IOPscience iopscience.iop.org

Home Search Collections Journals About Contact us My IOPscience

Defect-induced anomalies in STM images of semiconductors

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

1995 J. Phys.: Condens. Matter 7 5155

(http://iopscience.iop.org/0953-8984/7/27/004)

View the table of contents for this issue, or go to the journal homepage for more

Download details:
IP Address: 171.66.16.151
The article was downloaded on 12/05/2010 at 21:36

Please note that terms and conditions apply.

J. Phys.: Condens. Matter 7 (1995) 5155–5162. Printed in the UK

Defect-induced anomalies in STM images of semiconductors

D Lawunmi and M C Payne

Cavendish Laboratory, University of Cambridge, Madingley Road, Cambridge, CB3 0HE, UK

Received 5 December 1994, in final form 10 March 1995

Abstract. Transport mechanisms by which charge can flow from the surface of the semiconductor to the bulk of the semiconductor are studied. It is shown that these processes can have a significant effect on the apparent band gap of a semiconductor and can cause the STM current to perturb regions of the surface that have a large lateral displacement from the region of the surface into which the tunnel current was originally injected.

The scanning tunnelling microscope (STM) operates by bringing a sharp metallic tip close enough to a surface for a tunnel current to pass between the two. The tip is moved across the surface and a topograph is generated either from vertical displacements of the tip required to maintain a constant current flowing between the tip and the surface or from the variation of the tunnel current with the position of the tip as the tip is scanned across the surface. For a review of the STM see [1]. In the case of the semiconductor surfaces, the tunnel current flows predominantly into states localized at the surface of the semiconductor. The charge that is injected into these localized states must be transferred into the bulk of the semiconductor to complete the current path through the device. Although the currents used in the STM are extremely small, typically of the order of 1 nA, the current densities generated in the region analysed by the STM are extremely high, of the order of 10^9 A m^{-2}. In this letter we consider the mechanisms that allow current flow between the surface states and the bulk states and discuss the implications of this for analysing STM images of semiconductors.

The mechanisms that are expected to provide the most significant routes for charge from the surface to enter the bulk of the semiconductor are tunnelling and recombination processes such as two-body Auger processes and phonon emission processes [2]. Tunnelling processes will be considered first and recombination processes will be considered later.

The tunnel current density for tunnelling from a localized surface state into the bulk of the semiconductor will be calculated numerically using the tunnelling formalism originally developed by Oppenheimer [3] with wavefunctions for the bulk of the semiconductor and the localized surface state calculated using the WKB approximation. The probability per unit time for tunnelling from a surface state into the bulk of the semiconductor may be written as

$$P_{tunn} = \sum_n \frac{2\pi}{\hbar} |\langle\Psi_n|qV(z)|\Psi_s\rangle|^2 \, \delta(E_s - E_n) \tag{1}$$

where the sum is over bulk states, Ψ_n and Ψ_s are the wavefunctions of bulk states and the surface state, respectively, E_s is the energy of the surface state measured relative to the band edge at the surface as illustrated in figure 1 and E_n is the energy of bulk state n, q is

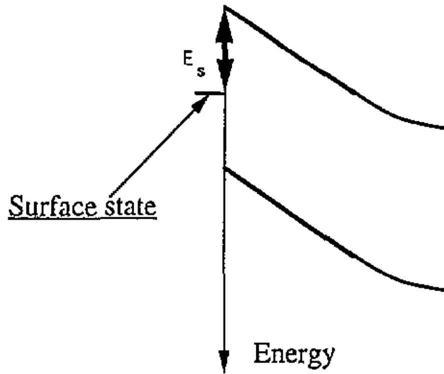

Figure 1. Schematic illustration of band bending at a semiconductor surface showing the value of E_s, which gives the energy of the surface state relative to the semiconductor band edge at the surface.

the charge on the electron and $V(z)$ is the potential in the surface depletion region of the semiconductor. In the depletion approximation $V(z)$ is given by [4]

$$V(z) = (qN_D w^2/2\varepsilon 0\varepsilon)[1 - z/w]^2 \tag{2}$$

where w is the width of the surface depletion region, N_D is the doping density in the semiconductor and ε_r is the relative dielectric constant of the semiconductor. The surface state wavefunction, Ψ_s, will be modelled by the wavefunction for a Dirac δ function potential well at the surface [5]. The tunnelling probability depends very sensitively on the effective mass in the tunnelling direction. This varies with the semiconductor and the wavevector of the tunnelling electron parallel to the surface. In order to simplify the discussion we shall initially assume a single value for the effective mass of $0.26m_e$, where m_e is the mass of a free electron, which is appropriate for tunnelling from the (111) surface of silicon [6]. The variations of the tunnel current with the energy of the surface state, E_s, for systems with various doping densities are shown in table 1. The enormous variation of tunnel current with both N_D and E_s can be clearly seen. For the highest doping density of 10^{24} m^{-3} the current of 1 nA injected into the surface by the STM can be dissipated to the bulk by tunnelling from a single surface state provided that this state is within 0.4 eV of the band edge. However, this situation changes significantly on decreasing the doping density and at a doping density of 10^{22} m^{-3} only when the surface state is less than 0.1 eV away from the band edge at the surface does the tunnel current to the bulk reach the value of 1 nA required to sustain a typical STM current. As the wavevector parallel to the surface, $k_\parallel$, is a good quantum number for elastic tunnelling these values of E_s are measured from the band edge at the same value of $k_\parallel$. Hence, these values of E_s are measured with respect to the direct band gap rather than the indirect band gap. Thus, even in the case of the highest doping density, elastic tunnelling from the single surface state to the bulk is capable of sustaining a typical STM current only over a very small range of the possible values of E_s. However, scattering processes can change the value of $k_\parallel$ and surface states which have values of $k_\parallel$ far from either the valence band maximum (VBM) or conduction band minimum (CBM) can produce significantly larger inelastic tunnel currents than elastic tunnel currents through scatterings that change the value of $k_\parallel$ for the tunnelling electron so that it lies near to either the VBM or the CBM. Such inelastic tunnel currents are typically no more that 1%

of the maximum elastic current [7]. In order to estimate the maximum current generated by inelastic tunnelling from a single surface state the values of the currents shown in table 1 should be multiplied by a factor of 0.01 and E_s measured from either the VBM or the CBM. Hence, for systems with a very high doping density of 10^{24} m^{-3} inelastic tunnelling will be capable of dissipating an STM current of 1 nA from any surface state which is within 0.35 eV of the VBM or the CBM. However, for all values of doping density, when the surface state lies close to the middle of the band gap, the required current flow to the bulk cannot be supported by tunnelling from a single surface state. At lower doping densities this is true for almost all the surface states except those very close to the VBM or the CBM. Finally, we shall briefly discuss the effect of varying the effective mass of the tunnelling electron from the value of $0.26m_e$ that we have assumed up to this point. If the effective mass is larger than $0.26m_e$ there is a larger range of energies for which either elastic or inelastic tunnelling cannot dissipate a current of 1 nA from a single surface state. In contrasts, if the effective mass is smaller than $0.26m_e$ there is a smaller range of energies for which either elastic or inelastic tunnelling cannot dissipate a current of 1 nA from a single surface state to the bulk, and below a critical value of the effective mass, the value of which will depend on the doping density, there will be no surface states which cannot dissipate a current of 1 nA to the bulk by either elastic or inelastic tunnelling.

Table 1. Variation of current from a single surface state with E_s, which is defined in figure 1, for tunnelling from the (111) surface of silicon for various doping densities, N_D. Currents smaller than 10^{-20} are not shown.

	Tunnel current (A)		
E_s (eV)	$N_D = 10^{24}$ m^{-3}	$N_D = 10^{23}$ m^{-3}	$N_D = 10^{22}$ m^{-3}
0.05	3×10^{-2}	5×10^{-3}	2×10^{-5}
0.1	2×10^{-2}	5×10^{-5}	3×10^{-13}
0.2	5×10^{-4}	1×10^{-11}	—
0.3	1×10^{-6}	—	—
0.4	7×10^{-10}	—	—
0.5	6×10^{-14}	—	—
0.6	2×10^{-19}	—	—

We shall now discuss transfer of carriers between surface and bulk states by the most important recombination processes, namely phonon emission processes, Auger recombination and radiative processes. Phonon emission processes can be divided into two categories. Cascade processes occur when an electron which is initially in some excited state loses energy, mainly by a cascade of one-phonon transitions between states located in the band gap. Multi phonon transitions occur when a number of phonons are emitted during the movement of an electron or hole between two sates in the band gap, or between a band gap state and the bulk. Multi phonon processes are clearly essential when the energy difference between the two energy levels involved in a transition exceeds the maximum one-phonon energy. The Auger processes relevant to the present problem are caused by Coulomb interactions of an electron or hole in a surface state with electrons or holes in the bulk of the semiconductor. The Auger recombination rate increases with the density of carriers since the recombination rates for the processes which involve interactions between an electron initially in a trap state and an electron in the bulk conduction band or between a hole initially in a trap state and a hole in the bulk valence band vary as $T_1 n^2 p_t$ and $T_2 p^2 n_t$, respectively, where T_1 and T_2 are Auger recombination coefficients, n and p are

the densities of electrons and holes in the conduction and valence bands and p_t and n_t are the number of full and empty trap states [8]. The threshold density of carriers at which Auger processes becomes significant is about 10^{22} m^{-3} [9].

The cross-sections for bulk recombination for the recombination mechanisms have been collated by Stoneham [2]. Theoretical models have been developed for calculating these recombination rates [9]. However, the resulting expressions are extremely complex and the final values for the recombination rates depend on a large number of parameters, some of which have not been determined very accurately. Unfortunately, this can lead to a large margin of error in the estimates of the recombination rates. Owing to the sensitivity of the model calculations of the recombination rates and the limited knowledge of many of these parameters, estimates of recombination rates based on Stoneham's values will be used to analyse the rate at which charge leaves a surface state of the semiconductor and enters the bulk.

The rate at which electrons or holes are captured from the bulk bands can be determined directly from the capture cross-section of the surface state, σ. The probability per unit time that a surface state captures an electron or hole from the bulk bands is proportional to the density of electrons (or holes) in these bands and can be expressed as [10]

$$P_{e,cap} = n\sigma_e v_e \tag{3a}$$

$$P_{h,cap} = p\sigma_h v_h \tag{3b}$$

where $P_{e,cap}$ is the probability per unit time that an electron is captured from the conduction band, and $P_{h,cap}$ is the probability per unit time that a hole is captured from the valence band, σ_e and σ_h are the capture cross-sections for electron and holes, respectively, and v_e and v_h are the thermal velocities of electrons and holes. The probabilities for electron and hole emission from the surface state to the bulk, $P_{e,em}$ and $P_{h,em}$, are related to the capture probabilities through detailed balance and are given by

$$P_{e,em} = \sigma_e v_e n_i \exp\{(E_s - E_i)/kT\} \tag{4a}$$

$$P_{h,em} = \sigma_h v_h n_i \exp\{-(E_s - E_i)/kT\} \tag{4b}$$

where n_i is the intrinsic carrier density, E_s is the energy of the surface state and E_i is the intrinsic energy level at the surface of the semiconductor.

In order to determine the current flows that can be sustained by the recombination mechanisms we shall again use the silicon(111) 2×1 surface as our model system. The Fermi level is pinned at this surface at an energy 0.4 eV above the valence band edge [11] giving carrier densities at the surface at room temperature of the order 10^{19} m^{-3} for the majority carriers (holes) and 10^{13} m^{-3} for the minority carriers (electrons) and a thermal velocity of these carriers of the order of 10^5 m s^{-1}. The current flows sustainable by the different recombination mechanisms will be compared by calculating the number of surface states required to sustain a current of 1 nA from the surface to the bulk. The number of surface states, n_s, required to sustain a current, I, flowing between the tip and the surface is

$$n_s = I/(Pq) \tag{5}$$

Table 2. Typical cross-sections for dominant capture mechanisms and the number of surface states required to dissipate a current of 1 nA from the surface to the bulk for capture and emission processes.

	Radiative	Auger	Phonon
Typical cross-section (Å^2)	10^{-5}–10^{-4}	10^{-1}	10^{-2}–10^{4}
Number of surface states for			
capture of majority carrier	10^{11}–10^{10}	10^{7}	10^{8}–10^{2}
capture of minority carrier	10^{17}–10^{16}	10^{13}	10^{14}–10^{8}
Emission from			
midgap state	10^{15}–10^{14}	10^{11}	10^{12}–10^{6}
quarter-gap state	10^{10}–10^{9}	10^{6}	10^{7}–10
band edge state	10^{5}–10^{4}	10	10^{2}–1

where P is the probability for the relevant electron or hole capture or emission process.

Table 2 shows the numbers of surface states required to sustain a current of 1 nA by capture of majority and minority carriers and by emission processes from surface states whose energies are at the middle of the band gap at the surface (midgap) a quarter of the way across the band gap at the surface (quarter-gap) and at the energy of a band edge at the surface (band edge). It can be seen from this table that none of the capture processes are capable of sustaining a current of 1 nA from a single surface state. These capture rates are relatively small because of the low carrier density in the region of the surface. However, as this carrier density is essentially unaffected by changes in the bulk doping density if the Fermi level is pinned at the surface we expect that this conclusion will apply to many semiconductor systems. Emission processes from states whose energies lie in a broad range around the middle of the band gap also cannot sustain a current of 1 nA from a single surface state. However, as the energy of the surface state approaches a band edge it is possible for emission processes via the phonon mechanism to sustain a current of 1 nA from a single surface state provided that the cross-section for the phonon recombination mechanism does not lie towards the bottom of the possible range of values.

The analysis presented above shows that in the case of a semiconductor surface for which the surface states lie relatively close to the middle of the gap neither tunnelling nor recombination occurs sufficiently rapidly to dissipate a typical STM current from a single surface state. However, the current will only continue to flow in such systems if the current out of the surface state is the same as the current injected into the state by the STM. If the carriers are not transferred directly to the bulk fast enough to produce a current of 1 nA, then the only possible mechanism for sustaining a continued current flow in the STM is for the carriers injected into the surface state to spread laterally across the surface. Then the required current can be transferred from the surface to the bulk using one of the mechanisms described above because the current density is reduced if the tunnelling or recombination occurs over an area that is larger than that of a single surface state. There is some intriguing experimental evidence supporting the idea of lateral spreading on the surfaces of semiconductors. Areas on the 2×1-reconstructed (100) surface of silicon with sizes of the order of 2000 Å^2 which appeared as diffuse bright circles in the STM images within which the 2×1 reconstruction could be clearly resolved have been observed [12].

The weakness of the analysis presented above is that we have assumed that the rates of surface recombination processes are similar to those of bulk recombination process. To show that these results are representative and to provide a quantitative value for the extent of the lateral spreading we shall consider the case of the 2×1 reconstruction on the (111) surface of silicon for which the surface recombination parameters have been obtained experimentally

by Halas and Bokor [13]. Bokor and Halas measured the decay of the number of excess carriers on the 2×1-reconstructed Si(111) surface following the application of a short laser pulse. They used these measurements to parametrize equations for the carrier transfer between the surface states and the bulk bands. Their equations describing the current flows between carriers in the surface and the bulk bands at time t are

$$J_n(t) = k_1\{n(t)[N_t - \pi^*(t)] - n_i\pi^*(t)\exp[(E_{\pi^*} - E_i)/kT]\} \tag{6a}$$

$$J_p(t) = -k_2 p(t)[N_t - \pi(t)] + \pi(t)/\tau_h \tag{6b}$$

where J_n is the electron current into the conduction band, J_p is the hole current into the valence band, π and π^* are the populations of electrons and holes in the bonding and antibonding surface states, E_{π^*} is the energy of the π^* band minimum, N_t is the total density of bonding and antibonding surface states, $1/\tau_h$ is the rate at which holes are emitted from the π band to the valence band ($P_{h,em}$ of equation (4b)) and k_1 and k_2 are equal to the product of the capture cross-section and thermal velocities (σv of equation (3)) for electrons and holes, respectively. The rapid communication with the bulk states for the π states required Bokor and Halas to include capture and emission processes for this band explicitly rather than relating their rates through detailed balance. The values of the parameters extracted from the experimental measurements are $k_1 = 3.33 \times 10^{-16}$ m^3 s^{-1}, $k_2 = 3.33 \times 10^{-13}$ m^3 s^{-1} and $\tau_h = 200$ ps. The extent of the lateral spreading expected for the Si(111) 2×1 surface can be determined by using these parameters to calculate the number of states required to sustain a surface to bulk current of 1 nA. For the case of holes injected into the π band we find that a current of 1 nA can be dissipated from a single surface state due to the very large value of $1/\tau_h$. However, for electrons injected into the π^* band 10^{13} surface states are required to dissipate the current to the bulk. The difference between the numbers of states required to dissipate the current for the π and π^* bands reflects the energies of the states in question. The π band is very close in energy to the valence band edge. This proximity to a band edge increases the rates of all the mechanisms that can transfer carriers from surface to bulk states. In contrast the π^* band lies close to the middle of the gap and the transfer rates from surface to bulk are much slower; consequently a much larger degree of lateral spreading is required to dissipate the current injected into these states. The results presented previously show that it is possible for a single state to sustain a current flow from 1 nA from the surface to the bulk, hence, once the tip is near such a site, the electrons injected into the surface can be rapidly transferred to the bulk via recombination at the defect site or tunnelling.

Lateral spreading of the current introduces a degree of non-locality in the imaging and hence is an important factor to consider when analysing STM images of semiconductors. Furthermore, the movement of the charge introduced into the surface will be affected by the surface conductivity and hence the resulting surface charge distribution can be highly anisotropic. The lateral spreading also profoundly changes the nature and the magnitude of the voltage drops that occur in STM experiments. Flores and Garcia [14] made a theoretical study of the voltage drops that occur during an STM analysis of a semiconductor surface by applying point contact theory [15]. They used a formula which gives the resistance as

$$R = \rho\frac{l}{s} \tag{7}$$

where ρ is the resistivity of the sample, l is the effective length of the sample and s represents the cross-sectional area of the current lines spreading in the vicinity of the semiconductor

surface. Flores and Garcia modelled an experimental study of the 7×7 reconstruction of the (111) surface of silicon [16]. They calculated the spreading resistance for a current flowing through an area of 13 Å^2. Their results for the resistance were in good agreement with the experimental measurements obtained by Binnig *et al* [16]. However, their analysis neglected the presence of a depletion region between the surface and the bulk of the semiconductor. The results in this paper show that for dangling bond states on the 7×7 surface, which are typical midgap states, there would have to be considerable lateral spreading of carriers across the surface before the current could be dissipated into the bulk. Once in the bulk the spreading resistance model could be applied to calculate the voltage drop for this section of the current path. But the cross-sectional area for the current flow is much greater than the 13 Å^2 assumed by Flores and Garcia and, hence, when the voltage drop across the bulk of the semiconductor is calculated using [7] the spreading resistance is negligible.

The experiments performed by Binnig *et al* [16, 17] show that semiconductor surfaces with a low bulk doping density are more difficult to analyse with an STM than are the corresponding semiconductor surfaces where the bulk has a high doping density. In order to obtain conduction when performing an STM analysis of semiconductors with low doping density Binnig *et al* [17] found that they had to perform their experiments at high temperatures ($T = 440$ K). The high temperature increases the density of carriers that are able to participate in electrical conduction, hence it also increases the phonon and the Auger rates of interaction between surface states and bulk states. Increasing the temperature or the doping density of a sample will tend to reduce the number of surface states that are needed to interact with the bulk in order to establish an effective electrical contact between the surface and the bulk of the semiconductor.

If an inordinate number of surface states are required to establish electrical contact between the surface and the bulk of the semiconductor, the conduction mechanism at the surface may not be rapid enough to connect the site at which the STM is injecting charge into the surface to enough surface states to allow the injected charge to enter the bulk at the same rate at which it enters the surface. If the dominant transport mechanisms that allow charge to flow from the surface to the bulk are very slow they can have a deleterious effect on the establishment of an efficient electrical contact between the semiconductor surface and bulk regions. This can result in the STM failing to function properly on surfaces where the rate at which charge from the surface enters the bulk is very slow.

In summary we have shown that the transport mechanisms by which charge flows from the semiconductor surface to the semiconductor bulk can have a profound effect on STM images of semiconductor surfaces. We have shown that non-localities can be introduced into the STM image by lateral spreading of charge across the surface.

Acknowledgments

One of us (DL) would like to thank Professor Volter Heine and Miss M Carlton for their support and encouragement.

References

[1] Chen C J 1993 *Introduction to Scanning Tunnelling Microscopy* (Oxford: Oxford University Press) ch 1
[2] Stoneham A M 1975 *Theory of Defects in Solids* (Oxford: Oxford University Press) p 521
[3] Oppenheimer J R 1928 *Phys. Rev.* **31** 66
[4] Sze S M 1981 *The Physics of Semiconductor Devices* 2nd edn (New York: Wiley) p 372

[5] Freeman L B and Dahlke W E 1969 *Solid State Electron.* **13** 1483
[6] Crowell C R 1969 *Solid State Electron.* **12** 55
[7] Jaklevic R C and Lambe J 1966 *Phys. Rev. Lett.* **17** 1139
Sah C T 1991 *Fundamentals of Solid-state Electronics* (Singapore: World Scientific) p 289
[8] Robbins D J and Landsberg P T 1980 *J. Phys. C: Solid State Phys.* **13** 2425
[9] Ridley B K 1982 *Quantum Processes in Semiconductors* (Oxford: Oxford Science Publications) p 275
[10] Lannoo M and Bourgoin J 1983 *Point Defects in Solids* vol 2 (Berlin: Springer) p 155
[11] Himpsel F J, Hollinger G and Pollak R A 1983 *Phys. Rev.* B **28** 7014
[12] Pethica J 1994 private communication
[13] Halas N J and Bokor J 1989 *Phys. Rev. Lett.* **62** 1679
[14] Flores F and Garcia N 1984 *Phys. Rev.* B **30** 2289
[15] Smith R A 1978 *Semiconductors* 2nd edn (Cambridge: Cambridge University Press) p 228
[16] Binnig G, Rohrer H, Gerber Ch and Weibel E 1983 *Phys. Rev. Lett.* **50** 120
[17] Binnig G, Rohrer H, Gerber Ch, Salvan F and Bario A M 1984 private communication to Flores and Garcia [14]