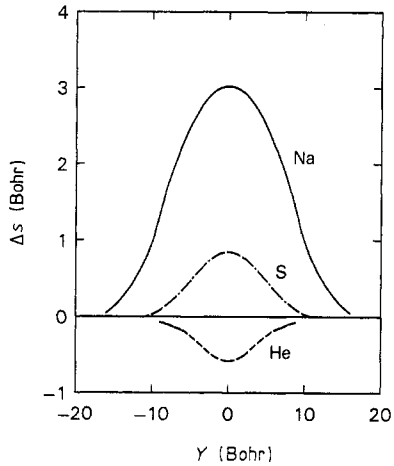

Figure 2. Calculated vertical tip displacement for a Na tip scanning over a Na, S and He atom, respectively.

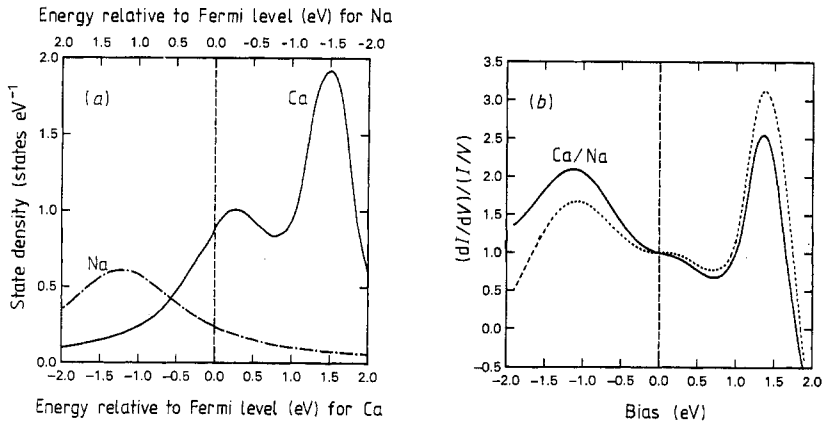

Figure 3. (*a*) Density of states induced by the adsorption of a Na and a Ca atom on a jellium electrode. The horizontal axis is inverted for the Na atom, which is interpreted as the tip. (*b*) $(\mathrm{d}I/\mathrm{d}V)/(I/V)$ as a function of applied bias for the Na–Ca tunnel junction. The full curve is an exact calculation; the broken curve is the result of a simple model (see [7]).

for the Ca/Na tunnelling system. The peaks observed in this spectrum have a close relation to the peaks seen in figure 3(*a*). There are two important implications. First, the tunnelling signal contains information that is to a large degree representative of the electronic structure of the tunnelling electrodes. Secondly, the tip electronic structure is just as prominent as the sample electronic structure. This second fact is one of the major reasons for irreproducibility of spectroscopic data. A recent experimental study has shown how adsorption of a contaminant (in particular, a strongly electronegative atom such as S or O) can seriously affect the STM image [8].

3. Variation of I, $\mathrm{d}I/\mathrm{d}V$ and $\mathrm{d}\ln I/\mathrm{d}\ln V$ with V

As outlined above, spectroscopic information can be obtained by measuring the variation of tunnelling current with voltage. In its simplest form one would measure the

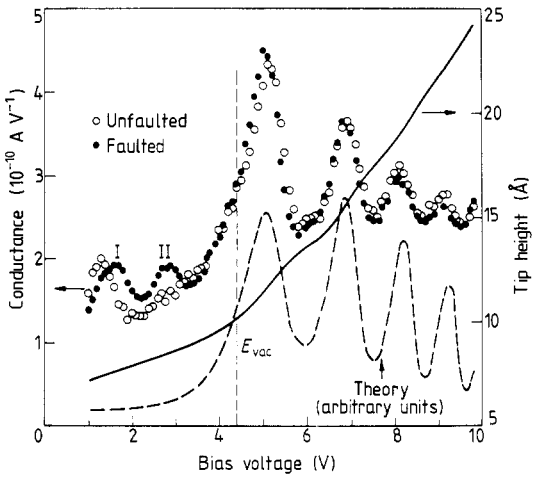

Figure 4. dI/dV spectra measured in the two halves of the Si(1 1 1)-(7 × 7) unit cell. Also shown is the variation of the sample–tip spacing with bias voltage, while the DC tunnelling current is held constant. Also shown are the tip height (full curve) and a calculation of the contribution of standing-wave states to the tunnelling conductance (broken curve).

tunnelling current I as a function of voltage V (I–V curves). Experimentally, the question is how to do that. Various schemes have been employed. Becker *et al* [9] and Binnig *et al* [10] first measured dI/dV by adding a small AC component to the DC bias voltage. The frequency ω was chosen sufficiently high so that the feedback circuit could not respond to it. The tunnelling current will contain an in-phase modulation with frequency ω that is the derivative of I with respect to V, at the DC bias voltage V at which the feedback circuit operates. By sweeping V, dI/dV can be measured as a function of V. Care has to be taken to pick up only the in-phase modulation of the tunnelling current with a lock-in amplifier. Displacement currents can give rise to strong phase-shifted signals.

There are three problems with this approach. First, the tip–sample separation varies with V, because I is kept constant by the feedback circuit. With increasing voltage and distance, the lateral resolution is degraded. Secondly, the DC tunnelling resistance is not constant during the measurement, again because I is kept contant. At decreasing bias voltage the DC conductance is increased (by decreasing the sample–tip separation) and diverges to infinity as zero bias voltage is approached. Likewise dI/dV diverges. In practice, useful data have not been obtained below 1 V, which for many studies is the most interesting region. Finally, this method requires that a tunnelling current can be maintained over the voltage range of interest. This is not always possible. If there is a band gap (which is quite common on semiconductor surfaces), a stable tunnelling current can often not be maintained if the bias voltage is set within the gap, and the tip will crash into the sample. Nonetheless, this method has been used to study the electronic structure of surfaces. Here we show results obtained by Becker *et al* [11] on the Si(1 1 1)-(7 × 7) surface (figure 4). Two dI/dV curves are shown obtained in the two halves of the (7 × 7) unit cell. As noticed before, no data were obtained below 1 V. Between 1 and 3 V, differences are seen between the two curves due to differences in the electronic structure in the two halves of the unit cell. Also shown is the variation of tip height above the surface, which increases with increasing voltage, as expected. The regular series of peaks observed at higher voltages are not due to surface electronic structure, but are due to

standing-wave-like states between tip and sample. These were observed earlier by Binnig *et al* and by Becker *et al* [9, 10]. The broken curve shows a theoretical prediction of these states, in good agreement with the data.

The problems mentioned above can be solved by measuring I–V curves at constant sample–tip separation. In order to achieve this, one has to break the feedback circuit. This was first done by Feenstra *et al* in their study of the Si(1 1 1)-(2 × 1) surface [12]. In this scheme the sample–tip distance does not change with voltage, and the tunnelling current is allowed to vanish because the feedback circuit is inactive. By digital recording of I–V curves, $\mathrm{d}I/\mathrm{d}V$ can be constructed afterwards. The problem of the dependence of the tunnel signal on the value of the DC tunnelling conductance can be resolved by normalising the dynamic conductance $\mathrm{d}I/\mathrm{d}V$ to the DC conductance I/V. Feenstra *et al* showed that the quantity $(\mathrm{d}I/\mathrm{d}V)/(I/V)$ (or $\mathrm{d}\ln I/\mathrm{d}\ln V$) corresponds closely to the sample density of states. As mentioned above, the theoretical studies by Lang support this view. The data by Feenstra *et al* on the Si(1 1 1)-(2 × 1) surface are shown in figure 5, together with the predicted density of states for the π-bonded chain model. There is a close correspondence between theory and experiment. (Also shown are data obtained on a Ni sample, which show no band gap and no surface states.) A note of warning is needed here: $\mathrm{d}\ln I/\mathrm{d}\ln V$ is not related *directly* to the density of states of sample and tip. Remember that the tunnelling current is also determined by the tunnelling transmission probability density. In simple terms, in addition to a finite density of states there must be a significant overlap between the corresponding sample and tip wavefunctions. If the sample has a large density of states, but these states do not overlap with the tip, these states are inaccessible in a tunnelling experiment. One example may be the d band on Ag. In a photoemission experiment, the d band would be the dominant feature in the electron energy distribution curve. However, the d orbitals are small and localised closely to the atomic core. In a tunnelling experiment the d electrons are invisible. Measurement of I–V curves at constant sample–tip separation has found wide application in spatially resolved studies of surface electronic structure. Two of the most notable cases are the study of localised atomic scale defects by Hamers *et al* [13], and the study of the interaction of NH_3 with Si(1 1 1) by Avouris *et al* [14]. Results by Hamers are shown in figure 6, obtained on the Si(1 1 1)-($\sqrt{3} \times \sqrt{3}$)Al surface. Images in figures 6(*a*) and (*b*) are obtained with +2 and −2 V bias, respectively, and show the ($\sqrt{3} \times \sqrt{3}$) lattice, with some defects. The defects appear as depressions in figure 6(*a*) and as elevations in figure 6(*b*). Figure 6(*c*) shows $\mathrm{d}\ln I/\mathrm{d}\ln V$ curves taken on a regular lattice position and on a defect position. In a detailed comparison with theory, Hamers *et al* showed that the regular lattice positions correspond with an Al atom adsorbed on top of three first-layer Si atoms. The three Al valence electrons pair with the three Si dangling bonds. These back-bonds are imaged in figure 6(*b*). An empty p_z orbital protrudes into the vacuum and is imaged in figure 6(*a*). The defects are due to Si atoms substituting for Al atoms. Si has one more valence electron than Al and the p_z orbital is therefore not empty, giving rise to the extra electronic state at −0.5 eV seen in figure 6(*c*). Hamers *et al* further showed that this defect is similar in nature to the well known P_b centre, observed at the Si/SiO_2 interface [13].

Avouris *et al* studied the reaction of NH_3 with the Si(1 1 1)-(7 × 7) surface [14]. Upon adsorption of NH_3 they found remarkable changes in the STM images, due to saturation of dangling bonds with H and NH_2. Curves of $\mathrm{d}\ln I/\mathrm{d}\ln V$ were measured on a dense grid, allowing a detailed study of the interaction (figure 7). It was found that the so-called restatom sites (dangling bonds in the second layer of the crystal) were the most reactive of all. Reaction with H gives rise to significant charge transfer to neighbouring adatoms, which react next.

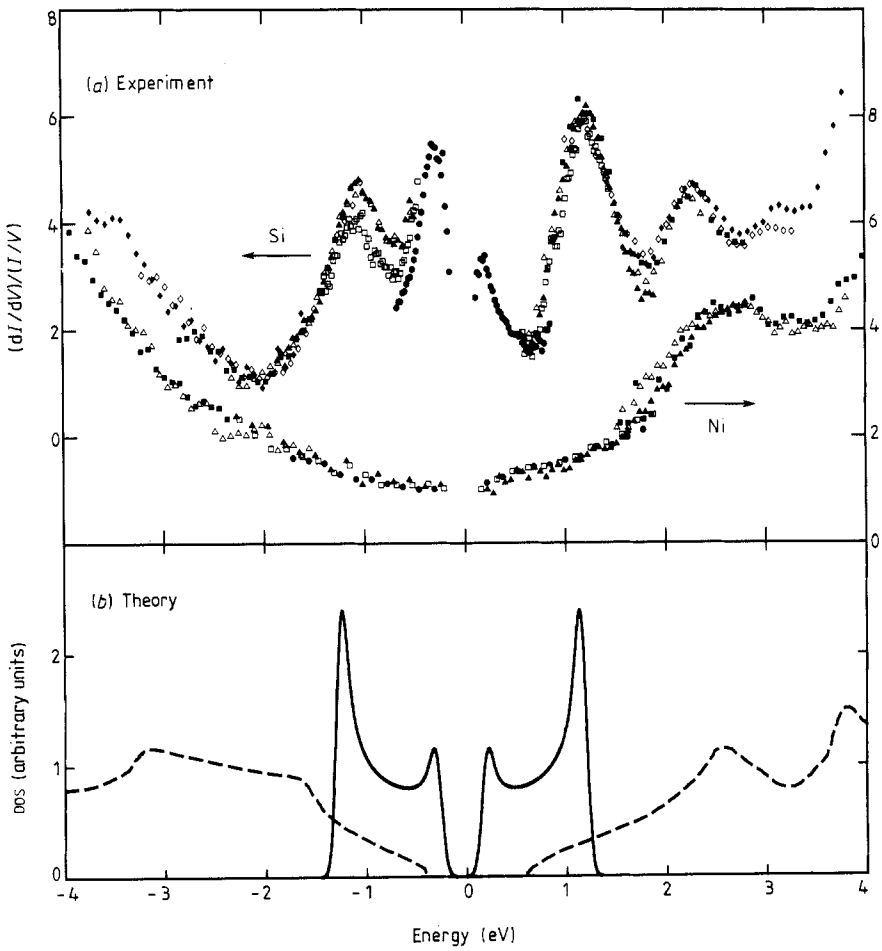

Figure 5. (*a*) $(\mathrm{d}I/\mathrm{d}V)/(I/V)$ for Si(1 1 1)-(2 × 1) and for a Ni surface. Notice the presence of strong electronic states on the Si surface. (*b*) Theoretical density of states for the π-bonded chain model. Broken curves represent bulk states; full curves are surface states.

Spatially resolved I–V curves have been measured to elucidate the electronic structure of many other semiconductor and metal surfaces. Here we have insufficient space to review this work, and now move on to the next subject, spectroscopic imaging.

4. Surface states in real space

With the STM capable of atomic-resolution images and of spectroscopy of surface electronic structure, the imaging of surface states in real space has provided a major, non-trivial challenge. Again, various approaches have been taken to this problem. We will review them here and comment on the merits and problems associated with each of these.

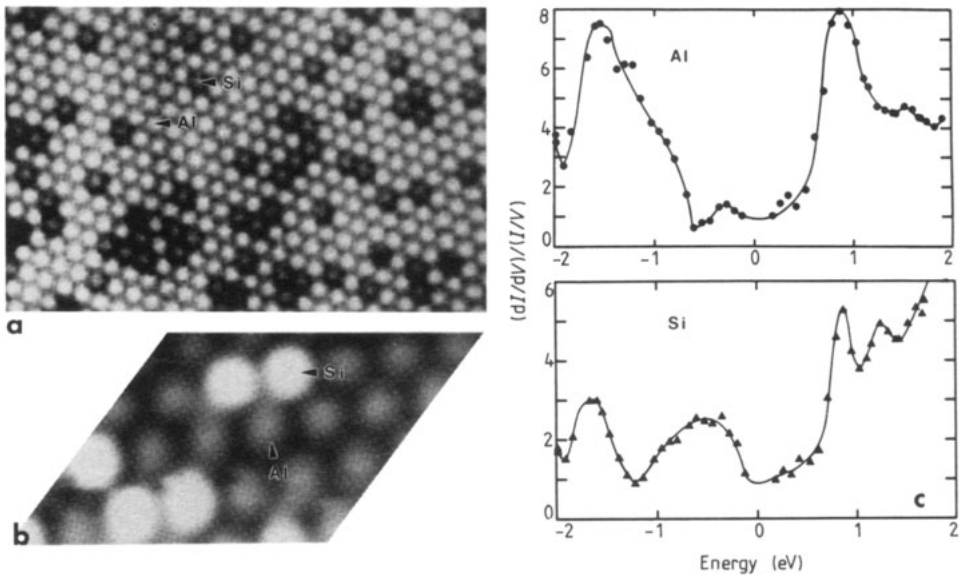

Figure 6. Si(1 1 1)-($\sqrt{3} \times \sqrt{3}$)Al: (*a*) empty-state image; (*b*) filled-state image; (*c*) (d*I*/d*V*)/(*I*/*V*) taken on an Al adatom (top) and a Si adatom (bottom).

4.1. Constant-current topograph (CCT)

Measuring topographs at constant tunnelling current at different bias voltages is the simplest way to try to obtain real-space surface state images, and examples have already been shown in figures 1 and 6. Here we show still another example, obtained on the GaAs(1 1 0) surface [15]. Figure 8(*a*) shows a tunnelling image obtained at +1.9 V, figure 8(*b*) was obtained at −1.9 V, representing empty and filled states, respectively. On the GaAs(1 1 0) surface the filled states are localised on the As atoms, the empty states are localised on the Ga atoms. Ga and As form zig-zag chains, alternating between Ga and As. This separation is clearly seen in figure 8: figure 8(*a*) is an image of the Ga states, figure 8(*b*) of the As states. Figure 8(*c*) shows a schematic top view of the surface atoms.

This method relies on the fact that at any bias voltage only the electronic states between the Fermi levels of the tip and the sample contribute to the tunnelling current. Thus, one obtains implicitly an image of those states, as is apparently the case in figure 8. Figure 1 appears to be somewhat more problematic, since both *geometric* and *electronic* structures contribute to both images. This would also be true if one images a region containing an atomic step, where in addition to the electronic structure one would observe the purely geometric features associated with the step.

It is tempting to think of figure 8 as an image of the Ga and As atoms as *chemically* distinct species. This, however, is a rather dangerous line of thought. The GaAs(1 1 0) surface has been studied very extensively and its geometric and electronic structures are known in great detail from a variety of experimental and theoretical studies. Thus the identification of the electronic states with the atomic species is simple, but still depends on external knowledge. Suppose, however, that the surface were to contain an antisite defect, giving rise to local changes in surface electronic and geometric structures. These changes would be easily observed with the STM, but identification of the chemical identity

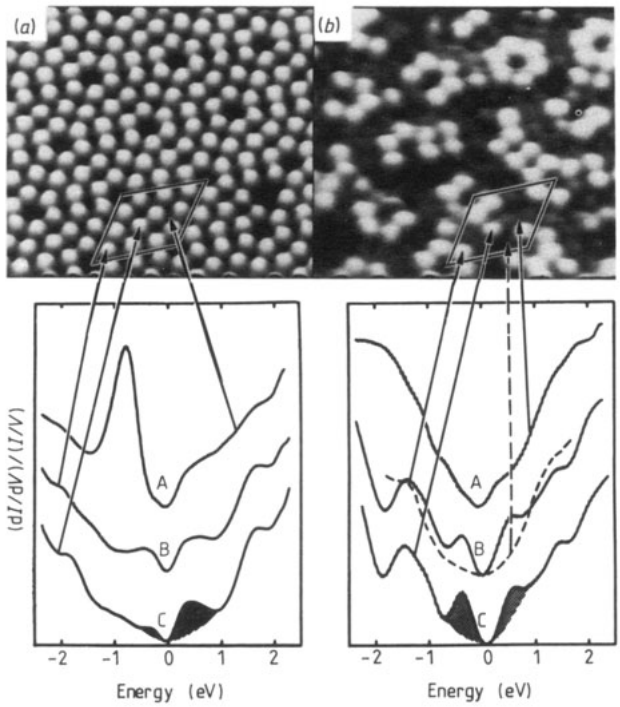

Figure 7. (*a*) Empty-state topograph of the Si(1 1 1)-(7 × 7) surface. Spatially resolved tunnelling spectra are shown below for the restatom site (A), an adatom next to the corner hole (B) and a centre adatom (C). (*b*) Empty-state topograph of the Si(1 1 1)-(7 × 7) surface after reaction with NH_3. Curve A was taken over a reacted restatom, curve B (broken) over a reacted corner adatom, curve B (full) over an unreacted corner adatom and curve C over an unreacted centre adatom. The removal of surface states on the reacted sites and changes in density of states in neighbouring unreacted sites are apparent from these measurements.

associated with the observed features would be extremely difficult and would have to relay on the detailed modelling of the defect, followed by extensive electronic structure calculations to be matched with the experimental data.

Voltage-dependent CCT have been used successfully by Stroscio *et al* [16] to observe band bending on GaAs around atomic-scale defects resulting from impurity adsorption. Such band bending can be observed as the presence of 'lakes' or 'hills' around these defects resulting from Fermi-level pinning at the defect site. The diameter of the lakes and hills corresponds to the Debye screening length.

STM images have also been recorded with much success on charge-density-wave compounds, such as $TaSe_2$, and other materials [17]. The charge density wave gives rise to a very strong modulation of the electron charge density in the vacuum region and can even completely obscure the underlying atomic corrugation.

The use of voltage-dependent topographs has one very attractive feature: it is very simple experimentally and it often provides a quick way to assess if there are any interesting differences between different polarities. As a spectroscopic tool, however, it is rather limited.

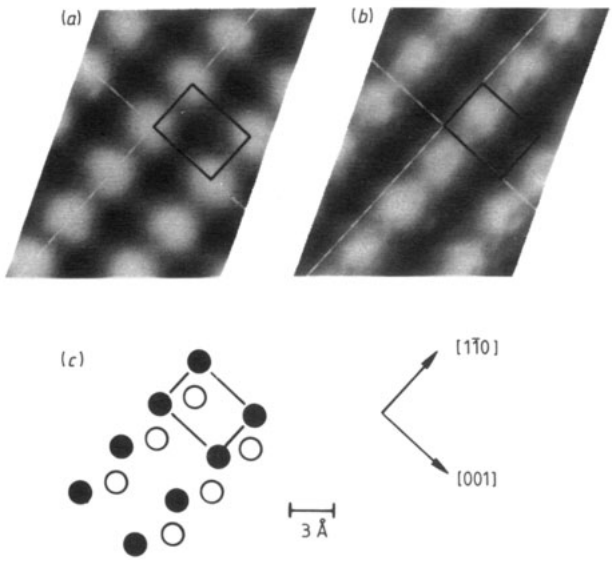

Figure 8. (*a*) Empty-state and (*b*) filled-state topographs on the cleaved GaAs(1 1 0) surface. (*c*) A schematic top view of the surface atoms.

4.2. Current imaging tunnelling spectroscopy (CITS)

Binnig *et al* obtained the first spectroscopic images by measuring dI/dV at a fixed bias voltage V in a raster scan [1]. A small voltage modulation was superimposed on the DC bias at a frequency ω, which was too fast for the feedback circuit to follow. The current modulation at frequency ω was measured with a lock-in amplifier and used to form an image of dI/dV. This method suffers from the same problems as outlined above for measuring dI/dV curves. A second disadvantage is that spectroscopic information is obtained at a single bias voltage only. If one wants spectroscopic information over a larger range of bias voltages, the experiment has to be repeated for every bias voltage. Problems associated with drift of the sample during the measurement and with variations of the geometric and electronic structures of the tip tend to make this experiment very troublesome. A further disadvantage is that the tip follows a different contour for each bias voltage, because different sample and tip states contribute to the tunnelling current at different voltages. This may affect the lateral resolution, in particular for features close to the resolution limit of the experiment. In addition, it has been shown that, in the absence of any electronic structure variation across the sample, the dI/dV image is not featureless, but presents an inverted topographic image (for a detailed discussion, see [6]). This is due to the fact that the transmission probability density T (given for flat electrodes in equation (2)) depends on the local radius of the curvature of the electrodes.

An alternative approach was taken by Hamers *et al* in their study of the Si(1 1 1) surface [18]. The reasoning goes as follows. If one could measures a full I–V curve in each pixel of a topographic image, one would have sufficient information to construct real-space images of the surface states. In order to do this, the feedback circuit has to be broken in each pixel and an I–V curve recorded. This was in effect what the authors did. The feedback circuit was operated at a bias voltage of +2 V. This voltage was present on the sample for about 0.1 ms. During this time the feedback circuit was activated by

opening a gate. In the next 0.4 ms the feedback circuit was inactive and the position of the tip held fixed. The bias voltage was ramped from +2 to −2 V, and the tunnelling current was measured at a number of voltages during this ramp by opening a sequence of gates, each connected to a boxcar averager. The signal from each boxcar averager was read by the computer through a number of analogue-to-digital converters (ADC). The entire cycle was repeated at a rate of 2 kHz. Thus, each boxcar averager measured an image of the tunnelling current at a different voltage, while the STM progressed along the raster scan with the tunnelling current stabilised at the feedback voltage of +2 V. The *I–V* curves could be seen in real time on an oscilloscope and displayed very strong variations with position. The variation of tunnelling current at a specific bias voltage can be shown as an image (current image, CI).

The feedback bias of +2 V has been chosen in this study after Tromp *et al* had shown earlier that the topographic image at this voltage closely follows the expected geometric contours for this surface, as calculated using atomic charge superposition methods, which ignore electronic structure contributions to the STM image [4]. In the first step of the analysis the tunnelling conductance (I/V) was plotted at a number of different positions inside the (7 × 7) unit cell. This is shown in figure 9(*a*). Strong differences are seen between different locations. For instance, the strong increase in conductance at −0.8 V seen on the restatoms (●) is absent on the adatoms (□), even though they are separated by only 4.43 Å. The steps in DC conductance are seen to correspond directly with the surface states long known from photoemission and inverse photoemission experiments [19]. Thus, this measurement allowed the authors for the first time to identify these states with specific features of the structure of the (7 × 7) surface. Real-space images of these states were subsequently obtained by taking the difference between current images just above and below the observed onsets in conductance. Such difference images, between 0 and −0.35 V and between −0.6 and −1.0 V, are shown in figure 10, together with a ball-and-stick model of the surface and a topographic image obtained at +2 V. The filled state closest to the Fermi level (known as the 'metallic edge' in photoemission jargon) is seen to reside on the adatoms (yellow in figure 10(*a*)) whereas the strong state at −0.8 V resides on the restatoms (blue). Together with the corner-hole dangling bond, these states form all the dangling bonds in the (7 × 7) unit cell, a total of 19.

The current images contain information not only on the electronic structure at the voltage at which they are obtained, but also on the electronic states contributing to the tunnelling current at the feedback voltage. Current images obtained at, say, +1 V will look different when the feedback voltage is −2 V instead of +2 V, because the tip follows a different contour, as shown in figure 1. Thus, in general, it is difficult (if not impossible) to interpret a current image directly. The problem is eased substantially by taking the *difference* between two current images, as we did in figure 10. One can think of this as subtracting out the common feedback in the two current images. One may improve this further by normalising this difference image to the DC conductance and thus creating a $\mathrm{d}\ln I/\mathrm{d}\ln V$ image. The images shown in figures 10(*c*) and (*d*) are $\mathrm{d}I/\mathrm{d}V$ images. The $\mathrm{d}\ln I/\mathrm{d}\ln V$ images created from the same data would be qualitatively similar due to the very pronounced nature of the surface states. In other cases we have noticed that the $\mathrm{d}I/\mathrm{d}V$ images may contain spurious contrast that is removed in the $\mathrm{d}\ln I/\mathrm{d}\ln V$ image. However, even in a $\mathrm{d}\ln I/\mathrm{d}\ln V$ image derived from a CITS experiment it may be impossible to remove spurious contrast completely, due to lateral variations in the transmission tunnelling probability. The major advantages of the CITS technique are that all spectroscopic information is obtained simultaneously, thus eliminating problems

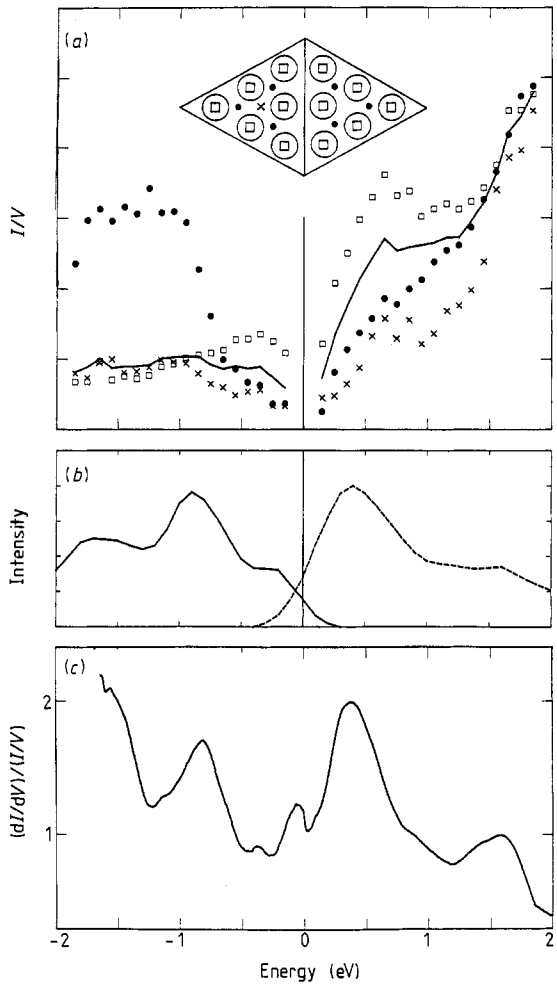

Figure 9. (*a*) Tunnelling conductance (I/V) measured in several locations inside the (7 × 7) unit cell: (□) adatom; (●) restatom; (×) centre position. (*b*) Surface states on Si(1 1 1)-(7 × 7) observed with photoemission and inverse photoemission [19]. (*c*) Area-averaged tunnelling spectra.

associated with sample drift and tip instabilities. Also, spectroscopic images at different bias voltages are obtained along identical spatial contours of the tip, thus removing ambiguities arising from possible variations in lateral resolution with tip–sample distance.

The CITS method described above has been used to study a number of other surfaces such as the Si(001)-(2 × 1) surface [3, 20] and the Si(1 1 1)-($\sqrt{3} \times \sqrt{3}$)Ag surface [21]. The technique asks much of the laboratory automation system used to drive the microscope. Presently it is possible to record a 256-point I–V curve in each pixel of a 250 × 250 topographic image. In each pixel the computer triggers a transient recorder, which acquires the I–V information, provides hardware averaging of multiple I–V curves at each location, and then sends the data to a second computer, which stores the results

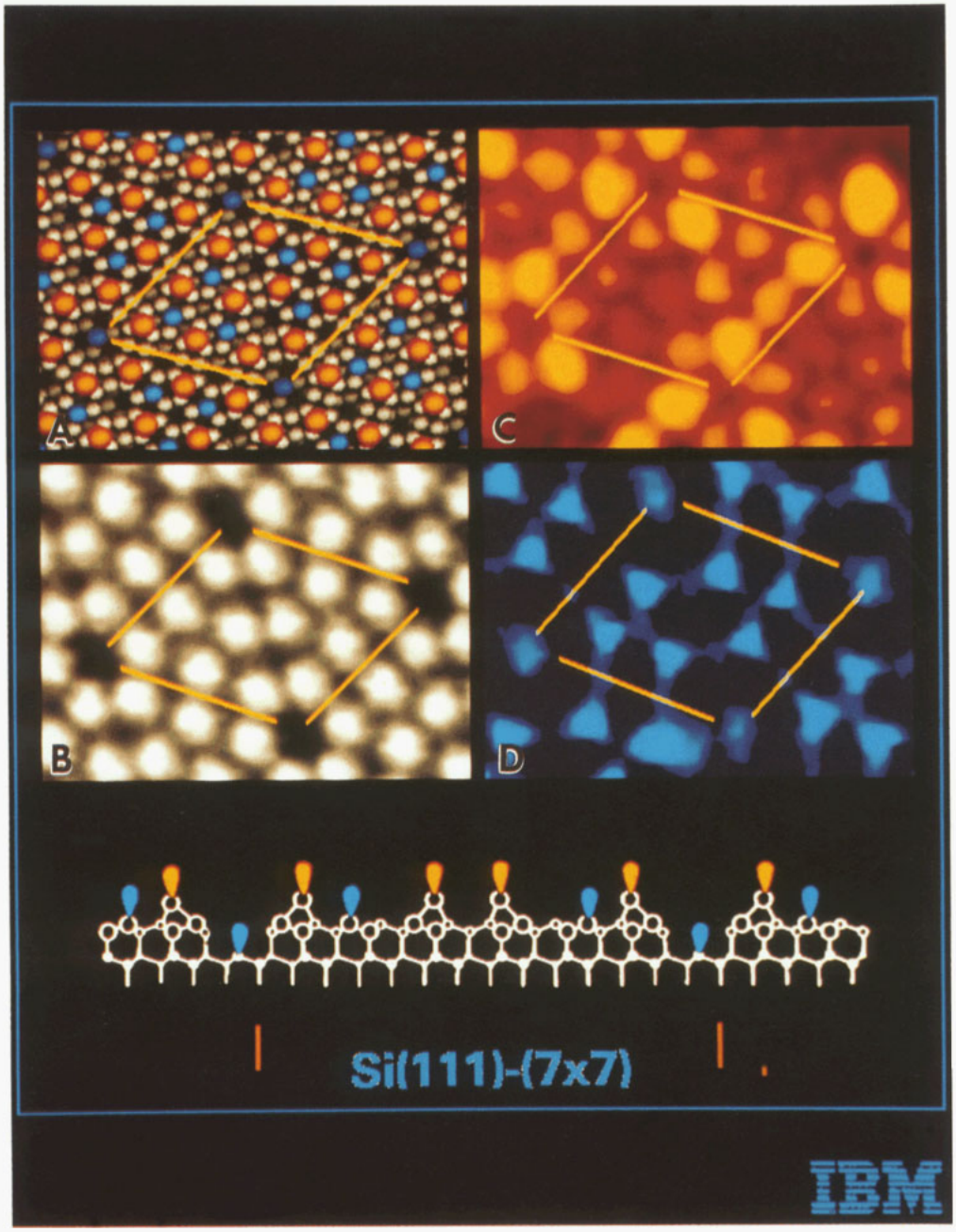

Figure 10. (*a*) Computer-generated top view of the Si(111)-(7 × 7) structure. Adatoms are orange, the restatoms in both halves of the unit cell and inside the corner hole are blue. Notice the stacking fault in the right side of the unit cell (see also (*e*)). (*b*) Empty-state topograph. (*c*) Adatom state, between E_F and −0.35 eV. (*d*) Restatom state between −0.6 and −1 eV. (*e*) Schematic cross section of surface.

directly on an optical disc. The data can be analysed either on-line or off-line on a larger mainframe computer.

5. Discussion

In the above we have reviewed some ways to obtain spectroscopic information with the STM, focusing primarily on the various ways researchers have tried to obtain reliable data. I think that everyone would agree that the best way to measure an I–V curve (or one of its derivatives) is to break the feedback circuit and then take the data. This requires that the microscope is stable during the period that the feedback circuit is open. This is usually not a problem.

Separating geometric from electronic information when one tries to obtain real-space images of surface states is somewhat more problematic. I believe that the differential CITS method, used to obtain real-space images of the Si(1 1 1)-(7 × 7) surface states as shown in the previous section, is better than any other method. It may not be ideal, in the sense that it may not be able to separate completely electronic and geometric structure information if the electronic structure information is weak. The method is definitely better than just measuring constant-current topographs at different bias voltages, because such images necessarily 'integrate' between 0 V and the feedback voltage. The CITS technique allows one to take a slice from the density of states between two arbitrary voltages and is therefore much more versatile.

In the last two years or so many researchers have taken a hybrid approach: they measure topographic images at one or more feedback voltages and supplement this with detailed d ln I/d ln V curves at specific locations. In this way one does not obtain real-space images of the surface electronic structure, but one does get the spatially resolved electronic data needed to answer most of the relevant questions. There are two questions remaining now, begging to be asked and answered. The first question is: Is the electronic structure of the tip not important? The answer is *yes*. Lang's theory shows clearly that the experiment does not know which is tip and which is sample and all wavefunctions are equal. It was recently shown how the presence of impurities on the tip may gravely affect the experimental results [8]. However, on well prepared, clean and stable metal tips (such as W or PtIr), the tip wavefunctions are apparently sufficiently featureless so as to be undiscernible in the data. The most important rule, then, is to reproduce all data, first on the same sample with the same tip, and secondly on different samples with different tips. During a number of experiments it will become clear which results form the common denominator and what is due to such things as double, triple or multiple tips and tip contaminants.

The second and most crucial question of all is: Suppose I have done everything right, and after much hard work I have obtained all the atomically resolved spectroscopic information that can be had from my carefully prepared sample. Now, is there a sure and simple way to understand what I have? The disappointing answer is, of course, *no*.

Figure 11 shows an image of the Si(1 1 1)-($\sqrt{3} \times \sqrt{3}$)Ag surface obtained at a bias voltage of −1 V [21]. The unit cell is outlined and contains two maxima, with the maxima arranged in a hexagonal network. The question one would like to answer is simple: Are these maxima due to Ag atoms or to Si atoms? When we first published this image we argued on spectroscopic grounds that these have to be Si atoms [21] . The argument is as follows. The surface is semiconducting, i.e. the surface states are all fully occupied. If the two maxima correspond to Ag atoms, then the unit cell has three Si dangling

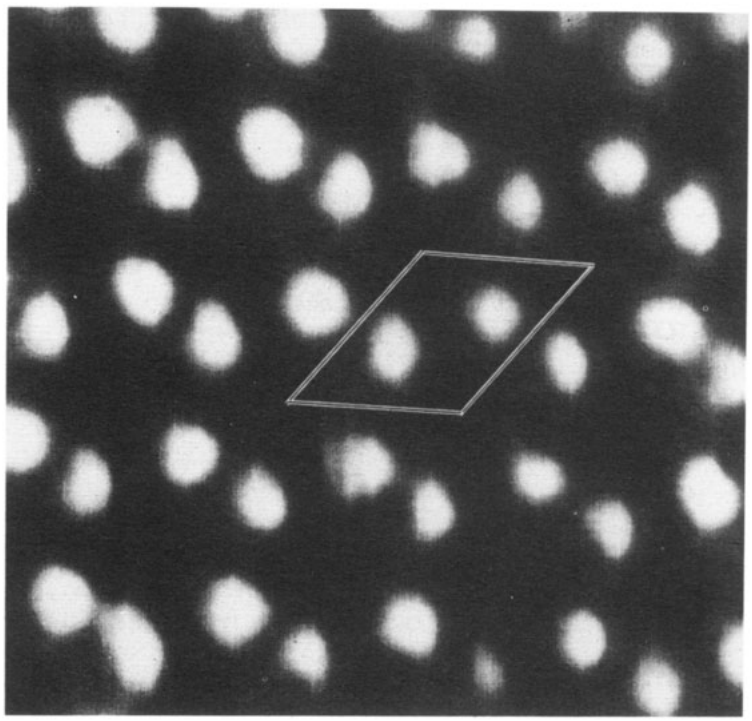

Figure 11. Filled-state topograph of the Si(1 1 1)-($\sqrt{3} \times \sqrt{3}$)Ag surface.

bonds, which may hybridise with the two Ag valence electrons, which leads to an odd number of electrons, which in turn must lead to a metallic surface because at least one electron per ($\sqrt{3} \times \sqrt{3}$) unit cell will remain unpaired. Therefore there must be three Ag atoms per unit cell, which are probably subsurface and invisible to the STM, and the two bumps we see are Si atoms. In the same issue of *Physical Review Letters* our colleagues at the IBM Almaden Research Center argued on the basis of essentially identical data that the maxima must be Ag atoms [22].

The simple fact is, no matter how many *I–V* curves one measures, the STM is not going to tell us whether the bumps are Ag or Si. Both x-ray diffraction results [23] and medium-energy ion scattering [24] results have since shown that the unit cell contains three Ag atoms, confirming the spectroscopic argument given above. In the structure consistent with the x-ray diffraction data the STM maxima correspond with Si atoms. In order to understand fully the STM results, a calculation of the electronic structure is needed, with the correct geometric structure as input. The x-ray diffraction and ion scattering results have also shown that the Si substrate is reconstructed very extensively, with the outer two double layers of the crystal severely distorted. Such an electronic structure calculation may not be forthcoming very soon.

5.1. STM as a crystallographic tool

The value of STM as a crystallographic tool is very apparent if one looks at the images shown in this paper. In one glance one obtains a detailed picture of the symmetries of

the surface unit cell. However, to claim that the classical surface crystallographic tools (such as low-energy electron diffraction (LEED), medium-energy ion scattering (MEIS), x-ray scattering—to mention just a few) have been rendered obsolete by the scanning tunnelling microscope would be a mistake. As explained above, the STM senses primarily *electronic* information about the surface, with high lateral resolution. However, it senses only those electronic states that have significant overlap with the tip, i.e. the tails of the wavefunction, protruding 5 Å or more into vacuum. In many cases (and in particular on many reconstructed semiconductor surfaces) much of the interesting structure is *below* the surface, where the STM does not see. In addition, one has to be careful that the apparent location of an electronic orbital is not the same as the location of the atomic core. The Si(0 0 1)-(2 × 1) surface, imaged at +1 eV, looks just like a (1 × 1) surface [20], because the empty orbitals stick out sideways from the dimers like rabbit ears. So, STM images may be helpful in determining the structure of a surface (as it was in the case of the Si(1 1 1)-(7 × 7) surface), but no single structure has been determined completely by the use of STM only.

5.2. *STM as an electron spectroscopy*

This paper deals with the contribution of the electronic structure of the sample surface (and of the tip) to the STM experiment. It is fair to say that STM is first and foremost a spectroscopic technique with unprecedented spatial resolution. However, also here one must be fully aware of its limitations. In measuring an I–V curve between −2 and +2 V in a couple of milliseconds, it is natural to think that one is performing the equivalent of a combined photoemission and inverse photoemission experiment in one broad sweep, with a much simpler instrument. The first important difference between the conventional spectroscopies and STM is that STM sees only those electronic states that protrude into the vacuum and overlap with the tip wavefunctions. Electronic states localised between, say, the first and second layer of a sample (such as for example the bonding orbitals between a simple adsorbate—like CO— and a surface) are invisible. In fact, it is taken for granted that many adsorbates go undetected in STM, or are detected only indirectly through the disappearance or alteration of the electron states characteristic of the clean sample (adsorption of H on Si(1 1 1) is a good example) [14]. Thus, there is a large and practically important class of spectroscopic features that the STM does not see. Secondly, there are those electronic states that are localised too close to the core. Attempts have been made to observe the d electrons in for instance Ag [21]. In photoemission the d bands are often the strongest feature in the spectrum and allow chemical identification of the adsorbate. STM, unfortunately, is blind to these important orbitals. As a result, attempts to make a direct chemical identification of an adsorbate using STM have been unsuccessful so far. This does not imply that in specific cases it may not be possible to assign a certain bump in an STM image, or a peak in a $\mathrm{d}\ln I/\mathrm{d}\ln V$ spectrum, to an adsorbate that has been carefully introduced, but no such assignments are automatic. Consider the following thought experiment. We take a clean surface and obtain an image. Next we adsorb oxygen on the surface and obtain a second image. The second image reveals a number of bumps that were not observed on the clean surface. Are these bumps the O atoms? It may be safe to say that the new bumps are induced by the adsorption of O. But are they O? The O may go subsurface and modify the wavefunctions of the surface atoms to which it bonds. A real example is the beautiful work by Stroscio *et al* on O adsorption on GaAs [16]. A wealth of phenomena is observed as a result of O adsorption. There is band bending around the adsorption site, which depends on the

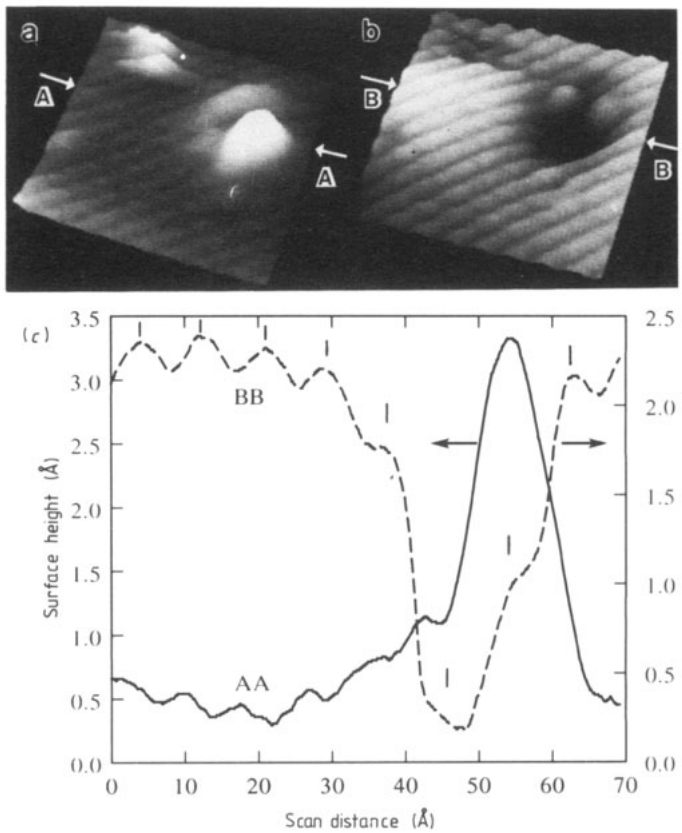

Figure 12. (*a*) Filled-state and (*b*) empty-state topographs of an oxygen-induced defect on GaAs(1 1 0). (*c*) Surface height contours along AA and BB. Positions on the Ga lattice are shown by the tick marks.

doping of the substrate. An atomic size protrusion (or depression, depending on the bias voltage) is seen at the adsorption site (see figure 12). Is this protrusion the O atom? Maybe. It may also be a Ga atom or an As atom. The adsorbed O atom may give rise to an antisite defect. It may release a Ga or an As atom from the surface, which diffuses somewhere else. There are probably other possibilities. The point is, one cannot tell. One way out of this dilemma is to call theory to the rescue. If one has a clue about the detailed atomic structure one can use band-structure theory to calculate the electronic structure and compare this with the experiment. This may resolve the issue. An example is the adsorption of Al on Si(1 1 1), where such calculations are available [13]. Another example is the adsorption of B on Si(1 1 1), also in a $(\sqrt{3} \times \sqrt{3})$ lattice. Again one observes bumps, as in the Al case [25, 26]. But, are the bumps B? It turns out that the bumps are not B but Si. The B is hiding below the surface, just underneath the Si, if the surface is prepared in a certain way. Without the help of theory it would be difficult to figure out what is going on in this case.

One of the most important features in angle-resolved photoemission and inverse photoemission experiments is the ability to determine the band structure of the surface under study. The dispersion of the electron states when plotted as a function of wavevector parallel to the surface is one of the most powerful aspects of these techniques,

allowing detailed comparison between experiments and band structures calculated for atomic models of the surface. STM is more like angle-integrated photoemission, and has no ability to resolve the wavevector of the electron in its initial state. A dispersion surface state with high density of states at the centre and the boundary of the Brillouin zone may give rise to two peaks in the tunnelling signal. States with $k_{\|} = 0$ decay slowest into vacuum and give rise to a stronger signal. However, in the d ln I/d ln V curves we just see one or more peaks, without any information about the tunnelling transmission probability or the initial-state wavevector. With additional information from angle-resolved photoemission experiments one may be able to make more definitive assignments of the initial-state wavevector, but tunnelling spectroscopy by itself does not provide such information.

6. Conclusions

One may conclude from the above that the strongest point of scanning tunnelling microscopy, its capability to perform spectroscopic studies of the wavefunctions protruding far into the vacuum with high spatial resolution, is at the same time its severest limitation. It is incredibly exciting to see atomically resolved images of an elusive surface structure appear on the computer monitor, as if by magic. On the other hand, it is a sobering experience to realise that the beauty of the image is sometimes more apparent than its meaning. Scanning tunnelling microscopy and spectroscopy are still in their infancy. While some of the problems discussed in this paper may be resolved in the future, at least some of them appear to be of a fundamental nature. In spite of these limitations, STM has proved to be one of the most exciting and revealing surface science techniques invented in the 1980s. There can be little doubt that STM will continue to make many unique and important contributions to our understanding of the physics and chemistry of surfaces in the years to come.

Acknowledgments

I am deeply indebted to Joe Demuth, Bob Hamers and Evert van Loenen with whom I have collaborated in a number of STM experiments. I have also learned a lot from discussions with Norton Lang and Randy Feenstra.

References

[1] Binnig B and Rohrer H 1986 *IBM J. Res. Dev.* **30** 355
[2] Kaiser W J and Jaklevic R C 1986 *IBM J. Res. Dev.* **30** 411
[3] Tromp R M, Hamers R J and Demuth J E 1986 *Science* **234** 304 and references therein
[4] Tromp R M, Hamers R J and Demuth J E 1986 *Phys. Rev.* B **34** 1388
[5] Tersoff J and Hamann D R 1983 *Phys. Rev. Lett.* **50** 1998; 1985 *Phys. Rev.* B **31** 805
[6] Hamers R J 1989 *Ann. Rev. Phys. Chem.* **40** 531
[7] Lang N D 1989 *Comment Condens. Matter. Phys.* **14** 253 and references therein
[8] Tromp R M, van Loenen E J, Demuth J E and Lang N D 1988 *Phys. Rev.* B **37** 9042
[9] Becker R S, Golovchenko J A, Hamann D R and Swartzentruber B S 1985 *Phys. Rev. Lett.* **55** 987
[10] Binnig G, Frank K H, Fuchs H, Garcia N, Reihl B, Rohrer H, Salvan F and Williams A R 1985 *Phys. Rev. Lett.* **55** 991
[11] Becker R S, Golovchenko J A, Hamann D R and Swartzentruber B S 1985 *Phys. Rev. Lett.* **55** 2032

[12] Feenstra R M, Thompson W A and Fein A P 1986 *Phys. Rev. Lett.* **56** 608; Stroscio J A, Feenstra R M and Fein A P 1986 *Phys. Rev. Lett.* **57** 2579
[13] Hamers R J and Demuth J E 1988 *Phys. Rev. Lett.* **60** 2527
[14] Avouris Ph and Wolkow R 1989 *Phys. Rev.* B **39** 5091
[15] Feenstra R M, Stroscio J A, Tersoff J and Fein A P 1987 *Phys. Rev. Lett.* **58** 1192
[16] Stroscio J A, Feenstra R M and Fein A P 1987 *Phys. Rev. Lett.* **58** 1668
[17] Coleman R V, Drake B, Hansma P K and Slough G 1985 *Phys. Rev. Lett.* **55** 394
[18] Hamers R J, Tromp R M and Demuth J E 1986 *Phys. Rev. Lett.* **56** 1972
[19] Himpsel F J and Fauster Th 1984 *J. Vac. Sci. Technol.* A **2** 815
[20] Hamers R J, Tromp R M and Demuth J E 1986 *Phys. Rev.* B **34** 5343
[21] van Loenen E J, Demuth J E, Tromp R M and Hamers R J 1987 *Phys. Rev. Lett.* **58** 373
[22] Wilson R J and Chiang S 1987 *Phys. Rev. Lett.* **58** 369
[23] Vlieg E, van der Gon A W, van der Veen J F, Macdonald J E and Norris C 1989 *Surf. Sci.* **209** 100
[24] Copel M and Tromp R M 1989 *Phys. Rev.* B **39** 12688
[25] Bedrossian P, Meade R D, Mortensen K, Chen D M, Golovchenko J A and Vanderbilt D to be published
[26] Avouris Ph to be published