Two-sphere model of photon emission from the scanning tunneling microscope

A. Downes, M. E. Taylor, and M. E. Welland
Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, United Kingdom
(Received 27 February 1997; revised manuscript received 21 October 1997)

Photon emission from the scanning tunneling microscope (STM) has been shown to produce spectra that are dependent on the dielectric function of the surface. This presents the exciting possibility of chemical information from the STM, previously impossible albeit with a few exceptions. However, the spectra are also critically dependent on the geometry of the tunnel junction—changes in either tip or sample curvature can have dramatic effects on the spectra. A model is presented here that allows for curvature in the sample as well as the tip, and the results compare favorably with existing experimental data. It predicts that under certain conditions, spectra can be made almost invariant to changes in the geometry, leaving them dependent solely on the sample material—making the STM a chemical probe. The model also predicts that gold and silver particles may be differentiated by comparing their variations of photon intensity with bias, and this prediction is confirmed experimentally. Finally, a strategy for chemical identification of any metal surface is presented. [S0163-1829(98)02711-8]

I. INTRODUCTION

The invention of the scanning tunneling microscope[1] (STM) revolutionized surface science because of its ability to produce real-space images of conducting surfaces with atomic resolution. One of the major drawbacks with the technique however, is that ordinarily it is unable to give any chemical information as tunneling occurs almost exclusively by free electrons. In 1988 it was shown that light can be emitted from metallic films,[2] and since then it has been shown that metallic surfaces produce emission spectra during tunneling which are dependent on their dielectric function ε (Ref. 3). If this phenomenon could be understood and controlled it would provide a powerful technique that would complement the STM's superior resolution with chemical analysis.

Methods of photon measurement fall into three basic categories. First, the intensity of the light can be measured simultaneously with the topography to give a photon map. Second, the tip may be positioned anywhere on the surface while tunneling, and a spectrum of the emitted light acquired. Finally, the variation of the intensity with applied bias may be investigated.

Berndt, Gimzewski, and Johansson[4] took emission spectra for a variety of tip and sample materials, and it was seen that a change of tip or sample material produced an altogether different spectrum. The mechanism of photon emission was given as follows: a localized surface plasmon is excited in the region beneath the tip, which radiates photons. Dielectric losses mean that the spectra contain some information about the dielectric function of both tip and sample material. For example, a gold surface has a cutoff point in its spectrum below ≈ 510 nm, which corresponds to $\varepsilon = -2$ for gold. If the tip is spherical, the plasmon decouples for energies above where $\varepsilon = -2$ *for any surface*. Also, for metals such as copper, spectra have dips that correspond to increased losses relating to interband transitions. So for flat surfaces at least, it seems that identifying a gold surface from a silver one, for example, is simple enough.

Photon maps of surfaces with more topographic variation show complicated behavior. It has been noted that the more curved a part of the surface is, the higher the photon counts in that region, and more recently it has been shown that the spectra vary with both the curvature of the tip[5] and sample.[6] If chemical information is to be obtained from surfaces with curvature, we need to know exactly how the spectra and intensity change with tip and sample materials, and their curvature.

A detailed theory of photon emission from flat surfaces has already been published.[7,8] The geometry was modeled as a sphere above a plane to represent a tip above a surface, and the predicted spectra were close enough to experimental ones to say that the model works well for flat surfaces. However, despite this good agreement, the analysis of the results still left questions unanswered for these planar surfaces—it was not made clear how the intensity or spectra vary with tip radius, nor whether this variation may change with tip material. The variation of spectra with bias was not explained fully, and more importantly, it is not easy to understand the mechanism, nor which parts of the mechanism change with materials or tip curvature. In all, there was a need for a theory that was modular enough so that it is clear which parts of the mechanism change when the materials and geometry are changed. Such a model forms the basis for this paper.

The major question that this model should try to answer is whether the spectra can be controlled so that geometric variations do not cause changes in emission spectra, producing spectra that are specific to a given material. The increase in interest of nanometer-sized particles and quantum dots also suggests that the model's ability to investigate the optical properties of small particles is of great importance.

II. THE MODEL

The problem of finding the spectral distribution of photon emission from the STM breaks down into a number of smaller tasks. First, the potential of the localized plasmon is found by solving Laplace's equation in the appropriate region. Second, the radiation rate for an excited plasmon hav-

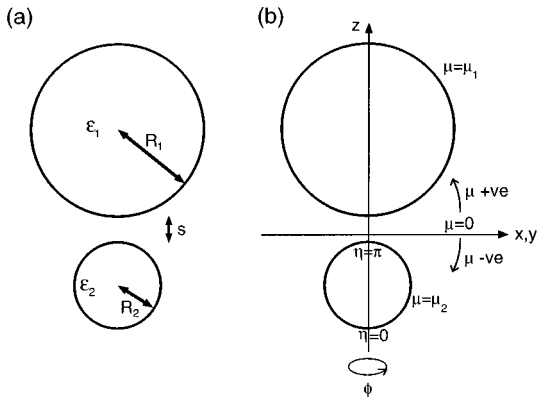

FIG. 1. (a) Model of the STM geometry. Top and bottom spheres represent the tip and sample with optical properties $\varepsilon_1(\omega)$ and $\varepsilon_2(\omega)$, respectively. (b) Representation in bispherical coordinates. $\mu=\mu_1$ on the surface of the top sphere, and $\mu=\mu_2$ on the bottom sphere's surface.

ing this potential is determined. Third, the rate of plasmon excitation by tunneling electrons is calculated. Finally, the rate of plasmon deexcitation by interaction with electrons (dielectric loss) is found; when all these rates of excitation and deexcitation are known, they are combined in an equilibrium to determine the rate of photon emission. As the plasmon is damped by the deexcitation, it is spectrally broadened. Actually, many plasmon modes at different energies are excited, and the spectral distribution is found by a superposition of all the broadened modes.

In order to describe a geometry for which surfaces may have curvature, a two-sphere geometry was chosen to simulate the tunnel junction in the STM. Figure 1 illustrates that the two spheres have variable radii and separation, and for convenience bispherical coordinates were used.[9] The three Cartesian coordinates x, y, and z are related to the bispherical coordinates of μ, η, and ϕ by the following transformation:

$$x=\frac{a\ \sin\ \eta\ \cos\ \phi}{\cosh\ \mu-\cos\ \eta},\quad y=\frac{a\ \sin\ \eta\ \sin\ \phi}{\cosh\ \mu-\cos\ \eta},$$
$$z=\frac{a\ \sinh\ \mu}{\cosh\ \mu-\cos\ \eta}. \tag{1}$$

The sphere-plane geometry can also be analyzed as a subset of the two-sphere geometry by putting $\mu_2=0$.

Laplace's equation ($\nabla^2\psi=0$, ψ being the plasmon potential) must be satisfied in both the spheres and in the vacuum gap outside them if the wavelength of these charge oscillations is assumed to be much larger than the size of the system—i.e., retardation effects are ignored. As the tip and sample radii of interest are $\leqslant 100$ Å and the measured wavelengths are several hundred nanometers, it is a reasonable approximation to consider a nonretarded potential. There is cylindrical symmetry in this case, and a potential that satisfies Laplace's equation in these coordinates with no ϕ dependence has been shown to be of the form[10]

$$\psi=F\sum_n\ \{A_n\ \exp[-(n+\tfrac{1}{2})\mu]$$
$$+B_n\ \exp[(n+\tfrac{1}{2})\mu]\}P_n(\cos\ \eta). \tag{2}$$

However, in the upper sphere the first term would become infinite as μ tends to infinity, and similarly the second term is ignored for the bottom sphere as μ becomes negative infinity. So the potential in all three regions (top sphere, gap, bottom sphere) can be written as

$$\psi_1=F\sum_n\ \{A_n'\ \exp[-(n+\tfrac{1}{2})\mu]\}P_n(\cos\ \eta),\quad \mu\geqslant\mu_1,$$

$$\psi_2=F\sum_n\ \{A_n\ \exp[-(n+\tfrac{1}{2})\mu]$$
$$+B_n\ \exp[(n+\tfrac{1}{2})\mu]\}P_n(\cos\ \eta),\quad \mu_1\geqslant\mu\geqslant\mu_2,$$

$$\psi_3=F\sum_n\ [B_n'\ \exp(n+\tfrac{1}{2})\mu]P_n(\cos\ \eta),\quad \mu\leqslant\mu_2, \tag{3}$$

where $\mathbf{F}=\sqrt{\cosh\ \mu-\cos\ \eta}$, and P_n are Legendre polynomials.

Two sets of boundary conditions must be applied. The potentials must be matched at both surfaces, i.e.,

$$\psi_1(\mu_1)=\psi_2(\mu_1)\quad\text{and}\quad\psi_2(\mu_2)=\psi_3(\mu_2), \tag{4}$$

and the normal component of the electric displacement is continuous across both surfaces, i.e.,

$$\underline{D}_1^{\perp}(\mu_1)=\underline{D}_2^{\perp}(\mu_1)\quad\text{and}\quad\underline{D}_2^{\perp}(\mu_2)=\underline{D}_3^{\perp}(\mu_2), \tag{5}$$

where $\underline{D}=\varepsilon_0\varepsilon_r\underline{E}$ and $\underline{\mathbf{E}}=-\nabla\psi$.

The implementation of these two boundary conditions leads to two equations relating A_n and B_n, and gives rise to a recurrence relation (due to the removal of the Legendre polynomials) with terms in $n+1$, n, and $n-1$ that is represented here in matrix form:

$$P_{n-1}\begin{pmatrix}A_{n-1}\\B_{n-1}\end{pmatrix}+Q_n\begin{pmatrix}A_n\\B_n\end{pmatrix}+R_{n+1}\begin{pmatrix}A_{n+1}\\B_{n+1}\end{pmatrix}=\underline{0}, \tag{6}$$

where

$$Q_n=\begin{bmatrix}\chi_1[(2n+1)\cosh\ \mu_1-\sinh\ \mu_1] & -[(2n+1)\cosh\ \mu_1+\chi_1\ \sinh\ \mu_1]\exp[(2n+1)\mu_1]\\ [\chi_2\ \sinh\ \mu_2-(2n+1)\cosh\ \mu_2]\exp[-(2n+1)\mu_2] & \chi_2[(2n+1)\cosh\ \mu_2+\sinh\ \mu_2]\end{bmatrix},$$

$$R_{n+1}=(n+1)\begin{bmatrix}-\chi_1\ \exp(-\mu_1) & \exp[(2n+2)\mu_1]\\ \exp[-(2n+2)\mu_2] & -\chi_2\ \exp(\mu_2)\end{bmatrix},$$

$$P_{n-1}=n\begin{bmatrix}-\chi_1\ \exp(\mu_1) & \exp(2n\mu_1)\\ \exp(-2n\mu_2) & -\chi_2\ \exp(-\mu_2)\end{bmatrix},$$

and $\chi_k(\omega)=[1-\varepsilon_k(\omega)]/[1+\varepsilon_k(\omega)]$. Equation (6) can be rewritten as an infinite matrix that is equivalent to an infinite series of 2×2 matrix equations for all n, or as one terminated after a large number of terms (up to $n_{\max}$).

Whenever the determinant of this large matrix is equal to zero, we find an energy at which solutions to Eq. (6) exist—the only variables in Eq. (6) that are functions of energy are the dielectric functions (relative permittivities), ε_1 and ε_2. A total of $n_{\max}$ solutions are found, at energies that asymptotically rise to where the real part of $\varepsilon_2=-1$ in most systems considered (see Sec. IV). These solutions are considered as separate modes, each mode having its own potential, so all calculations are performed for a given mode.

It is convenient now to normalize the values of A_n and B_n, as they represent the magnitude of the potential. This is done by equating the electrostatic energy in all space to that of the ground-state energy of the mode in question, i.e.,

$$\tfrac{1}{2}\hbar\omega_{\text{mode}}=\tfrac{1}{2}\varepsilon_0\int_V \underline{E}\cdot\underline{E}^* d\tau, \tag{7}$$

which gives the normalization factor to obtain the actual A_n and B_n. The magnitude of the potential of the dominant mode in most cases is found to be in the range 0.1–1 V.

Now that we have the potential of a particular mode in all regions ψ, we can calculate the power radiated from this oscillating dipole. The cylindrically symmetric potentials in Eq. (2) have no ϕ dependence, meaning that the charge distributions on the spheres cannot give rise to any net dipole moment in the x-y plane. Starting from the normalized potentials ψ, the charge densities on spheres 1 and 2, respectively, are

$$\sigma_1=\varepsilon_0[E_{1n}-E_{2n}]|_{\mu=\mu_1} \quad \text{and} \quad \sigma_2=\varepsilon_0[E_{2n}-E_{3n}]|_{\mu=\mu_2}, \tag{8}$$

where E_{mn} is the component of the electric-field normal to the sphere, with ψ_m coming from Eqs. (3), (5), or (6). The total dipole moment in the z direction for spheres 1 and 2 is

$$P_z=\int_{s_1}\sigma_1 z\, ds_1+\int_{s_2}\sigma_2 z\, ds_2. \tag{9}$$

Using classical electrodynamics, the power $\mathcal{P}_{\text{rad}}$ radiated from these oscillating dipoles is

$$\mathcal{P}_{\text{rad}}=\frac{\omega^4}{32\pi^2\varepsilon_0 c^3}\int |P_z|^2 d\Omega\, \sin^2\theta$$
$$=\frac{4\omega^4 a^4\varepsilon_0}{3c^3}\left[\sum_n (2n+1)(B_n-A_n)\right]^2. \tag{10}$$

If there exists a metallic plane underneath the two spheres, then the integral in Eq. (10) must include a factor of $|1+T|^2$, where T is the Fresnel coefficient for a dipole above a reflecting plane.[10]

This value for the power is converted into a rate of photon emission—i.e., photons emitted per second—by dividing by $\hbar\omega_{\text{mode}}$ to give a rate which we will call ν_{rad}. Both spheres should be charge neutral, and this can be used to test the accuracy of the calculations leading to A_n and B_n, as the expressions for the charges on the spheres are

$$q_1=\int_{s_1}\sigma_1 dS \quad \text{and} \quad q_2=\int_{s_2}\sigma_2 dS$$

or

$$q_1=2\pi\varepsilon_0 a\sum B_n \quad \text{and} \quad q_2=2\pi\varepsilon_0 a\sum A_n. \tag{11}$$

The errors in these sums are always found to be insignificant for each mode.

The matrix element for inelastic electron tunneling (considered to be from tip to sample in Fig. 2, but it may be chosen to be in the opposite direction), with the plasmon as the perturbation, from perturbation theory is

$$M=\int dz\ \chi_L^*\psi_n(z)\chi_R, \tag{12}$$

with the electron wave functions as

$$\chi_L=C\ \exp[-\alpha_L(z-s/2)], \quad \chi_R=C'\ \exp[-\alpha_R(z+s/2)],$$

and the plasmon potential coming from Eq. (3). C and C' are normalization constants that do not appear in the final expression, and M is taken over both the gap and the electrodes, for a square barrier.

We use the approximations $\eta=\pi$ and $\mu=2z/a$—i.e., tunneling takes place at the very bottom of the top sphere, and as it does so over a few angstroms this is a fair approximation. The tunneling gap must be much smaller than the radius of either sphere.

The rate of inelastic tunneling from Fermi's golden rule will then be

$$\nu_{\text{inel}}=\sum_L\sum_R 2\frac{2\pi}{\hbar}|M|^2\delta(E_L-E_R-\hbar\omega)f_L(1-f_R), \tag{13}$$

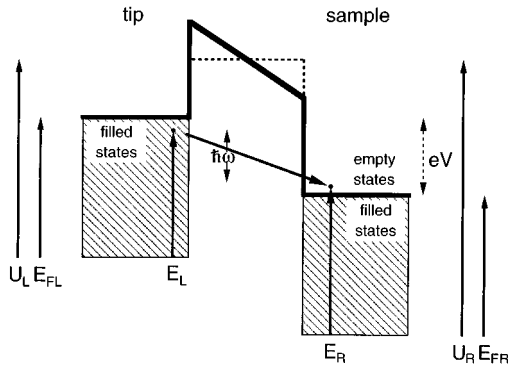

FIG. 2. Inelastic energy diagram. Electrons lose an energy $\hbar\omega_{\text{mode}}$ when tunneling across the gap. E_L and E_R are the energies of a general pair of filled (starting) and empty (final) states. Fermi energies E_{FL} and E_{FR} are marked for left and right, respectively.

where f_L and f_R are the Fermi factors ($=1$ when an energy level is occupied—not corrected for thermal broadening) and the factor of 2 is due to the inclusion of both possible spin states.

We now have a value for the rate of inelastic tunneling for a given mode (and an overall elastic current is calculated for the same electron wave functions and barrier). Next we must evaluate a loss rate that relies on there being an imaginary part of the dielectric function. The power lost per unit volume from the oscillation,

$$\mathcal{P}_{\rm diel} = -\tfrac{1}{2}\ \mathrm{Re}\{\dot{\underline{D}}^* \cdot \underline{E}\}, \tag{14}$$

where $\dot{\underline{D}} = -(i\omega)\varepsilon_r \varepsilon_0 \underline{E}$ and $\varepsilon_r = \varepsilon_{\rm real} + i\varepsilon_{\rm im}$.

Integrating this power density over a volume gives

$$\mathcal{P}_{\rm diel} = \tfrac{1}{2}\omega\varepsilon_0 \int_{\rm spheres} \varepsilon_{\rm im} \underline{E}^* \cdot \underline{E}\, d\tau, \tag{15}$$

where $\varepsilon_{\rm im}$ is the imaginary part of the dielectric function for whichever sphere. This expression is evaluated from the potentials, then divided by $\hbar\omega_{\rm mode}$, to give the rate of dielectric loss, $\nu_{\rm diel}$.

In order to obtain a steady state we must construct a rate equation balancing excitation (inelastic tunneling) and deexcitation (radiation and dielectric loss):

$$(1+\overline{n_{\rm occ}})\,\nu_{\rm inel} = \overline{n_{\rm occ}}\,\nu_{\rm rad} + \overline{n_{\rm occ}}\,\nu_{\rm diel}, \tag{16}$$

where $\overline{n_{\rm occ}}$ is an averaged occupation number for the plasmons. The enhancement factor for bosons (i.e., the plasmons) is present on the left-hand side of Eq. (16). $\overline{n_{\rm occ}}$ is found to be much smaller than 1 here.

So the rate of inelastic events

$$(1+\overline{n_{\rm occ}})\,\nu_{\rm inel} \quad \text{or} \quad \frac{\nu_{\rm inel}(\nu_{\rm rad}+\nu_{\rm diel})}{\nu_{\rm rad}+\nu_{\rm diel}-\nu_{\rm inel}}, \tag{17}$$

and the rate of photon emission events is

$$\overline{n_{\rm occ}}\,\nu_{\rm rad} \quad \text{or} \quad \frac{\nu_{\rm inel}\nu_{\rm rad}}{\nu_{\rm rad}+\nu_{\rm diel}-\nu_{\rm inel}}. \tag{18}$$

All these rates are for a given mode. The total rates can be found simply by adding the rates for all modes found, as they are independent.

Due to the uncertainty principle, each mode can radiate a photon with different energy to that of the plasmon mode. Experimental results show broadened features,[4] and often the modes are so broadened they are indistinguishable as separate peaks. The finite lifetime originates from the presence of the dielectric loss and radiative processes that damp the oscillation. So the width of the Lorentzian,

$$\Delta E_{\rm mode} = \hbar\nu_{\rm diel} + \hbar\nu_{\rm rad}. \tag{19}$$

The response away from the eigenfrequency means that the rate Eq. (16) should be replaced by

$$g(\omega-\omega_{\rm mode})(1+\overline{n_{\rm occ}})\,\nu_{\rm inel}(\omega) = g(\omega-\omega_{\rm mode})\overline{n_{\rm occ}}\,\nu_{\rm rad}(\omega) + g(\omega-\omega_{\rm mode})\overline{n_{\rm occ}}\,\nu_{\rm diel}(\omega), \tag{20}$$

where g is a Lorentzian with a width of $\Delta E_{\rm mode}$ (full width at half maximum). All the rates, $\nu_{\rm ine}$, $\nu_{\rm rad}$, and $\nu_{\rm diel}$ must be evaluated at all frequencies ω.

An improvement on the Lorentzian broadening can be made by using an energy-dependent or ''variable'' broadening. The revised line shape is the power spectrum of a damped oscillator whose loss term is a function of energy, in this case due to the energy-dependent dielectric functions. The emission spectrum is just a superposition of all the broadened modes.

A bulk plasmon occurs at an energy where the real part of the dielectric function, $\varepsilon_{\rm real}=0$, and a surface plasmon where $\varepsilon_{\rm real}=-1$. A plasmon on a single sphere occurs at $\varepsilon_{\rm real}=-2$, so we can expect the spheres to be decoupled for energies above that at which $\varepsilon_{\rm real}=-2$. This is confirmed by the drop in the field enhancement factor of Johansson and Monreal.[8] However, because the use of broadening would still predict emission above this energy, it is necessary to force the spectra to zero at and above this energy. This is achieved by the use of an image peak of opposite sign symmetrically above the decoupling energy. Both this modification, and the broadening approximations effectively make A_n and B_n energy dependent, as they are in Johansson's model.

The imaginary part of the dielectric function has two components: interband transitions (exciting other electrons) and Drude losses (due to the relaxation time of electrons constituting the plasmon). In small particles, the interband transitions are unaffected, but the Drude losses increase due to surface scattering. This results in a doubling of $\varepsilon_{\rm im}$ for a silver particle of diameter ≈ 20 Å. $\varepsilon_{\rm real}$ is not measurably affected. So smaller particles should give rise to broader features in the spectra, as the broadening is proportional to $\varepsilon_{\rm im}$ in Eq. (15), although the positions of the modes will not change as $\varepsilon_{\rm real}$ does not change. If the particles are nonspherical the energy at which the spheres decouple will be different. For prolate spheroids this decoupling will occur when $\varepsilon_{\rm real}$ is below -2, and for oblate spheroids when it is between -1 and -2. So the cutoff in spectra is expected to be at higher energies (shorter wavelengths) for flat grains or blunt tips and at lower energies for sharp tips or high aspect ratio structures.

Johansson's model[7,8] should also be considered so that it may be compared with ours. It starts with the reciprocity theorem of electrodynamics, to relate the electric field some place far away from the tip to the inelastic tunnel current under the tip. It achieves this by swapping the ''detector'' (far from the tip) and ''source'' (under the tip) in a similar way to how the radiation pattern of an antenna can be swapped for the response of a receiver. Like an antenna, an enhancement in the field occurs under the tip because of the favorable geometry and materials. This response function, or enhancement factor, is solved from potentials of the same form as Eq. (3), and the recurrence relations for the sphere-sphere case in Eq. (6) effectively reduce to those quoted by Johansson for the sphere-plane case by substituting $\mu_2=0$. The coefficients corresponding to A_n and B_n can be calculated at any energy in Johansson's model. It was quoted that at lower energies, A_0 and B_0 were the dominant ''modes'' and at higher energies higher numbered ''modes'' dominated, although in our model these ''modes'' are merely components of each mode. A similar finding is also made for

this model, that for the lowest-energy mode the dominant components are A_0 and B_0, and, for example, the fourth-lowest energy mode will be dominated by A_3 and B_3. The inelastic current matrix is calculated for a trapezoidal barrier by Johansson; however, the inelastic matrix uses a response function that is constant throughout the gap, whereas in reality it usually changes sign. For our model, this perturbation is taken as varying in the gap and an approximate contribution from the electrodes is also included in the matrix. In Johansson's model, at any frequency the electric field and subsequently the photon emission are calculated at the detector. In our two-sphere model, modes at a variety of frequencies are responsible for the photon emission, and their potentials and energies are calculated explicitly. Other advantages of this model are its two-sphere capability, and the separate calculation of the three rates of excitation and deexcitation. Also, all these rates and other physical parameters such as dipoles and charge densities and plasmon production rates can be evaluated for each mode and each sphere (or plane) separately, if required. Furthermore, the explicit nature of the calculations involving the potentials are easier to understand than the use of the enhancement factor, which combines all the geometric and material properties. However, the convolution of the modes with the line shape and the $\varepsilon_{\text{real}}=-2$ cutoff is less preferable to Johansson's approach for numerical accuracy. In short, Johansson's model treats the phenomenon as enhancement, then emission; ours treats it as excitation (inelastic tunneling), then deexcitation (photon emission or dielectric loss).

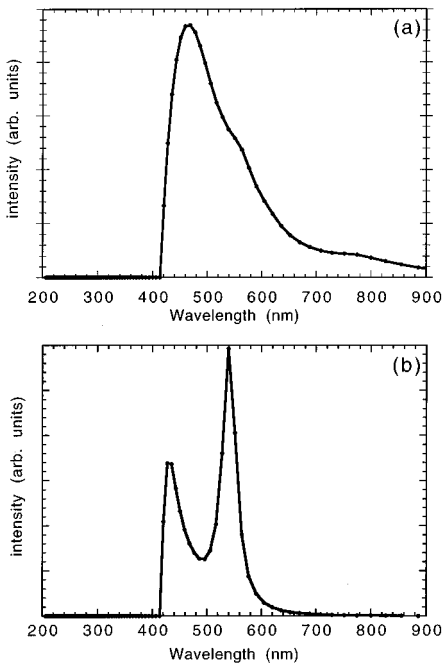

FIG. 3. Predictions of emission spectra from our model to compare with Ref. 10 for (a) W and (b) Ag tips with radii of 300 Å on a Ag(111) surface. Tip-surface separations are (a) 5 Å and (b) 7.5 Å. These spectra do not include the response of the detector (plotted in Ref. 4).

III. COMPARISON OF THEORY WITH EXPERIMENTS

The model is able to predict spectra for any combination of tip and sample materials, provided their dielectric functions, Fermi energies, and work functions are known. The tip-sample separation, tip bias, and tip and sample radii are all variables, although in the sphere-plane case we use $\mu_2=0$. Before considering the two-sphere case we first consider the sphere-plane case so as to afford comparison with previous experimental results.[4,11]

Figure 3 shows predictions from our model for both W and Ag tips on a flat Ag surface. The conditions are chosen to compare with the spectra of Berndt, Gimzewski, and Johansson and the predictions by Johansson.[4] There is good qualitative agreement with experiment from both models, although the positions of peaks tend to be predicted at shorter wavelengths in both cases. It is important to note that in the W-Ag case the spectral peak does not correspond to the lowest-energy mode. In fact, the lowest-energy mode here occurs at 3.29 eV ($=377$ nm), which is above the maximum energy of the tunneling electrons (3 eV). The spectral peak is a result of the convolution of this lowest-energy mode, which has very high broadening (3.3 eV), and the density of states for inelastic tunneling. This basically forces the spectrum to zero at the maximum and minimum electron energies (3 and 0 eV here) producing a peak at some intermediate energy that is unrelated to the mode itself. For Ag-Ag, however, the peaks in the spectrum do correspond to the lowest-energy modes as they occur at lower energies and have reduced broadening. Quantum efficiencies were calculated as 2.6×10^{-4} (W-Ag) and 9.1×10^{-4} (Ag-Ag) photons emitted per tunneling electron, which agree to the right order of magnitude with those determined experimentally. If a W tip is used, the modes are situated close to the energy where $\varepsilon_{\text{real}}=-1$ for the sample (3.68 eV for Ag), and changing the tip to a noble metal (Ag, Au, Cu) will cause the modes to be shifted down in energy. Increasing the tip radius also causes the energy of the modes to fall, so for Ag-Ag the peaks are redshifted. For the W-Ag case, however, the positions of the very broad modes above the maximum electron energy shift only slightly, causing minor changes to the emission spectrum.

Similar calculations for the W-Cu case contain dips in the emission below ≈600 nm in both of the models and the real data.[3,4] We predict only one mode at 385 nm, and the spectral features are a result of changes in the imaginary part of the dielectric function of Cu (Fig. 4), as the mode is highly broadened. The cutoff for both Cu and Ag surfaces at 3 V bias is limited by the quantum cutoff (the maximum electron energy of 3 eV), and not the $\varepsilon_{\text{real}}=-2$ cutoff. However, when the bias increases, the spectra change so as to be limited by this $\varepsilon_{\text{real}}=-2$ cutoff (≈3.5 eV for both Cu and Ag), and the peaks in the spectrum are shifted up in energy until the electron energy rises above this cutoff. The variation of spectra with bias for the W-Au surface can be seen in Fig. 5. At 2 V bias, the spectra are limited by the minimum electron energy, 2 eV or 620 nm, but when more than 3 V bias is applied the spectra become similar, being dominated by the $\varepsilon_{\text{real}}=-2$ cutoff. This can be seen in both experimental data[4,11] and our model's predictions.

The variation of photon intensity with *tip* radius was calculated to be $\eta\propto R^{2.4}$ for the W-Au junction with a tip bias of $+3.0$ V. This exponent varies between 2.0 and 2.5 for all

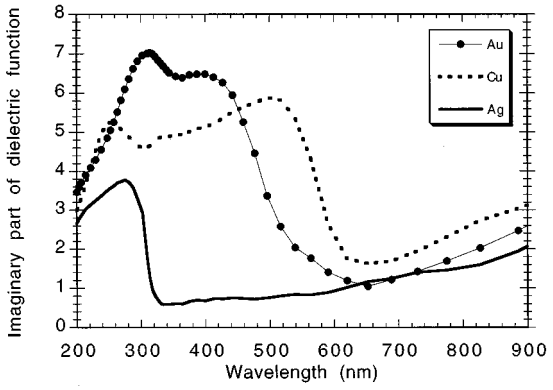

FIG. 4. Imaginary part of the dielectric function used in our model for the noble metals Au, Cu, and Ag. Calculated from Ref. 12.

combinations of metals in sphere-plane junctions. The dependence of the individual rates of excitation and deexcitation on tip radius was calculated to be $\nu_{\rm inel}\propto R^{-0.4}$, $\nu_{\rm rad}\propto R^{2.4}$, $\nu_{\rm diel}\propto R^{0.4}$. Although these combine together in Eq. (16), as $\eta\propto R^{1.6}$, the missing $R^{0.8}$ comes partly from the increasing proportion of the broadened mode able to emit light—i.e., that which is below the $\varepsilon_{\rm real}=-2$ cutoff, and partly from the increasing number of other modes appearing below the cutoff. An advantage of the model is that we are able to determine the relative sizes of the *individual* rates of excitation and deexcitation *for each mode*. These were $\nu_{\rm inel}=7\times10^{10}\ {\rm s}^{-1}$, $\nu_{\rm rad}=3\times10^{13}\ {\rm s}^{-1}$, and $\nu_{\rm diel}=5\times10^{15}\ {\rm s}^{-1}$ for the dominant and lowest-energy mode of the W-Ag junction modeled in Fig. 3. This results in a natural broadening of $\hbar(\nu_{\rm diel}+\nu_{\rm rad})=3.3$ eV. For Ag-Ag the broadening of the two lowest-energy modes is 0.12 and 0.27 eV. For the W-Ag junction [Fig. 3(a)] we calculated that the inelastic current accounts for 10% of the total current, and that the plasmon to

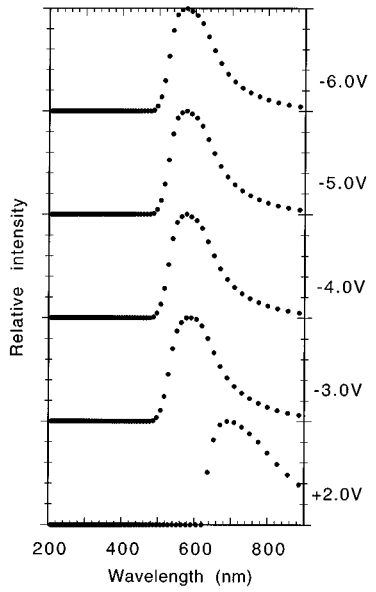

FIG. 5. Predictions of the dependence of emission spectra on tip bias from our model to compare with those in Ref. 11, for a 300-Å radius W tip and a Au surface. The tip-surface separation is constant at 5 Å.

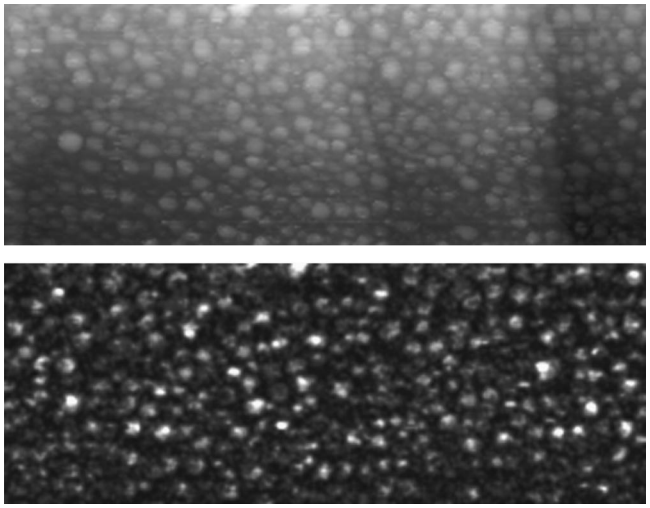

FIG. 6. Top: topography of Ag spheres with sizes of up to 17 Å. The images are of size $2000\times750\ {\rm Å}^2$. Bottom: photon map of the same area—actually a map of quantum efficiency (the photon flux divided by the current) to remove small variations in tunneling current. The images were taken at a tip bias of -3.1 V, tunneling current 10 nA, and the photon counts per pixel are 0–140, corresponding to a quantum efficiency of $0-2.7\times10^{-5}$. The tip radius was deduced to be 73 Å from the apparent lateral broadening of the spheres in the topography. The spheres are formed by evaporation of Ag in ~1 mbar of argon in a preparation chamber, and the STM chamber is kept at a pressure of $<10^{-10}$ mbar. The silicon substrate, originally the Si(111)-7×7 surface, is nonemitting, and two monatomic steps can be seen running almost vertically in the topography.

photon decay has a probability of 0.25% (all other events being dielectric losses). In marked comparison, the Ag-Ag junction [Fig. 3(b)] has only 0.4% inelastic current, but a probability of 24% decay to photons.

We now consider the sphere-sphere case, which is similar to the sphere-plane case in that the $\varepsilon_{\rm real}=2$ cutoff and $\varepsilon_{\rm im}$ variations will be present in the spectra, but different in that the modes occur at lower energies (so that the spectra are more redshifted). Additionally, a sphere has a larger dipole than that of a flat surface (where it is simply the image dipole of the tip), unless the sample sphere has a far smaller radius than the tip. The only change to the model is the use of $\mu_2=0$, the two cases being otherwise identical in their methods and assumptions.

Some of the motivation for the model arose from our experimental results of photon emission from Ag clusters. A photon map is presented in Fig. 6, of Ag spheres formed by evaporation in 1 mbar argon. In the size range 5–20 Å we found experimentally that the quantum efficiency $\eta\propto R_{\rm tip}R_{\rm sample}$ for a W tip of radius 73 Å. Our model, however, predicts that $\eta\propto(R_{\rm tip})^{0.5}(R_{\rm sample})^{3.5}$. This apparent discrepancy can be explained by simply noting that many instabilities occurred during experiments. This almost certainly is a consequence of momentary contacts between tip and sample so that the tip picked up Ag particles.[13] Assuming then a Ag tip, the model's prediction for was $\eta\propto R_{\rm tip}R_{\rm sample}$, in complete agreement with results. Although there is no simple rule for any combination of tip and sample materials, in general increasing the tip *or* sample radius for any metal will cause an increase in emission (until retardation effects be-

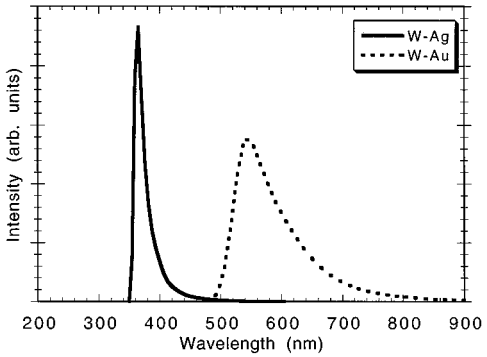

FIG. 7. Predictions of emission spectra for W tips and (a) Ag, (b) Au particles. In both cases, the tip bias is -4 V, tip radius 73 Å, sample radius 10 Å, and separation 5 Å. Overall efficiencies are W-Ag, 2.6×10^{-7}; W-Au, 3.0×10^{-7}. Predictions for a larger radius particles produce very similarly shaped spectra, as do those for Ag tips.

come important). A further consequence of increasing tip or sample radius is that modes will be redshifted.

The spectra predicted for two-sphere W-Ag and W-Au junctions are shown in Fig. 7, for a 73 Å radius tip and 10 Å radius particles at -4 V bias. The two spectral peaks, at around 370 nm (violet) for Ag and 540 nm (green) for Au, can be seen to be entirely separable, and remain so for particles with a radius of less than a few hundred angstroms. When a large radius Ag tip is used on either a flat or spherical Ag surface, peaks such as those in Fig. 3(b) occur. If the sample radius increases these peaks will be redshifted. However, for a *small* radius Ag tip (or larger radius W tip) the peaks stay close to $\varepsilon_{\text{real}}=-1$ and the amount of redshifting is greatly reduced. The spectra are then less influenced by changes in geometry. The variations of quantum efficiency with bias for the W-Ag and W-Au junctions of Fig. 7 are shown in Fig. 8. Ag and Au spheres are predicted to have these characteristic curves, whether the tip material is Ag, Au, or W. Changes in tip or sample radius also had little effect.

In order to compare the model with more experimental data, we deposited Ag and Au as 1-nm spheres on Si(111)

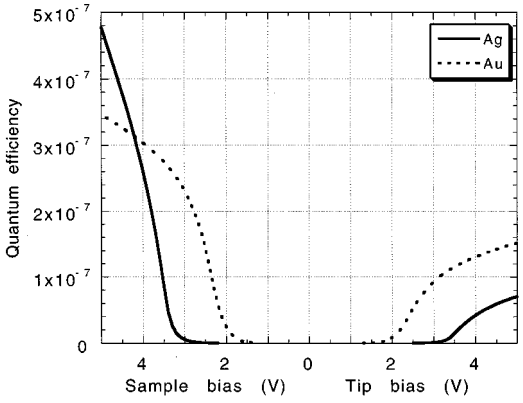

FIG. 8. Predictions of efficiency-voltage characteristics for Ag and Au particles and a W tip. In both cases the tip radius is 73 Å, sample radius 10 Å, and separation 5 Å. Again, Ag tips or larger particles produce similarly shaped curves.

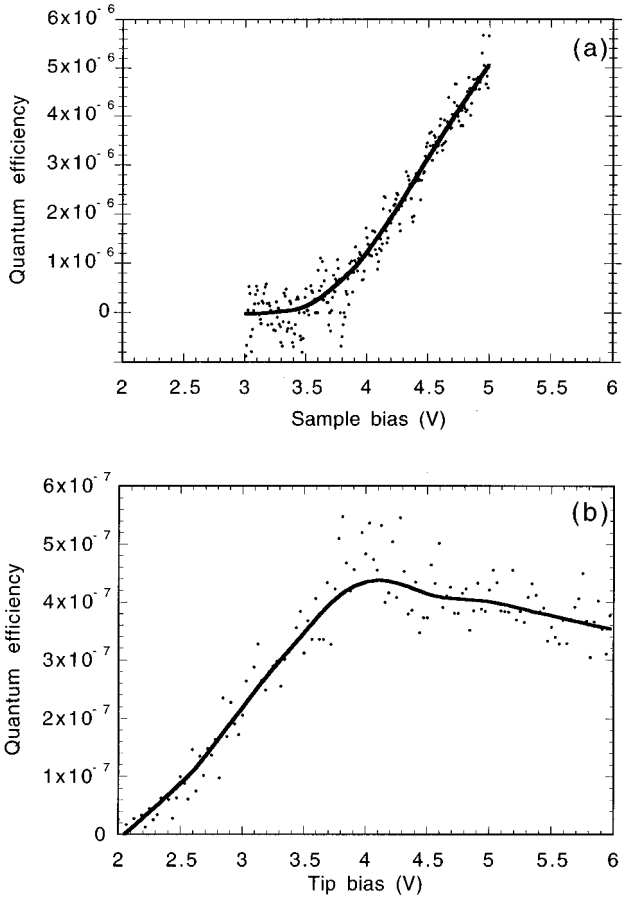

FIG. 9. Quantum efficiency as a function of tip bias for 10-Å-diam particles, taken while the feedback loop was held, for a W tip. (a) Ag particle, tip radius calculated as 65 Å from apparent broadening of particles in images (left-hand scale). (b) Au particle, tip radius unknown (right-hand scale). The high level of noise in both cases is due to the division of two noisy signals—the photon counts and tunnel current. A similar shaped curve to (a) was seen for a deposit of size 100 Å, formed when a voltage pulse was applied to a tip believed to be coated with a number of silver particles.

surfaces. The metals were evaporated in ~1 mbar Ar, and the Si substrate was nonemitting. These results are shown in Fig. 9 for Ag and Au spheres, which show excellent agreement with the model's predictions (Fig. 8). This implies a method to perform chemical mapping, whereby photons would be collected at ≈-2.5 and ≈-3.5 V. The first map would show only Au particles and the second, both Ag and Au particles.

IV. STRATEGY FOR CHEMICAL IDENTIFICATION

Having demonstrated that Ag and Au particles can be distinguished from each other, we can use the model to predict whether other metals can be distinguished by means of their spectra. With an appropriate bias and choice of tip, there are three possible features specific to the metal that occur in the spectrum. First, each metal will have a cutoff in the spectrum at a characteristic energy where $\varepsilon_{\text{real}}=-2$ (provided the surface is either relatively planar or spherical). Sec-

TABLE I. Optical data for common metals (calculated from Ref. 12), and their methods of identification from emission spectra.

Metal	Energy of $\varepsilon_{real}=-2$	Value of ε_{im} at 2.5 eV	Features in ε_{im} in optical spectrum (1.4–4 eV)	Method of identification
Ag	3.5 eV	1.4		cutoff, broadening (see Ni)
Al	9.0 eV	9	peak 1.5 eV	features
Au	2.5 eV	2.6	rises above 2.0 eV	cutoff
Co	4.8 eV	13		cutoff in UV
Cr	6.2 eV	22		cutoff in UV (see Ir)
Cu	3.5 eV	6	jumps above 2.0 eV	features
Fe	5.2 eV	16		cutoff in UV (see W)
Ir	6.2 eV	16	shoulder at 3.9 eV	cutoff in UV, features (see Cr)
Li	3.4 eV	0.9	levels off above 2.3 eV	cutoff, features
K	2.3 eV	0.11		cutoff
Mg	6.0 eV	3.1	[a]	[a]
Mn	3.8 eV	14		(cutoff, see Ti)[b]
Mo	1.4 eV	25		cutoff, broadening (see Zn)
Na	3.3 eV	0.26	shoulder 1.9–2.5 eV	cutoff, features
Ni	3.4 eV	10		cutoff, broadening (see Ag)
Pt	4.5 eV	13		cutoff in UV
Ta	1.9 eV	11	large, broad peak 2–4 eV	cutoff, features
Ti	3.8 eV	9	(small shoulder at 1.4 eV)	(cutoff, features, see Mn)[b]
V	4.2 eV	12	broad peak at 1.6 eV	features, cutoff in UV
W	5.3 eV	18	small peak at 3.4 eV, small dip at 1.6 eV	features, cutoff in UV (see Fe)
Zn	1.4 eV	5		cutoff, broadening (see Mo)

[a]Insufficient data.

[b]Unlikely to be identified due to another element with a similar cutoff energy and broadening.

ond, some metals will have dips or peaks in ε_{im} that will cause spectra to have characteristic peaks or dips at certain energies. Third, metals could be distinguished by comparing the broadening of the spectra. Table I shows cutoff energies, features in ε_{im}, and relative broadening for a number of common metals.

It should be noted however, only features in the sample's ε_{im} below its cutoff energy can show up in the spectrum. Also, if a W tip is used on Ag particles then the spectrum will stop at 3.5 eV ($\varepsilon_{real}=-2$ for Ag), whereas a spectrum from an Al particle will stop at 5.3 eV ($\varepsilon_{real}=-2$ for the W tip) rather than 9.0 eV ($\varepsilon_{real}=-2$ for the Al sample). In other words, a tip will only be able to identify cutoff energies below that of its own. If the tip is used on a sample with a higher-energy cutoff then the resulting emission spectra are dominated by the tip's dielectric function rather than the sample's. This can be seen in the results of Berndt and Gimzewski[14] of the W-Ti and W-Fe systems. The first spectrum was featureless (following the dielectric function of Ti) but the second had more structure in the range 600–800 nm (following the dielectric function of W, which has variations in that region).

In order to identify the widest possible range of metals, the ideal tip would be Al and have a radius of ~ 50 Å. In this way spectra would always be limited to the sample's $\varepsilon_{real}=-2$ energy, and would be broad enough to show up variations in ε_{im}. Although its good topographic resolution would be another benefit, its major drawback would be its low efficiency, and it would take a sophisticated collection system and spectrometer capable of working from ~ 1.3–6.3 eV to identify all metals. A higher efficiency alternative would be a ~ 100 Å Ag tip, though it would be able to identify many fewer metals due to its lower cutoff energy (but it would not require ultraviolet detection). With either of these two tips, the modes would be kept close to the surface plasmon frequency (at $\varepsilon_{real}=-1$) and as such the spectra should be largely invariant to changes in geometry. And whichever tip is used, the bias must be large enough to show up the cutoff energy of the material, and both a high tunnel current and good collection/detection system should be used to ensure a low-noise spectrum.

V. CONCLUSION

Early studies of photon emission from metals concentrated on the localized plasmon and how the peak(s) in the emission were a result of some sort of resonance phenomenon. Photon maps of granular films were unpredictable[15] and the whole process seemed interesting but little more than an extension of investigations into photon emission from solid-state tunnel junctions. Later, from work on single-crystal metals,[4,11] it became clear that the dielectric function of the sample played an important role in the emission spectrum. Although many features in these spectra corresponded to the dielectric functions, interest was focused more on the positions of the peaks, and the fact that they varied with geometry suggested that it would be difficult to extract chemical information.

The theory presented in this paper is modular enough to show which parts of the mechanism are dominated by geometry and materials, and it is able to cope with sample curvature, which is essential. The predictions of the theory compare favorably with experimental results in the sphere-plane case, and in the sphere-sphere case it agrees well with our experimental findings, where chemical identification of particles was achieved even without a spectrum.

To identify *any* metal particle or part of a surface, it is necessary to take the emission spectrum. For a certain tip radius and material, the spectrum will be dominated by the dielectric function of the sample. Then almost all metals can be identified from either their $\varepsilon_{\text{real}}=-2$ energy (where the spectrum is forced to zero) or features at certain energies relating to changes in ε_{im}, provided the bias is high enough. For nonspherical tips or surfaces, the spectrum will be forced to zero where $\varepsilon_{\text{real}}$ is different to -2, so metals may be confused, but this uncertainty in energy should be small, as $\varepsilon_{\text{real}}$ changes rapidly.

In order to identify a wide range of metals, a spectrometer with extended ultraviolet sensitivity is required. Also, as emission can be an inefficient process for many metals, both photon collection and detection efficiencies should be high, otherwise a long collection time may mean smaller particles will not be identified due to drift in the STM. Yet, the exciting promise of such a powerful technique that combines the superior resolution of the STM with simultaneous chemical analysis should be incentive enough for many to implement such an add-on optical system.

ACKNOWLEDGMENT

This work was carried out as part of an ESPRIT basic research program (Project No. 8523, PRONANO).

[1] G. Binnig and H. Rohrer, Helv. Phys. Acta **55**, 726 (1982).
[2] J. K. Gimzewski, J. K. Sass, R. R. Schlittler, and J. Shott, Europhys. Lett. **8**, 435 (1989).
[3] R. Berndt, J. K. Gimzewski, and P. Johansson, Phys. Rev. Lett. **67**, 3796 (1991).
[4] R. Berndt, J. K. Gimzewski, and P. Johansson, Phys. Rev. Lett. **71**, 3493 (1993).
[5] T. Umeno, R. Nishitani, A. Kasuya, and Y. Nishina, Phys. Rev. B **54**, 13 499 (1996).
[6] K. Ito, S. Ohyama, Y. Uehara, and S. Ushioda, Surf. Sci. **324**, 282 (1995).
[7] P. Johansson, R. Monreal, and P. Apell, Phys. Rev. B **42**, 9210 (1990).
[8] P. Johansson and R. Monreal, Z. Phys. B **84**, 269 (1991).
[9] G. R. Dean, Phys. Rev. **35**, 459 (1912).
[10] R. W. Rendell, D. J. Scalapino, and B. Mühlschlegel, Phys. Rev. Lett. **41**, 1746 (1978).
[11] R. Berndt and J. K. Gimzewski, Surf. Sci. **269/270**, 556 (1991).
[12] J. H. Weaver and H. P. R. Frederikse, in *CRC Handbook of Chemistry and Physics*, edited by D. R. Lide (CRC Press, Boca Raton, FL, 1995), Sec. 12, p. 126.
[13] It was seen that particles which were not bonded well to the surface could be removed when the bias reached the value used, -3.0 V.
[14] R. Berndt and J. K. Gimzewski, Int. J. Mod. Phys. B **7**, 516 (1992).
[15] A. W. McKinnon, M. E. Welland, T. M. H. Wong, and J. K. Gimzewski, Phys. Rev. B **48**, 15 250 (1993).