Pramāṇa, Vol. 19, No. 3, September 1982, pp. 231–236. © Printed in India.

On the theory of electron tunnelling

D K ROY, N S T SAI and K N RAI*
Department of Physics, Indian Institute of Technology,
Hauz Khas, New Delhi 110 016, India
*On leave from Department of Physics, Bhagalpur University, Bhagalpur 812 007, India

MS received 1 December 1981; revised 28 June 1982

Abstract. A generalized expression for the differential tunnelling current density based on the problem of electron energy distribution introduced during the process has been presented. This is directly applicable to junction devices for the evaluation of their tunnelling I—V characteristics.

Keywords. Electron tunnelling; electron energy distribution.

1. Introduction

The first ever experimental evidence for particle tunnelling across potential barriers (Lilienfeld 1922) stimulated earlier workers to formulate its essential laws in the following words: (i) the particle energy must be conserved during tunnelling and (ii) the tunnelling current density is to be obtained by multiplying the incident current density by the tunnelling probability. No concept of tunnelling time is obviously deducible from these. They were employed subsequently by Fowler and Nordheim (1928), Gamow (1928) and Zener (1934) to their respective problems and they reported favourable agreement of their predictions with experiments. But, the finite limit of resolution of the field emission microscope introduced by Muller 1937 could not be understood on these premises. The field ionization of hydrogen atoms was, however, explained differently by Oppenheimer (1928) on the basis of time-dependent perturbation.

The theoretical understanding of this phenomenon took a dramatic turn with the introduction of the Esaki (1958) and Josephson (1962) effects. However, none of these could be understood thoroughly on the notions of the tunnelling problem mentioned above. Esaki's tunnelling current density expression could be explained only on the perturbation treatment of tunnelling of Bardeen (1961) which was a refinement over the ideas of Oppenheimer (1928). Cohen *et al* (1962) subsequently improved upon this treatment on the basis of the second quantized formulations later to be used by Josephson (1962) to predict his effects.

The theoretical formulations of the tunnelling problem that came to be established, therefore, were: (i) a time-independent approach based on the principle of conservation of energy and (ii) a time-dependent treatment depending upon the perturbation

*A list of symbols appear at the end of the paper.

theory. The predictions of these approaches agree since in the time-dependent perturbation theory a small change in the electron energy is allowed. If this change is presumed to be negligible compared to the original electron energy the latter would obviously merge with the former. Hence, it had customarily been a matter of convenience to theoreticians to adopt either of them to investigate a particular problem on tunnelling.

But it is not difficult to visualize that none of these techniques is capable enough to analyze the former in its true perspective. When electron waves happen to negotiate a barrier, their energies have at least to be uncertain by the heights of the barrier after the event and these are considerably larger relative to original electron energies. Hence, the ideas of insignificant or no change in the electron energies as embodied in the above formalisms of the tunnelling problem does not seem to be tenable. This was noticed by Roy (1977) who also suggested simultaneously that owing to larger barrier heights involved in tunnelling, the electron energy distribution introduced during the process is not to be ignored. The subsequent analyses by Roy *et al* (1977) lead to a definition of tunnelling time as well as provided an appropriate expression for the tunnelling current density.

2. Tunnelling current density

During the transit of electrons through the barrier, its potential energy is not only a function of position but also of time. Since the potential energy is a scalar quantity it should be expressed as,

$$V(x, t) = V_1(x) + V_2(t), \tag{1}$$

for $x_1 \leqslant x \leqslant x_2$ and $0 \leqslant t \leqslant \tau$ where x_1 and x_2 are the classical barrier turning points and τ is the electron tunnelling time. Since, $V(x, t)=0$ for $t<0$ and $t>\tau$, one must then naturally conclude that $V_1(x)=-V_2(t)$. Thus, although $V_2(t)$ happens to be a function of time, its extreme values depend upon $V_1(x)$. Also, since it exists only during tunnelling, one may define the tunnelling time as the time during which the electron potential energy remains non-zero. Again, as this function also controls the flow rate of the probability density the tunnelling current density is also determined by it. Substituting (1) into the Schrodinger's equation and separating the variables (Roy *et al* 1977), one obtains,

$$\psi(x, t) = a(t)\,\psi_l(x) \exp\left(\frac{-iE_l t}{\hbar}\right) + b(t)\,\psi_r(x) \exp\left(\frac{-iE_r t}{\hbar}\right), \tag{2}$$

where the time-dependent coefficients $a(t)$ and $b(t)$ are related to $V_2(t)$. It is to be noted that this wave function has been obtained as a general solution of the problem and not on the principles of the time-dependent perturbation theory. Since, the intensity of transmitted electrons is relatively weaker compared to the incident ones, the assumption $b(t) \lessdot a(t)$ sounds reasonably good. Also, since the transmitted intensity is measured relative to the incident one which also remains almost unattenuated in intensity during the process, there is no loss in generality in presuming $a(t) \simeq 1$ and $\dot{a}(t)=0$. The tunnelling current density is, however, to be estimated by solving the equation of continuity. The conventional current density operator which is also based on the latter has, however, failed to yield a true complexion of the tunnelling current density (Singh 1980). The tunnelling current density spectrum generated by a plane electron wave of energy E_l incident upon a potential barrier has been obtained as a direct solution of the equation of continuity in the following form:

$$J_t = J_{01} \frac{\sin\left[\left(\frac{E_l - E_r}{\hbar}\right)\tau\right]}{\left[\frac{(E_l - E_r)}{\hbar}\tau\right]} + J_{02} \sin\left[\frac{(E_l - E_r)}{\hbar}\tau + \theta\right], \tag{3}$$

where $$J_{01} = \frac{2q\tau\,|T_{lr}|^2}{\hbar^2 s}, \tag{4}$$

and $$J_{02} = \frac{2qT_{lr}}{\hbar s}\int_{x_1}^{x_2} \psi_r \psi_l^* \,\mathrm{d}x. \tag{5}$$

T_{lr} represents the standard tunnelling matrix element and $s=\int_{x_1}^{x_2}|\psi_r|^2\,\mathrm{d}x$. In equation (3), θ ac-counts for the additional phase difference introduced between $a(t)$ and $b(t)$ due to effects other than time. The second term in (3) has been interpreted as representing the Josephson effect (Roy and Sai 1982).

In metals and semiconductors, owing to the existence of a band of energy levels, a group of incoherent electron waves (varying randomly in phase) fall upon a potential barrier. Regaıding such an incident electron wave group of energy spread dE_l, the net differential tunnelling current density caused by them may be expressed as,

$$dJ_t(E_l) = \rho_l(E_l) f_l(E_l) dE_l \left[\sum_{\phi=-\infty}^{+\infty} \left\{J_{01}\frac{\sin\phi}{\phi} + J_{02}\sin(\phi+\theta)\right\}\right], \tag{6}$$

where $\rho_l(E_l), f_l(E_l)$ respectively denote the density of states and the Fermi distribution functions at the incident end and $\phi[=(E_l - E_r)\tau/\hbar]$ represents the phase of a transmitted electron wave measured relative to the phasor $(E_l\,\tau/\hbar)$. The phase difference amongst transmitted waves would vary ıandomly because the incident waves are incoherent and the barrier cannot alter their relative phase differences. Therefore, it appears more appropriate to state that the phase difference amongst the electron waves in a group are conserved during tunnelling. The limits of ϕ in (6) take into account the complete incoherence of the incident or the transmitted waves in question. To transform the summations in (6) into appropriate integrals we may note that since $\phi=(E_l - E_r)\tau/\hbar$,

$$\delta\phi = \frac{E_l\,\Delta t}{\hbar} = \frac{\Delta E_r\,\tau}{\hbar}, \tag{7}$$

Equation (7) clearly expresses that the phase difference introduced amongst waves of energy E_l due to their incidence upon the barrier at different times having a spread of $\triangle t$ can be reflected as a spread $\triangle E_r$ in their energy at the transmitted end. This explains the possibility of observing the energy distribution effect during tunnelling. By virtue of (7), the limiting phase difference between waves may then be expressed as,

$$d\phi = \epsilon\tau/\hbar \tag{8}$$

where ϵ measures the minimum energy separation between the levels of the band at the transmitted end. Around E_r, this may be expressed as,

$$\epsilon = \frac{1}{\Omega\rho_r(E_r)[1 - f_r(E_r)]}, \tag{9}$$

where Ω is the volume of the material to which tunnelling occurs and $\rho_r(E_r)$, $f_r(E_r)$ respectively denote the density of states and the Fermi distribution functions there. Since, from (8) $\hbar d\phi/\epsilon\tau = 1$, we have upon multiplying (6) by it,

$$dJ_t(E_l) = \rho_l(E_l) f_l(E_l) dE_l \; \frac{\hbar}{\epsilon\tau}\left[J_{01} \int_{-\infty}^{+\infty} \frac{\sin\phi}{\phi}\, d\phi + J_{02} \int_{-\infty}^{+\infty} \sin(\phi + \theta)\, d\phi\right], \tag{10}$$

where ϵ has been extracted out of integrals on the presumption that it remains relatively constant compared to the variations in ϕ. The second integral in (10) under this assumption would clearly vanish. The net differential tunnelling current density across the barrier would then be given by,

$$dJ_t(E_l) = \frac{\pi\hbar}{\epsilon\tau} J_{01}\, \rho_l(E_l) f_l(E_l)\, dE_l \tag{11}$$

This expression is directly applicable to junction devices for the evaluation of their I-V characteristics.

Some special cases of (11) are worth mentioning here. If at the transmitted end, only vacant energy levels exist at $E_r (= E_l)$, by virtue of (9), equation (11) may be expressed as,

$$dJ_t(E_l) = \frac{\pi\hbar\Omega}{\tau} J_{01}\, \rho_l(E_l)\, \rho_r(E_l) f_l(E_l)\, \{1 - f_r(E_l)\}\, dE_l. \tag{12}$$

Incidentally, this happens to be exactly in the form of Esaki's tunnelling current density expression from the left to the right of a degenerate *p-n* junction. On the other hand if at the transmitted end, the energy levels exist only around $E_r (\neq E_l)$, the current density expression in that case would reduce to,

$$dJ_t(E_l) = \frac{\pi\hbar\Omega}{\tau} J_{01} \rho_r(E_r)\, \{1 - f_r(E_r)\}\, \rho_l(E_l) f_l(E_l) dE_l. \tag{13}$$

Equations (11), (12) or (13) have to be integrated over appropriate ranges of E_l in order to determine the net tunnelling current density.

It is interesting to compare (11) with the corresponding conventional expression,

$$dJ_t(E_l)_{\text{conv}} = J_{\text{conv}}\, Z\, \rho_l(E_l)\, f_l(E_l)\, \mathrm{d}E_l \tag{14}$$

where Z is the electronic tunnelling probability and $J_{\text{conv}} = q \times$ thermal velocity of electrons. Next, comparing (11) with (14), one obtains,

$$\frac{dJ_t(E_l)}{dJ_t\,(E_l)_{\text{conv}}} = \frac{\pi\hbar}{\epsilon\,\tau}\,\frac{J_{01}}{J_{\text{conv.}}Z} \tag{15}$$

For a rectangular barrier the right side of (15) works out to be $(V_0/2\epsilon)\,(V_0/E)^{1/2}$ where V_0 is the height of the barrier and E represents the electron energy (Roy 1979). Taking $\epsilon \sim E \sim 1$ meV and $V_0 \sim 1$ eV, the above ratio works out to be in the vicinity of 10^5. Experimentally reported tunnelling current densities have also been found to difffer by this order of magnitude (Fischer and Giaever 1961).

List of Symbols

$a(t), b(t)$	time dependent coefficients
E	electron energy
E_l, E_r	electron energies on the left and the right of the barrier
$f_l(E_l), f_r(E_r)$	Fermi distribution functions
i	$\sqrt{-1}$
J_t	tunnelling current density
q	electronic charge
S	$\int_{x_1}^{x_2} \lvert\psi_r\rvert^2\, \mathrm{d}x$
T_{lr}	tunnelling transition matrix element
$V(x, t)$	electron potential energy function
$V_1(x)$	position-dependent electron potential energy
$V_2(t)$	time-dependent potential energy
$\psi(x, t)$	electron wave function
ψ_l, ψ_r	barrier wave functions on the left and the right sides
Ω	volume on the side to which electron tunnelling occurs
$\rho_l(E), \rho_r(E)$	density of state functions on the left and the right hand ends

References

Bardeen J 1961 *Phys. Rev. Lett.* **6** 57

Cohen M H, Falicov L M and Phillips J C 1962 *Phys. Rev. Lett.* **8** 316

Esaki L 1958 *Phys. Rev.* **109** 603
Fischer J C and Giaever I 1961 *J. Appl. Phys.* **32** 172
Fowler R and Nordheim L 1928 *Proc. Phys. Soc.* (*London*) **A119** 173
Gamow G 1928 *Z. Phys.* **51** 204
Josephson B D 1962 *Phys. Lett.* **1** 251
Lilienfeld J E 1922 *Z. Phys.* **23** 505
Muller E W 1937 *Z. Phys.* **106** 541
Oppenheimer J R 1928 *Phys. Rev.* **31** 66
Roy D K 1977 *Tunnelling and negative resistance phenomena in semiconductors* (Oxford: Pergamon)
Roy D K and Sai N S T 1982 *Indian J. Pure. Appl. Phys.* **20** 300
Roy P N 1979 *Tunnelling in junctions* Ph.D. Thesis, Bhagalpur University, Bhagalpur
Roy P N, Singh P N and Roy D K 1977 *Phys. Lett.* **A63** 81
Singh P N 1980 Tunnelling in Schottky barrier and metal-insulator semiconductor junctions Ph.D Thesis, Bhagalpur University, Bhagalpur
Zener C 1934 *Proc. R. Soc.* (*London*) **A145** 523