

On Electron Transport Across Interfaces Connecting Materials with Different Effective Masses

Anatoly A. Grinberg and Serge Luryi, *Fellow, IEEE*

Abstract—We reconsider the problem of electronic transport across a heterojunction interface connecting two materials with different effective masses. Thermionic current does not remain invariant when the position X_{int} of the interface is shifted relative to that of the potential energy maximum, X_{max} , even if the shift is smaller than the carrier mean free path. We analyze every situation that arises when $X_{\text{max}} = X_{\text{int}}$, as well as the limiting case $X_{\text{max}} \rightarrow X_{\text{int}}$. Besides the effective mass discontinuity at the interface, we allow for a band-edge discontinuity within a potential barrier of arbitrary shape. In most practical situations, the effective Richardson constant governing thermionic emission over a heterointerface barrier is determined by the effective mass in the material that contains X_{max} . This statement is rigorously true when the effective mass at X_{max} is lower than that in the other material, otherwise it is an approximation dependent on the value of the interface potential. The issue clarified in this work has relevance to theoretical constructs involving model boundary conditions at heterojunction interfaces.

Index Terms—Heterojunction, thermionic emission, transmission.

I. INTRODUCTION

THEORY of electronic transport across an interface of different semiconductors is still being developed. In general, this is a difficult problem. For example, considering the interface between the crystals of Si and Ge, one has to take into account the different location in the Brillouin zone of the conduction band minima in these semiconductors. Large additional crystal momentum is therefore required to match electronic wavefunctions and transmission becomes forbidden except at very high energies [1]. As far as we are aware, there is no satisfactory theory which would describe carrier transport across a heterointerface in terms of the bulk parameters of constituent crystals.

Fortunately, for many practical heterojunction pairs, where the conduction band minima are similar, it is often a reasonable approximation to use the simplest theory based on the conservation of the lateral momentum and the total energy of particles transmitted through the interface. The only parameters of such a theory are the effective carrier masses in the two crystals and the band discontinuity at the interface.

In this approximation, the heterostructure transport problem is similar to that of electronic transmission across a

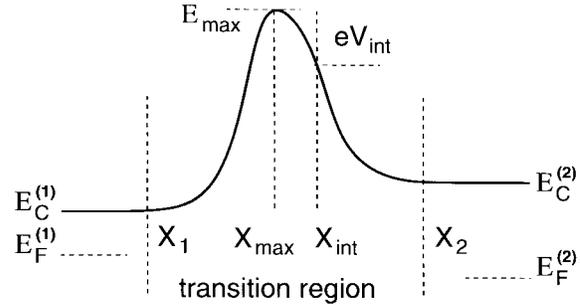


Fig. 1. Electron transport for an arbitrary-shape continuous barrier. Effective mass is m_1 to left and m_2 to the right of plane X_{int} . Thickness of the transition region is less than the mean free path for all scattering processes other than electron collision with the barrier itself.

potential barrier. In principle, this problem could be adequately described using a transmission coefficient which would incorporate on equal footing both the quantum mechanical tunneling processes and the classical kinematic restrictions at the interface between two media of different effective mass. Thus, it should be always possible to write the current per unit area from side 1 to side 2 of a potential barrier (Fig. 1) in the form

$$J_{1 \rightarrow 2} = e \int \frac{2 d\mathbf{k}}{(2\pi)^3} \frac{1}{\hbar} \frac{\partial E}{\partial k_z} \mathcal{D}(\mathbf{k}) f_1[E(\mathbf{k})] (1 - f_2[E(\mathbf{k})]) \quad (1)$$

where $f_i(E)$ are electron distribution functions in the planes X_i ($i = 1, 2$) and $\mathcal{D}(\mathbf{k}) = \mathcal{D}(k_x, k_y, k_z)$ is the quantum mechanical transmission (QMT) coefficient, which is defined as the ratio of the transmitted to incident flux for an electron of wavenumber \mathbf{k} incident on the barrier in plane X_1 . Such a procedure, however, is exceedingly difficult to carry out for a heterointerface that has an arbitrary position within a potential barrier of general shape.

We shall confine our consideration to purely thermionic transport, neglecting the quantum mechanical processes of tunneling and above-barrier reflections and focusing our attention on the consequences of classical kinematic restrictions. It is clear that thermionic currents in both directions at a heterointerface must be described by the same Richardson constant because in equilibrium the two currents must balance. What is the value of the common effective Richardson constant when the effective masses in the two materials are different?

This problem was first discussed by Stratton [2] in the instance of field emission from a semiconductor into vacuum. Stratton concluded that the thermionic current is given by a conventional Richardson formula in which, however, the constant A^* corresponds to the lower of the two effective

Manuscript received August 1, 1997; revised December 15, 1997. The review of this paper was arranged by Editor J. R. Hauser.

A. Grinberg is with Bell Laboratories, Lucent Technologies, Murray Hill, NJ 07974 USA.

S. Luryi is with the Department of Electrical Engineering, State University of New York at Stony Brook, Stony Brook, NY 11794 USA.

Publisher Item Identifier S 0018-9383(98)04501-8.

masses, m_1 and m_2 , viz

$$A^* = \frac{em^*k^2}{2\pi^2\hbar^3}, \quad m^* = \min\{m_1, m_2\}. \quad (2)$$

Similar conclusions were reached by Grinberg [3] for semiconductor heterostructures. These results were supposed to be valid for any two media where constant energy surfaces are spherical but the effective masses are different. However, applying essentially the same model to metal-semiconductor junctions in the limit of a large Schottky barrier height, Crowell [4] had shown that the common Richardson constant is determined by the effective mass in the semiconductor, irrespective of whether it is larger or smaller. Clearly, special assumptions must be made for (2) to be valid. Elucidation of these assumptions, and derivation of general results for thermionic emission are the main goals of the present work.

A number of workers, who considered this problem, have seemingly circumvented the question of choice of the appropriate Richardson constant by assuming that the potential maximum is shifted away from the interface, e.g., into vacuum in the field emission problem (or into semiconductor in the Schottky barrier problem) due to the image force or a distributed surface charge. However, the problem does not go away if the maximum is shifted by a distance less than the electronic mean free path, since there will be electrons that are reflected from the interface and go back over the potential maximum. Because of this the thermionic transport cannot be correctly described without specifying the relative position of the potential maximum and the heterointerface. This previously overlooked point is the central focus of our work. Depending on where X_{int} is positioned in the barrier, certain groups of electrons incident on the barrier from the heavier mass side may or may not be reflected by the interface. This purely kinematic condition determines an effective Richardson constant, which thus becomes dependent on X_{int} .

The problem is less apparent but equally critical when one includes not only thermionic emission but also tunneling processes. If all the information about the electron transmission through the barrier and the heterointerface remains hidden in the QMT, one obtains expressions that are formally correct but hardly useful, since they are no more amenable to actual calculation of the current than is the basic equation (1).

In the often cited treatment by Wu and Yang [5] the current is expressed in the form of quadratures with the QMT coefficient in the integrand. These expressions are derived from (1) by assuming a parabolic and isotropic spectra $E(\mathbf{k})$ in both materials and transforming the integration variables from k_x , k_y , and k_z to energy variables

$$E_{\perp} = (\hbar^2/2m^*)k_z^2 \quad (3)$$

$$E_{\parallel} = (\hbar^2/2m^*)(k_x^2 + k_y^2) \quad (4)$$

where $m^* = m_1$ on side 1 and $m^* = m_2$ on side 2. Further development of this approach should involve specification of the limits for integration in variables E_{\perp} and E_{\parallel} . However, since Wu and Yang [5] included tunneling processes into consideration, the limits corresponding to thermionic emission were not explicitly defined. Moreover, since they had considered a potential barrier of arbitrary shape, without specifying

the exact position of the heterointerface within the barrier the information about this position was implicitly hidden within the QMT. In this situation, the integration limits are meaningless—even when tunneling is included—because the resultant quadrature expressions can be viewed as formally correct only to the extent that the QMT coefficient implicitly includes all the unspecified details about the positions of the heterointerface and the potential maximum in the transition region.

In light of this, developing a general thermionic theory including tunneling, does not appear promising for heterosystems. The problem reduces, in general, to a difficult calculation of the QMT coefficient, followed by a straightforward numerical evaluation of the integral (1). On the other hand, restricting the consideration to classical thermionic emission, one can make further progress, as discussed below. In particular, we show that the low-mass rule [2], [4], expressed by (2), is valid only for a special case of a continuous potential, when the potential energy maximum is located in the plain where the effective mass suffers a discontinuity. Moreover, for the general case we show that the low-mass rule (2) remains universally valid when the potential energy maximum is located in the lower-mass semiconductor. In contrast, no universal relation can be established for the case when the top of the barrier falls within the heavy-mass semiconductor, as in this case the result depends on the actual potential and the band discontinuity at the interface. A simple rule emerges, however, when the energy separation between the top of the barrier and the interface potential is much larger than kT . In this practically most relevant case the Richardson constant is determined by the effective mass in the material that contains the barrier top.

II. CONTINUOUS POTENTIAL

Let us first consider a heterosystem with a potential barrier of general shape, Fig. 1, which, however, includes no band discontinuity (the latter case will be treated in Section III). We assume that outside the transition region marked by planes X_1 and X_2 the system is in quasi-equilibrium. The heterointerface is assumed located at $x = X_{\text{int}}$ where $X_1 \leq X_{\text{int}} \leq X_2$ and the potential maximum at $x = X_{\text{max}}$ where $X_1 \leq X_{\text{max}} \leq X_2$.

Leaving aside (till Section IV) the question of validity of the thermionic approximation, we calculate the thermionic current assuming that the carrier mean free path λ is much larger than the thickness of the transition region, $X_2 - X_1 \ll \lambda$.

Electronic energies in planes X_1 and X_2 referenced to an arbitrary common level (Fig. 1) will be denoted by superscripts (1) and (2), respectively

$$E^{(1)} = E_{\perp}^{(1)} + E_{\parallel}^{(1)} + E_C^{(1)} \quad (5a)$$

$$E^{(2)} = E_{\perp}^{(2)} + E_{\parallel}^{(2)} + E_C^{(2)} \quad (5b)$$

where E_{\perp} and E_{\parallel} denote, respectively, the kinetic energies of the carrier motion perpendicular and parallel to the interface.

Let us denote by $J_{1 \rightarrow 2}$ the current density of electrons in plane X_1 flying toward the interface with “allowed” energies, i.e., those energies that enable an electron to reach plane X_2 . Formally, in (1) we assign the value $\mathcal{D} = 1$ to classically allowed trajectories and $\mathcal{D} = 0$ otherwise. The similar quantity

for allowed electrons in plane X_2 will be denoted by $J_{2 \rightarrow 1}$. Using (1) and (3)–(5), the current densities $J_{1 \rightarrow 2}$ and $J_{2 \rightarrow 1}$ can be brought into the form

$$J_{1 \rightarrow 2} = -A_1 T^2 \int f_1(E^{(1)}) dE_{\perp}^{(1)} dE_{\parallel}^{(1)} [1 - f_2(E^{(2)})] \quad (6a)$$

$$J_{2 \rightarrow 1} = A_2 T^2 \int f_2(E^{(2)}) dE_{\perp}^{(2)} dE_{\parallel}^{(2)} [1 - f_1(E^{(1)})] \quad (6b)$$

where $A_i = em_i k^2 / 2\pi^2 \hbar^3$ ($i = 1, 2$) are Richardson's constants of the two materials.

In (6) and throughout this paper, carrier energies are expressed in units of kT .

The main problem in formulating the thermionic current is to specify the range of integration in (6), i.e., to formally separate the classically allowed and the forbidden electronic trajectories.

In each of (6) one has $E^{(1)} = E^{(2)}$ due to conservation of energy. Since neither of the quantities $E_{\parallel}^{(1)}$ and $E_{\parallel}^{(2)}$ vary in the course of particle motion through the barrier, except on crossing the interface, the conservation of momentum parallel to the interface implies

$$m_1 E_{\parallel}^{(1)} = m_2 E_{\parallel}^{(2)}. \quad (7)$$

From the conservation of energy and (7) it follows that

$$E_{\perp}^{(1)} = E_{\perp}^{(2)} + (1 - \theta^{-1}) E_{\parallel}^{(2)} + E_C^{(2)} - E_C^{(1)} \quad (8a)$$

$$E_{\parallel}^{(1)} = \theta^{-1} E_{\parallel}^{(2)} \quad (8b)$$

where $\theta \equiv m_1/m_2$. The transformation law for the elementary area in the planes of variables $\{E_{\parallel}^{(1)}, E_{\perp}^{(1)}\}$ and $\{E_{\parallel}^{(2)}, E_{\perp}^{(2)}\}$ follows from (8)

$$dE_{\parallel}^{(1)} dE_{\perp}^{(1)} = \theta^{-1} dE_{\parallel}^{(2)} dE_{\perp}^{(2)}. \quad (9)$$

Using (8) and (9), we can express the integrands in (6) in terms of one or the other set of variables. Written in the same variables, the integrals in (6a) and (6b) are taken between identical limits. Indeed, for any allowed electron trajectory, originating in plane X_2 and labeled by $\{E_{\parallel}^{(2)}, E_{\perp}^{(2)}\}$, the corresponding trajectory, labeled by $\{E_{\parallel}^{(1)}, E_{\perp}^{(1)}\}$ according to (8) and originating from plane X_1 , will be also allowed. Therefore, combining (6a) and (6b) we can write the current density $J \equiv J_{1 \rightarrow 2} + J_{2 \rightarrow 1}$ in either of the two equivalent forms:

$$J = A_1 T^2 \int [f_2(E^{(1)}) - f_1(E^{(1)})] dE_{\parallel}^{(1)} dE_{\perp}^{(1)} \quad (10a)$$

$$J = A_2 T^2 \int [f_2(E^{(2)}) - f_1(E^{(2)})] dE_{\parallel}^{(2)} dE_{\perp}^{(2)}. \quad (10b)$$

To evaluate the current we must specify the range of integration (10) either in variables $\{E_{\parallel}^{(1)}, E_{\perp}^{(1)}\}$ or $\{E_{\parallel}^{(2)}, E_{\perp}^{(2)}\}$.

Consider first what happens to the normal contribution to kinetic energy when a particle crosses the effective mass boundary, $x = X_{\text{int}}$. From the conservation of energy in the transition from the plane $X_{\text{int}}^{(-)} \equiv X_{\text{int}} - \delta x$ to the plane $X_{\text{int}}^{(+)} \equiv X_{\text{int}} + \delta x$, we have

$$E_{\perp}^{(-)} - E_{\perp}^{(+)} = (\theta - 1) E_{\parallel}^{(1)} = (1 - \theta^{-1}) E_{\parallel}^{(2)} \quad (11)$$

where $E_{\perp}^{(-)} \equiv E_{\perp}[X_{\text{int}}^{(-)}]$ and $E_{\perp}^{(+)} \equiv E_{\perp}[X_{\text{int}}^{(+)}$. It follows from (11) that if on crossing a boundary the effective mass of

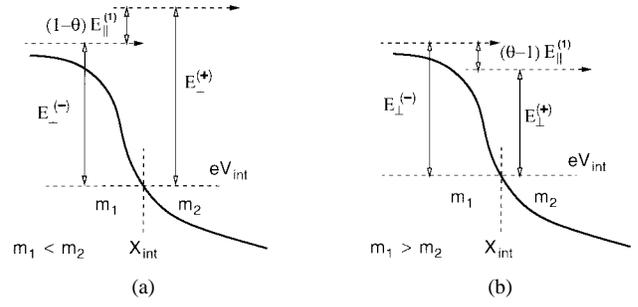


Fig. 2. Illustration of electron kinematics for a continuous potential at the interface.

a particle increases, then so does the normal part, E_{\perp} , of its kinetic energy—for any nonvanishing value of in-plane kinetic energy, $E_{\parallel} \neq 0$, see Fig. 2.

To carry the analysis further, we need to specify the relative positions of the planes X_{int} and X_{max} . Although (10) allow us to evaluate the current in either set of variables $\{E_{\parallel}^{(1)}, E_{\perp}^{(1)}\}$ or $\{E_{\parallel}^{(2)}, E_{\perp}^{(2)}\}$, it is more convenient to proceed in that set which corresponds to the effective mass of the region that includes X_{max} . This gives rise to a simpler configuration of allowed energies and hence simpler integration domain.

Consider first the case when $X_{\text{int}} > X_{\text{max}}$. An obvious necessary condition that an electron originating at X_1 could overcome the barrier is

$$E_{\perp}^{(1)} > E_{\text{max}} - E_C^{(1)} \quad (12)$$

where $E_{\text{max}} = eV(X_{\text{max}})$. This, however, is not a sufficient condition. Approaching X_{int} from the left, the normal energy of this electron will be

$$E_{\perp}^{(-)} = E_{\perp}^{(1)} - (eV_{\text{int}} - E_C^{(1)})$$

and immediately upon crossing the boundary it will be given by (11)

$$E_{\perp}^{(+)} = E_{\perp}^{(-)} - (\theta - 1) E_{\parallel}^{(1)}.$$

For the particle to be able to cross X_{int} the quantity $E_{\perp}^{(+)}$ should be positive, whence we arrive at another inequality

$$E_{\perp}^{(1)} > [eV_{\text{int}} - E_C^{(1)}] + (\theta - 1) E_{\parallel}^{(1)} \quad (13)$$

which must be satisfied together with (12). Evidently, for $\theta < 1$ the inequality (12) is stronger than (13) and hence the latter does not affect the allowed energies, cf., Fig. 3(a). In contrast, for $\theta > 1$ it depends on $E_{\parallel}^{(1)}$ which of the two inequalities is more restrictive. The allowed energies for $\theta > 1$ are shown in Fig. 3(b).

Having defined the domain of integration for $X_{\text{int}} > X_{\text{max}}$, we can write the current density in the general form

$$J = A_1 T^2 \int_{E_{\text{max}} - E_C^{(1)}}^{\infty} dE_{\perp}^{(1)} \cdot \left(\int_0^{\infty} dE_{\parallel}^{(1)} [f_2(E^{(1)}) - f_1(E^{(1)})] - \mathcal{H}(\theta - 1) \cdot \int_{\mathcal{E}(E_{\perp}^{(1)})}^{\infty} dE_{\parallel}^{(1)} [f_2(E^{(1)}) - f_1(E^{(1)})] \right) \quad (14)$$

for $X_{\text{int}} > X_{\text{max}}$

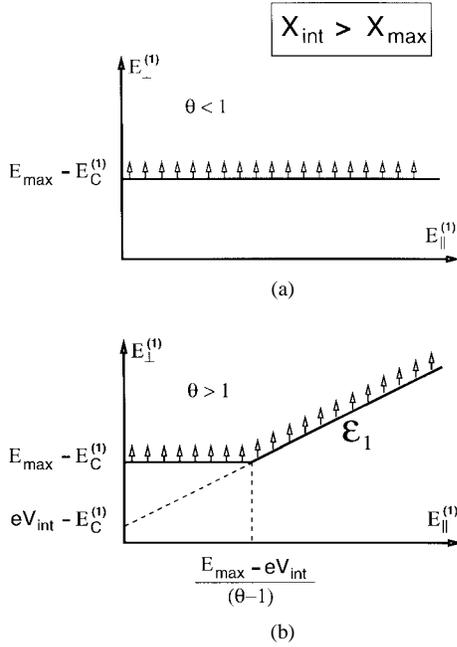


Fig. 3. Allowed energies for $X_{\text{int}} > X_{\text{max}}$ for (a) $m_1 \geq m_2$ and (b) $m_1 \leq m_2$. Small arrows indicate the domain of integration. They originate from the domain boundary. Line \mathcal{E}_1 is described by the equation $E_{\perp}^{(1)} = eV_{\text{int}} - E_C^{(1)} + (\theta - 1)E_{\parallel}^{(1)}$.

where the lower limit $\mathcal{E}(E_{\perp}^{(1)})$ in the second integral over $E_{\parallel}^{(1)}$ is given by

$$\mathcal{E}(E_{\perp}^{(1)}) = \frac{E_{\perp}^{(1)} - [eV_{\text{int}} - E_C^{(1)}]}{\theta - 1} \quad (15)$$

and $\mathcal{H}(\theta)$ is the step function (i.e., the second integral is subtracted only when $m_1 > m_2$).

The case $X_{\text{int}} < X_{\text{max}}$ is best considered in variables $\{E_{\parallel}^{(2)}, E_{\perp}^{(2)}\}$ and can be derived from the previous case by a reflection transformation, $x \rightarrow -x$. Expressions for the current density are obtained from (14) and (15) by transposing the indices $1 \leftrightarrow 2$, replacing $\theta \rightarrow \theta^{-1}$, and changing the sign of the current, $J \rightarrow -J$. The result is

$$J = A_2 T^2 \int_{E_{\text{max}} - E_C^{(2)}}^{\infty} dE_{\perp}^{(2)} \cdot \left(\int_0^{\infty} dE_{\parallel}^{(2)} [f_2(E^{(2)}) - f_1(E^{(2)})] - \mathcal{H}(1 - \theta) \cdot \int_{\mathcal{E}(E_{\perp}^{(2)})}^{\infty} dE_{\parallel}^{(2)} [f_2(E^{(2)}) - f_1(E^{(2)})] \right) \quad (16)$$

for $X_{\text{int}} < X_{\text{max}}$

where the lower limit $\mathcal{E}(E_{\perp}^{(2)})$ in the second integral over $E_{\parallel}^{(2)}$ is given by

$$\mathcal{E}(E_{\perp}^{(2)}) = \frac{E_{\perp}^{(2)} - [eV_{\text{int}} - E_C^{(2)}]}{\theta^{-1} - 1} \quad (17)$$

The range of allowed energies for $X_{\text{int}} < X_{\text{max}}$ in the variables $\{E_{\parallel}^{(2)}, E_{\perp}^{(2)}\}$ is shown in Fig. 4. Fig. 5 illustrates the

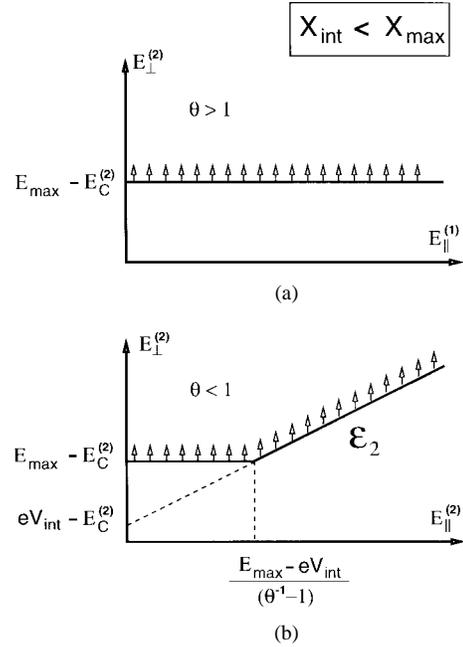


Fig. 4. Allowed energies and the integration domain in variables $\{E_{\parallel}^{(2)}, E_{\perp}^{(2)}\}$ for $X_{\text{int}} < X_{\text{max}}$ and (a) $m_1 \geq m_2$, (b) $m_1 \leq m_2$. Line \mathcal{E}_2 is described by the equation $E_{\perp}^{(2)} = eV_{\text{int}} - E_C^{(2)} + (\theta^{-1} - 1)E_{\parallel}^{(2)}$.

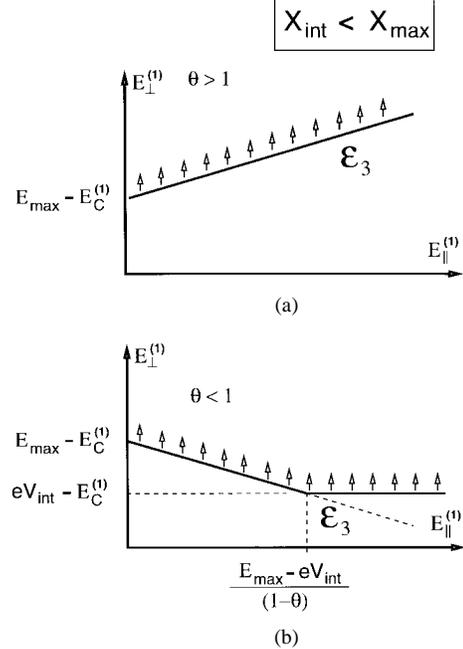


Fig. 5. Same as in Fig. 4 but in variables $\{E_{\parallel}^{(1)}, E_{\perp}^{(1)}\}$. Lines \mathcal{E}_3 are described by the equation $E_{\perp}^{(1)} = E_{\text{max}} - E_C^{(1)} + (\theta - 1)E_{\parallel}^{(1)}$.

same range in variables $\{E_{\parallel}^{(1)}, E_{\perp}^{(1)}\}$ and shows that in these variables the domain of integration has a more complicated form.

It follows from (14)–(17) that for $X_{\text{int}} = X_{\text{max}}$ the current is governed by the Richardson constant (2) corresponding to the lower effective mass. Thus, for $m_1 < m_2$, the second

term in (14) vanishes and one obtains the usual thermionic expression with A_1 . For $m_1 > m_2$, the vanishing second term is in (16) and A_2 becomes the effective Richardson constant.

The same rule applies to the case $X_{\text{int}} \neq X_{\text{max}}$, provided the potential energy maximum is located in the lower mass semiconductor.

In contrast, when the potential energy maximum falls in the heavier mass semiconductor, the effective Richardson constant depends on the potential energy value at the heterointerface relative to the maximum ($E_{\text{max}} - eV_{\text{int}}$). Taking the distribution functions in the nondegenerate Boltzmann form, we find from (14) for $m_1 > m_2$ and $X_{\text{int}} > X_{\text{max}}$

$$J = A_2 T^2 \eta e^{E_F^{(2)} - E_{\text{max}}} (1 - e^{E_F^{(1)} - E_F^{(2)}}) \quad (18)$$

where η is given by

$$\eta = \theta - (\theta - 1) \exp\left(-\frac{E_{\text{max}} - eV_{\text{int}}}{\theta - 1}\right), \quad \theta > 1. \quad (19)$$

Similarly, from (17) we find for $m_1 < m_2$ and $X_{\text{int}} < X_{\text{max}}$

$$J = A_1 T^2 \eta e^{E_F^{(1)} - E_{\text{max}}} (e^{E_F^{(2)} - E_F^{(1)}} - 1) \quad (20)$$

with η given by

$$\eta = \theta^{-1} - (\theta^{-1} - 1) \exp\left(-\frac{E_{\text{max}} - eV_{\text{int}}}{\theta^{-1} - 1}\right), \quad \theta < 1. \quad (21)$$

Equations (18)–(21) imply that when the potential energy at the interface is far below E_{max} , then the common Richardson constant is governed by the effective mass in the material where the maximum is located. Indeed, from (19) it follows that

$$\eta \rightarrow \theta \quad \text{for} \quad \frac{E_{\text{max}} - eV_{\text{int}}}{\theta - 1} \gg 1$$

hence, the effective Richardson constant ηA_2 in (18) tends to A_1 . Same conclusion, naturally, follows from (20) where $\eta A_2 \rightarrow A_1$ when in (21)

$$\eta \rightarrow \theta^{-1} \quad \text{for} \quad \frac{E_{\text{max}} - eV_{\text{int}}}{\theta^{-1} - 1} \gg 1.$$

III. DISCONTINUOUS POTENTIAL

Consider now the case when the potential suffers a discontinuity at the heterointerface,

$$V(x) \rightarrow \begin{cases} V_{\text{int}}^{(-)} & \text{for } x \rightarrow X_{\text{int}}^{(-)}; \\ V_{\text{int}}^{(+)} & \text{for } x \rightarrow X_{\text{int}}^{(+)} \end{cases} \quad (22)$$

where $eV_{\text{int}}^{(+)} - eV_{\text{int}}^{(-)} \equiv \Delta E_C|_{X_{\text{int}}} \neq 0$ (see Fig. 6). This is the case most relevant to practical semiconductor heterojunctions, such as those used in emitter-base junctions of bipolar transistors [Fig. 6(a)]. The situation shown in Fig. 6(b) arises when one introduces interfacial dipoles by epitaxy in order to

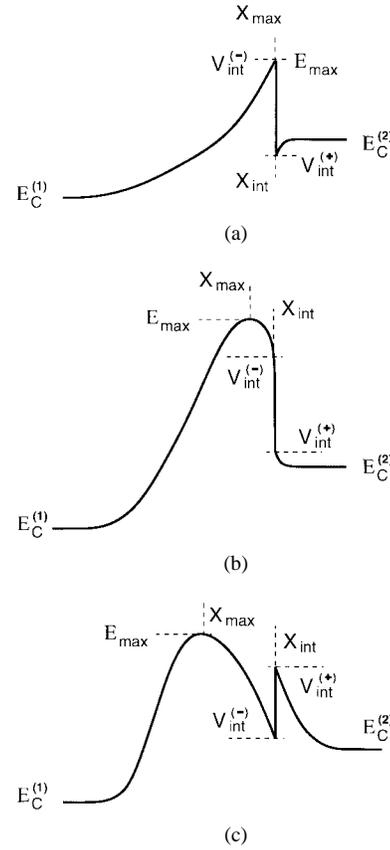


Fig. 6. Potential diagrams of heterojunctions with a conduction band discontinuity at the interface for $X_{\text{int}} \geq X_{\text{max}}$.

enhance a band-edge discontinuity [6]. For completeness, we shall also consider the situation, described in Fig. 6(c).

Without a loss of generality we can restrict our consideration to the case $X_{\text{int}} \geq X_{\text{max}}$, as illustrated in Fig. 6. In this case the convenient variables are $\{E_{\parallel}^{(1)}, E_{\perp}^{(1)}\}$. As above, the case $X_{\text{int}} < X_{\text{max}}$ is obtained by reflection $x \rightarrow -x$; the corresponding expressions in variables $\{E_{\parallel}^{(2)}, E_{\perp}^{(2)}\}$ result by transposing the indices $1 \leftrightarrow 2$ with the obvious replacements:

$$\begin{aligned} J &\rightarrow -J; \\ \theta &\rightarrow \theta^{-1}; \\ V_{\text{int}}^{(+)} &\leftrightarrow V_{\text{int}}^{(-)}. \end{aligned} \quad (23)$$

Consider first the case when X_{max} does not coincide with the interface, viz. $X_{\text{int}} > X_{\text{max}}$, as in Fig. 6(b) and (c). Instead of Fig. 2 we now have the band diagram shown in Fig. 7. Evidently, (11) must be replaced by

$$E_{\perp}^{(-)} - E_{\perp}^{(+)} = (\theta - 1)E_{\parallel}^{(1)} + [eV_{\text{int}}^{(+)} - eV_{\text{int}}^{(-)}]. \quad (24)$$

On the other hand, (12) remains in force. Following steps similar to those leading to (13), we find that allowed energies must satisfy besides (12) an inequality of the form

$$E_{\perp}^{(1)} > [eV_{\text{int}}^{(+)} - E_C^{(1)}] + (\theta - 1)E_{\parallel}^{(1)} \quad (25)$$

which is identical to (13) except that V_{int} is replaced by $V_{\text{int}}^{(+)}$. Accordingly, the same replacement must be made in Fig. 3(b)

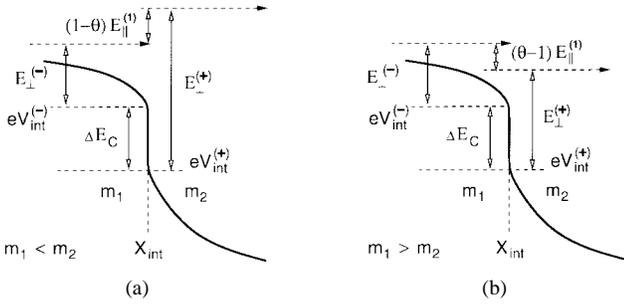


Fig. 7. Illustration of electron kinematics for a potential with a conduction band discontinuity at the interface.

and hence the lower integration limit in (14) will be given by the following expression [instead of (15)]:

$$\mathcal{E}(E_{\perp}^{(1)}) = \frac{E_{\perp}^{(1)} - [eV_{\text{int}}^{(+)} - E_C^{(1)}]}{\theta - 1}. \quad (26)$$

As before, if $m_1 < m_2$, i.e., when the potential maximum is in the lower mass semiconductor, the second integral in (14) is eliminated and we obtain the conventional thermionic expression with $A^* = A_1$. The lower-mass rule remains valid in this case.

In contrast, for $m_1 > m_2$, i.e., when the potential maximum is in the heavy-mass region, the effective Richardson constant becomes a function of $[E_{\text{max}} - eV_{\text{int}}^{(+)}]$, similar to the case of continuous potential considered in Section II. For a nondegenerate Boltzmann distribution, (14) now leads to an equation of the form (18) but with η given by

$$\eta = \theta - (\theta - 1) \exp\left(-\frac{[E_{\text{max}} - eV_{\text{int}}^{(+)}]}{\theta - 1}\right), \quad \theta > 1. \quad (27)$$

Another way of arriving at the same result (27), starting from equations of Section II that have been derived for a continuous potential, is by the following argument. We note that, given the quasi-Fermi levels in the two materials, (18) and (19) are controlled only by the interface potential V_{int} and the barrier height E_{max} , and depend neither on the shape of the barrier nor the peak position, provided the latter stays in the heavy mass region, $X_{\text{max}} < X_{\text{int}}$. The same is true for the more general equation (14), given the distribution functions f_1 and f_2 in the two reservoirs.

The result of the integration (14) is therefore invariant under an arbitrary continuous deformation of the potential profile, subject to the condition $X_{\text{max}} < X_{\text{int}}$ and fixed values of E_{max} and V_{int} relative to $E_C^{(1)}$. Such a transformation is illustrated in Fig. 8(a). In the course of deformation, the potential $V(x)$ may acquire steep walls where its derivative may diverge, $V'(x) \rightarrow \infty$, so that the potential at X_{int} would become multivalued in the limit. In this case, the relevant value of $V(X_{\text{int}})$ in (17) is evidently $V[X_{\text{int}}^{(+)}] \equiv V_{\text{int}}^{(+)}$.

Consider now the case $X_{\text{max}} = X_{\text{int}}$, which can be regarded as the limiting case of $X_{\text{max}} \rightarrow X_{\text{int}}$. Evidently, we can continuously deform the potential diagram of Fig. 6(b) into Fig. 6(a), as is illustrated in Fig. 8(b), whereas the diagram

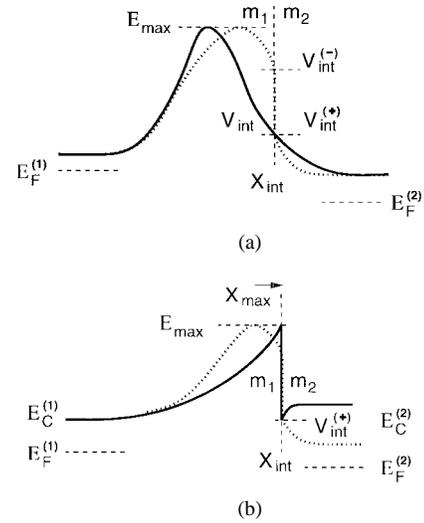


Fig. 8. Continuous deformations of the potential profile, which leave thermionic current invariant.

in Fig. 6(c) cannot be so deformed.¹ Hence, when making the limiting transformation $X_{\text{max}} \rightarrow X_{\text{int}}$ from the side $X_{\text{max}} < X_{\text{int}}$, we must assume that $V_{\text{int}}^{(+)} < V_{\text{int}}^{(-)}$. In such a transformation $eV_{\text{int}}^{(-)} \rightarrow E_{\text{max}}$, cf., Fig. 8(b). Hence, the thermionic expression that results for a nondegenerate distribution in the case $m_1 > m_2$ is of the form (18) with η given by (27). For $m_1 < m_2$ the result is of Stratton's form (2).

Thus, the conclusion we derived at the end of Section II for a continuous potential, remains valid for an arbitrary potential with a band-edge discontinuity at the interface: *the Richardson constant is governed by the effective mass of the material containing the potential maximum*. This statement is rigorously true when the effective mass in this material is lower; otherwise it is an approximation dependent on the value of the interface potential. This approximation becomes accurate (with an exponentially negligible error) when one of the following inequalities is satisfied:

$$\frac{E_{\text{max}} - eV_{\text{int}}^{(+)}}{\theta - 1} \gg 1, \quad \text{for } X_{\text{max}} < X_{\text{int}}; \quad (28)$$

$$\frac{E_{\text{max}} - eV_{\text{int}}^{(-)}}{\theta - 1} \gg 1, \quad \text{for } X_{\text{max}} > X_{\text{int}}. \quad (29)$$

IV. CONDITIONS FOR VALIDITY OF THE THERMIONIC APPROXIMATION

The effective mass discontinuity gives rise to an interesting question regarding the validity of thermionic approximation. The invariant deformation of the potential in the region between X_{max} and X_{int} —including re-positioning of the point X_{max} —discussed in the preceding Section, is permissible only if the distances involved are much shorter than the electronic mean free path λ . This may appear to be an unnecessary restriction, because it certainly goes beyond the usual condition for the validity of thermionic approximation. As is well known, see, e.g. [7], this approximation is valid provided λ exceeds

¹In treating the case $X_{\text{max}} \neq X_{\text{int}}$ we have naturally assumed that $E_{\text{max}} > V_{\text{int}}^{(+)}$, $V_{\text{int}}^{(-)}$, irrespective of whether ΔE_C is positive [as in Fig. 6(c)] or negative [Fig. 6(b)].

the distance in which the barrier falls by kT from its maximum value (the Bethe condition). This condition may be insufficient, however, to ensure the validity of (14).

Indeed, consider electrons whose E_{\perp} is large enough to overcome the potential barrier but not large enough to cross the interface into the region of lower mass. Such electrons suffer an internal reflection at the interface and return back to the quasi-equilibrium ensemble from which they had originated—unless they are allowed to scatter in between X_{int} and X_{max} . If scattering is allowed, even elastic, then (13) is no longer a restriction, because the normal component E_{\perp} can increase at the expense of E_{\parallel} . We therefore conclude that the validity of the “exotic” thermionic regime with $\eta \neq 1$ requires that the Bethe condition be supplemented by

$$\lambda \gg |X_{\text{max}} - X_{\text{int}}|. \quad (30)$$

In contrast, the inequality (30) is not required when X_{max} is displaced from X_{int} in the lighter mass direction. Physically, the difference is associated with the fact that particles incident on the heterointerface from the lighter-mass side are not reflected back, as discussed above in connection with (12) and (13).

V. BOUNDARY CONDITIONS AT A HETEROINTERFACE CARRYING CURRENT

Boundary conditions at an abrupt heterointerface are often required in theoretical descriptions of such problems as minority carrier injection in heterojunction bipolar transistors (HBT) [8]–[12]. If the quasi-Fermi level of minority carriers were continuous at the interface, the ratio of electron concentration on the emitter side ($n^{(-)}$) to that on the base side ($n^{(+)}$) in the vicinity of the discontinuity, would be described by an equilibrium “barometric” formula

$$\frac{n^{(-)}}{n^{(+)}} = \frac{N_C^{(E)}}{N_C^{(B)}} \times e^{-\Delta E_C} \quad (31)$$

where $N_C^{(E)}$ and $N_C^{(B)}$ are the conduction band densities of states in the emitter and base, respectively. However, the quasi-Fermi level is discontinuous at an abrupt interface under current, see, e.g., the discussion in [10], and hence the ratio $n^{(-)}/n^{(+)}$ should be determined from kinetic considerations.

In general, the thermionic model does not prescribe a local rule for boundary conditions at heterointerfaces, because the effective common Richardson constant governing thermionic fluxes at the interface depends on the interface potential relative to the potential maximum. However, in most practical cases, including the base-emitter junction in modern HBT, one of the inequalities (28) or (29) is well satisfied and hence to a good approximation the common Richardson constant corresponds to the material where the potential maximum is located.

In this case the appropriate boundary condition on the flux of electrons is of the form

$$J = -e v_R N_C^{(E)} \left(\frac{n^{(-)}}{N_C^{(E)}} - \frac{n^{(+)}}{N_C^{(B)}} e^{-\Delta E_C} \right) \quad (32)$$

where v_R is the Richardson velocity, $v_R = \sqrt{kT/2\pi m}$, and m is the effective mass of electrons in the emitter. Note that (32) differs from (7) of [10] where the lower-mass rule [2], [4] was assumed to be universally correct.

VI. CONCLUSION

Thermionic current does not remain invariant if the position of the interface between two media of different effective carrier mass is shifted relative to the potential maximum. The lower-mass rule, governing the effective Richardson constant, fails when the maximum of potential energy is located in the heavy-mass semiconductor. Such a situation is realized, e.g., when one introduces interfacial dipoles by epitaxy in order to enhance a band-edge discontinuity [6]. Another common realization of this situation occurs in III-V compound-semiconductor HBT's at the junction between a wide-gap emitter and a narrow-gap base, since the wider gap material usually has the heavier effective mass.

In most practical cases, however, the potential energy maximum E_{max} is separated from the potential energy eV_{int} at the interface by several kT . In this situation, the effective Richardson constant is governed by the m^* of the material containing the potential maximum.

Thermionic emission is very important for the operation of modern III-V HBT's. It is the dominant mechanism of minority carrier injection into the base. However, existing experimental data are not sufficient for a meaningful comparison between predictions of different theories. Such a comparison would require the design of special test structures.

The issue clarified in this work has a practical relevance mostly to theoretical constructs involving model boundary conditions at heterojunction interfaces.

REFERENCES

- [1] S. M. Sze, *High-Speed Semiconductor Devices*. New York: Wiley, 1990, Section 7.2.4.
- [2] R. Stratton, “Theory of field emission from semiconductors,” *Phys. Rev.*, vol. 125, pp. 67–82, 1962.
- [3] C. R. Crowell, “Richardson constant and tunneling effective mass for thermionic and thermionic-field emission in Schottky barrier diodes,” *Solid-State Electron.*, vol. 12, pp. 55–59, 1969.
- [4] A. A. Grinberg, “Thermionic emission in heterosystems with different effective electronic masses,” *Phys. Rev. B*, vol. 33, pp. 7256–7258, 1986.
- [5] C. M. Wu and E. S. Yang, “Carrier transport across heterojunction interfaces,” *Solid-State Electron.*, vol. 22, pp. 241–248, 1979.
- [6] F. Capasso, A. Y. Cho, K. Mohammed, and P. W. Foy, “Doping interface dipoles: Tunable heterojunction band-edge discontinuities by molecular beam epitaxy,” *Appl. Phys. Lett.*, vol. 46, pp. 664–666, 1985.
- [7] E. H. Rhoderick, “Comments on the conduction mechanism in Schottky diodes,” *J. Phys. D*, vol. 5, pp. 1920–1929, 1972.
- [8] K. Horio and H. Yanai, “Numerical modeling of heterojunctions including the thermionic emission mechanism at the heterojunction interface,” *IEEE Trans. Electron Devices*, vol. 37, pp. 1093–1098, Apr. 1990.
- [9] D. Schroeder, “The inflow moments method for the description of electron transport at material interfaces,” *J. Appl. Phys.*, vol. 72, pp. 964–970, 1992.
- [10] A. A. Grinberg and S. Luryi, “On the thermionic-diffusion theory of minority transport in heterostructure bipolar transistors,” *IEEE Trans. Electron Devices*, vol. 40, pp. 859–866, May 1993.
- [11] H. Hjelmgren and T.-W. Tang, “Thermionic emission in a hydrodynamic model for heterojunction structures,” *Solid-State Electron.*, vol. 37, pp. 1649–1657, 1994.
- [12] J. Schneider, E. Koenig, U. Erben, U. Seiler, and H. Schumacher, “Thermionic-diffusion model for abrupt HBT's including self-heating inside the multilayer nonplanar device structure,” *Solid-State Electron.*, vol. 39, pp. 377–384, 1996.



Anatoly A. Grinberg received the Ph.D. degree in theoretical physics from the Lebedev Institute of Physics, Moscow, Russia, in 1961.

From 1958 to 1980, he worked at the Ioffe Institute of Physics and Technology, Saint Petersburg, Russia, as a Senior Staff Scientist. In 1980, he came to the United States and joined the Department of Physics of New York University. Beginning in 1982, he has been associated with the Department of Electrical Engineering, University of Minnesota, and since 1985, also with Lucent Technologies (formerly AT&T Bell Laboratories), Murray Hill, NJ, where, since 1988, he has been a Member of Technical Staff. His research has included transport phenomena, linear and nonlinear optics of semiconductors, properties of the 2-DEG, and modeling of the high-speed devices.



Serge Luryi (M'81–SM'85–F'90) received the Ph.D. degree in physics from the University of Toronto, Toronto, Ont., Canada, in 1978.

From 1980 to 1994, he was a Member of Technical Staff at Lucent Technologies (formerly AT&T Bell Laboratories), Murray Hill, NJ. During this period, he served as a Group Supervisor in several device research departments, dealing with VLSI, quantum phenomena, and optoelectronics. He has published over 140 papers and filed 28 U.S. patents in the areas of high-speed electronic and photonic devices, material science, and advanced packaging. In 1994, he joined the faculty of the State University of New York at Stony Brook where is currently a Leading Professor and Chairman of the Department of Electrical Engineering.

Dr. Luryi served on the Editorial Board of IEEE TRANSACTIONS ON ELECTRON DEVICES from 1986 to 1990, first as an Associate Editor and then the Editor-in-Chief. In 1989, he was elected Fellow of the IEEE "for contributions in the field of heterojunction devices" and in 1993, Fellow of the American Physical Society "for theory of electron transport in low-dimensional systems and invention of novel electron devices." In 1990, Bell Laboratories recognized him with the Distinguished Member of Technical Staff Award. In 1995, he organized and served as the Director of an advanced research workshop (NATO ARW, Ile de Bendor, France) on the "Future Trends in Microelectronics: Reflections on the Road to Nanotechnology." A second workshop in this series, "FTM: Off the Beaten Path" took place in June 1998; see <http://www.ee.sunysb.edu/~serge/ARW-2/home.html>.