PHYSICAL REVIEW B VOLUME 39, NUMBER 2 15 JANUARY 1989-I

Sample-dependent resolution in scanning tunneling microscopy

J. Tersoff
IBM Research Division, Thomas J. Watson Research Center, Yorktown Heights, New York 10598

(Received 21 July 1988)

In any microscopy or spectroscopy, quantitative analysis of data requires some knowledge of the instrumental resolution function. Analysis in fact often involves image deconvolution, at least implicitly. However, for scanning tunneling microscopy (STM), the very definition of resolution becomes problematic and has until now been addressed only for metals, the simplest case. Here a natural general expression for the resolution of STM is developed. This definition gives a resolution function which may be strongly sample dependent in the case of semiconductors or semimetals, and which has an anomalous line shape and linewidth in certain cases. Thus image deconvolution is not generally possible, even in principle, without an understanding of the sample electronic structure.

I. INTRODUCTION

In any microscopy or spectroscopy, the image or spectrum is viewed as an "ideal" image containing the desired information, convoluted with an instrumental resolution function or line shape. The resolution is therefore one of the most crucial properties characterizing the instrument. Quantitative analysis of data can depend critically on knowledge of the instrumental resolution function. Under favorable circumstances, it is even possible to deconvolute the image, reducing or eliminating the distortion associated with the instrumental resolution. In addition, the first step in improving the instrumental resolution is understanding the factors which determine the resolution.

The purpose of this paper is to present a unified discussion of the resolution of scanning tunneling microscopy (STM), in which the very different cases of metals[1-3] (normal resolution) and certain semiconductors and semimetals[4] (anomalous resolution) are seen to be particular instances of the general *sample-dependent* resolution. In particular, the case of anomalous resolution, first mentioned in Ref. 4, is here discussed in detail for the first time. In order to isolate the role of the sample in determining the effective resolution of STM, we restrict ourselves to a simple model for the tip[1,2] throughout most of this paper.

STM itself has been described in detail by its inventors.[5] In the "constant current" mode, the operating voltage between surface and tip is fixed, and the height of the tip is constantly adjusted so that the tunneling current remains equal to some specified value as the tip is scanned laterally along the surface. While other modes of imaging exist, this mode is by far the most generally useful, since in certain limits the path of the tip corresponds directly to a topograph of the surface.

The central result here is that, with the most natural definition of the resolution function, the resolution depends critically upon the electronic structure of the sample. If states from throughout the surface Brillouin zone (SBZ) contribute to the tunnel current, as for simple or noble metals, then the resolution assumes a well-defined limit[1-3] which is approximately sample independent, and which corresponds closely to one's intuition. However, if only states from a small part of the SBZ contribute to the current, as is often the case for semiconductor surfaces at low tunnel voltage, then the resolution function can exhibit very anomalous behavior.

The most extreme situation arises if only states near the edge of the SBZ contribute to the current. In that case there is a striking apparent enhancement of the ability to resolve the unit cell. However, the resolution for features within the unit cell is not enhanced. Thus the "line shape" of the instrument in this case depends upon the lattice constant of the surface.

This very peculiar situation is perhaps best appreciated by considering the surface as imaging the tip, instead of vice versa. While the two perspectives must yield identical predictions for the image, the former view makes it easy to appreciate how the apparent instrumental line shape could depend on the surface periodicity and electronic structure. From that viewpoint, perhaps the remarkable thing is that, for metals, this dependence conveniently disappears.

II. DEFINING THE RESOLUTION OF STM

A. Resolution of linear measurements

Defining the resolution of STM raises tricky issues for two reasons. First, STM is inherently nonlinear, so the usual definition of resolution in terms of convolution with an instrumental function cannot be applied directly. Second, resolution can only be defined relative to what the instrument should ideally measure. In many other microscopies and spectroscopies, one knows exactly what the instrument is intended to measure, and the resolution is a measure of the deviation from this ideal. In STM, the ideal image is not so well defined, making the definition of the resolution correspondingly ambiguous. Choosing a convention for the "ideal" STM image is therefore an unavoidable first step in defining the resolution.

More formally, one defines resolution by assuming that there exists some ideal image or spectrum $I_0(x)$, which would be seen in the case of perfect instrumental resolution. The variable x here may represent one or more dimensions of space, time, energy, etc., depending upon the microscopy or spectroscopy in question. The actual measured image $I(x)$ is then

$$I(x)=\int I_0(x-y)F(y)dy \ . \tag{1}$$

Equation (1) should be viewed as the definition of the resolution function $F(x)$.

The resolution [i.e., the width of $F(x)$] should not be equated with the size of the smallest distinguishable feature in the image. The latter size is actually determined by the signal-to-noise ratio, as discussed elsewhere.[6] In principle, if $F(x)$ is known, and in the absence of noise, the actual image $I(x)$ contains no less information than the ideal image I_0. However, the convolution (1) tends to suppress the components of $I(x)$ with high spatial frequency, and even a little noise or uncertainty in $F(x)$ can then render deconvolution impossible, so that information is truly lost.

It is often convenient to Fourier transform (1) to obtain

$$\tilde{I}(q)=\tilde{I}_0(q)\tilde{F}(q) \ . \tag{2}$$

Here $\tilde{I}(q)$ and $\tilde{F}(q)$ are the Fourier transforms of $I(x)$ and $F(x)$, e.g.,

$$\tilde{I}(q)=\int I(x)\exp(iqx)dx \ ,$$

neglecting normalization. An obvious advantage of (2) is that $\tilde{F}(q)$ can be determined directly as $\tilde{I}(q)/\tilde{I}_0(q)$.

It is important to recognize that $F(x)$ is a well-defined function, independent of the specific image, only if the measured image I (the "output") is a *linear* function of the ideal image I_0 (the "input"). While the meaning of the ideal image I_0 may be obvious in the case of an optical microscope, it is not so for STM.

The task here is to determine $F(x)$. To do so, however, we must first identify a natural choice for the definition of I_0. Moreover, it is essential to make this choice in such a way that the linearity condition mentioned above applies.

B. Form of the STM image

We now consider the form of the STM image, in order to motivate a choice of I_0 and to set the groundwork for evaluating the resolution function F. In general, the resolution must depend on the tip shape and electronic structure. For example, several authors[7,8] have presented evidence that experimental STM images sometimes are grossly distorted by the tip shape, in particular when there are multiple distinct tunneling sites (in effect, multiple tips).

Here we are primarily concerned with the dependence of the resolution on the sample, so for simplicity we restrict ourselves to the idealized tip model of Tersoff and Hamann,[1,2] in which tunneling matrix elements are approximated by those for an s-wave tip wave function. In addition, we neglect such extrinsic factors as finite instrumental response time, vibrational noise, and mechanical interactions between the tip and surface. (It should, however, be stressed that these extrinsic effects may often be the factors which actually limit the STM resolution in practice.)

The resulting image corresponds to a contour of constant surface local density of states $\rho(\mathbf{r},E_F)$, where

$$\rho(\mathbf{r},E)\equiv \sum_{\nu,k_\parallel} |\psi_{\nu,k_\parallel}(\mathbf{r})|^2\delta(E-E_{\nu,k_\parallel}) \ . \tag{3}$$

More specifically, the center of curvature of the tip follows one of these contours, the particular contour depending upon the voltage, current, and other factors. Here $E_{\nu,k_\parallel}$ is the energy of eigenstate $\psi_{\nu,k_\parallel}$, $k_\parallel$ is the surface wave vector (in the case of a periodic surface), and the index ν runs over the remaining quantum numbers.

The approximations involved in this model are expected to be rather accurate in most cases, especially for monatomic tips.[2] Moreover, the model has proven adequate for the quantitative interpretation of STM images,[1,9] and has been tested by comparison with more exact calculations in simple cases.[10] We therefore accept without further discussion that the STM image does in fact correspond to a contour of $\rho(\mathbf{r},E_F)$, to sufficient accuracy for the present purpose.

For finite voltage, assuming a structureless density of states for the tip, it is merely necessary to integrate $\rho(\mathbf{r},E)$ over the appropriate range of energy. In principle, ψ should be calculated in the presence of the electric field, but for small voltages this complication can be neglected in practice.[11] Thus while, for convenience, we will generally discuss tunneling at small voltage, the discussion and results carry over directly to imaging at larger voltages.

Let us rewrite $\rho(\mathbf{r},E_F)$ (or the corresponding integral in the case of finite voltage) as $\rho(\mathbf{x},z)$, where we separate lateral and vertical position as $\mathbf{r}=(\mathbf{x},z)$, suppressing the energy argument for notational simplicity. The STM image $z(\mathbf{x})$ in the constant-current (topographic) imaging mode is implicitly defined by

$$\rho(\mathbf{x},z)=\rho_T \ , \tag{4}$$

where ρ_T is proportional to the tunneling current at which the microscope is operated.[2]

Unfortunately, this imaging process is inherently nonlinear, whereas resolution is only well defined for a linear measurement. It is therefore useful to work in the limit of weak corrugation, so that the imaging process can be linearized. We then write

$$z=z_0+\zeta(\mathbf{x},z_0) \ , \tag{5}$$

where z_0 is an average tip height defined below, which depends on ρ_T but is not easily accessible experimentally, and ζ is the small corrugation which constitutes the image, and which depends on the tip height z_0 or equivalently on ρ_T. Expanding $\rho(\mathbf{x},z)$ about $z=z_0$, (4) and (5) give

$$\zeta(\mathbf{x},z_0)\simeq[\rho_T-\rho(\mathbf{x},z_0)]\Big/\frac{d}{dz}\rho(\mathbf{x},z_0) \ . \tag{6}$$

Because of the approximately exponential decay of the wave function, for weak corrugation one can write[2]

$$d\rho(\mathbf{x},z_0)/dz\simeq-\rho(\mathbf{x},z_0)/\lambda \ . \tag{7}$$

The decay length λ is discussed below, as is the range of validity of this expression. Also, we define the characteristic tip height z_0 by the condition that the lateral average of $\rho(\mathbf{x},z_0)$ is

$$A^{-1}\int \rho(\mathbf{x},z_0)d\mathbf{x}\equiv\tilde{\rho}_0(z_0)=\rho_T \;, \tag{8}$$

where A is the area of integration. Thus z_0 represents (within the model of Ref. 2) the average height of the *effective center of curvature* of the tip.

To lowest order in the small quantity $[\rho(\mathbf{x},z_0)-\rho_T]/\rho_T$ one can then write

$$\zeta(\mathbf{x},z_0)\simeq\lambda[\rho(\mathbf{x},z_0)-\rho_T]/\rho_T \;. \tag{9}$$

Equation (9) is a very simple but central result. It states that, in the limit of weak corrugation, the image $\zeta(\mathbf{x},z_0)$ is simply proportional to the fractional variation of $\rho(\mathbf{x},z_0)$ about its mean value in the plane $z=z_0$.

In analogy with (9), we therefore propose to identify I_0 with the fractional variation of $\rho(\mathbf{x},z_0)$ about its mean value, evaluated in some *reference plane* which is taken here as the origin. Thus

$$I_0(\mathbf{x})\equiv\lambda[\rho(\mathbf{x},0)-\tilde{\rho}_0(0)]/\tilde{\rho}_0(0) \;. \tag{10}$$

Note that this convention does not require that the corrugation be weak in the reference plane. The implicit dependence of λ on z due to the spatial variation of the potential is neglected, since this variation is actually rather weak except very near the surface.

There remains an ambiguity in the choice of the reference plane $z=0$. A reasonable choice might be the average classical turning point for electrons at the Fermi level, or a point slightly further away from the surface. The treatment here is most rigorous when the plane is chosen relatively far from the surface. However, this gain in rigor is balanced by a loss of substantive content, since more of the smoothing of the wave functions is then included in I_0 instead of in the resolution function, where it intuitively belongs. In the limit that the reference plane is taken as the plane z_0 in which the tip moves, the nominal resolution becomes perfect (within the present model for the tip), merely because the convention for the ideal image corresponds to the actual image. Conversely, if the reference plane is chosen very near the surface, then I_0 will reflect topographic features of the surface in fine detail.

C. Form of the resolution function

It is now convenient to assume that the surface is periodic, and to work with the Fourier-transformed quantities. The periodicity can later be taken to be arbitrarily large, to include nonperiodic surfaces. We then write

$$\rho(\mathbf{x},z)=\sum_{G}\tilde{\rho}_G(z)\exp(i\mathbf{G}\cdot\mathbf{x}) \;, \tag{11}$$

where G denotes the *surface* reciprocal lattice vectors, a tilde indicates a reciprocal-space quantity, and $\tilde{\rho}_0$ was defined above to be just the $G=0$ term of (11). Now (9) and (10) may be rewritten as

$$\tilde{\zeta}_G(z_0)\simeq\lambda\tilde{\rho}_G(z_0)/\tilde{\rho}_0(z_0) \;, \tag{12}$$

$$I_0(G)=\lambda\tilde{\rho}_G(0)/\tilde{\rho}_0(0) \;. \tag{13}$$

Combining (12) and (13) with (2) gives the desired expression for the resolution function:

$$F(G)\simeq\tilde{\rho}_G(z_0)\tilde{\rho}_0(0)/\tilde{\rho}_0(z_0)\tilde{\rho}_G(0) \;. \tag{14}$$

This formula may, of course, be Fourier transformed to give an explicit line shape $F(\mathbf{x})$.

Equation (14) is a central result here, providing an explicit statement of how the resolution of STM depends both on the tip height, and on the sample electronic structure. Any more specific statement requires detailed knowledge of the specific sample, i.e., of $\tilde{\rho}_G(z)$.

While (14) was derived under the assumption of weak corrugation, this restriction represents an unavoidable limitation, reflecting the limited applicability of the very concept of resolution to the highly nonlinear STM measurement. For practical purposes, (14) can no doubt be usefully applied well beyond the range of strict validity, as a working definition of resolution. Additionally, if the corrugation ζ is weak even in the reference plane, then (14) can be rewritten entirely in terms of observable images,

$$F(G)\simeq\tilde{\zeta}_G(z_0)/\tilde{\zeta}_G(0) \;.$$

D. Role of the tip

The STM tip can, of course, have a major effect on the image. This can be included in a straightforward way into the definition of the resolution. Specifically, from (2), $\tilde{F}(q)=\tilde{I}(q)/\tilde{I}_0(q)$. If we take $\tilde{I}(q)$ as the image obtained with the actual tip, but continue to define the ideal image $\tilde{I}_0(q)$ as that for a point tip via Eq. (4), then in analogy with (14) one obtains

$$F(G)\simeq\tilde{j}_G(z_0)\tilde{\rho}_0(0)/\tilde{j}_0(z_0)\tilde{\rho}_G(0) \;, \tag{15}$$

where $\tilde{j}_G$ is the two-dimensional Fourier transform of $j(x,z)$, the actual tunneling current (for the given tip) as a function of tip position. However, it is difficult to go further without specific assumptions about the tip wave functions.

III. EVALUATING THE RESOLUTION FUNCTION

A. General approach

In order to evaluate the resolution (14) in any specific case, one must in principle know the surface local density of states (3), i.e., $\rho_G(z)$. This can, of course, be computed numerically, at least in principle. In order to draw somewhat more general conclusions, however, we follow Ref. 2 in making use of the known asymptotic behavior of the wave functions. Specifically, by neglecting the variation in the potential $V(\mathbf{r})$ over the region of interest, the wave functions can be expanded in generalized (complex) plane waves, i.e.,

$$\psi_{k_\parallel}=\sum_{G}a(k_\parallel,\mathbf{G})\exp[i(\mathbf{k}_\parallel+\mathbf{G})\cdot\mathbf{x}]\exp(-\kappa_G z) \;, \tag{16}$$

where $\kappa_G^2=|\mathbf{k}_\parallel+\mathbf{G}|^2+\kappa^2$, and $\hbar^2\kappa^2/2m=V-E_{k_\parallel}$. Here V is the average potential in the region of interest, and the index ν of (3) has been suppressed for brevity. One can then combine (3) and (16) to express $\tilde{\rho}_G(z)$ in terms of products of such complex plane waves. In particular, from (16) one can, following Ref. 2, identify λ in (7) as 2κ. Moreover, one finds that (7) is accurate so long as either the corrugation is weak ($\zeta \ll \lambda$), or the lateral length scale of interest is long ($G \ll 2\kappa$).

B. Results for metallic surfaces

In principle, a precise evaluation of the resolution from (14) requires a detailed knowledge of the surface electronic structure. However, for metals, it has previously been noted[1-3] that one can derive some results concerning the resolution from fairly general considerations. In particular, an ansatz based on the superposition of atomiclike densities gives a description which corresponds remarkably well the exact asymptotic behavior.[2,12] These previous results are combined and restated here for completeness.

The conclusion to be drawn from from the discussion below is that for metals, the instrumental resolution for STM may be adequately described by a Gaussian line shape. Thus the effect of finite resolution in this case corresponds closely to one's intuitive expectations, and so the interpretation of images is generally not complicated by the sort of subtleties discussed later for the case of semiconducting surfaces.

Consider a spherical atomiclike charge density ρ_a of the form

$$\rho_a(r)=C\exp(-2\kappa r)/r \ .$$

(The denominator is for analytical convenience only,[2] since the behavior at large r is dominated by the exponential.) We imagine that $\rho(\mathbf{r},E_F)$ can be mimicked by a sum of such atoms, perhaps with a different coefficient C for each inequivalent atom. Then an arbitrary sum of such "atoms," restricted to the half-space $z<0$, gives a charge density which in the half-space $z>0$ can be Fourier transformed to yield[2]

$$\tilde{\rho}_G(z)=B_G\exp[-(4\kappa^2+G^2)^{1/2}z] \ , \qquad (17)$$

where B is a constant for each G, depending only on κ and on the arrangement of atoms, and the values of their coefficients C. [For the present discussion, it is actually unimportant whether realistic values of B_G can be obtained by superposition of spherical charges; our concern is simply to motivate the model form (17) of the exponential dependence[2,12] on z and G.]

Combining (14) and (17), one may immediately obtain an explicit model form for the resolution function,

$$\tilde{F}(G)\simeq\exp[2\kappa z_0-(4\kappa^2+G^2)^{1/2}z_0] \ . \qquad (18)$$

Since κ is proportional to the square root of the work function, which itself does not vary too much among metals of interest here, (18) represents a resolution function which is nearly independent of the specific sample.

The "justification" for this ansatz is relatively simple.[2,12] It can be shown[2] that (a) the term $\tilde{\rho}_0(z)$ is asymptotically dominated by states near $k_\parallel=0$ (if such states exist at the relevant energy), giving a behavior $\rho_0(z)=B_0\exp(-2\kappa z)$; (b) the term $\tilde{\rho}_g(z)$, where g is the smallest G, is asymptotically dominated by states near the SBZ edge, $k_\parallel=g/2$, giving a behavior

$$\tilde{\rho}_g(z)=B_g\exp[-(4\kappa^2+g^2)^{1/2}z] \ ;$$

and (c) the terms $\tilde{\rho}_G(z)$, where $G\gg g$, are approximately independent of $k_\parallel$, with a behavior

$$\tilde{\rho}_G(z)=B_G\exp[-(4\kappa^2+G^2)^{1/2}z] \ .$$

Thus (17) is asymptotically exact for the large-G components of ρ_G, regardless of the detailed electronic structure, and also for the lowest two G components ($G=0$ and $G=g$) if both $k_\parallel=0$ and $k_\parallel=g$ contribute to the charge density from states at E_F (or in the energy range relevant for the specified voltage).

In general, for metals the observable G components obey $G\ll 2\kappa$, i.e., typically $G\le 0.8$ Å^{-1}, while $2\kappa\sim 2$ Å^{-1}. (There may, however, be cases where this is less well obeyed.[13]) Then (18) may be expanded as

$$\tilde{F}(G)\simeq\exp(-G^2z_0/4\kappa) \ . \qquad (19)$$

This simplified model resolution function is simply a Gaussian with rms width $(z_0/2\kappa)^{1/2}$.

It is worth stressing that the approximate applicability of (17) only requires that a substantial range of $k_\parallel$ contribute to the tunneling current. For example, the "most typical" k point (in the sense of "special point" schemes) is $k_\parallel=g/4$. That point alone contributes[12] terms which go asymptotically as

$$\exp[-2(\kappa^2+g^2/16)^{1/2}z]$$

and

$$\exp\{-[\kappa+(\kappa^2+9g^2/16)^{1/2}]z\}$$

for $\tilde{\rho}_0$ and $\tilde{\rho}_g$, respectively. For $G\ll 2\kappa$ (generally valid for observable features on metals), these decay constants are a good approximation to the asymptotic values from states at the zone center and edge. The decay constant for $\tilde{\rho}_0$ differs from the $k_\parallel=0$ value by only $g^2/16\kappa$, and that for $\tilde{\rho}_g$ differs from the $k_\parallel=g/2$ value by only $g^2/32\kappa$. Thus that one point alone gives very similar asymptotic behavior to the full SBZ sampling for all G components. If, for example, there are no states near $k_\parallel=0$ which contribute to the tunneling current, or none near $g/2$, the characteristic metallic behavior (18) will hardly be affected.

C. Results for semiconducting and semimetallic surfaces

Equation (18) gives an explicit result for the resolution function, i.e., the instrumental line shape, for the case of STM of metal surfaces. In the analysis below, we show that (18) is approximately valid in general, even for semiconducting surfaces, *except* for the lowest Fourier component of the image. However, the lowest component $F(g)$ (where g is the smallest reciprocal lattice vector) may in general have a value which deviates grossly from

(18), and which depends sensitively upon the electronic structure of the surface.

This effect distorts the image in a manner which is simple, but which can have profound consequences for the interpretation of STM images. For example, in the extreme case that $F(g)$ is greatly enhanced relative to the higher components, the image will have a simple sinusoidal[14] variation, with the periodicity of the surface, giving the misleading appearance of a single topographic feature per unit cell. This effect has been observed[9,15,16] on Si(111)-(2×1), GaAs(110), and $1T$-TaS_2. Fortunately, in each case a prior knowledge of the electronic structure permitted a correct interpretation.

We now proceed with the analysis of STM resolution for semiconducting surfaces. The results for metal surfaces, Eqs. (17)–(19), are based on the assumption that the entire SBZ (or at least some "typical" $k_{\parallel}$) contributes to the tunneling. This assumption is certainly appropriate for metals, and in addition it may often apply well to semiconductors, especially at moderately large tunneling voltage or for large unit cells or disordered surfaces. Moreover, the assumption is not critical when the structure of interest is well localized within the unit cell ($G \gg g$). Thus Tromp *et al.*[17] were successful in interpreting images of Si(111)-(7×7) based on this atom-superposition ansatz.

In contrast, in STM of semiconducting surfaces at the lowest possible tunneling voltage, tunneling takes place into or out of states at the band edge. In that case only one or a few states (or pockets of states) contribute to (3). Consider, in particular, tunneling to states which are quasi-two-dimensional. These could be either surface states, as for Si(111)-(2×1),[15] or states of a quasi-two-dimensional material such as $1T$-TaS_2 or graphite.[4] Then for semiconductors (and often for semimetals), the states at the band edge (or Fermi level) generally fall at either the center or edge of the surface Brillouin zone. The case of tunneling to zone-edge states is particularly interesting, and has been discussed in Ref. 4.

For tunneling to states at the edge of the surface Brillouin zone, $k_{\parallel}=g/2$, where g is again the smallest G. Substituting this into (16), and assuming reflection symmetry, one finds that

$$\psi = \sin(k_{\parallel}x)\exp[-(\kappa^2+k_{\parallel}^2)^{1/2}z]$$

plus higher Fourier components.

Substituting into (3) gives

$$\tilde{\rho}_0 = 2\tilde{\rho}_g = \tfrac{1}{2}\exp[-2(\kappa^2+k_{\parallel}^2)^{1/2}z] \; .$$

Thus the g Fourier component decays no more rapidly than the zero Fourier component, and substituting into (14) gives $F(g)=1$, independent of z. For $G>g$, however, $\tilde{\rho}_G$ decays faster with increasing G, as expected, so for $G>g$, $\tilde{F}(G)$ still decreases with increasing z as for a metal. In fact, it is easy to verify explicitly that (18) is still valid even for the second-lowest Fourier component $F(2g)$ in this case, at least for $G \ll 2\kappa$.

This is a very peculiar result. It implies that, for large z (i.e., large tunneling distance or tip radius), the ability to resolve structure within the unit cell decreases and is lost, just as for a metal surface, since this structure corresponds to $G>g$. But because $F(g)=1$, the unit cell itself is well resolved even if it is very small (large g), and even if the tip is relatively blunt or far from the surface, as long as the model of Ref. 2 is applicable.

This enhanced resolution of the unit cell is particularly striking for graphite,[4] where the 2-Å unit cell is easily resolved, even though such small structures are not resolved on metal surfaces. In fact, it may well be the case for most semiconductor surfaces that the resolution is enhanced over that expected for metals by electronic structure effects. This has been explicitly verified in the case of GaAs(110) by a reexamination of earlier[9] results. In that case, the realistic buckled geometry shows a much larger enhancement of $F(g)$ than does they hypothetical ideal (unbuckled) surface, consistent with the fact that the latter is metallic.

Of course, it the states sampled have $k_{\parallel}$ slightly displaced from $g/2$, as will generally be the case (e.g., because of the finite applied voltage), the enhancement of $F(g)$ will be correspondingly reduced. A quantitative analysis of this reduction is straightforward, but not particularly interesting.

Other cases may also be interesting, although less dramatic. For example, if only states near $k_{\parallel}=0$ contribute to tunneling, the resolution $F(g)$ of the lowest Fourier component is *suppressed* relative to the metallic case, rather than enhanced. However, no striking cases of this effect have yet come to light.

IV. CONCLUSIONS

Because of its inherently nonlinear character, the resolution of STM cannot be treated in complete generality. This difficulty is compounded by the lack of an obvious definition of the "ideal" STM image, with respect to which the resolution can be defined. Nevertheless, a useful working definition of the instrumental resolution may be formulated.

The two crucial steps are, first, restricting ourselves (at least formally) to the small-corrugation regime, where the imaging process can be linearized, and second, making a natural and intuitive (although, strictly speaking, nonunique) definition of the ideal STM image, corresponding to the hypothetical limit of ideal resolution.

With the resulting definition of the instrumental resolution function, embodied in Eq. (14), we have considered the dependence of the resolution upon the sample electronic structure. For features that are well localized within the unit cell, the resolution is well described by (18) for any sample. However, the behavior of the fundamental Fourier component of the image may depend dramatically on the sample electronic structure.

For metals, the validity of (18) extends to all Fourier components, giving a well-defined instrumental line shape which is essentially independent of the sample. However, for semiconductors and semimetals, the resolution $F(g)$ of the fundamental Fourier component corresponding to the unit cell may be drastically enhanced (or suppressed), distorting the image. Since one makes assumptions (often unconsciously) about the resolution function when interpreting images, the anomalous line shape can be misleading, making correct interpretation more difficult.

Based on heuristic considerations, and on an analysis of previous results[9] for GaAs(110), we speculate that this enhancement of the resolution of the unit cell is a common phenomenon for semiconducting surfaces with small unit cells. In these cases, considerable care is necessary in interpreting the STM images in a quantitative way, precisely because of the nonintuitive form of the effective instrumental line shape.

---

[1] J. Tersoff and D. R. Hamann, Phys. Rev. Lett. **50**, 1998 (1983).

[2] J. Tersoff and D. R. Hamann, Phys. Rev. B **31**, 805 (1985).

[3] E. Stoll, Surf. Sci. **143**, L411 (1984).

[4] J. Tersoff, Phys. Rev. Lett. **57**, 440 (1986).

[5] G. Binnig and H. Rohrer, Helv. Phys. Acta **55**, 726 (1982); Surf. Sci. **152/153**, 17 (1985).

[6] P. K. Hansma and J. Tersoff, J. Appl. Phys. **61**, R1 (1987).

[7] S. Park, J. Nogami, and C. F. Quate, Phys. Rev. B **36**, 2863 (1987).

[8] H. A. Mizes, S. Park, and W. A. Harrison, Phys. Rev. B **36**, 4491 (1987).

[9] R. M. Feenstra, J. A. Stroscio, J. Tersoff, and A. P. Fein, Phys. Rev. Lett. **58**, 1192 (1987).

[10] N. D. Lang, Phys. Rev. Lett. **56**, 1164 (1986).

[11] N. D. Lang, Phys. Rev. B **34**, 5947 (1986).

[12] J. Tersoff, M. J. Cardillo, and D. R. Hamann, Phys. Rev. B **32**, 5044 (1985).

[13] Recently, the resolution of atoms on Au(111) was reported by V. M. Hallmark, S. Chiang, J. F. Rabolt, J. D. Swalen, and R. J. Wilson, Phys. Rev. Lett. **59**, 2879 (1988). In this case, G is no longer much smaller than 2κ, and Eq. (18) cannot be reduced to (19). In any case, it was not anticipated theoretically that Au(111) could be resolved with such a significant corrugation amplitude, under the assumption of a significant tunneling gap and no mechanical interaction between surface and tip. The mechanism underlying this unexpected observation has not yet been established.

[14] Actually, the image corrugation in this case is sinusoidal only in the weak-corrugation limit. A full nonlinear treatment is given in Ref. 4, for the case that higher Fourier components are negligible compared with $F(g)$.

[15] J. A. Stroscio, R. M. Feenstra, and A. P. Fein, Phys. Rev. Lett. **57**, 2579 (1986).

[16] R. V. Coleman, B. Drake, P. K. Hansma, and G. Slough, Phys. Rev. Lett. **55**, 394 (1985).

[17] R. M. Tromp, R. J. Hamers, and J. E. Demuth, Phys. Rev. B **34**, 1388 (1986).