Modelling of complicated nanometre resonant tunnelling devices with quantum dots

1991 J. Phys.: Condens. Matter 3 2651

(http://iopscience.iop.org/0953-8984/3/16/005)

J. Phys.: Condens. Matter 3 (1991) 2651–2664. Printed in the UK

Modelling of complicated nanometre resonant tunnelling devices with quantum dots

M Sumetskii
Leningrad Electrical Engineering Institute for Communications, Moyka Embankment 61, Leningrad 191065, USSR

Received 15 June 1990, in final form 26 November 1990

Abstract. The semiclassical expression for the transmission probability of a complicated multidimensional resonant tunnelling structure is obtained. This expression generalizes the well known Breit–Wigner formula and allows an analytical study of various structures containing electrodes and quantum dots. It is proved that the resonant conductance of a structure with N quantum dots cannot exceed N (in units of $2e^2/h$). It is also shown that the resonant conductance of an arbitrary structure containing a quantum dot with non-degenerate level connected in series with all others cannot exceed unity. The results obtained are applied to the analytical calculation of devices with a few quantum dots and chain and closed chain devices. The effect of the disappearance of the resonant peak corresponding to a well defined level is demonstrated. For a large number of identical dots in some examples considered, the conductance has a double-peak form in the individual energy band. This result is shown to have a clear physical meaning.

1. Introduction

Progress in the fabrication technology of nanometre structures allows one at present to make nanometre resonant tunnelling devices not only in one [1] but also in two or three dimensions [2, 3]. Such devices, created on the basis of semiconductor heterostructures combined with metallic or semiconductor dots and strips, have prospects for future electronics. Therefore, theoretical investigation of their electrophysical properties and modelling of new, more complicated, resonant tunnelling devices is of considerable interest.

The structure of the devices that are investigated in the present paper is shown in order of complication in figure 1(b)–(f). Electrodes and quantum dots are shown cross-hatched. The simplest artificial resonant tunnelling structure in this sequence (shown in figure 1(a)) was created for the first time in [2] and theoretically investigated in several papers [4–7]. As for other more complicated structures, they are likely to be created in the not-too-distant future, owing to the high-speed development of nano-technology. For example, the techniques of growing nanometre semiconductor structures localized in more than one dimension [8], and techniques of local band bending in semiconductor films by superimposed metallic strips under voltage [3], should be useful for multidimensional band engineering. Such technology does not need atomic size accuracy, since the wavelength of electrons in semiconductors has a nanometre order. Nanometre metallic dots and strips can be superimposed on a semiconductor film also

0953-8984/91/162651 + 14 $03.50

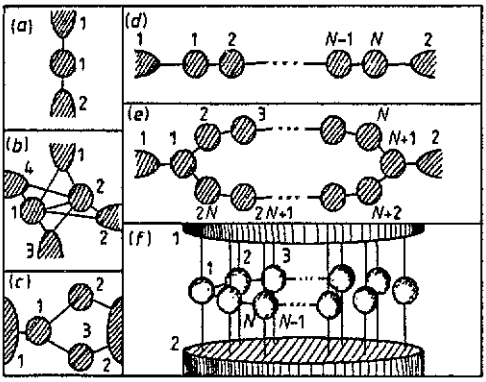

Figure 1. Different configurations of quantum dots and electrodes considered.

by scanning tunnelling microscope (STM) technology. Another method of using the STM in nanometre device creation consists of making shallow indentations by displacement of substrate atoms [9, 10]. Such indentations, when covered by conducting material and subjected to a voltage, can be used for electrostatic multidimensional band bending in a semiconductor film disposed under the surface of the substrate. The formation of both polymeric molecules and a few atoms with resonant levels must also be mentioned as possible nanometre resonant tunnelling structures [11].

The three-dimensional resonant tunnelling system of model zero-radius quantum wells displaced in a rectangular potential barrier was studied theoretically in [12]. Resonant tunnelling through two-dimensional arrays of model quantum dots was investigated numerically in [13, 14]. In [13] interesting predictions were made for the possible upper value of conductance as a function of the concrete configuration of the system. The expression for the transmission probability obtained in the present paper permits a study of rather complicated structures in analytical form. In section 2 we give the general semiclassical expression for the transmission probability, generalizing the Breit–Wigner formula for the case of many quantum dots with interacting resonant levels and many electrodes. In section 3 the general inequalities for the upper value of the conductance are proved. Sections 4 and 5 are devoted to the investigation of devices comprising a few quantum dots and chain and closed chain devices. In section 6 we discuss the results obtained.

2. The initial equations

Assume that the energy of incident electrons in an electrode is close to the defined number of levels $\{E_j\}$ of the quantum dots $\{j\}$. Thus only these levels will be taken into account in the resonant conductance calculations. Suppose also that the temperature of the system is zero and the process of resonant tunnelling is perfectly stationary and coherent.

The algebraic expressions obtained below for the penetrability of resonant tunnelling structures contain the energy of incident electrons E, the energy levels E_i in the ith quantum dot, the partial width $\Gamma_m^{(n)}$ of decay from quantum dot m to the neighbouring electrode n, and the overlap integrals δ_{ij} for neighbouring dots i and j. These expressions

are essentially the generalization of the well known Breit–Wigner formula for the conductance from electrode 1 to electrode 2:

$$G_{12} = \frac{\Gamma_2^{(1)}\Gamma_2^{(2)}}{(E - E_1)^2 + (\Gamma_1^{(1)} + \Gamma_1^{(1)})^2} \tag{1}$$

which can be directly applied only to case (*a*) in figure 1. We measure the conductance in units of $2e^2/h$. The general form of expression (1) is independent of the dimensions of the system [4, 15].

Let $\mathbf{T}_{mn}$ be the transmission matrix from electrode m to electrode n. Consider the total transmission probability from electrode m to electrode n:

$$G_{mn} = \mathrm{Tr}(\mathbf{T}_{mn}\mathbf{T}_{mn}^{\dagger}) \tag{2}$$

where the trace and the product of the operators means integration over the initial and final states of scattering electrons. In the present paper we are dealing with multi-dimensional systems having a finite number of resonant levels. In [16] expression (2) was shown to be the conductance of such strongly localized systems having two probes (electrodes). In the principal examples of our paper (sections 4.3 and 5) two-electrode structures are investigated. It should be noted that expression (2) generally defines the total transmission probabilities, which play a fundamental role in many-probe system theory [17]. Below, we shall consider systems with many electrodes, but confine ourselves only to calculating G_{mn} and investigating its properties, having in mind that for the two-electrode case G_{mn} defines the conductance of the system.

The expression for the values G_{mn} is obtained by combining the techniques of calculation of a model system with zero-radius potential wells [12, 18] and semiclassical techniques [15, 19] based on the existence of the most probable tunnelling paths (MPTP) coupling electrodes and quantum dots. Consider the Hamiltonian

$$H = \frac{p^2}{2m} + V(r) + \sum_{j=1}^{N} \hat{u}(|r - r_j|) \qquad p = \frac{i\hbar\nabla}{\hbar} \tag{3}$$

which describes a system of N zero-radius wells at points r_j displaced in the relatively smooth barrier $V(r)$, which has a jump only at the surfaces of electrodes.

The formula for the transmission probability G_{mn} from electrode m to electrode n can be obtained by straightforward generalization of methods [15] to the case of many quantum wells in the underbarrier region. As a result we have

$$G_{mn} = \mathrm{Tr}(\mathbf{\Gamma}_m\mathbf{R}\mathbf{\Gamma}_n\mathbf{R}^{\dagger}) = \sum_{j,k=1}^{N} |r_{jk}|^2\Gamma_j^{(m)}\Gamma_k^{(n)} \tag{4}$$

where the trace means summing over all quantum dots and

$$\mathbf{R} = \|r_{ij}\| = (E\mathbf{1} - \mathbf{E} - i\tfrac{1}{2}\mathbf{\Gamma})^{-1} \qquad \mathbf{\Gamma} = \begin{vmatrix} \Gamma_1 & 0 & \dots & 0 \\ 0 & \Gamma_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \Gamma_N \end{vmatrix} = \sum_{m=1}^{M} \mathbf{\Gamma}_m$$

$$\mathbf{E} = \begin{vmatrix} E_1 & \delta_{12} & \dots & \delta_{1N} \\ \delta_{21} & E_2 & \dots & \delta_{2N} \\ \vdots & \vdots & & \vdots \\ \delta_{N1} & \delta_{N2} & \dots & E_N \end{vmatrix} \qquad \mathbf{\Gamma}_m = \begin{vmatrix} \Gamma_1^{(m)} & 0 & \dots & 0 \\ 0 & \Gamma_2^{(m)} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \Gamma_N^{(m)} \end{vmatrix}. \tag{5}$$

Here M is the total number of electrodes in the system, $\mathbf{I}$ is the unit matrix and $\mathbf{\Gamma}$ is the matrix of total width of the levels. The symmetric matrix $\mathbf{E}$ is the energy matrix, so that the equation $\det(E\mathbf{I} - \mathbf{E}) = 0$ defines eigenvalues of the system for $\mathbf{\Gamma} = \mathbf{0}$.

The deduction of (4) and (5) is based on the semiclassical assumption that there exists only one MPTP s_{ij} connecting dot (electrode) i with dot (electrode) j. Thus both the partial widths $\Gamma_i^{(j)}$ and the overlap integrals δ_{ij} can be calculated in a small neighbourhood of the corresponding MPTP. First, normalized eigenfunctions ψ_i must be calculated at quantum dot i without accounting for level broadening and splitting. Then, in order to calculate the partial width $\Gamma_i^{(j)}$ one must continue the eigenfunction ψ_i along the MPTP s_{ij}. The flux of the corresponding outgoing wave at electrode j, averaged over a plane P transverse to the tunnelling direction, will give the value of $\Gamma_i^{(j)}$:

$$\Gamma_i^{(j)} = \frac{\hbar^2}{2m}\left|\int_P \mathrm{d}P\left(\psi_i \frac{\partial \psi_i^*}{\partial s} - \psi_i^* \frac{\partial \psi_i}{\partial s}\right)\right|$$

where m is the electron mass. Similarly, δ_{ij} is defined as the overlap integral between functions ψ_i and ψ_j, continued along the MPTP s_{ij}, over a plane P' transverse to s_{ij}:

$$\delta_{ij} = \frac{\hbar^2}{2m}\left|\int_{P'} \mathrm{d}P'\left(\psi_i \frac{\partial \psi_j^*}{\partial s} - \psi_j^* \frac{\partial \psi_i}{\partial s}\right)\right|.$$

We propose below that the parameters $\Gamma_i^{(j)}$ and δ_i are known, having in mind that the method of their calculation was developed earlier [19, 20].

In the multidimensional case the formulae (4) and (5) can be obtained, in analogy to (1) [15], by averaging the outgoing flux of the wavefunction over the momentum components of incident electrons in an emitter electrode (the energy E fixed), and by integrating it over the plane transverse to the tunnelling direction in a collector electrode. Such averaging assumes that the characteristic distance a of the model potential inside electrodes and in the underbarrier region, except in quantum dots, satisfies the semiclassical condition $ka \gg 1$ (k is the characteristic absolute value of the wavenumber). The order of interference terms neglected in (4) and (5) is not more than $(ka)^{-1}$.

Note that equations (4) and (5) only contain structure parameters which are sufficiently common, have a clear physical meaning and are valid under broader conditions for quantum wells than assumed in (3). In analogy to the Breit–Wigner formula (1), these equations hold when the energy interval under consideration (in particular the values $E_i - E_j$), $\Gamma_m^{(n)}$ and δ_{ij} are small compared with characteristic values of energy in the individual quantum wells.

3. Some general relations and inequalities for resonant tunnelling probability

Now we are able to consider some general situations and to provide some important inequalities for conductance as a consequence of the expressions (4) and (5).

3.1. General inequality

Obviously the transmission probability (4) can reach the value N for independent quantum dots when $\delta_{ij} = 0$. It can be proved that $G_{mn} \leqslant N$ always. We prove an inequality

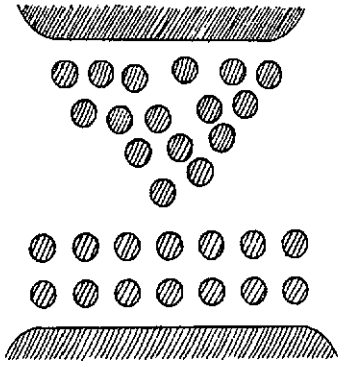

Figure 2. The case when two parts of the structure contact via one quantum dot only.

that is slightly stronger. Let us choose two arbitrary but non-intersecting sets of electrodes namely S_1 and S_2. Then

$$\sum_{m\in S_1}\sum_{n\in S_2} G_{mn} \leqslant N. \tag{6}$$

The proof of (6) is given in appendix 1.

3.2. Well defined levels

The levels of the structure are well defined if their widths are small compared to their separation. In this case the value N in inequality (6) must be changed to unity.

Assume that unitary transformation **U** diagonalizes energy matrix **E** so that

$$\mathbf{U}^*\mathbf{E}\mathbf{U} = \begin{vmatrix} \tilde{E}_1 & 0 & \dots & 0 \\ 0 & \tilde{E}_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \tilde{E}_N \end{vmatrix}.$$

If E is very near to non-degenerate level $\tilde{E}_k$ (resonant approximation), then all elements of **R** are relatively small except r_{kk} and we have

$$\sum_{m\in S_1}\sum_{n\in S_2} G_{mn} = \sum_{m\in S_1} g_{kk}^{(m)} \sum_{n\in S_2} g_{kk}^{(n)}/[(E-\tilde{E}_k)^2 + \tfrac{1}{4}(g_{kk})^2] \leqslant 1 \tag{7}$$

where g_{kk} and $g_{kk}^{(m)}$ are elements of matrices $\mathbf{U}^*\boldsymbol{\Gamma}\mathbf{U}$ and $\mathbf{U}^*\boldsymbol{\Gamma}_m\mathbf{U}$, respectively. Thus in the case of well defined levels ($g_{kk} \ll |E_k - E_j|$), the resonant conductance is given by the usual Breit–Wigner formula and cannot exceed unity.

It is interesting to note that according to (7) the effect of the disappearance of a resonant peak corresponding to a definite level can be observed. It occurs when one of the sums in the numerator in (7) vanishes. An example of such an effect will be considered in section 4.3.

3.3. Connection via one quantum dot

The value N in (6) must be changed to unity also in another remarkable situation. Suppose that the structure consists of two non-intersecting sets of dots, D_1 and D_2, which are connected only via one dot, as shown, e.g. in figure 2. Such a configuration is similar

to that appearing in the theory of the tunnelling microscope when the STM tip contacts the surface investigated via only one atom. Let S_1 and S_2 now be the sets of electrodes connected only with dots D_1 and D_2 respectively. Suppose that the connecting dot mentioned has number $j = 0$ and $\mathbf{R}^{(k)} = \|r_{ij}^{(k)}\|$ is the matrix defined similar to $\mathbf{R}$ in (5) for the structure $\{D_k, S_k\}$ assumed independent (i.e. for $\delta_{0j} = 0$). Then it can be shown that

$$r_{jk} = \sum_{l\in D_1} r_{jl}^{(1)}\delta_{0l} \sum_{p\in D_2} r_{kp}^{(2)}\delta_{0p} \times \left(E - E_0 - \sum_{l,p\in S_1} r_{lp}^{(1)}\delta_{0l}\delta_{0p} - \sum_{l,p\in S_2} r_{lp}^{(2)}\delta_{0l}\delta_{0p}\right)^{-1}. \tag{8}$$

It is remarkable that $\mathbf{R}$ is expressed only via the elements of matrices $\mathbf{R}^{(k)}$ and the parameters of the connecting dot.

Instead of (6) we now have:

$$\sum_{m\in S_1}\sum_{n\in S_2} G_{mn} \leqslant 1. \tag{9}$$

The proof of this important inequality is given in appendix 2. It is valid also when electrode S_1 contacts other elements of the system via one quantum dot only.

4. Systems with a few quantum dots

Let us apply the result of section 2 to the cases of one, two and three quantum dots in the underbarrier region. In analogy with classical electrical engineering theory, we are going to consider the simplest parallel and series connections of quantum dots. In the quantum case there is no principal difference between parallel and series connections, and in general these notions are not applicable. Nevertheless, in the semiclassical situation considered the existence of MPTP makes these notions physically reasonable.

4.1. One quantum dot

The formulae (4) and (5) for $N = 1$ give the result found in [21]:

$$G_{mn} = \Gamma_1^{(m)}\Gamma_1^{(n)}/(|E - E_1|^2 + \tfrac{1}{4}\Gamma_1^2).$$

4.2. Two quantum dots

Assuming $N = 2$ in (3)–(5) we have (see figure 1(*b*)):

$$G_{13} = [|E - E_1 - i\tfrac{1}{2}\Gamma_1|^2\Gamma_2^{(1)}\Gamma_2^{(3)} + |E - E_2 - i\tfrac{1}{2}\Gamma_2|^2\Gamma_1^{(1)}\Gamma_1^{(3)} + \delta_{12}^2(\Gamma_1^{(1)}\Gamma_2^{(3)} + \Gamma_1^{(3)}\Gamma_2^{(1)})]/|(E - E_1 - i\tfrac{1}{2}\Gamma_1)(E - E_2 - i\tfrac{1}{2}\Gamma_2) - \delta_{12}^2|^2.$$

If the values $\Gamma_j^{(1)}$ and $\Gamma_2^{(3)}$ are relatively small, then we come to the case of two successive quantum dots, which is essentially one-dimensional and was studied in several papers (see [12, 22]). Let the system be symmetric, so that $\Gamma_j^{(1)} = \Gamma_j^{(3)} = \Gamma$ and $E_1 = E_2 = E_0$, and for simplicity $\Gamma_j^{(4)} = \Gamma_j^{(2)} = 0$ (parallel connection). Then

$$G_{13} = 2\Gamma^2[(E - E_0)^2 + \Gamma^2 + \delta_{12}^2]/[(E - E_0)^4 + 2(E - E_0)^2(\Gamma^2 - \delta_{12}^2) + (\Gamma^2 + \delta_{12}^2)^2].$$

It is easy to see now that two different situations appear.

(i) $\Gamma^2 < 3\delta_{12}^2$. Then $G_{13}(E)$ has two maxima for

$$E^{(1,2)} = E_0 \pm (\Gamma^2 + \delta_{12}^2)^{1/4}[2\delta_{12} - (\Gamma^2 + \delta_{12}^2)^{1/2}]^{1/2}$$

so that

$$G_{13}(E^{(j)}) = \Gamma^2/2\delta_{12}[(\Gamma^2 + \delta_{12}^2)^{1/2} - \delta_{12}] \leqslant \tfrac{8}{9}.$$

For $\Gamma \to 0$, the values $E^{(1,2)}$ tend to the eigenvalues of the system $E_0 \pm \delta_{12}$. The minimum of G_{13} is found for $E = E_0$.

(ii) $\Gamma^2 \geqslant 3\delta_{12}^2$. Only one maximum is present in this case for $E = E_0$ when

$$G_{13}(E_0) = 2\Gamma^2/(\Gamma^2 + \delta^2) \leqslant 2.$$

4.3. Three quantum dots

The general formula for G_{mn} in the case $N = 3$ (figure 1(*c*)) is rather cumbersome. Therefore, let us simplify this case by making the following assumptions:

$$\begin{aligned} &\delta_{23} = 0 \qquad \delta_{13} = \delta_{12} = \delta \qquad \Gamma_1^{(2)} = \Gamma_2^{(1)} = \Gamma_3^{(1)} = 0 \\ &\Gamma_1^{(1)} = \Gamma_2^{(2)} = \Gamma_3^{(2)} = \Gamma \qquad E_1 = E_2 = E_3 = E_0. \end{aligned} \tag{10}$$

The MPTP not neglected in calculations according to (10) are shown in figure 1(*c*). Using the language of classical electrical engineering we may say that (10) describes quantum dot 1 connected in series with dots 2 and 3, which are parallel to each other. Then

$$G_{12} = 2\Gamma^2\delta^2/[(E - E_0)^4 + 2(E - E_0)^2(\tfrac{1}{4}\Gamma^2 - 2\delta^2) + (\tfrac{1}{4}\Gamma^2 + 2\delta^2)^2].$$

Two situations appear again.

(i) $\Gamma^2 < 8\delta^2$. Then G_{12} has two maxima for

$$E^{(1,2)} = E_0 \pm (2\delta^2 - \tfrac{1}{4}\Gamma^2)^{1/2}$$

with the value of G_{12} independent of Γ and δ,

$$G_{12}(E^{(j)}) = 1$$

and minimum

$$G_{12}(E_0) = 2\Gamma^2\delta^2/(\tfrac{1}{4}\Gamma^2 + 2\delta^2)^2 \leqslant 1. \tag{11}$$

(ii) $\Gamma \geqslant 8\delta^2$. Only one maximum is present for $E = E_0$ (see (11)).

Note that the considered structure of three quantum dots for $\Gamma \to 0$ has three eigenvalues (E_0 and $E_0 \pm 2^{1/2}\delta$) but only two peaks appear in $G_{12}(E)$, unlike the case of three series-connected quantum dots. Thus the example considered demonstrates the

effect of the disappearance of a peak corresponding to the level E_0. This effect was mentioned above (section 3.2).

5. Devices with many quantum dots: chain and closed chain

In this section devices with N quantum dots are considered. The expression for the conductance is obtained for finite N, and then the limiting case $N \gg 1$ is examined.

5.1. Chain of quantum dots

Consider a chain of quantum dots whose beginning and end are connected with electrodes 1 and 2 respectively (figure 1(*d*)). Under the assumptions made we can put $\delta_{jk} = 0$ for $|j-k| > 1$ (tight-binding approximation). Then it can be obtained that

$$G_{12} = \Gamma_1^{(1)}\Gamma_N^{(2)}\left(\prod_{j=1}^{N-1}\delta_{j,j+1}\right)^2 / |\det(E\mathbf{I} - \mathbf{E} - \tfrac{i}{2}\Gamma)|^2. \tag{12}$$

It is clear that this expression, as in the one-dimensional case, cannot exceed unity. This fact also follows from the inequality (9).

Let all the dots be identical, $E_j = E_0$, and all the overlap integrals for neighbouring dots be equal to each other, $\delta_{j,j+1} = \delta$. Let the level widths of all the dots except the first and the last be equal to zero and $\Gamma_1 = \Gamma_1^{(1)} = \Gamma_N = \Gamma_N^{(2)} = \Gamma$. Then after long algebra we have the result obtained in [22]:

$$G_{12} = 2\Gamma^2\delta^2 \sin^2\alpha / \{d^+ + d^- \cos[(2N-2)\alpha + \beta]\} \tag{13}$$

with

$$\alpha = \cos^{-1}[(E-E_0)/2\delta] \qquad \beta = 2\tan^{-1}[\delta^2 \sin(2\alpha)/(\delta^2\cos(2\alpha) + \tfrac{1}{4}\Gamma^2)]$$

$$d^{\pm} = (\delta^2 + \tfrac{1}{4}\Gamma^2)^2 \pm \Gamma^2\delta^2 \sin^2\alpha.$$

This expression is valid for all N, even for $N = 1$. Formally it is valid for $E - E_0 < 2\delta$, but it can be extended to all E by analytical continuation†. For large N the value of G_{12} outside the allowed band $(E_0 - 2\delta, E_0 + 2\delta)$ is exponentially small.

The maxima of expression (13) are reached when $\cos[(2N-2)\alpha + \beta] = 1$. All of them are equal to unity.

Consider the case $N \gg 1$. Then averaging (13) over $N\alpha$ we have

$$\langle G_{12}\rangle = \Gamma\delta \sin\alpha/(\delta^2 + \tfrac{1}{4}\Gamma^2). \tag{14}$$

The maximum of $\langle G_{12}\rangle$, equal to unity, is reached for $\Gamma = 2\delta$ and $E = E_0$ (see figure 3(*a*)).

5.2. Closed chain in series between two electrodes

The case of a closed chain device with identical quantum dots ($E_j = E_0$) demonstrates different features of conductance from the previous case. We again assume that $\delta_{j,j+1} = \delta$. The total number of dots in the closed chain considered is now $2N$. Electrodes 1 and

† As well as the formulae for G_{12} in sections 5.2 and 5.3.

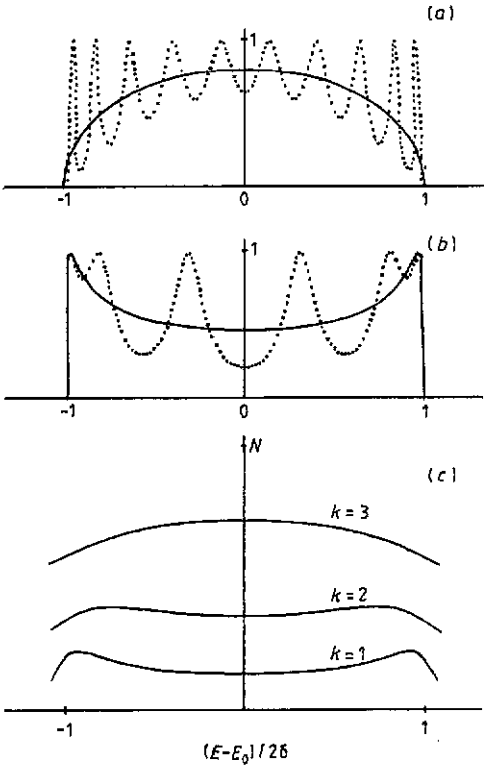

Figure 3. The dependence of conductance (dots) and average conductance (curves) on energy E in the interval $(E_0 - 2\delta, E_0 + 2\delta)$ for the structures shown in (*a*) figure 1(*d*), (*b*) figure 1(*e*) and (*c*) figure 1(*f*). In cases (*a*) and (*b*) we put $\Gamma/\delta = 1$ and the total number of dots in the structure equal to 10. In case (*c*) the conductance is calculated according to formula (19) and $\Gamma/\delta = 2\exp(k-3)$, $k = 1, 2, 3$.

2 contact only with dots 1 and $N+1$ respectively (figure 1(*e*)); moreover, the values $\Gamma_1^{(1)} = \Gamma_{N+1}^{(2)} = \Gamma$ and all other values $\Gamma_i^{(j)} = 0$. Then the conductance is

$$G_{12} = 16\Gamma^2\delta^2 \sin^2\alpha/[(f^+)^4 + (f^-)^4 - 2(f^+f^-)^2\cos(2N\alpha)] \tag{15}$$

with

$$f^{\pm} = 2\delta\sin\alpha \pm \Gamma/2 \qquad \alpha = \cos^{-1}[(E - E_0)/2\delta].$$

The maxima $G_{12} = 1$ are now reached if the condition $\cos(2N\alpha) = 1$ holds.

For $N \gg 1$ the average value of Γ_{12} over $N\alpha$ has the form

$$\langle G_{12}\rangle = 2\Gamma\delta\sin\alpha/(4\delta^2\sin^2\alpha + \tfrac{1}{4}\Gamma^2). \tag{16}$$

Thus, unlike the previous case, the expression (16) for $\Gamma < 4\delta$ has two maxima equal to unity and a minimum between them (see figure 3(*b*)). For very small Γ the value $\langle G_{12}\rangle$ is a small constant $\Gamma/2\delta$ inside the allowed energy band and has sharp peaks with maximum equal to unity on its boundaries. For $\Gamma \geqslant 4\delta$ the form of $\langle G_{12}\rangle$ is similar to the one shown in figure 3(*a*).

5.3. Closed chain in parallel between two electrodes

Let us now examine a system of N identical dots forming a closed chain lying in a surface parallel to the surfaces of electrodes 1 and 2 (figure 1(*f*)). Assuming that $\delta_{j,j+1} = \delta$, $\Gamma_j = \Gamma_j^{(1)} + \Gamma_j^{(2)}$ and $\Gamma_j^{(1)} = \Gamma_j^{(2)} = \Gamma$, we have for $N \geqslant 3$

$$G_{12} = \frac{\Gamma^2|q|^2}{d^2|(q^2-1)(q^N-1)|^2}\left\{N\left[\frac{(|q|^{2N}-1)(|q|^2+1)}{|q|^2-1} + 2\,\mathrm{Re}\left(q\,\frac{q^N-(q^*)^N}{q-q^*}\right)\right]\right\} \tag{17}$$

with

$$q = -\cos\alpha - \mathrm{sgn}(E-E_0)r^- + \mathrm{i}(\Gamma/2\delta + r^+)$$

$$r^{\pm} = 2^{-1/2}\{[(\sin^2\alpha + \Gamma^2/4\delta^2)^2 + \Gamma^2/\delta^2\cos^2\alpha]^{1/2} \pm (\sin^2\alpha + \Gamma^2/4\delta^2)\}^{1/2}$$

$$\alpha = \cos^{-1}[(E-E_0)/2\delta].$$

For independent quantum dots, when $\delta \ll \Gamma$, this formula gives the Breit–Wigner result, summing the conductances of individual dots:

$$G_{12} = N\Gamma^2/[(E-E_0)^2 + \Gamma^2].$$

The procedure of averaging $G_{12}(E)$ for $N \gg 1$ now depends on the relation between δ, Γ, N and $E - E_0$, since the width of a level in every dot of the considered system is finite. Actually, if $\Gamma \ll 2\delta \sin\alpha$ then

$$q = \exp[-\mathrm{i}\alpha + \Gamma/(2\delta\sin\alpha)]$$

and the absolute value of q is near to unity. Nevertheless for large N the value of q^N in (17) may be large or may have an order of unity. Let us consider the situations when $N \gg 1$. Then the last item in (17) is relatively small.

(i) If Γ is so small that $|q|^{2N} - 1 \ll 1$ (or $N\Gamma \ll \delta\sin\alpha$) then

$$\langle G_{12}\rangle = N\Gamma/(2\delta\sin\alpha). \tag{18}$$

Under the assumptions made, this expression is much smaller than unity. The function $\langle G_{12}(E)\rangle$ has a minimum in the centre of the allowed band and maxima near its boundaries where (18) fails.

(ii) In the opposite case $|q|^N \gg 1$ (or $N\Gamma \gg \delta\sin\alpha$) it is easy to find from (17) without averaging that

$$G_{12} = N\Gamma^2|q|^2(|q|^2+1)/\delta^2(|q|^2-1)|q^2-1|^2. \tag{19}$$

The function $G_{12}(E)/N$ for different Γ/δ is shown in figure 3(*c*). If $\Gamma < 2\delta$ then two peaks exist. Unlike in figure 3(*b*) the corresponding maxima vanish for small Γ/δ. For $\Gamma \ll \delta\sin\alpha$, we obtain that G_{12} is equal to $N\Gamma/(2\delta\sin\alpha)$ and coincides with the average value (18) of case (i).

(iii) Suppose an intermediate case when $\Gamma \ll \delta\sin\alpha$ and $|q|^N \gtrsim 1$ 1 (or $N\Gamma \gtrsim \delta\sin\alpha$). Then the average value $\langle G_{12}\rangle$ is defined by (18) again and may have an order much smaller than N. The maxima of G_{12} are now found from

$$\max G_{12} = \frac{s(\mathrm{e}^s+1)}{\mathrm{e}^s-1} > 2 \qquad s = \langle G_{12}\rangle = \frac{N\Gamma}{2\delta\sin\alpha} \ll N.$$

If s (or Γ) is small, then max G_{12} tends to its minimum value equal to 2. This does not contradict the general result (7) because most of the levels of a closed chain (except one or two of them) are doubly degenerate in the approximation considered.

The different behaviour of the average conductance dependence via E in the cases considered (single- or double-peak form) has a clear physical meaning discussed in the next section.

6. Results and discussion

With the help of the semiclassical expressions (4) and (5) obtained, we have found some remarkable features of the conductance and investigated analytically several models of resonant tunnelling devices.

In general, the resonant tunnelling conductance of a structure with N quantum dots cannot exceed N, but the concrete configuration of the quantum dots restricts the possible upper bound of the conductance to a smaller value. According to inequality (9) the conductance of the structure containing a dot connected in series with all others cannot exceed unity. It seems that a more general inequality can be proved: if there are k such dots, then the conductance cannot exceed k.

In the case of well defined levels, when their widths are small compared to their separation, the resonant conductance is given by the usual Breit–Wigner formula and cannot exceed unity. The unitary transformation that diagonalizes the energy matrix $\mathbf{E}$ may give a very small (or zero in the approximation considered) value of one of the sums in the numerator of (7). Then the effect of the disappearance of the resonant peak corresponding to a given level appears. Such an example was considered in section 4.3 for the case of three quantum dots. It can be shown that the indicated effect cannot appear for the case of two quantum dots.

The comparison of the results in sections 5.2 and 5.3 shows once more the non-triviality of inequality (9). Owing to double degeneracy of most of the levels in a closed chain, it is natural to suppose that for a well defined level the maximum value of the conductance is equal to 2. In section 5.3 such a result really occurs but, owing to (9), it cannot occur in section 5.2. The latter can be understood in detail if we consider, in addition to section 3.2, the situation when one of the levels is doubly degenerate (distinguish this situation from the case when the level is degenerate in the individual quantum well and the expressions (4) and (5) fail). We shall not dwell here on this consideration.

One can see that inequalities (6), (7) and (9) fail when the semiclassical condition for the width of potential barriers is not fulfilled. In reality, if the potential barriers vanish, then the motion in the potential well becomes more and more delocalized and in general an arbitrary flux of electrons can be passed.

In section 4 we have pointed out the limiting situation of identical quantum dots with $N \gg 1$ when an allowed energy band appears (chain and closed chain devices). It was shown that the averaged conductance dependence in the energy band might have single or double maxima. In order to explain the double-peak behaviour of the conductance, let us consider the case when Γ is small and the levels of the structures are well defined (see section 3.2). Let us calculate their widths. For the simple chain of section 5.1 we have the formula obtained in [22]:

$$g_{kk} = 4\Gamma \sin^2 \alpha_k/(N+1) \qquad \alpha_k = \alpha|_{E=\tilde{E}_k}. \tag{20}$$

For the case of the closed chain of section 5.2:

$$g_{kk} = 2\Gamma/N. \tag{21}$$

And for the case of the closed chain of section 5.3:

$$g_{kk} = 2\Gamma. \tag{22}$$

Here k is the number of the level. It is important that the widths (21) and (22), unlike (20), are independent of the number of eigenvalues of the chain. One can understand this fact if it is mentioned that the decay channels in the simple chain considered are longitudinal to the motion of electrons in the chain. On the contrary, in closed chains they are all transverse to the indicated motion.

According to (7) for all of the cases the conductance is expressed by the Breit–Wigner formula

$$G_{12}^{(k)} = \gamma_k (g_{kk})^2 / [4(E - \bar{E}_k)^2 + (g_{kk})^2]$$

where $\gamma_k = 1$ for sections 5.1 and 5.2, and $\gamma_k = 2$ for section 5.3.

For large N the eigenvalues of the chain, according to (13) and (15), are given by

$$E_k = \bar{E}_0 + 2\delta \cos(\pi k/N) \qquad (\text{or } \alpha_k = \cos^{-1}(\pi k/N)).$$

Hence the density of states (degeneracy of the levels in closed chain not taken into account) is

$$\frac{dk}{dE} = \frac{N}{\pi}\left[1 - \left(\frac{E - E_0}{2\delta}\right)^2\right]^{-1/2} = \frac{N}{\pi \sin \alpha}. \tag{23}$$

The value (23) is very large near the boundaries of the allowed energy band. And this is just the cause of double-peak appearance in average conductance. Actually it is easy to find that for small Γ

$$\langle G_{12} \rangle = \frac{dk}{dE} \int G_{12}^{(k)} \, dE = \frac{\pi}{2} \gamma_k g_{kk} \frac{dk}{dE}.$$

Thus, two factors are multiplied here: the density of states and the average intensity of energy peaks of the individual levels. In the case of a simple chain, the value g_{kk} vanishes quicker than dk/dE grows near the boundaries of the allowed band, and so the conductance vanishes there. But for the closed chain the widths (21) and (22) are independent of the eigenvalue number and a peak arises near each of the band boundaries.

Finally, let us compare the results obtained for a chain with the results for a one-dimensional double-barrier structure with a wide rectangular well. In the latter case the density of states near the bottom E_c of the well is proportional to $(E - E_c)^{-1/2}$, which is perfectly similar to (23). But the average intensity (the width) of the levels is proportional to $(E - E_c)^{1/2}$. Thus in this case the conductance near the bottom of the well neither vanishes nor has any peculiarity.

Appendix 1

In order to prove (6) first let us prove the inequality

$$\mathbf{R}^{-1}(\boldsymbol{\Gamma}^{(1)})^{-1}(\mathbf{R}^{\dagger})^{-1} - \boldsymbol{\Gamma}^{(2)} \geqslant \mathbf{0} \tag{A1.1}$$

with

$$\boldsymbol{\Gamma}^{(k)} = \sum_{m \in S_k} \boldsymbol{\Gamma}_m$$

where '⩾' for matrices means 'positive definite' [23]. Actually, if x is an arbitrary complex vector then using (5) we have

$$x^*[\mathbf{R}^{-1}(\mathbf{\Gamma}^{(1)})^{-1}(\mathbf{R}^\dagger)^{-1} - \mathbf{\Gamma}^{(2)}]x$$

$$= x^*[(E\mathbf{I} - \mathbf{E})(\mathbf{\Gamma}^{(1)})^{-1}(E\mathbf{I} - \mathbf{E}) + \tfrac{1}{4}\mathbf{\Gamma}(\mathbf{\Gamma}^{(1)})^{-1}\mathbf{\Gamma} - \mathbf{\Gamma}^{(2)}]x$$

$$\geqslant x^*[\tfrac{1}{4}\mathbf{\Gamma}[\mathbf{\Gamma}^{(1)})^{-1}\mathbf{\Gamma} - \mathbf{\Gamma}^{(2)}]x \geqslant 0. \tag{A1.2}$$

Here the first inequality is a consequence of the positive definiteness of matrix $(E\mathbf{I} - \mathbf{E})(\mathbf{\Gamma}^{(1)})^{-1}(E\mathbf{I} - \mathbf{E})$ (matrix $(\mathbf{\Gamma}^{(1)})^{-1}$ being positive definite and $E\mathbf{I} - \mathbf{E}$ being symmetric). The second inequality is easy to verify, the matrices $\mathbf{\Gamma}$ and $\mathbf{\Gamma}^{(k)}$ being diagonal, positive definite and satisfying the inequality $\mathbf{\Gamma} \geqslant \mathbf{\Gamma}^{(1)} + \mathbf{\Gamma}^{(2)}$.

Applying the transformation $\mathbf{R}(\ldots)\mathbf{R}^\dagger$ to (A1.1) we have

$$(\mathbf{\Gamma}^{(1)})^{-1} - \mathbf{R}\mathbf{\Gamma}^{(2)}\mathbf{R}^\dagger \geqslant \mathbf{0}. \tag{A1.3}$$

The trace of the product of positive definite matrices is positive [23]. Let us multiply (A1.3) by positive definite matrix $\mathbf{\Gamma}^{(1)}$. As a result we have

$$\mathrm{Tr}(\mathbf{I} - \mathbf{\Gamma}^{(1)}\mathbf{R}\mathbf{\Gamma}^{(2)}\mathbf{R}^\dagger) \geqslant 0. \tag{A1.4}$$

This inequality is obviously equivalent to (6).

Appendix 2

Let us prove inequality (9). Substitute (8) into (4). It is not difficult to show that (9) holds for all $E - E_0$ if

$$\left[\mathrm{Im}\left(\sum_{l,p\in D_1} r^{(1)}_{lp}\delta_{0l}\delta_{0p} + \sum_{l,p\in D_2} r^{(2)}_{lp}\delta_{0l}\delta_{0p}\right)\right]^2$$

$$\geqslant \left|\sum_{j,l\in D_1} \Gamma^{(1)}_j r^{(1)}_{jl}\delta_{0l}\right|^2 \left|\sum_{k,p\in D_2} \Gamma^{(2)}_k r^{(2)}_{kp}\delta_{0p}\right|^2 \tag{A2.1}$$

where

$$\mathbf{\Gamma}^{(k)} = \sum_{m\in S_k} \Gamma_m.$$

Matrices $\mathrm{Im}\,\mathbf{R}^{(k)}$ are negative definite since for arbitrary real vector x we have

$$X\,\mathrm{Im}\,\mathbf{R}^{(k)}x = -\mathrm{Im}(x(\mathbf{R}^{(k)})^*x) = -\mathrm{Im}(x\mathbf{R}^{(k)}(\mathbf{R}^{(k)})^{-1}(\mathbf{R}^{(k)})^*x)$$

$$= -(y\mathbf{\Gamma}^{(k)}y + z\mathbf{\Gamma}^{(k)}z] \leqslant 0 \tag{A2.2}$$

where $y = \mathrm{Re}(\mathbf{R}^{(k)}x)$, $z = \mathrm{Im}(\mathbf{R}^{(k)}x)$. To obtain (A2.2) we used

$$\mathbf{\Gamma}^{(k)} = 2\,\mathrm{Im}[(\mathbf{R}^{(k)})^{-1}]. \tag{A2.3}$$

According to (A2.2) each of the two sums on the left side of (A2.1) is negative. After

applying the inequality for arithmetic and geometrical averages to them we see that in order to prove (A2.1) it is sufficient to prove the inequality

$$-2\sum_{l,p\in D_k}\mathrm{Im}(r_{lp}^{(k)})\delta_{0l}\delta_{0p}\geqslant\sum_{j,l,p\in D_k}\Gamma_j^{(k)}r_{jl}^{(k)}(r_{jp}^{(k)})^*\delta_{0l}\delta_{0p} \tag{A2.4}$$

for $k=1,2$. But (2.4) is an equality due to (A2.3) and to the easily verified relation

$$2\,\mathrm{Im}\,\mathbf{R}^{(k)}+\mathrm{Re}(\mathbf{R}^{(k)}\Gamma^{(k)}(\mathbf{R}^{(k)})^*)=\mathbf{0}. \tag{A2.5}$$

References

[1] Capasso F, Mohammed K and Cho A Y 1986 *IEEE J. Quantum Electron.* **22** 1853
[2] Reed M A, Randall J N, Aggarwal R J, *et al* 1988 *Phys. Rev. Lett.* **60** 535
[3] Chou S Y, Allee D R, Pease R F W and Harris J S Jr 1989 *Appl. Phys. Lett.* **55** 176
[4] Kalmeyer V and Laughlin R B 1987 *Phys. Rev.* B **35** 9805
[5] Büttiker M 1988 *IBM J. Res. Dev.* **32** 63
[6] Bryant G W 1989 *Phys. Rev.* B **39** 3145
[7] Liu H C and Aers G C 1989 *J. Appl. Phys.* **65** 4908
[8] Tsuchiya M, Gaines J M, Yan R H, Simes R J, Holtz P O, Colden L A and Petroff P M 1989 *Phys. Rev. Lett.* **62** 466
[9] Li Y Z, Vazquez L, Piner R, Andres R P and Reinfenberger R 1989 *Appl. Phys. Lett.* **54** 1424
[10] van Loenen E J, Dijkkamp D, Hoeven A J, Lenssinck J M and Dieleman J 1989 *Appl. Phys. Lett.* **55** 1312
[11] Sautet P and Joachim C 1988 *Chem. Phys. Lett.* **153** 511; 1988 *Phys. Rev.* B **38** 12238
[12] Lifshits I M and Kirpichenkov V Ya 1979 *Zh. Eksp. Teor. Fiz.* **77** 989 (*Sov. Phys.–JETP* **50** 499)
[13] Xue W and Lee P A 1988 *Phys. Rev.* B **38** 3913
[14] Aers G C and Liu H C 1990 *Solid State Commun.* **73** 19
[15] Sumetstkii M Yu 1986 *Pis. Zh. Eksp. Teor. Fiz.* **44** 287 (*JETP Lett.* **44** 369)
[16] Büttiker M, Imry Y, Landauer R and Pinhas S 1985 *Phys. Rev.* B **31** 6207
Landauer R 1970 *Phil. Mag.* **21** 863
Thouless D J 1981 *Phys. Rev. Lett.* **47** 972
Azbel M Ya 1981 *J. Phys. C: Solid State Phys.* **14** L225
Imry Y 1986 *Directions of Condensed Matter Physics* ed G Grinstein and G Mazenko (Singapore: World Scientific) vol 1, p 101
[17] Büttiker M 1988 *IBM J. Res. Dev.* **32** 317
[18] Demkov Yu N and Ostrovskii V N 1975 *Method of Zero-Radius Potential in Atomic Physics* (Leningrad: Leningrad University Press) (Engl. transl. Plenum Press, New York, 1988)
[19] Sumetskii M Yu 1988 *Zh. Eksp. Teor. Fiz.* **94** 7 (*Sov. Phys.–JETP* **67** 438)
[20] De Vega H J, Gervais J-L and Sakita B 1978 *Nucl. Phys.* B **139** 20
[21] Büttiker M 1988 *Phys. Rev.* B **38** 12724
[22] Malov V V and Iogansen L V 1980 *Opt. Spektrosk.* **48** 146 (*Opt. Spectrosc.* **48** 81)
[23] Bellman R 1960 *Introduction to Matrix Analysis* (New York: McGraw-Hill)