Consequences of Wannier - Stark quantization on the impact ionization rate in insulators and large-bandgap semiconductors

1998 J. Phys.: Condens. Matter 10 L607

(http://iopscience.iop.org/0953-8984/10/35/002)

J. Phys.: Condens. Matter **10** (1998) L607–L613. Printed in the UK PII: S0953-8984(98)95053-6

LETTER TO THE EDITOR

Consequences of Wannier–Stark quantization on the impact ionization rate in insulators and large-bandgap semiconductors

B K Ridley

Department of Physics, University of Essex, Colchester CO4 3SQ, UK

Received 12 June 1998

Abstract. At very high electric fields the formation of Wannier–Stark states converts the electron gas into a quasi-2D system. Comparison of the bulk and quasi-2D impact ionization rates within the framework of a simple bandstructure model shows that the 2D rate is significantly smaller. It is concluded that in materials where Wannier–Stark states can be established, breakdown by impact ionization is inhibited.

The possibility that Bragg reflection could cause the motion of an electron in a periodic potential to become oscillatory at high electric fields is well known [1–3]. The condition for this is that a full traversal of the first Brillouin zone in the direction of the field occurs before the electron is scattered, i.e. $eFa\tau/\hbar \geqslant 1$, where F is the field, a is the lattice constant and τ is the scattering time. The conduction band converts to a ladder of Wannier–Stark (WS) states whose properties continue to attract attention [4, 5]. A study of the transport of electrons in a single-band WS ladder has been made by a number of authors [6–8] in which it has been shown that the motion of electrons perpendicular to the field was describable in terms of a very high temperature, and that the nearly flat distribution that this implied led to a negative differential resistance (NDR). Following the analysis of Esaki and Tsu [9] of transport in a superlattice, there is now considerable experimental and theoretical interest in the investigation of WS states in layered structures (see e.g. [10]). As far as this author is aware the question of how impact ionization across the forbidden gap in bulk material is affected has not been addressed. A recent estimate of the scattering time in GaN yielded 1.4×10^{14} s^{-1} implying that a field of at least 4 MV cm^{-1} would be required to produce a WS ladder [11]. At fields of this magnitude the question of impact ionization cannot be avoided.

A comprehensive treatment of this topic would require an account of the detailed valence band and conduction band structure, plus the structure of higher lying conduction bands, and it would be intensively numerical. Nevertheless, the essential physics of the situation can be illustrated by adopting a very simple model for the bandstructure, with an analysis limited to the case of a single conduction band and a single valence band. For concreteness we consider the case of an electron in the conduction band exciting an electron in the valence band. In order to illustrate, in the simplest manner, the point that the formation of WS states inhibits impact ionization, the two bands are taken to be of simple tight-binding form for a cubic lattice. We will assume that communication between bands other than via impact ionization can be ignored. This simple model has the virtue of focusing attention

0953-8984/98/350607+07$19.50

on the effect of the difference in effective dimensionality associated with WS localization, and on the effect of an effective increase of threshold energy.

The quantum-mechanical bulk impact-ionization rate for an electron in a parabolic band with energy E close to threshold has the well known quadratic energy dependence [12, 13]. It is derived assuming that collisions with like-spins give negligible rates through interference, and may be presented in the form [14]

$$W_{3D} = W_o\left(\frac{\varepsilon_o}{\varepsilon_\infty}\right)^2 \frac{m_c^*}{m_o}\frac{I_c^2 I_v^2}{(1+2\mu)^{3/2}}\left(\frac{E-E_T}{E_g}\right)^2 \tag{1}$$

where $W_o = (e^2/4\pi\varepsilon_o)^2 m_0/\overline{h}^3$, ε_∞ is the high-frequency permittivity, m_c^* is the electron effective mass, $\mu = m_c^*/m_h^*$, m_h^* is the hole effective mass, I_c and I_v are cell-periodic overlap integrals, E_T is the threshold energy and E_g is the bandgap. The assumption of parabolicity restricts the validity of this expression to narrow-bandgap semiconductors, but a small modification can be made to increase its general validity. The modification is based on the assumption that the electron after impact ionizing, and the electron knocked out of the valence band, end up in low-energy parts of the conduction band where the band is parabolic to a good approximation. The electron producing impact ionization, on the other hand, is usually in a high-energy region of the band where the energy $E(\boldsymbol{k})$ is non-parabolic. The modification of equation (1) consists of simply refraining from using the parabolic relation between E and $\boldsymbol{k}$ (the wavevector of the impacting electron) in the derivation without changing anything else. The more generally applicable equation that is thereby obtained is

$$W_{3D} = W_o\left(\frac{\varepsilon_o}{\varepsilon_\infty}\right)^2 \frac{m_c^*}{m_o}\frac{I_c^2 I_v^2 (1+2\mu)^{5/2}}{(1+\mu)^4}\left(\frac{E(k)-(\mu/1+2\mu)(\overline{h}^2k^2/2m_c^*)-E_g}{(\overline{h}^2k^2/2m_c^*)}\right)^2. \tag{2}$$

Note that the impact-ionization process is forbidden if

$$E(k) < \frac{\mu}{1+2\mu}\frac{\overline{h}^2k^2}{2m_c^*} + E_g$$

which may be true in the higher energy regions of the conduction band. Indeed, there are crystallographic directions that do not have states appropriate for initiating impact ionization. Figure 1(a) illustrates this point. The energy bandstructure in the tight-binding model for a cubic lattice has the form:

$$E(k) = E_0 \sum_{\alpha=x,y,z} \sin^2\left(k_\alpha a/2\right) \tag{3}$$

where a is the lattice constant. With $E_g = E_0$, electrons travelling in the $\langle 100\rangle$ direction cannot impact ionize. In large-bandgap semiconductors the condition $E_g > E_0$ can arise, and this may totally inhibit impact ionizing collisions initiated by electrons in the lowest conduction band whether WS states are formed or not. In our discussion we will assume that this is *not* the case, since we are interested in contrasting the rates of impact ionization with and without WS states.

It may be noted that any problems computing I_c and I_v can be bypassed since we are interested only in relative magnitudes and we can assume that the relevant quantities do not change significantly in the transition from 3D to quasi-2D.

As far as the author is aware there is no expression equivalent to equation (1) for the impact-ionization rate for the quasi-2D case. However, the derivation of the 2D rate is straightforward and the result can readily be applied to the situation involving WS

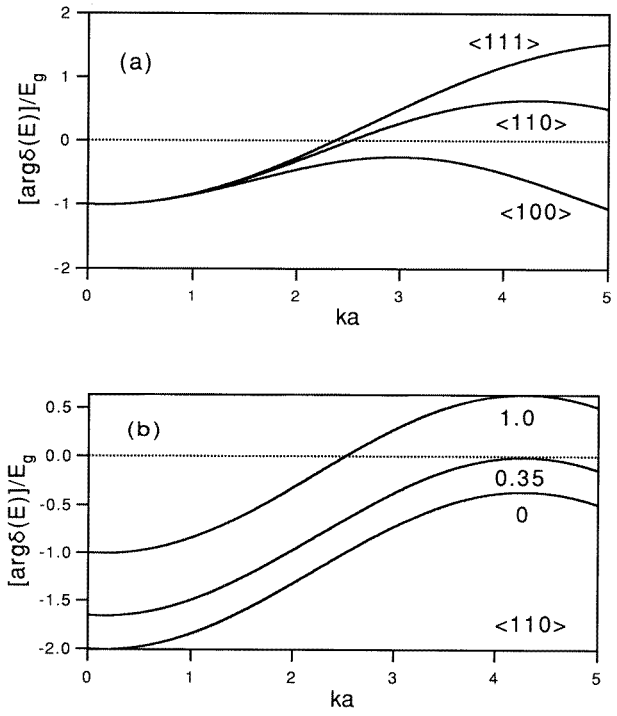

Figure 1. The argument of the energy-conserving delta function as a function of the wavevector of the impacting electron along principal cubic crystallographic directions assuming a cubic tight-binding bandstructure and $E_g = E_0$. (a) Bulk material. (b) Wannier–Stark confinement along a $\langle 100 \rangle$ direction, k along a $\langle 110 \rangle$ direction. The figures on the curves refer to the field reductions in the effective energy gap in units of E_0.

localization. The rate, again neglecting like-spin collisions and assuming that the electrons after impact ionization are in the parabolic region of the conduction band, is given by

$$W_{2D} = W_0 \left(\frac{\varepsilon_0}{\varepsilon_\infty} \right)^2 \frac{m_c^*}{m_0} \frac{2\pi I_c^2 I_v^2 (1+2\mu)}{(1+\mu)^2} \times \sum_{n_2,n_3,n_4} \left| F_{n_1,n_2}^{n_3,n_4}(k) \right|^2 \left(\frac{E(k) - (\mu/1+2\mu)(\hbar^2 k^2/2m_c^*) - E_g^*}{(\hbar^2 k^2/2m_c^*)} \right) \tag{4}$$

where

$$F_{n_1,n_2}^{n_3,n_4}(k) = \int\int \psi_{n_4}(z_2)\psi_{n_3}(z_1)\psi_{n_2}(z_2)\psi_{n_1}(z_1)\, \mathrm{e}^{-k|z_2-z_1|}\, \mathrm{d}z_1\, \mathrm{d}z_2/a^2 \tag{5}$$

$$E_g^* = E_g + \tfrac{1}{2}\left(E_{CB} + E_{VB}\right) - eF\left(z_{02}(n_4) + z_{01}(n_3) - z_{02}(n_2) - z_{01}(n_1)\right). \tag{6}$$

The initial states are designated by n_1 (conduction band) and n_2 (valence band). The final states n_3 and n_4 are both in the conduction band. The field is in the z-direction and $z_{01}(n_1)$ etc are the positions of the centres of the WS states. The form factor contains the Bloch oscillator wavefunctions of the states involved. E_{CB} and E_{VB} are the widths of the conduction and valence bands in the direction of the field. The wavevector $\boldsymbol{k}$, lying in the plane normal to the field, is that of the impacting electron.

The effective energy gap, E_g^*, is increased over the actual bandgap, E_g, by the sum of half the bandwidths of the conduction and valence bands, since the impacting electron and the target electron are both in WS states, but it also depends on the potential gradient. The effective enhancement of the bandgap can be nullified by transitions that send electrons down the potential gradient, but that will be at the expense of decrease of overlap of the initial and final wavefunctions.

The major differences from the 3D case are:

(1) The appearance of a form factor which quantifies the overlap and separation of the electrons involved.

(2) The effective energy gap is increased by the sum of the half-bandwidths of the conduction and valence bands and is dependent on the field.

(3) The energy dependence at threshold is linear instead of quadratic. If the field, and therefore the confinement, is along a $\langle 100 \rangle$ direction the optimum direction for the impacting electron is along a perpendicular $\langle 110 \rangle$ direction. If $E_g = E_0$ as before, and if E_0 is also the valence band energy, (i.e. $E_0 = E_{CB} = E_{VB}$) the effective bandgap without help from the applied field is now $2E_g$. Figure 1(b) illustrates that even along the $\langle 110 \rangle$ direction there is insufficient energy to impact ionize, and that some help from the field is needed, though this will be at the expense of a reduced overlap of initial and final states.

The possible transitions can be categorized as follows. The transitions of the impacting electron can be *intrasubband* ($n_3 = n_1$) or *intersubband* ($n_3 \neq n_1$) (figure 2); and they can be fully local, $z_{02}(n_2) = z_{01}(n_2)$ and $z_{02}(n_4) = z_{01}(n_3)$ (a in figure 2), half-local (b), or completely non-local (c). In fully local intrasubband transitions the impacting and target electrons before and after the interaction occupy WS states centred on the same unit cell. In this case the effective energy gap consists of the sum of the conduction and valence band half gaps and does not involve the potential gradient. In half-local intrasubband transitions the impacting and target electrons occupy WS states centred on the same unit cell, but the excited electron is found in a WS state centred elsewhere. In non-local intrasubband transitions the target and excited electron occupy WS states spatially separated from that of the impacting electron. In local intersubband transitions impacting and target electrons occupy WS states centred on the same unit cell and move to the same WS state after excitation (b in figure 2). In intersubband transitions the effective energy gap always involves the potential gradient. In half-local intersubband transitions spatial coincidence occurs only before or after impact. In non-local intersubband transitions there is no spatial coincidence (c in figure 2).

Generally, the wavevector of the impacting electron is large, and hence the form factors for most half- and completely non-local transitions will be negligible. Even the least partially non-local transition will have a form factor less than that for a local transition by a factor $\exp(-ka)$. Little error will be incurred by neglecting contributions from all transitions that have a non-local component. We therefore take into account only local transitions (a and b in figure 2).

Our model proceeds to make the simplifying assumption that $m_c^* = m_h^* = m_0$. In order to obtain an estimate of the form factor we exploit our simple model for the bandstructure in the field direction, namely

$$E = E_0 \sin^2 Ka/2 \qquad -\frac{\pi}{a} \lesssim K \lesssim \frac{\pi}{a} \tag{7}$$

where E_0 is the bandwidth, and this allows us to express the WS eigenfunction in the form of a Bessel function [8],

$$\psi_n(z) = J_n(u) \qquad \text{where} \qquad u = E_0/2eFa \qquad n = \big(z - z_o\big)/a \tag{8}$$

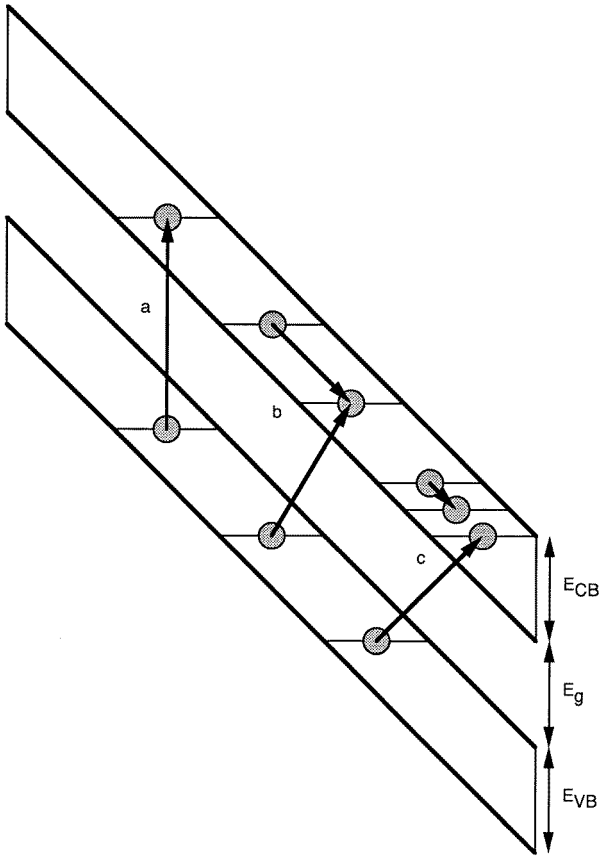

Figure 2. Types of impact ionizing transitions between Wannier–Stark states. Intrasubband transitions (impacting electron stays in initial WS state). Intersubband transitions (impacting electron changes WS state). Local transitions are depicted in a and b, non-local in c.

where z_0 is the position of the centre of the Bloch oscillation. Unless the field is enormous the argument u is very large and the asymptotic form of the Bessel function can be used. Normalization is approximated by ignoring the exponential-like tails for large n and imposing a cut-off at $|z| = \pi ua/2$. For simplicity we assume that $E_0 = E_{CB} = E_{VB}$. Then for local transitions such that $z_{02}(n_2) = z_{01}(n_1)$, $z_{02}(n_4) = z_{01}(n_3)$ and $d = z_{01}(n_3) - z_{01}(n_1)$, we obtain for $u \gg 1$

$$F^{n_3 n_4}_{n_1 n_2}(k) = \frac{2\cos^2(\pi d/2a)}{\pi uka}\left[1 - \frac{|d|}{\pi ua} - \frac{1}{\pi uka}(1 - \exp\{-\pi uka(1 - |d|/\pi ua)\})\right]. \tag{9}$$

This expression gives unity when both $|d| = 0$ and $k = 0$, as it should (figure 3(a)). Now $d = na$, where n is an integer, and so the form factor is zero unless n is even.

For the purpose of evaluating the sum in equation (4) we assume that there is zero overlap for $|d| > \pi ua$ and that the effective energy gap (equation (6)) is given by

$$E^*_g = E_g + E_0(1 - d/\pi ua) \qquad -\pi ua \lesssim d \lesssim \pi ua \tag{10}$$

so that the minimum effective energy gap is E_g. In other words we explicitly ignore Franz–Keldysh, sub-energy gap transitions. Figure 3(b) shows the dependence of $|F^{n_3 n_4}_{n_1 n_2}(k)|^2$ on d,

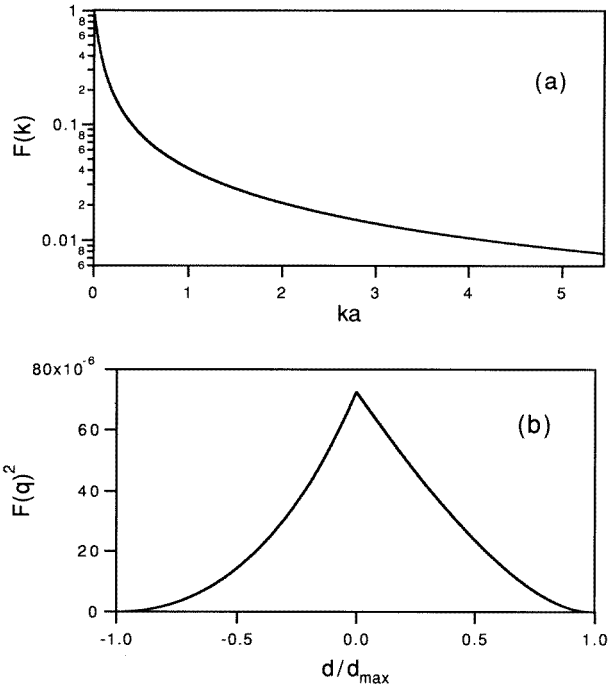

Figure 3. The form factor (a) as a function of normalized wavevector, ka ($d = 0$), (b) its square as a function of the spatial separation, d, of initial and final states, including the variation of k at threshold with d.

taking into account the dependence of k on d. In the example shown in figure 3(b), we have taken $E_B = E_g = 3.4$ eV, $a = 2.25$ Å and $F = 5$ MV cm^{-1}. Note that the rate vanishes when the effective bandgap, E_g^*, equals the bulk bandgap, *whatever the initial energy*. This is because the wavefunction overlap vanishes. Thus, finite rates are always associated with effective bandgaps that are larger than the actual bandgap. This means that they are always smaller than the bulk rate at a given field. They are also smaller when the comparison is made at the respective threshold energies because of the poor overlap, even taking into account all allowable transitions. A measure of the contribution made by the form factor is the sum; in our example, $\sum(F(k))^2 \approx 10^{-4}$. This is a measure of the reduction of impact ionization rate at threshold energy caused by the formation of WS states.

The conclusion that can be drawn is that impact ionization is inhibited if WS states form. This suggests that breakdown is dependent on the field direction relative to the crystallographic direction. Orientation of the field along a direction in which the dimension of the Brillouin zone is small encourages the formation of WS states (given the reasonable assumption that the scattering rate is largely independent of field orientation) and such an orientation therefore inhibits breakdown. In cubic crystals the dimension of the Brillouin zone is smallest along the ⟨100⟩ direction. A further factor in the case of large-bandgap semiconductors, even in the absence of WS states, is the necessity for the impacting electron to transfer to an upper band in order to acquire sufficient energy to impact ionize, and this is another factor encouraging inhibition. One therefore may speculate that electrical breakdown in materials possessing energy gaps of the order of or greater than typical conduction bandwidths will be associated with a mechanism other than impact ionization.

The author is indebted to the US Office of Naval Research (N00014-96-1-0998, MURI N00014-96-1-1223) and to EPSRC (GR/L 56725) for supporting this work.

References

[1] Bloch F 1928 *Z. Phys.* **52** 555
[2] Houston W V 1940 *Phys. Rev.* **57** 184
[3] Wannier G H 1960 *Phys. Rev.* **117** 432
Wannier G H 1962 *Rev. Mod. Phys.* **34** 645
[4] Avron J E 1982 *Ann. Phys.* **143** 33
[5] Zak J 1996 *J. Phys.: Condens. Matter* **8** 8295
[6] Bychkov Yu A and Dykhne A M 1965 *Sov. Phys.–JETP* **21** 779
[7] Bryskin V V and Firsov Yu A 1972 *Sov. Phys. Solid State* **13** 2729
Bryskin V V and Firsov Yu A 1972 *Sov. Phys.–JETP* **34** 1272
[8] Levinson I B and Yasevichyore Ya 1972 *Sov. Phys.–JETP* **35** 991
[9] Esaki L and Tsu R 1970 *IBM J. Res. Dev.* B **12** 61
[10] Mourokh L G and Smyrnov A Y 1998 *J. Phys.: Condens. Matter* **10** 3213
[11] Ridley B K 1997 *Proc. MRS Spring Meeting (San Francisco, April, 1997)* (Pittsburgh, PA: Materials Research Society)
[12] Keldysh L V 1960 *Sov. Phys.–JETP* **10** 509
[13] Robbins D J 1980 *Phys. Status Solidi* b **97** 9387
Robbins D J 1980 *Phys. Status Solidi* b **98** 1
[14] Ridley B K 1993 *Quantum Processes in Semiconductors* (Oxford: Oxford University Press) 3rd edn