

Theory of the scanning tunnelling microscope

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

1991 J. Phys.: Condens. Matter 3 4313

(<http://iopscience.iop.org/0953-8984/3/24/001>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 171.66.16.147

The article was downloaded on 11/05/2010 at 12:10

Please note that [terms and conditions apply](#).

Theory of the scanning tunnelling microscope

J B Pendry, A B Prêtre and B C H Krutzen

The Blackett Laboratory, Imperial College of Science, Technology and Medicine,
Prince Consort Road, London SW7 2BZ, UK

Received 1 February 1991

Abstract. The formula for tunnelling between two objects is fundamental to understanding the scanning tunnelling microscope. Although an exact expression for the one-electron tunnelling current is known, its derivation has previously been complex. Here we present a new derivation which is relatively simple and based on physical arguments. We go on to use our methods to show for the first time the connection between the exact formula and the widely used transfer Hamiltonian method of Bardeen. Finally we emphasize the importance of choosing a formula that minimizes the computational effort required to raster the STM tip across a surface.

1. Introduction

The general problem of tunnelling between two objects has been addressed by numerous authors. Perhaps the most complete and rigorous treatment is that of Feuchtwang [1, 2] which pre-dates the scanning tunnelling microscope by ten years. The transfer Hamiltonian approach of Bardeen [3] has been much used, and more recently Tersoff and Hamann [4] have reformulated the problem. Lucas *et al* [5] and others have developed the theory further and applied it to calculations of STM current. A concise review of the theory can be found in the article by Lucas [6].

The original formulation of tunnelling theory by Feuchtwang is rigorous, but involves four long papers of complex algebra. The purpose of this paper is,

- (i) to present a concise and transparent derivation of the tunnelling formula based on physical arguments;
- (ii) to solve the long standing problem of deriving the Bardeen transfer Hamiltonian approximation from the exact formula;
- (iii) to present a fast algorithm for rastering the STM tip over the surface, valid in a wide range of circumstances.

2. Deriving the current formula

We consider an interface between two systems, A and B, each of which is connected to its own reservoir at infinity. When the Fermi energies are equal in the two reservoirs, current from A to B balances current from B to A at each energy. If A is raised by δV relative to B, then in a range of energies from E_F to $E_F + \delta V$ the current is entirely due to electrons in the A reservoir heading towards the interface. Transmission of

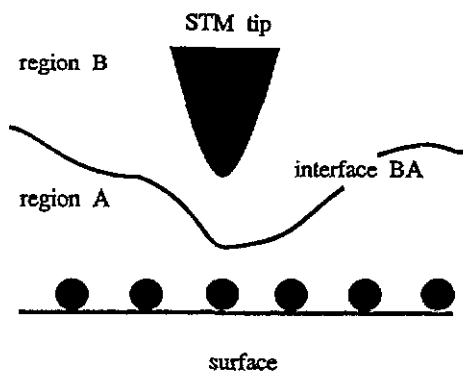


Figure 1. Schematic diagram of the surface and tip.

these electrons across the interface and ultimately into the B reservoir constitutes the tunnelling process. Figure 1 illustrates the situation. Electron-electron interactions are not considered here.

The problem is to identify which part of the wavefunctions at the interface has come from the A reservoir, and which from the B reservoir. We address this problem by erecting an impenetrable barrier, at the interface between A and B, which will later be removed in a systematic way allowing identification of the sources of the wavefunctions, and hence the tunnelling current. We define $\phi_A(E)$ and $\phi_B(E)$ to be eigenstates of this separated system. Note that $\phi_A = 0$ everywhere on side B and conversely $\phi_B = 0$ everywhere on side A. Now restore the connection between sides A and B. Suppose this involves a change in potential of v_{AB} . In a finite elements approach to calculations this potential can be thought of as a set of bonds linking together the finite elements in the two halves of the problem. We shall discuss a particular form of v_{AB} appropriate to continuum models in section 3. We can sum the perturbation series for those waves which start in reservoir A and penetrate into B:

$$\delta\phi_{BA} = G_B v_{BA} \phi_A + G_B v_{BA} G_A v_{AB} G_B v_{BA} \phi_A + \dots = G_B t_{BA} \phi_A \quad (1)$$

where G_A and G_B are the causal Green functions confined to one or other sides of the system,

$$G_A = \sum_{\phi_A} \frac{\phi_A \phi_A^*}{E - E_A + i\epsilon} \quad (2)$$

$$G_B = \sum_{\phi_B} \frac{\phi_B \phi_B^*}{E - E_B + i\epsilon} \quad (3)$$

and t_{BA} is the t -matrix which formally sums this series:

$$t_{BA} = v_{BA} (1 - G_A v_{AB} G_B v_{BA})^{-1}. \quad (4)$$

Next we observe that $\delta\phi_{BA}$ obeys the Schrödinger equation within B. For purposes of proof we add an infinitesimal imaginary component to the energy which will ultimately tend to zero,

$$(E + i\epsilon)\delta\phi_{BA} = -\frac{\hbar^2}{2m}\nabla^2\delta\phi_{BA} + V\delta\phi_{BA} \quad (5)$$

and if the potential is assumed to be real and local

$$(E - i\epsilon)\delta\phi_{BA}^* = -\frac{\hbar^2}{2m}\nabla^2\delta\phi_{BA}^* + V\delta\phi_{BA}^* \tag{6}$$

Multiply the first of these equations by $\delta\phi_{BA}^*$ and the second by $\delta\phi_{BA}$, subtract, and integrate over region B

$$\int_B 2i\epsilon\delta\phi_{BA}\delta\phi_{BA}^* d^3\mathbf{r} = \int_B -\frac{\hbar^2}{2m}(\delta\phi_{BA}^*\nabla^2\delta\phi_{BA} - \delta\phi_{BA}\nabla^2\delta\phi_{BA}^*) d^3\mathbf{r} \tag{7}$$

The infinitesimal imaginary part to the energy introduces absorption into the problem so that $\delta\phi_{BA}$ decays slowly into region B and the integral on the left hand side converges. The integral on the right hand side we can transform by Gauss's theorem to give an integral over the interface dividing regions A and B,

$$\int_B 2i\epsilon\delta\phi_{BA}\delta\phi_{BA}^* d^3\mathbf{r} = \int_{BA} -\frac{\hbar^2}{2m}(\delta\phi_{BA}^*\nabla\delta\phi_{BA} - \delta\phi_{BA}\nabla\delta\phi_{BA}^*) \cdot d^2\mathbf{S} \tag{8}$$

and identify the familiar expression for the electrical current,

$$\mathbf{J}_{BA} = \frac{i\hbar e}{2m}(\delta\phi_{BA}^*\nabla\delta\phi_{BA} - \delta\phi_{BA}\nabla\delta\phi_{BA}^*) \tag{9}$$

and so retrieve

$$\int_B 2i\epsilon e\delta\phi_{BA}\delta\phi_{BA}^* d^3\mathbf{r} = \int_{BA} i\hbar\mathbf{J}_{BA} \cdot d^2\mathbf{S} \tag{10}$$

where \mathbf{J}_{BA} is the current due to ϕ_A flowing across the interface between A and B and represents the contribution of a single 'mode'. For example in the case of a pure planar interface the momentum parallel to the surface would define a mode. Summing over all contributions to \mathbf{J}_{BA} we calculate the differential conductance of the STM, that is to say the rate of change of tunnelling current with applied voltage, V ,

$$\frac{dj_T}{dV} = \sum_{J_{BA}} \int_{BA} e\mathbf{J}_{BA} \cdot d^2\mathbf{S} \sigma_A(E) = \frac{2e\epsilon^2}{\hbar} \sum_{J_{BA}} \int_B \delta\phi_{BA}\delta\phi_{BA}^* d^3\mathbf{r} \sigma_A(E) \tag{11}$$

where $\sigma_A(E)$ is the density of states in mode ϕ_A . Substituting for $\delta\phi_{BA}$

$$\frac{dj_T}{dV} = \frac{2e\epsilon^2}{\hbar} \sum_{J_{BA}} \langle \phi_A | t_{BA}^\dagger G_B^\dagger G_B t_{BA} | \phi_A \rangle \sigma_A(E) \tag{12}$$

where the integration signs are subsumed in operator notation. Two further substitutions are required,

$$G_B^\dagger G_B = \sum_{\phi_B} \frac{\phi_B \langle \phi_B | \phi_B \rangle \phi_B^*}{(E - E_B - i\epsilon)(E - E_B + i\epsilon)} = \sum_{\phi_B} \frac{\phi_B \phi_B^*}{(E - E_B)^2 + \epsilon^2} = -\epsilon^{-1} \Im G_B \tag{13}$$

and

$$\sum_{\phi_A} \phi_A \phi_A^* \sigma_A(E) = -\pi^{-1} \Im G_A \tag{14}$$

where $\Im G_A$ stands for the imaginary part of G_A , so that finally

$$\frac{dj_T}{dV} = \frac{2e^2}{\hbar\pi} \text{trace} \left(t_{BA}^\dagger (\Im G_B) t_{BA} (\Im G_A) \right). \quad (15)$$

Note that the four implicit integrations are over the *interface* dividing regions A and B. This is an exact expression for the current across an arbitrary boundary in terms of the Green functions of the separated regions evaluated on the interface dividing them, and the t -matrix connecting them. We note in passing that our theory implicitly goes beyond linear response. If a large voltage is applied between A and B, then the differential conductance must be integrated through the voltage drop, and some account should be taken of the changed self-consistent potential in the barrier. Both these possibilities can be accommodated within our formula.

There are numerous methodologies available for evaluating this expression. What is needed is some means of calculating the reflection coefficients of the surface and the tip, and hence the respective Green functions. Lucas in his review [6] describes several ways in which this can be done. For periodic surfaces standard methods are available, see [7, 8], but for the tip some form of finite elements approach, such as that described by Lucas *et al* [5], or the causal surface Green function method (CSGFM) [11], is probably more appropriate.

3. Comparison with other theories

First we shall show that our equation (15) is the same as equation (3.20) of Feuchtwang [2]. We start by expressing t_{BA} in terms of the causal Green function of the entire region A plus B with the barrier removed,

$$t_{BA} = v_{BA} + v_{BA} G v_{BA}. \quad (16)$$

The relationship of $\Im G_A$ and $\Im G_B$ to Feuchtwang's ρ is given by

$$\Im G_A(\mathbf{r}_1, \mathbf{r}_2) = -\frac{1}{2} \rho_1(\mathbf{r}_1, \mathbf{r}_2) \quad (17)$$

$$\Im G_B(\mathbf{r}_3, \mathbf{r}_4) = -\frac{1}{2} \rho_2(\mathbf{r}_3, \mathbf{r}_4). \quad (18)$$

It is now necessary to define the impenetrable barrier, v_{BA} . Figure 2 shows a sketch of v_{BA} in a direction normal to the interface. Three sorts of wavefunctions are shown: ϕ is a wavefunction in the absence of the barrier and shows no discontinuities at the interface, ϕ_A and ϕ_B are wavefunctions in the presence of the barrier, which is taken to be high enough, and thick enough, to stop the wavefunctions penetrating from one side to the other. Thus in the limit

$$a = 0 \quad \alpha = \infty \quad a^2 \alpha = \infty. \quad (19)$$

Within the barrier

$$\phi_A(x) = \phi_A(0) \exp(-\beta x) \quad (20)$$

$$\phi_B(x) = \phi_B(a) \exp(+\beta(x - a)) \quad (21)$$

where in the limit of a high barrier

$$\frac{\hbar^2 \beta^2}{2m} = \alpha. \quad (22)$$

Note that v_{BA} is the potential which removes this barrier so that using these definitions we can prove that in the limit of a high barrier

$$\int \phi_A v_{BA} \phi_B = 0 \quad (23)$$

$$\int \phi v_{BA} \phi_B = \frac{\hbar^2}{2m} \int_{BA} d^2 S \cdot \phi \left(-\vec{\nabla} \right) \phi_B = \frac{\hbar^2}{2m} \int_{BA} d^2 S \cdot \phi \left(\vec{\nabla} - \vec{\nabla} \right) \phi_B \quad (24)$$

$$\int \phi_A v_{BA} \phi = \frac{\hbar^2}{2m} \int_{BA} d^2 S \cdot \phi_A \left(\vec{\nabla} \right) \phi = \frac{\hbar^2}{2m} \int_{BA} d^2 S \cdot \phi_A \left(\vec{\nabla} - \vec{\nabla} \right) \phi. \quad (25)$$

With these substitutions we transform (15) to

$$\begin{aligned} \frac{dj_T}{dV} = \frac{2e^2}{\hbar\pi} \iiint \int_{BA} d^2 S_1 \cdot \vec{\nabla}_1 \rho_1(\mathbf{r}_1, \mathbf{r}_2) \vec{\nabla}_2 \cdot d^2 S_2 \\ \times d^2 S_3 \cdot \vec{\nabla}_3 \rho_2(\mathbf{r}_3, \mathbf{r}_4) \vec{\nabla}_4 \cdot d^2 S_4 \left(\frac{\hbar^2}{2m} \right)^2 G(\mathbf{r}_2, \mathbf{r}_3) \left(\frac{\hbar^2}{2m} \right)^2 G^\dagger(\mathbf{r}_4, \mathbf{r}_1) \end{aligned} \quad (26)$$

where we have recognized that the density matrices, ρ , vanish on the interface BA, but that their derivatives do not.

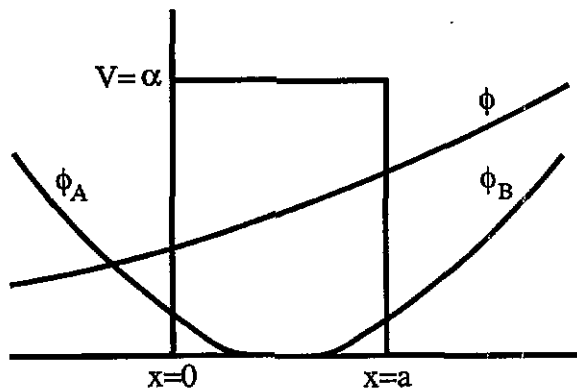


Figure 2. The impenetrable barrier separating regions A and B.

Bardeen's original formulation of the tunnelling problem [3] has been widely applied, but the relationship of this approximate theory to formally exact theories has not been defined. To quote Feuchtwang and Cutler [9] 'To date no consistent, let alone unique, solution of this troublesome problem has been found.'

The problem is that the impenetrable barrier defined by (19), or by any of the alternative definitions, is not a weak perturbation. However, since the STM current is

generally weak compared with what would flow through a point actually in contact with the surface, it can reasonably be assumed that some quantities in the theory can legitimately be treated as small. Derivation of the transfer Hamiltonian formula from (15) is surprisingly complex.

The trick is to recognize that the tunnelling current passes through regions which are remote from both surface and tip. Although the potentials inside the tip and inside the surface are strong and extend to infinity, their influence at the tunnelling interface is weak and can be treated in perturbation theory.

Removing the potentials inside the surface and inside the tip does not change the Green functions at the interface very much, but will remove the crucial imaginary part which is responsible for the tunnelling current. Of course one has to be careful about what is meant by removing the potential because both the tip and the surface have potentials that extend into the vacuum. These tails on the potential can be ignored if they are effectively weak and short range, but sometimes the tail of the surface potential can bind surface states that extend far into the vacuum: for example the Rydberg surface states that have been much studied in inverse photoemission (see Echenique and Pendry [10] for a review of these states). When the interface cannot be chosen out of the range of such surface states the tunnelling Hamiltonian method breaks down.

For the moment assume that the surface and tip potentials have been removed. This leaves us with three Green functions all of which are real because by removing both the tip and the surface we have removed all the electronic states at the tunnelling energy: G_0 is the causal Green function for the entire region A plus B with no barrier at the interface and no tip or surface potential, G_{0A} and G_{0B} are the causal Green functions for regions A and B respectively with the barrier in place but no tip or surface potential.

Now let us return to (15). Using (16) we can approximate

$$t_{BA} \approx t_{BA}^\dagger \approx v_{BA} + v_{BA} G_0 v_{BA} \quad (27)$$

so that (15) becomes

$$\frac{dj_T}{dV} \approx \frac{2e^2}{\hbar\pi} \text{trace} (v_{BA} G_0 v_{BA} (\Im G_B) v_{BA} G_0 v_{BA} (\Im G_A)) \quad (28)$$

where we have recognized that $\Im G_A$ and $\Im G_B$ vanish at the interface BA. Next we calculate the surface corrections to G_{0A} as a perturbation,

$$\Im G_A \approx \Im (G_{0A} + G_{0A} \delta t_S G_{0A}) = G_{0A} \Im (\delta t_S) G_{0A} \quad (29)$$

and similarly

$$\Im G_B \approx \Im (G_{0B} + G_{0B} \delta t_T G_{0B}) = G_{0B} \Im (\delta t_T) G_{0B} \quad (30)$$

so that our expression for the tunnelling conductance becomes

$$\frac{dj_T}{dV} \approx \frac{2e^2}{\hbar\pi} \text{trace} (v_{BA} G_0 v_{BA} G_{0B} \Im (\delta t_T) G_{0B} v_{BA} G_0 v_{BA} G_{0A} \Im (\delta t_S) G_{0A}) \quad (31)$$

which can be simplified by noting that in region A

$$G_0 = G_{0A} + G_0 v_{BA} G_{0A} \quad (32)$$

and in region B

$$G_0 = G_{0B} + G_{0B}v_{BA}G_0 \quad (33)$$

to give

$$\frac{dj_T}{dV} \approx \frac{2e^2}{\hbar\pi} \text{trace}(v_{BA}G_0\Im(\delta t_T)G_0v_{BA}G_{0A}\Im(\delta t_S)G_{0A}). \quad (34)$$

We note that (34) contains expressions of the form

$$\dots G_0v_{BA}G_{0A} \dots = \dots \int_{BA} G_0(\mathbf{r}', \mathbf{r}) \left(-\frac{\hbar^2}{2m} (\vec{\nabla} - \vec{\nabla}') \cdot d^2S \right) G_{0A}(\mathbf{r}, \mathbf{r}'') \dots \quad (35)$$

and that

$$G_{0A} = G_0 - G_0v_{BA}G_{0A} \quad (36)$$

hence

$$G_{0A}(\mathbf{r}, \mathbf{r}'') = G_0(\mathbf{r}, \mathbf{r}'') + f(\mathbf{r}, \mathbf{r}'') \quad (37)$$

where f as a function of \mathbf{r} is free of singularities in region A, obeys the same Schrödinger equation as does G_0 , and decays exponentially away from the interface, BA. Substituting for G_{0A}

$$\begin{aligned} \dots G_0v_{BA}G_{0A} \dots \\ = \dots \int_{BA} G_0(\mathbf{r}', \mathbf{r}) \left(-\frac{\hbar^2}{2m} (\vec{\nabla} - \vec{\nabla}') \cdot d^2S \right) (G_0(\mathbf{r}, \mathbf{r}'') + f(\mathbf{r}, \mathbf{r}'')) \dots \end{aligned} \quad (38)$$

The contribution from f vanishes, as can be proved by applying Gauss's theorem to

$$\int_{BA} G_0(\mathbf{r}', \mathbf{r}) (\vec{\nabla} - \vec{\nabla}') \cdot d^2S f(\mathbf{r}, \mathbf{r}'') = \int_{\infty A} G_0(\mathbf{r}', \mathbf{r}) (\vec{\nabla} - \vec{\nabla}') \cdot d^2S f(\mathbf{r}, \mathbf{r}'') \quad (39)$$

where the surface of integration on the right hand side has been removed far away from the interface BA. Since f decays exponentially away from BA, its contribution to the integral must vanish and (38) simplifies to

$$\begin{aligned} \dots G_0v_{BA}G_{0A} \dots = \dots \int_{BA} G_0(\mathbf{r}', \mathbf{r}) \left(-\frac{\hbar^2}{2m} (\vec{\nabla} - \vec{\nabla}') \cdot d^2S \right) G_0(\mathbf{r}, \mathbf{r}'') \dots \\ = \dots G_0I_{BA}G_0 \dots \end{aligned} \quad (40)$$

where we have used the operator I_{BA} to simplify our notation. Our expression for the tunnelling conductance becomes

$$\begin{aligned} \frac{dj_T}{dV} \approx \frac{2e^2}{\hbar\pi} \text{trace}(I_{BA}G_0\Im(\delta t_T)G_0I_{BA}G_0\Im(\delta t_S)G_0) \\ \approx \frac{2e^2}{\hbar\pi} \text{trace}(I_{BA}\Im(G_T)I_{BA}\Im(G_S)) \end{aligned} \quad (41)$$

where G_T and G_S are the Green functions of the system containing respectively only the tip and only the surface, calculated *without* the barrier at the interface. This is Bardeen's transfer Hamiltonian formula, valid only when an interface between tip and surface can be found on which the density of states is small.

4. Rastering the tip across the surface

Calculation of the tunnelling current is not a trivial computation, and if the entire calculation had to be repeated for every position of the tip the effort would be prohibitive. In this respect the transfer Hamiltonian formalism has a great advantage because once G_T and G_S have been calculated, at any subsequent position of the tip, R , the tunnelling conductance can be found by a simple convolution,

$$\frac{dj_T}{dV} \approx \frac{2e^2}{\hbar\pi} \iint_{BA} dS_1 dS_2 I_{BA}(r_1) \Im G_T(r_1 - R, r_2 - R) I_{BA}(r_2) \Im G_S(r_2, r_1). \quad (42)$$

For this reason the transfer Hamiltonian formalism should be used wherever its approximations are valid.

Suppose that it is not possible to define an interface on which the density of states of both the tip and the surface is small. Then the transfer Hamiltonian formalism is not valid. However, it may still be the case that we can avoid the immense labour of recalculating all the Green functions for each new tip position. The simplicity of (35) arises because one of the objects in the convolution, G_T , is invariant with respect to translation across the surface. It may sometimes happen that this desirable feature can be retained.

In general an STM tip is not brought closer than 5–10 Å from the surface, otherwise electrical and mechanical instabilities result. In the centre of a 10 Å gap between surface and tip, the dominant potential is in most instances that due to image forces. On the surfaces of insulators there may be uncompensated charges which invalidate this assumption, but for most conductors it is true. Furthermore at 5 Å from the surface the image force has lost all but its zero order Fourier components parallel to the surface, due to their exponential decay away from the surface. So at least the potential at a point mid-way between tip and surface is invariant to translation of the tip across the surface. In addition we know that the surface states of the image potential are not strongly modulated parallel to the surface. This has been demonstrated theoretically and confirmed by inverse photoemission experiments: a review of work in this field can be found in Echenique and Pendry [10]. This means that the Green functions evaluated at an interface half way between tip and surface will be approximately invariant to translation of the tip. So returning to (15) and remembering that region B is associated with the tip, we can write

$$\begin{aligned} & \iint_{BA} d^2 S_2 d^2 S_3 t_{BA}^\dagger(r_1, r_2) (\Im G_B(r_2, r_3)) t_{BA}(r_3, r_4) \\ & \approx \iint_{BA} d^2 S_2 d^2 S_3 t_{BA}^\dagger(r_1 - R, r_2) (\Im G_B(r_2, r_3)) t_{BA}(r_3, r_4 - R) \\ & = f_T(r_1 - R, r_4 - R). \end{aligned} \quad (43)$$

Hence on substituting into (15)

$$\frac{dj_T}{dV} = \frac{2e^2}{\hbar\pi} \iint_{BA} d^2 S_1 d^2 S_4 f_T(r_1 - R, r_4 - R) (\Im G_A(r_4, r_1)) \quad (44)$$

which has the desired form: f_T and G_A can be prepared independent of the position of the tip and then combined in this simple formula for any tip position.

5. Conclusions

We have given a simple proof of the exact formula for the STM current, and shown that it is equivalent to the Feuchtwang result previously derived by more complicated methods. The relationship to the transfer Hamiltonian formulation of tunnelling was demonstrated for the first time, under the assumption that the density of states in the tunnelling barrier can be assumed small. Finally we pointed out the importance of an efficient algorithm for rastering an STM tip over a surface, and showed that the transfer Hamiltonian satisfies this requirement. This efficiency could be retained under a more general set of assumptions than required for the validity of the transfer Hamiltonian, but in the most general instance a new calculation for each tip position will be required.

Acknowledgments

We thank BP for support of this work.

References

- [1] Feuchtwang T E 1974 *Phys. Rev. B* **10** 4121, 4135; 1975 *Phys. Rev. B* **12** 3979
- [2] Feuchtwang T E 1976 *Phys. Rev. B* **13** 517
- [3] Bardeen J 1961 *Phys. Rev. Lett.* **6** 57
- [4] Tersoff J and Hamann D 1985 *Phys. Rev. B* **31** 805
Tersoff J 1989 *Phys. Rev. B* **39** 1052
- [5] Lucas A A, Morawitz H, Henry G R, Vigneron J-P, Lambin Ph, Cutler P H and Feuchtwang T E
1987 *Phys. Rev. B* **37** 10708
- [6] Lucas A A 1990 *Europhys. News* **21** 63
- [7] Pendry J B 1976 *Surf. Sci.* **57** 679
- [8] Hopkinson J F L, Pendry J B and Titterton D J 1981 *Comput. Phys. Commun.* **19** 69
- [9] Feuchtwang T E and Cutler P H 1987 *Phys. Scr.* **35** 132
- [10] Echenique P M and Pendry J B 1990 *Prog. Surf. Sci.* **32** 111
- [11] Pendry J B, Prêtre A B and Rous P J 1991 *Surf. Sci.* to appear