Quantitative Determination of Nanoscale Electronic Properties of Semiconductor Surfaces by Scanning Tunnelling Spectroscopy

Randall M. Feenstra  
*Carnegie Mellon University, feenstra@andrew.cmu.edu*

S. Gaan  
*Carnegie Mellon University*

Follow this and additional works at: [http://repository.cmu.edu/physics](http://repository.cmu.edu/physics)

Published In  
Quantitative Determination of Nanoscale Electronic Properties of Semiconductor Surfaces by Scanning Tunnelling Spectroscopy

R. M. Feenstra\(^1\) and S. Gaan

Dept. Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Email: feenstra@cmu.edu

Abstract. Simulation of tunnelling spectra obtained from semiconductor surfaces permits quantitative evaluation of nanoscale electronic properties of the surface. Band offsets associated with quantum wells or quantum dots can thus be evaluated, as can be electronic properties associated with particular point defects within the material. An overview of the methods employed for the analysis is given, emphasizing the critical requirements of both the experiment and theory that must be fulfilled for a realistic determination of electronic properties.

1. Introduction

The scanning tunneling microscope (STM) enables direct imaging of structures on semiconductor surfaces with lateral resolutions on the order of 0.3 nm and vertical resolution of 0.01 nm or better [1]. It is well known however that the images actually reveal the electronic charge density (integrated over an energy window with width on the order of an eV on either side of the Fermi energy) [2,3]. Hence, deriving atomic structures and/or compositions is challenging. Comparison of the measured results to theoretical predictions, over a range of sample-tip bias voltages, has proven to be an effective means of determining structural models. Spectroscopic measurements, i.e. scanning tunneling spectroscopy (STS), provide important additional input for this type of analysis [4,5,6]. For semiconductor surfaces with complex reconstructions, the aforementioned methods provide a powerful means of determined the structure of the surfaces. (For \textit{elemental} semiconductors, structures can sometimes be deduced directly from STM images, but for \textit{compound} semiconductor the added complexity demands the use of theoretical predictions of structures [7]).

A somewhat different class of problems that can be investigated by STM/S is associated with surfaces for which the arrangement of surface atoms is quite simple and for which there is \textit{not} a high density of surface states throughout the band gap, e.g. the nonreconstructed (110) surfaces of cleaved III-V semiconductors [8] or passivated surfaces such as H-terminated Si(001) [9]. In many cases, large well-ordered regions of such surfaces can be prepared, and then the presence of isolated defects, often intentionally introduced into the structure, can be achieved. Also, heterostructures of e.g. III-V semiconductor can be cleaved and viewed in cross-section, thus revealing quantum wells, quantum dots, or other similar nanostructures [10].

In these sort of surfaces with relatively simple structure and low density of surface states, an important effect that occurs in STM/S is tip-induced band bending – the extension of the electric field...
across the vacuum gap into the semiconductor as illustrated in figure 1. This tip-induced band bending can have dramatic effects on the spectra, not only producing shifts in the position of spectral lines or edges, but also greatly affecting the magnitude of the tunneling current or conductance over certain voltage ranges. In order to obtain accurate energy values from spectra (e.g. for spectral lines, band offsets, etc.) it is necessary to quantify the effects of tip-induced band bending on the spectra.

Figure 1. (a) Schematic diagram of energy bands with tip-induced band bending, showing the valence band maximum \( E_V \) and the conduction band minimum \( E_C \). The sample Fermi level is denoted by \( E_F \) with the tip Fermi level at \( E_F + eV \) where \( V \) is the applied sample-tip bias voltage. The band bending at the surface is denoted by \( U_0 \). Quantum effects within the semiconductor are illustrated in (b) and (c) for wavefunction tailing (arrow) through a depletion region and for localized accumulation state formation, respectively.

Early studies of semiconductor surfaces permitted computation of tip-induced band bending in a qualitative or semi-quantitative manner [8,11-13]. In the past decade a realistic method for obtaining the band bending in three dimensions (3D) [14] has been developed, and associated computer codes are available [15]. With those electrostatic potentials, computations of tunnel current are relatively straightforward within an effective mass (envelope function) treatment. Two methods have been developed for such computations, one employing a planar tunneling geometry [16] and the other treating a nonplanar tip with a plane wave expansion method [17]. The former yields a detailed spectrum of states with arbitrarily fine energy spacing, although energy levels are approximate since the planar geometry has been assumed (a proper nonplanar tip is used in computing the potential, but not for the current). The latter method permits a correct treatment of the tip geometry, but because of computational limitations in the plane wave expansion, only a small (≈10 nm on a side) region in the semiconductor around the point opposite the tip apex can be treated. Nevertheless, the results of the two methods are in good agreement for situations where the radial dependence of the potential is not too large (i.e. when no region of laterally localized potential exists in the semiconductor).

In this paper we provide an example of computed tunneling spectra for a semiconductor system, namely InAs quantum dots in GaAs. We demonstrate how, through fitting of the computed spectra to experiment, the tip-related parameters such as tip radius, tip-sample separation, and tip-sample contact potential (work function difference) can be determined. Furthermore, additional features in the observed spectra, discrete states in quantum dots in the present case, can be derived from the observed spectra. We also briefly discuss limitations in the computational methods, particularly those associated with non-equilibrium occupation of carriers in the semiconductor.
2. Experimental Methods

In addition to the usual requirements in any STM/S experiment of stable operation of the microscope and clean metallic probe tips, comparison of experiment to theory for detailed semiconductor spectroscopy has several additional experimental requirements. First, it is crucial that the spectra have high dynamic range, i.e. extending over 3 – 4 orders of magnitude at least, in order to permit a comprehensive comparison with computed results. As illustrated many years ago, band edges are in general not well defined for spectra that extend only over 1-2 orders of magnitude [8]. One convenient means of achieving this high dynamic range, and simultaneously acquiring the data in a short period of time (a few seconds), is to employ variable tip-sample spacing during the measurement [18]. The separation is varied as \( s = s_0 + \Delta s \) with \( \Delta s = a|V| \), using a value for \( a \) of about 0.1 nm/V. For comparison of theory to experiment, the data is normalized to constant separation using a multiplicative factor of \( \exp(2\kappa\Delta s) \), where a voltage-averaged value of \( \kappa \) obtained from experiment is employed. Ideally the \( \kappa \) value would be close to a theoretically ideal value of \( \approx 10 \text{ nm}^{-1} \), although in some cases lower \( \kappa \) values are found in the experiments. One possible reason for this nonideality are transport limitations in the current, further discussed in section 4. In any case, so long as the \( \kappa \) value does not deviate too much from an ideal one, then this method of normalization seems to be adequate. (A normalized conductance of the form \( (dI/dV)/(V^2) \) is very convenient for qualitative display of experimental data [6,18], but this normalization itself can somewhat distort the data so that for detailed quantitative analysis we find that the conductance at constant separation to be a better quantity).

A second important requirement for the experimental data is that spatially resolved spectra must in general be acquired, both at some typical location of the bare semiconductor surface and at specific locations near the region of interest on the surface. The former spectra are needed in order to enable determination of the tip-related parameters – tip radius, tip-sample separation, and tip-sample contact potential. With those, then spectra acquired near the region of interest can be used to determine values of additional parameters associated with that region. The simultaneous need for high dynamic range and spatially resolved spectra is a significant demand in STM/S work, since thermal drift in the instrument and/or unintentional changes in the tip structure can lead to incomplete data sets (use of a low temperature STM can aid in minimizing these instrumental difficulties).

3. Computational Methods

The geometry of the problem consists of a tip-vacuum-semiconductor junction. To obtain the electrostatic potential a finite-element method has been developed that can efficiently solve the 3D problem including possible occupation of electronic states on the surface or in the bulk of the semiconductor [12]. The solution is found iteratively, thus permitting evaluation of Poisson's equation including a possible high degree of nonlinearity in the charge density. In the absence of any electrostatic potential in the semiconductor, the charge density for a bulk band is denoted by \( \rho(E_F) \) and for a surface state by \( \sigma(E_F) \) where \( E_F \) is the Fermi energy (the temperature dependence of \( \rho \), \( \sigma \), and \( E_F \) are fully included in our computations but are not indicated here for ease of notation). In the presence of an electrostatic potential energy \( U(x,y,z) \) (i.e. \( -e \) times the electrostatic potential where \( e \) is the fundamental charge), the charge densities are given in a semi-classical approximation simply by \( \rho(E_F - U(x,y,z)) \) for a bulk band or \( \sigma(E_F - U(x,y,0)) \) for a surface band. The electrostatic potential energy is given by Poisson's equation

\[
\nabla^2 U = \frac{\rho(E_F - U(x,y,z))}{e \varepsilon \varepsilon_0} \tag{1}
\]

where \( \varepsilon_0 \) is the permittivity of vacuum, \( \varepsilon \) is the relative permittivity (dielectric constant) of the semiconductor, and with \( \rho = 0 \) in the vacuum. The boundary condition at the surface is given by
with the semiconductor spanning the half-space \( z < 0 \). For the finite-element evaluation, in the vacuum a generalization of prolate spheroidal coordinates \((\xi, \phi, \eta)\) are used so as to match the shape of the probe tip [19], and in the semiconductor cylindrical coordinates \((r, \phi, z)\) are used. In both cases the spacing between grid points is not constant but increases both with the radius \( r \) in the semiconductor (or with \( \xi \) in the vacuum) and with the depth \( z \) into the semiconductor. In this way, solutions out to very large (effectively infinite) values of those coordinates can be obtained, with \( U \rightarrow 0 \) at those large values. On the surface of the probe tip \( U = eV + \Delta U \) where \( V \) is the sample-tip voltage and

\[
\Delta U = \phi_m - \chi - (E_C - E_F)
\]

is the contact potential (work function difference) between tip and sample, where \( \phi_m \) is the tip work function, \( \chi \) is the semiconductor electron affinity, and \( E_C - E_F \) is the separation of conduction band (CB) edge deep inside the semiconductor from the Fermi energy.

The iterative method that has been developed permits solution of equations (1) and (2) for any specified set of bulk and/or surface bands. Also, solutions are possible not only for the semi-classical case described in those equations but also for a fully quantum mechanical treatment of particular states, as needed e.g. for a self-consistent treatment of accumulation or inversion situations in the semiconductor [20] (in which case the bulk charge density is expressed as functional of \( U(x, y, z) \)).

An example of a computed electrostatic potential is shown in figure 2, for the case of an InAs quantum dot (QD) embedded within GaAs [17]. The presence of the QD does not affect the potential (assuming equal dielectric constants for the InAs and GaAs), but it does affect the tunnel current by introducing electron and hole states within the GaAs band gap. The tip apex in figure 2 is positioned 4 nm from the center of a lens-shaped QD (shown in grey), and a strip of charge density associated with a surface step is located 6 nm on the other side of the QD (the physical step is not represented in the theory, only the charge density associated with the step). A constant state-density throughout the bandgap arising from the step is assumed, with value \( 2.5 \text{ nm}^{-1}\text{eV}^{-1} \) and charge-neutrality level 0.25 eV above the VB maximum [17].

The above method for evaluating the electrostatic potential provides an exact solution. In contrast, evaluation of the tunnelling current requires some approximations. It should be noted that quantum effects within the semiconductor as illustrated in figures 1(b) and (c) can be very important in determining the current, so a treatment of the problem using e.g. only a transmission factor for the

![Figure 2. Electrostatic potential energy for a 2-nm radius-of-curvature probe-tip located 1.0 nm from a semiconductor surface. The sample-tip voltage is set at +1.0 V and the contact potential between tip and sample is -0.87 eV, so the electrostatic potential energy of the tip relative to a point deep inside the semiconductor is +0.13 eV. Contours are shown for potential energies (eV) of 0.155 (red), 0.232 (green), 0.310 (cyan), 0.388 (blue), 0.465 (magenta). At distances further inside the semiconductor the potential falls gradually to 0 eV. (From Ref. [17]).](image)
tunneling through the vacuum would be very unrealistic. It is necessary to compute the wavefunctions through both the vacuum and the semiconductor region where the potential is varying. From the wavefunctions, the current can be obtained using the Bardeen formalism [21], written for the case of a sharp tip by using the Tersoff-Hamann approximation [3,17],

$$I = \frac{8\pi e h}{m} R^2 k_F \sum_{\mu} \left[ f(E_{\mu}) - f(E_{\mu} - eV) \right] \left| \Psi_{\mu}(0,0,s) \right|^2$$

(4)

where $m$ is the free-electron mass, $R$ is the tip radius-of-curvature, $k_F$ is its Fermi wavevector, $E_{\mu}$ is the energy of an eigenstate of the sample, $f(E_{\mu})$ is the Fermi-Dirac occupation factor for that state, $f(E_{\mu} + eV)$ is the occupation factor for the corresponding state in the tip, and $\Psi_{\mu}(0,0,s)$ is the wavefunction of the sample state evaluated at the position of the tip apex: $x = 0$, $y = 0$, and $z = s$ where $s$ is the tip-sample separation. The wavefunctions are evaluated within an effective mass (envelope function) approximation, writing Schrödinger's equation for a CB as

$$- \frac{\hbar^2}{2m_e} \nabla^2 \psi_{\mu} + \left( E_C + U(x,y,z) \right) \psi_{\mu} = E_{\mu} \psi_{\mu},$$

(5a)

for a valence band (VB) as

$$+ \frac{\hbar^2}{2m_h} \nabla^2 \psi_{\mu} + \left( E_V + U(x,y,z) \right) \psi_{\mu} = E_{\mu} \psi_{\mu},$$

(5b)

for the probe tip as

$$- \frac{\hbar^2}{2m} \nabla^2 \psi_{\mu} + \left( E_F + eV - W \right) \psi_{\mu} = E_{\mu} \psi_{\mu},$$

(5c)

and for the vacuum as

$$- \frac{\hbar^2}{2m} \nabla^2 \psi_{\mu} + \left( E_C + \chi - U(x,y,z) \right) \psi_{\mu} = E_{\mu} \psi_{\mu}$$

(5d)

where $E_V$ and $E_C$ are fixed locations of the band edges (i.e. corresponding to their values deep inside the semiconductor), $m_e$ and $m_h$ are effective masses, and $W$ is the width of the metallic band in the semiconductor below its Fermi energy. In a typical computation, a single CB and three VBs (light-hole, heavy-hole, and split-off) are used. At the semiconductor surface, the wavefunctions are taken to be continuous and their derivatives with respect to $z$ are related by

$$\frac{\partial \psi_{\mu}}{\partial z}(x,y,0_{+}) = \frac{m_i}{m_{\mu}} \frac{\partial \psi_{\mu}}{\partial z}(x,y,0_{-}).$$

(6)

where $m_i$ equals $m_e$ or $m_h$, as appropriate. At the surface of the probe tip, both the wavefunctions and their derivatives are continuous.

Two schemes have been employed for evaluating the wavefunctions. In the first, the tunnel current is written in a form appropriate for planar tunneling. It can be argued that this approximation is equivalent to a semi-classical treatment of the radial (and angular) components of the wavefunctions while maintaining a fully quantum-mechanical treatment of their $z$-component [16]. The $z$-components of the wavefunction are then obtained directly by numerical equation of the corresponding $z$-part of the Schrödinger equation. This approximation for the wavefunctions is valid so long as there are no regions of localized potential in the semiconductor and the probe tip is not too sharp (and actually it works quite well even for tip radii as small as 1 nm) [17].

Regions of localized potential are what occur e.g. around quantum dots, so that in such cases the planar geometry is not suitable and one must employ a type of plane wave expansion method: Plane waves in the $z$-direction are matched to decaying exponentials in the vacuum, with the wavefunctions thus obtained also incorporating the boundary condition of equation (6) [17]. Solution of the
Schrödinger equation is then accomplished with the usual eigenvalue method. Such methods are very computationally demanding, even for a modest number of plane waves (≈10) in each of the x, y, and z directions. Consequently, only a relatively small region of the semiconductor (≈10 nm in each direction) is sampled by the method. Nevertheless, this small region turns out to be sufficient for providing a reasonable description of a tunnelling spectrum, at least for temperatures that are not too low. Results using both of these methods, in the absence of a localized potential in the semiconductor, are in good agreement with each other [17].

Computations performed as described above yield self-consistency between the potential and the current for situations of semiconductor depletion, which is the usual case in most spectra. But, for situations of accumulation or depletion, the semi-classical charge density is not realistic. In those cases, a quantum-mechanical charge density must be formed from the wavefunctions, and the potential re-evaluated with this new charge density. A self-consistent solution is then obtained iteratively. This method provides a complete solution of the problem when planar evaluation of the current is used [20], but for the plane wave expansion method the new charge density is obtained only over a restricted region of the semiconductor. Some method to extend that charge density over the entire semiconductor would be needed, and a detailed formulation of that procedure has not been derived to date.

4. Comparison of theory to experiment

It should be emphasized that all of the results obtained from the computational methods described in section 3 depend entirely on the input parameters. The basic parameters for any STS tunnelling computation on a semiconductor surface are tip radius-of-curvature R, tip-sample separation s, tip-sample contact potential ΔU (i.e. work function difference), and an effective tunnelling area of the apex of the tip. This latter parameter could be deduced from the tip radius for a perfectly spherical tip apex, but we take it to be an independent parameter (which affects only the overall magnitude of a spectrum), to allow for a small protrusion on the tip apex. In general the values of these parameters are not something that are of interest per se, but rather, their values are needed only to obtain a spectrum that agrees overall with experiment. Then, in a typical study, there will be one or more additional parameters (such as band offsets) whose values are quantities of interest. Comparison between experiment and theory therefore reduces to a curve fitting process, often involving several spectra acquired/simulated at different points. With the four tip-related parameters, together with additional parameters (such as band offsets) associated with the point of interest or other surface phenomena (such a surface charge), many thousands of evaluations of the theoretical spectra are required.

Evaluation of a single current at one sample-tip voltage requires a time ranging from minutes to hours depending on the complexity of the problem, and typically there might be 30 such evaluation needed for a spectra (15 voltage points, and 2 evaluation at each in order to obtain the conductance dV/dI). Hence, significant computational resources (multiple processors) are required.

An example of a detailed comparison between experiment and theory is given in figure 3, showing spectra for the InAs QD already introduced in figure 2. A spectrum acquired 6 nm from the QD is shown in (a), revealing states associated with the bare GaAs surface. A spectrum containing features arising from the QD confined states is shown in (b), acquired 4 nm from the QD center (at distances nearer to the QD these confined-state features grow in size, but charging of the QD is a problem, so the analysis is restricted to only this QD spectrum for which the current through the QD states is relatively small) [17]. For the spectrum of the bare GaAs, this would normally be fit using four parameters, s, R, ΔU, and an overall magnitude for the spectrum. In the present case however, atomic steps existed on this cleavage surface (arising from strain in the heterostructure due to the QDs). Very little tip-induced band bending is apparent in the spectra of figure 3(a) (i.e. the apparent bandgap is close to the GaAs gap of 1.42 eV), revealing the influence of the states of the steps. These states are modelled using a energetically uniform state-density across the bandgap, with density σ and charge neutrality level E_N. Consequently, six parameters are used to fit the spectrum of the bare GaAs. This is a large number of parameters, so the model possibly overfits the experimental data, hence producing
rather high correlation between the $\Delta U$ and $E_N$ parameters and values for the $R$ parameter that are somewhat smaller than expected on the basis of other investigations [16]. But importantly, the added features found with the GaAs bandgap region for the QD spectrum are not at all dependent on these parameter values of $s$, $R$, $\Delta U$, $\sigma$, or $E_N$, so those parameters only serve to provide a "background" description of the bulk GaAs part of the spectra from which the QD features can then be separated.

Figure 3. Tunnelling spectra, acquired (a) 6 nm from an InAs quantum dot (QD) and (b) 4 nm from the QD. Experimental results are shown by lines and theoretical fits by circles. The curves (a) and (b) are shifted from each other for ease of viewing. Values for the fitting parameters are listed. The location of states from the theory is shown at the bottom of the figure, giving the voltages at which the tip Fermi energy is aligned with QD states (for the VB, only light-hole states are shown) and with the hatched regions showing unconfined VB and CB states in the near-surface region. (From Ref. [17]).

To describe the QD, a minimum of two additional parameters are needed: the offsets between the InAs and GaAs CBs and VBs, $\Delta E_C$ and $\Delta E_V$, respectively. With those two parameters the theory yields confined electron and hole states for the QD. However, it turns out that for the simple effective mass theory used here that the density of such states (i.e. number per unit energy) is significantly underestimated [22], so that even if the lowest energy state is correctly found in the theory then the next higher energy state would be incorrect. To account for this failing of the theory, one additional parameter was employed, a scale factor $\alpha$ by which the overall size of the QD was increased. The larger QD then increases the density of states such that it can match the experiment. The final fit is shown in figure 3, where the values of all nine parameters (the eight listed there, plus one overall magnitude) is found by simultaneous fitting to both experimental spectra. In this way, the energies of the confined electron and hole states are determined, as schematically indicated in the lower part of the figure. These energies, in conjunction with the size, shape, and composition of the QDs determined from the STM images and considerations of elastic strain relaxation in the material [17], thus constitute a complete determination of the electronic and structural properties for this nanostructure. This information is useful since it can be used to distinguish between current predictive theoretical models for the confined state energies associated with QDs [17].

5. Summary
In summary, we have discussed in this work the application of computational methods to the quantitative analysis of tunnelling spectra from semiconductor surfaces. The electrostatic potential in the semiconductor, arising from the voltage difference between tip and sample as well as from occupation of electronic states on the surface or in the bulk of the semiconductor, can be obtained using an iterative finite-element method. Tunnel currents are then computed within the effective mass (envelope function) approximation by using either a semi-classical treatment of the lateral parts of the
wavefunction or a completely quantum-mechanical treatment but one that is restricted to a relatively small region around the tunnel junction.

A number of limitations of the present theory exist, as discussed elsewhere [22]. One limitation in particular arises from the assumption that the tunneling process is the rate-limiting step in the transport of carriers from the probe-tip to the semiconductor, so that an "equilibrium" situation exists in which the occupation of carriers in the semiconductor is simply described by a constant Fermi energy. But in actuality, situations arise (at low temperature, or low doping, or for confined states, etc.) in which transport in the semiconductor itself is somewhat limited. In those cases, a much more complicated theory is needed to describe the distribution of carriers and resulting charge densities on the surface and in the bulk of the semiconductor. Such a theory could prove to be useful in enabling the determination of transport parameters for the carriers, even in nanoscale situations as occur in the STM geometry.

Acknowledgements
This work was supported by the U.S. National Science Foundation, grant DMR-0856240.

References