Tunneling through superlattices: the effect of anisotropy and kinematic coupling

S V Halilov, X Y Huang, M Hytha, R Stephenson, A Yiptong, H Takeuchi, N Cody and R J Mears

Mears Technologies, 189 Wells Avenue, Newton, MA 02459, USA

E-mail: sam_halilov@yahoo.com

Received 10 August 2012, in final form 23 October 2012
Published 12 November 2012
Online at stacks.iop.org/JPhysCM/24/495801

Abstract
The tunneling of carriers in stratified superlattice systems is analyzed in terms of the constituent effective mass tensor. The focus is on the effects on the tunneling which are caused by the side regions of an intervening barrier. Depending on the covalency and work function in the constituent layers of a superlattice, it is concluded that the kinematics in the regions on either side determined by the effective carrier mass and its interference with the band offset at heterojunctions leads to either a constructive or a destructive effect on the tunneling current. As an example, Si$_{1-x}$Ge$_x$/Si and Al$_x$Ga$_{1-x}$As/GaAs superlattices are demonstrated to reduce the tunneling current at certain fractional thicknesses and stoichiometries of the constituent slabs without affecting the lateral mobility. The findings show, in general, how manipulation of the carrier’s effective mass tensor through stoichiometric/structural modulation of the heterostructure may be used to control the tunneling current through a given potential barrier, given that the characteristic de Broglie wavelength exceeds all the constituent dimensions, thus offering a method complementary to high-$k$ technologies.

(Some figures may appear in colour only in the online journal)

1. Introduction

Superlattices are a major tool to control the transport properties of semiconductor devices. The growth of stratified structures has achieved a level where the quality of interfaces is high enough to assume overall periodicity along the layers (see e.g. [1, 2] and references therein). The effect of structural anisotropy on the carrier flow is well studied in the quasi-classical Boltzmann limit, even in those cases where inelastic scattering due to phonons and thermally activated carriers from structural defect levels is involved in shaping the total current. However, an accurate description of quantum transport and particularly tunneling is less straightforward and the mechanism of coupling between the various components of the carrier quasi-momentum in the presence of the lattice anisotropy is not well documented. While multiple-scattering methods (see e.g. [3] and references therein) provide, in principle, a suitable and accurate computational tool for evaluation and characterization of quantum transport in mesoscopic systems, in practice a much simpler multi-scale approach including the combination of an envelope-function technique with first-principles band-structure methods has been proven to be sufficient at nanometer scale and above and accurate enough to correctly reproduce transport and optical features (see e.g. [4]. Since the introduction of the envelope function and effective mass theory for crystals with slowly varying inhomogeneities [5, 6], the formalism of the single-band description has experienced multiple attempts to improve the accuracy of the method in structures with rapid inhomogeneities, leading eventually to the exact envelope-function theory (see e.g. [7] and references therein). It was shown that the accurate envelope-function equations are in general non-local, so that in the limit of slowly varying stoichiometry, the related non-local terms vanish and the ordinary local envelope-function approach is well justified. This local approximation has been proved to be fairly accurate in systems such as GaAs/GaAlAs quantum wells [8] or Si/Si$_{1-x}$Ge$_x$ strained layer heterostructures [9–11]. The
non-local part of the potential in the envelope-function equation was demonstrated to be essential only close to the interface [12]; its magnitude is likely related to the electron charge density in the vicinity of the interface [13]. In the Si/Si$_1-$$x$Ge$_x$ system, for example, the non-local \textit{ab initio} pseudopotential assumes the form of the bulk potential already one layer away from the interface [9], confirming that the associated electric dipole in the interface region leads basically to a rigid shift of band lineups. It was also pointed out [12] that the local effective mass equation for multilayer composite systems is justified if the zone center energies do not change significantly as a function of their position on the scale of the band gap. This is an essential point for simulation of tunneling through a multilayer structure which includes variable stoichiometry/chemical composition to justify the transmission/reflection matrix method in its simplest local envelope-function approximation.

The transfer matrix method has been applied to various types of heterostructures with a constant effective carrier mass focusing mostly on the modulation of the work function to control the resonant tunneling [14–19]. The advantage of the transfer matrix approach was demonstrated in modeling of resonant tunneling including spatially varying effective masses [4, 19], which necessarily is related to the continuity of the envelope functions and the current flow at every discontinuity. Again, correlation between the longitudinal and transverse carrier dispersions, i.e. the effect of $k_z$ on the longitudinal transport, has been established within the effective mass approximation [20–23].

Here, we concentrate on the effect of the actual variation of the mass tensor and the band offsets due to the modulated stoichiometry in the regions on either side of an intervening barrier on the tunneling current. It is assumed that the change of the characteristic de Broglie wavelength is small enough compared to the constituent slab dimensions. A setting of a field-effect transistor with a stratified structure can be considered as a framework guide. There is a growing interest in heterostructure field-effect transistors based on modulation-doped Si/SiGe quantum wells, due to the enhanced mobility of the carriers and the possibility of realizing quantum computation. Besides, the tunneling effects on the nanometer scale.

The layout of the paper is as follows. To demonstrate a recursive method and can easily be applied to systems with any number of chemically different constituent slabs. Thus, the purpose of the section is to offer a universal recipe on how to design the side regions of an intervening barrier in order to control the tunneling current through the kinematic coupling effect. Numerical applications including first-principle simulations are given in section 4, where it is demonstrated that the modulated channel structure in Si/Si$_1-$$x$Ge$_x$ and GaAs/(Ga$_{x}$Al$_{1-x}$)$_{1/2}$As layered systems has a diminishing effect upon the amplitude of the tunneling current for a certain aspect ratio between the constituent layers—a valuable observation with regard to power consumption reduction in gate-based devices. Finally, section 5 summarizes the key features and results of the current multi-scale approach in quantization of the tunneling effects on the nanometer scale.

2. The effect of the effective mass tensor on the tunneling in bulk systems

Our consideration of the tunneling is based on a quasi-classical picture of the electron propagation [25, 26], which in its simplest realization involves a potential barrier sandwiched by semi-infinite ordered conducting or semiconducting materials, the latter obviously serving as a source of carriers.

While electronic parameters such as the band dispersion around local minima and band lineups at interfaces can be derived from first-principle calculations, it is more suitable to describe the carrier transport on the scale of the entire multiple-barrier system in terms of the envelope-function technique. The latter is briefly detailed below for the triple-slab situation sketched in figure 1 before the model of the non-uniform anisotropic effective mass tensor is extended to materials with stratified structures to establish an additional kinematic effect upon tunneling. The system’s single-particle eigenfunction $\psi_{\epsilon}$ at an energy $\epsilon$ is approximated as a superposition of the constituent Bloch solutions $\psi_{nk}^{i}$:

$$\psi_{\epsilon}(r) = \sum_{\epsilon_{nk} = \epsilon} \Theta_{i} X_{\epsilon_{nk}^{i}}^{(i)}(k_{z}) \psi_{nk}^{i}(r),$$

where $\epsilon_{nk}$ stands for the $nk$-band eigenvalue of the $i$th slab, and $\Theta_{i} = 1$ if $r$ is within the $i$th slab, and vanishing otherwise.

Figure 1. Schematic cross section of the potential barrier sandwiched between conducting materials. The longitudinal and transverse effective mass tensor components in the $i$th slab are labeled as $m_{\|i}$ and $m_{\perp,i}$, respectively. The thick lines are for the right-running waves of amplitude $A_{il}$ and the thin lines are for the left-running waves of amplitude $A_{ir}$.
It will further be assumed that there is only a single band \( nk \) within each slab. The Fourier transforms of the expansion coefficients \( \chi_{n}^{(j)}(k_{z}) \) in equation (1),

\[
\chi_{n}^{(j)}(z) = \int dk_{z} e^{-ik_{z}z} \chi_{n}^{(j)}(k_{z}),
\]

(2)
can be considered as envelope wavefunctions obeying the following Wannier-type equation (see e.g. [27]):

\[
\left[ e_{n}^{(j)}(i\nabla_{z}) + v(z - \epsilon) \right] \chi_{n}^{(j)}(z) = 0,
\]

(3)
with \( v(z) \equiv 1/A \int dz \nu \{ r_{1}, z \} \) defined as an effective potential \( \nu (r) \) of ions, electrons and external sources averaged over the planes of periodicity. Expansion around the band extremum implies the following well-known procedure:

\[
e_{n}^{(j)}(i\nabla_{z}) = e_{0}^{(j)} + \frac{h^{2}\nabla^{2}_{z}}{2m_{e}},
\]

(4)
where the fact is used that the periodic \( u_{nj}(r) \) part of the bulk eigenvector \( \psi_{nj}(r) = u_{nj}(r) e^{ik_{n}r} \) does not vary rapidly with \( k_{z} \).

The approximations used above lead to the following boundary conditions at every interface [4] (L is left and R is right of the interface):

\[
\chi_{n}^{(L)}(z = 0) = \chi_{n}^{(R)}(z = 0),
\]

(5)
\[
\frac{1}{m_{e}} \nabla_{z} \chi_{n}^{(L)}(z = 0) = \frac{1}{m_{e}} \nabla_{z} \chi_{n}^{(R)}(z = 0),
\]

(6)
which imply continuity of the wave solution \( \psi_{n}(r) \) to equation (1) and its derivative along \( z \) corrected by a factor of the effective mass along the same dimension. No effects of scattering such as those due to phonons, interface roughness, impurities, defects, etc, are taken into account, i.e. the total energy \( \epsilon \) and parallel-to-interface momentum \( k_{||} \) is assumed as being conserved while propagating through the stratified structure. Within the effective mass tensor \( m_{eff} \equiv 3\hbar^{2}u_{nj}/\hbar \partial k_{n} \partial k_{n} \) approximation, a concept useful for relatively small deviations from the extremum of the band energy, the total energy is written as

\[
\epsilon = \frac{\hbar^{2}k_{z}^{2}}{2m_{e}} + \frac{\hbar^{2}k_{||}^{2}}{2m_{e}(k_{||})} + v(z),
\]

(7)
where \( m_{e} \equiv m_{ee} \) and \( m_{||} = k_{||} \cdot \hat{m} \cdot k_{||} \) denote projection of the tensor along \( z \) and along \( k_{||} \) in the \( xy \)-plane.

In terms of the transmission amplitude of the envelope function \( A_{j+}(\epsilon, k_{||}) \) at a given energy \( \epsilon \), the direct tunneling current density (current per unit area) is written [26, 28] as

\[
j_{dt} = \frac{2e}{h} \int \frac{d\epsilon |\nu_{T} - n_{T,\epsilon} + \nu_{Vg}|}{(2\pi)^{2}} T(\epsilon, k_{||}),
\]

(8)
where \( T \equiv |A_{j+}|^{2} \). \( V_{g} \) is the voltage bias across the barrier and the carrier state distribution is given by the Fermi–Dirac function \( n_{T,\epsilon} \). Further derivation of the one-dimensional tunneling probability for the situation depicted in figure 1 is based on the matching conditions (5) and (6) and the WKB approximation for the wavefunction. The WKB method applied to the envelope function is obviously justified since the change of the characteristic de Broglie wavelength within each constituent slab is smaller than the dimensions of the slabs in the superlattice. General solutions are written in terms of right- and left-running waves with amplitudes \( A_{j+} \) and \( A_{j-} \), respectively, within each region as \( \chi_{n}^{(j)}(z) = A_{j+} e^{i \int dz k_{n}(z)} + A_{j-} e^{-i \int dz k_{n}(z)}. \) As a result, the amplitude of the transmitted wave \( A_{j+} \), as a function of the quasi-momenta gains a prefactor which includes the effects of the effective mass tensor modulated both transversely and longitudinally (see e.g. [29]),

\[
|A_{j+}| = \frac{4k_{||}k_{z}m_{32}m_{32}e^{k_{||}t_{2}}}{(k_{z}m_{22} - k_{z}m_{32})(k_{z}m_{12} - k_{z}m_{32})},
\]

(9)
if a limit \( k_{z}t_{2} \gg 1 \) is assumed, \( t_{2} \) standing for the thickness of the barrier. The direct tunneling current (8) across the layered structure can then be expressed as an integral over the regions of \( k_{||} \) which correspond to positive-energy states on both sides of the potential barrier. Coupling between \( k_{||} \) and \( k_{z} \) established through the energy conservation (7),

\[
k_{z}^{2} = \Re c(2m_{32}[e - \nu(z) - \hbar^{2}k_{||}^{2}/2m_{||}])^{1/2},
\]

(10)
translates into a specific dependence of the tunneling probability on the variable effective mass in the entire structure,

\[
T(\epsilon, k_{||}) = \frac{\chi_{3}^{(5)}\chi_{3}^{(3)}}{\chi_{1}^{(5)}\chi_{1}^{(3)}} \frac{m_{34}k_{3}}{m_{32}k_{3}} |A_{j+}|^{2}.
\]

(11)
The integral in equation (8) is proportional to the overlap of the projections \( S_{\parallel}(k_{||}, k_{||}) \) of a constant-energy surface on the \( k_{||} \) plane in the Brillouin zone on each side of the barrier [28]. The effect of the effective mass anisotropy on the direct tunneling current is easily revealed by integration over \( k_{||} \) in the limit of a high barrier width \( k_{z}t_{2} \gg 1 \). To keep the notation short, the longitudinal and transverse components of the tensor will be relabeled as \( m_{32} \to m_{1}, m_{31} \to \mu_{1}, \) respectively. A more tractable picture can also be obtained by omitting the effects of the space charge, that is, no effect of the Airy-type wave solutions in the channel or the image forces at the interface will be considered. Without loss of generality, it is assumed that the applied gate bias voltage \( V_{g} \) is low enough to enable linearization of the tunneling, and the barrier potential \( v_{c}(z) \) is approximated by a rectangle of height \( v_{g} \). Further assumptions are justified in the low-field limit: the \( z \)-dependence of \( k_{z} \) is omitted, the quasi-momenta of equation (10) are replaced with \( k_{z} = \sqrt{2m_{1}(\epsilon - \hbar^{2}k_{||}^{2}/2\mu_{1}) - v_{c}} \), where the band alignment parameters \( v_{i} \) are such that \( v_{i} = 0 \) for \( i = 1, 3 \) and \( v_{2} \) determines the barrier height \( V_{g} \). Under the additional assumption of a cylindrical symmetry around the normal to the interfaces, i.e. small in-plane anisotropy, the current density reads

\[
j_{d} = \frac{2e}{h} \nu_{T}^{2} v_{g}^{3/2} \frac{\mu_{1} m_{3}^{3/2}}{m_{32} m_{12}^{3/2}} \times \exp(-2\nu_{T}^{2} t_{2}^{2} m_{23}^{3/2} v_{B}^{1/2}) \int d\epsilon \frac{\partial n_{T,\epsilon}(\epsilon)}{\partial \epsilon},
\]

(12)
The conduction band is shown in figure 2, where the electronic parameters of a component of the carrier effective mass tensor in the material adjacent to the insulator, obtained as a result of parallel momentum integration at room temperature in equation (8). The current is nearly linear with respect to \(\mu\) and changes as \(1/(m_{\parallel}^*)\) within a certain range, confirming the approximate expression (12). Note \(J_{g\ell}(m)\) is multiplied by a factor of 10.

Therefore, the effect of anisotropy of the mass tensor can be understood as follows: the tunneling current scales linearly with the transverse mass \(m_{\perp}\) in the layer with itinerant carriers and is inversely proportional to the square root of the longitudinal mass \(m_{\parallel}\).

In the case of a large magnitude of the voltage drop \(V_g \leq V_B\), the direct tunneling density can be written in a form similar to the Fowler–Nordheim expression, in which case a triangular shape of the potential drop \(V_g(z) = (z - z_i)V_B/\ell_2\) and crossing of the band edge energy with the slope of the potential drop within the barrier are taken into account. The high-field tunneling is governed by the variable tunneling path length \(z_i = V_B/\ell_2/V_g\) and acquires an effective mass prefactor similar to the one in equation (12).

To verify the mass tensor dependence of the tunneling current contained in the formula (12), equation (8) was integrated over \(k_B\) numerically. The results of the integration are shown in figure 2, where the electronic parameters of a GaAs/Al_{0.5}Ga_{0.5}As/metal triple-slab system have been used: \(m_2 = 0.4m_e\) for the longitudinal effective mass in the barrier region, \(m_1 = 1.0m_e\) on the metallic side, the cut-off energy in the conduction band is \(\epsilon = 0.02\) Ha, the band offset between the conduction bands in the channel and the gate insulator is \(V_B = 0.12\) Ha and insulator thickness is \(\ell_2 = 20\) Å. A strong linear dependence of the tunneling upon the transverse mass \(m_{\perp}\) in figure 2 is justified in a relatively narrow range of values, \(0 < \mu_1 < m_{\text{GaAs}}^* = 0.067m_e\), above which the current becomes saturated, whereas it changes as the inverse of the square root of the longitudinal mass, \(\sqrt{m_{\parallel}}\), in the entire range of values. The reference value for the tunneling current \(J_{BI}\) in GaAs has to be taken at \(\mu = m = m_{\text{GaAs}}^* = 0.067m_e\).

In general, the current is a total over all types of carriers with different effective masses in the conduction or valence band, which is the case in systems with multi-valley electronic structure. Therefore, the qualitative picture given so far for the tunneling has only limited value which is meaningful for amplitudes of the applied voltage \(V_g\) below the band offsets between various valleys in order to the mitigate statistical mixing of different carriers. It is a well-known fact that the gate tunneling in field-effect transistors scales exponentially with the insulator thickness \(\ell_2\), voltage drop \(V_g\) across the insulator, effective mass of the carrier \(m_2\) within the insulator and the area of the interface between the gate insulator and the channel, the latter obviously related to the channel length. On the other hand, the tunneling is very sensitive to the effective mass tensor \(m_{\alpha\beta}^*\) of the carrier along the entire tunneling path including, e.g., channel and gate materials, which is an observation less appreciated by a discussion of gate leakage in pertinent systems.

3. The lateral differential mass effect in stratified systems

In the general case of a stratified structure the exponential prefactors in the expressions for the tunneling current have to be taken into consideration to account for the effects of the anisotropic dispersion of the carrier. In terms of the mass components \(m_1, m_2\) parallel and perpendicular to the constituent slabs, respectively, a lower ratio \(m_1/m_2\) implies higher lateral mobility and a lower longitudinal tunneling current, thus leading to an enhanced performance of gate-controlled semiconductor devices. There is an additional kinematic effect of stratification on the tunneling probability which is considered below.

So far, the effective mass has been considered as a parameter predetermined by a given stoichiometry such as GaAs/GaAlAs heterostructure and no possibility for an actual variation of the mass in the channel has ever been assumed. Here, we focus on a situation where the effective mass tensor and band offset of the constituent channel slabs are considered on the same footing. An appropriate potential profile is illustrated in figure 3. The carrier supply region is modeled by a heterostructure with periodicity \(l = \ell_1 + \ell_2\) truncated at the potential barrier \(i = 0\), and its spatially modulated effective mass is approximated by bulk values \(m_{1,i,}\) \(m_{1,i,}\) for incident and reflected waves, respectively, is established again by using the recursive transmission/reflection matrix approach. This is an approximation obviously justified for stratified systems with the constituent slabs much thicker than the minimum dimension for which the concept of constituent slab effective mass is still meaningful, i.e. \(l_i \gg a_i\), where \(a_i\) stands for the lattice parameter of the \(i\)th slab. Furthermore, it is assumed that the majority of the carriers are in the vicinity of the equilibrium in the Brillouin zone. In order to capture the effect of the alternating effective mass in a feasible way, the entire heterostructure with \(i = 1, 2\) components is replaced with an effective bulk material, so that its effective mass can then be used instead of \(m_{1,1}\) in equation (9) obtained earlier for the transmission amplitude. Apparently, certain parallels can be made with band-structure

![Figure 2](image-url)
calculations on a one-dimensional crystal performed within the Kronig–Penney model [30], although the current model of the potential profile is essentially three dimensional and the carrier band dispersion is assumed to have an extremum in every heterostructure constituent domain.

From the matching conditions for the envelope functions and group velocity at each interface, and using the recursive effective mass method validated through the Bloch periodicity, \( \chi(z + l) = e^{i q \cdot l} \chi(z) \), where \( q \equiv \{ k_l, q_z \} \) stands for the wavevector of the superlattice. The dispersion relation of the heterostructure is determined through the following equation:

\[
\cos(q \cdot l) = F(E),
\]

\[
F(E) \equiv \cos(k_1 l_1) \cos(k_2 l_2) - \frac{1}{2} \left( \frac{k_{1m_2}}{k_{2m_1}} + \frac{k_{2m_1}}{k_{1m_2}} \right) \sin(k_1 l_1) \sin(k_2 l_2),
\]

(13b)

where \( k_i = \left[ 2m_i (E - V_i) \right]^{1/2} \). Thus, effective potential barriers \( V_i = k_i^2 / 2m_i + v_i \) seen by the carriers due to the differential transverse mass \( \Delta \mu \equiv \mu_2 - \mu_1 \) and the band offset between constituent slabs \( \Delta \nu = v_1 - v_2 \) reduce the three-dimensional propagation problem to a single-dimensional task, as long as \( k_1 \) is conserved during multiple scattering. The dispersion relation for the composite multilayer material is easily retrieved by solving equation (13a) analytically under some simplifying assumptions. Apparently, \( \delta \equiv V_1 - V_2 \equiv \frac{k_1^2}{2 \mu_1 \mu_2} \Delta \mu + \Delta \nu \)

(14)

serves as a primary parameter of the model. The longitudinal superlattice dispersion relation \( E_z(q_z) = E_0 + q_z^2 / 2M_z \) is readily established as a result of Taylor expansion in terms of \( \delta \), if one assumes the kinetic regime \( k_1 l_1, k_2 l_2 \ll 1 \), as well as small variations of the transverse effective mass and band discontinuities along the tunneling path

\[
|\delta| \ll \frac{1}{l_1 l_2 \sqrt{\mu_1 \mu_2}}.
\]

(16)

At this point, we introduce a ‘heteromass’ \( M_z \) as the effective longitudinal mass of the carriers in the entire superlattice to distinguish it from the actual effective masses of the carriers in every constituent slab. It is easily conjectured that

\[
M_z \equiv \frac{1}{l} \frac{\partial^2 E}{\partial q_z^2} \approx \frac{1}{l} \left( m_1 l_1 + m_2 l_2 \right) + \frac{|\delta| m_1 m_2 l_1 l_2}{3l^2} \left[ 3l_1 l_2 + l_2^2 + \frac{m_1}{m_2} l_2^2 \right] + \frac{m_1}{m_2} l_1^2 \right]
\]

(17)

becomes effectively heavier due to the variation of the effective transverse masses contained in \( \delta \) as long as \( k_\parallel \) is non-vanishing, which is always the case when the system is away from its equilibrium distribution. Note that this ‘fictitious’ heteromass can be orders of magnitude larger than the bare electron mass. It has to be emphasized that the well-documented resonance effect [8, 14–18] of the alternating potential profile on the tunneling probability associated currently with the band offset \( \Delta \nu \) should not be confused with the 3D effect of the coupling between the \( k_z \) and \( k_\parallel \) momenta as reflected in the lateral anisotropy parameter \( \delta \). This is the effect of the modulated lateral effective mass causing an extra effect on the 3D tunneling, whereas modulation of the potential due to the variable stoichiometry in a heterostructure appears in our simplified picture as mere a \( k_\parallel \)-independent shift in the heteromass \( M_z \). As follows from the structure of \( \delta \) in equation (14), both the effective mass anisotropy and the potential modulation effects on \( M_z \) can either be additive or destructive with the magnitude of the lateral momentum \( k_\parallel \) determining the overall strength of the interference. In general, the effects add to each other resulting in enhanced \( M_z \) if the electrons experience lighter (heavier) lateral mass \( \mu \) at higher (lower) work function \( \nu \) by traveling from slab 1 to slab 2, i.e., when \( \Delta \mu \Delta \nu > 0 \). Alternatively, in the case of \( \Delta \mu \Delta \nu < 0 \) the two effects interfere destructively leading to a complete annihilation at lateral momenta

\[
k_\parallel \sim \sqrt{2|\Delta \nu||\mu_1 \mu_2| \Delta \mu},
\]

(18)

given that the accelerating field is small enough to ensure that there is primarily a single-type valley population. An accurate determination of the quasi-particle population in the presence of an accelerating field is usually a matter of Monte Carlo simulations [31], which include various scattering mechanisms such as electron–phonon scattering, impact ionization, interface roughness, Coulomb scattering, etc. Usually, for fields up to \( \sim 10^7 \, \text{V m}^{-1} \), the energy gain of the carriers does not exceed a few tenths of an eV, which translates as \( k_\parallel < 0.1 \) in atomic units for the cut-off momentum. This observation also justifies our kinetic regime choice of a single-valley population picture adopted here for tunneling simulations.

Not only the presence of a heterojunction but also the fractional thickness of the constituent slabs has an influence...
on the timing of the carrier traversing the system. This pure wave effect is already contained in equation (11) for the tunneling probability as it scales as the inverse of $M_z$ ($m_{1z}$) and therefore is impacted by the dimensional parameters contained in equation (17). The appropriate condition can be derived from equation (17) at a given $k_l$, i.e. $\delta$, by varying $M_z$ with respect to the slab dimensions $l_i$. Analysis of the derivative confirms an effect of entanglement between the longitudinal and transverse flows of the carriers, for which $|\delta|(m_1 - m_2)^2$ serves as a coupling parameter. Obviously, the latter gets stronger for carriers with higher lateral momentum $k_l$; provided that the constituent mass tensors are anisotropic. Note that the effect is still there even if the constituent longitudinal masses are identical as long as the transverse mobility is varying across the heterostructure. As the mobility along a certain direction, on the other hand, is scaled as a scattering time averaged over all dimensions divided by the effective mass in the same direction, the lateral mass modulation translates into a transient time as the inverse of the group velocity for longitudinal transport. The expression for the $M_z$ considered as a function of the fractional thickness $l_1/l$ at a fixed superlattice period $l = l_1 + l_2$ a.u. has an obvious maximum at every non-vanishing $k_l$. However, the formula (17) is less accurate at higher values of $k_l$ when the carriers have higher energy population. In general, there is an effect of the superlattice period $l$ on the heteromass. This is due to the fact that the widths of the quantum wells and potential barriers scale with the structure period, which results in wider gaps in the heterostructure band structure and particularly in reduced dispersion of the bands. Thus, the scattering on the broader barriers will cause the longitudinal heteromass to grow, adding to the reduced overall tunneling current. This is confirmed by direct integration over $k_1$ in equation (8) for the direct tunneling current density $j_{th}$ through a multi-barrier set, after $m_{1z}$ has been replaced by $M_z(k_1)$ according to equation (17). The resonance of the band offset and the lateral modulation of the effective mass $\Delta \mu \neq 0$ in the hetero-channel upon the heteromass lead to an anti-resonance effect on the tunneling current evaluated at a fixed energy of the carriers $\varepsilon = 0.4$ eV. In order to capture this and related effects which are contributing to the magnitudes of the potential barriers and quantum wells in the heterostructure more accurately, one has to go beyond the approximation of small potential perturbations used above in the derivation of equation (17). Extension to situations with higher values of the structure period $l$, larger $k_l$ and therefore higher electric field in the channel can be made by treating the inverted band dispersion relation given by equation (13a) numerically. Writing the inverted dispersion $q_\parallel(E)$ as

$$q_\parallel = \frac{1}{l} \cos^{-1} F(E),$$

(19)

where $F(E)$ is given by the equation (13b), one easily verifies the following formula for the longitudinal heteromass $M_z$:

$$M_z = -\frac{1}{E^2} \left[ \frac{\partial F}{\partial E} \right]^3 \left( F \left[ \frac{\partial F}{\partial E} \right]^2 + (1 - F^2) \left[ \frac{\partial^2 F}{\partial E^2} \right] \right)^{-1}. $$

(20)

### Table 1. Components of the effective mass tensor in bulk Si, diamond structure, and in Si$_{1-x}$Ge$_x$ alloy, $x = 0.75$, in units of the free electron mass $m$, conduction states only. Directions are shown in the standard frame of silicon.

<table>
<thead>
<tr>
<th>System, val.</th>
<th>Eff. mass $m_{\alpha\beta}$, u. of $m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si$_{1-x}$Ge$_x$(L)</td>
<td>$m^* = 1.59$ [001], $\mu^* = 0.08$ [110]</td>
</tr>
<tr>
<td>Si (Δ)</td>
<td>$m^* = 0.98$ [001], $\mu^* = 0.19$ [110]</td>
</tr>
</tbody>
</table>

### Table 2. The same as table 1 for bulk GaAs, zinc-blend structure, and Al$_{0.5}$Ga$_{0.5}$As ternary alloy.

<table>
<thead>
<tr>
<th>System, val.</th>
<th>Eff. mass $m_{\alpha\beta}$, u. of $m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlGaAs (X)</td>
<td>$m^* = 1.3$ [001], $\mu^* = 0.23$ [110]</td>
</tr>
<tr>
<td>AlGaAs (L)</td>
<td>$m^* = 1.9$ [001], $\mu^* = 0.0754$ [110]</td>
</tr>
<tr>
<td>AlGaAs (T)</td>
<td>$m^* = 1.067$</td>
</tr>
<tr>
<td>GaAs (Γ)</td>
<td>$m^* = 0.067$</td>
</tr>
<tr>
<td>GaAs (L)</td>
<td>$m^* = 1.9$ [001], $\mu^* = 0.0754$ [110]</td>
</tr>
<tr>
<td>GaAs (X)</td>
<td>$m^* = 1.3$ [001], $\mu^* = 0.23$ [110]</td>
</tr>
</tbody>
</table>

It is assumed that the derivatives in equation (17) are taken at the minimum of $E(q_\parallel)$, i.e. where $q_\parallel(E) = 0$. Obviously, a condition $|F| < 1$ has to be observed in order to ensure the real solutions $E(q_\parallel)$. Numerically, the function $F(E)$ has been extended analytically in the complex plane in order to enable imaginary $k_1$ and $k_2$, i.e. in order to capture the situations when the carrier energy $E$ is below the barrier maximum or maximum $k_l^2/2\mu_1$ caused by the transverse degree of freedom.

### 4. Applications: Si$_{1-x}$Ge$_x$/Si and (Ga$_x$Al$_{1-x}$)As/GaAs heterostructures

Two systems, Si$_{1-x}$Ge$_x$/Si and (Ga$_x$Al$_{1-x}$)As/GaAs, have been chosen for illustrative purposes for the effects of the carrier mass anisotropy conjectured with heterojunctions upon tunneling transport. This particular choice is justified by the high promise which these systems have for semiconductor technology, primarily because of the enhanced mobility as compared to that of silicon. Yet, by miniaturization of the channel length, the route inevitably leads to problems of enhanced leakage, which is mostly contributed by the direct tunneling currents. As was shown above, evaluation of the effect of the structural anisotropy on the tunneling can be achieved by using the constituent mass tensor components and potential parameters such as the barrier height and band offsets only. Besides, only a minimum of information about the carrier thermal quantum statistics is needed as long as the consideration is limited by the kinetic approximation (small deviation from equilibrium). The case of positive interference ($\Delta \mu \Delta \nu > 0$) between the effect of the band offset and the effective mass effect is exemplified by Ge/Si, with the mass parameters given in table 1, whereas AlGaAs/GaAs, with the corresponding data shown in table 2, serves as a system where the two effects have negative interference ($\Delta \mu \Delta \nu < 0$). However, the actual type of interference depends in general on many aspects such as the stoichiometry and related effects of strain, as well as the geometrical parameters of the heterostructure. The strain effects are particularly strong in the
case of the Si$_{1-x}$Ge$_x$/Si system, due to a lattice mismatch, which amounts to a considerable 4% in the case of $x = 1$. As is well known [9–11], there is a certain change of the effective mass and the band lineup parameters correlated with the stoichiometry of Si$_{1-x}$Ge$_x$. The associated strain effect has been described earlier [9] in terms of the equilibrium elastic constants for bulk Ge ($i = 1$) and Si ($i = 2$) and the fractional thicknesses $l_i$. Further consideration is limited to the effects of strain on the conduction state parameters only, in which case there are six-fold degenerated $\Delta$-symmetry valleys in bulk Si with the lowest energy located close to the X-point, and a $\Lambda$-symmetry minimum at the L-point in bulk Ge. However, the $\Delta$-states in unstrained Ge are only a few tenths of eV away from the L-minimum, thus increasing the probability of the carrier population nesting with different symmetries at elevated temperatures and high fields. This statistical effect of mixed symmetries is out of the scope of the present work. According to the deformation potential approach [32], the $\Delta$-valleys under uniaxial strain $\epsilon_{xy}$ become inequivalent: the bands along [100] and [101] split off from the one along [001]. This results in four-fold and two-fold degenerated states with overall splitting between them which scales as the relative change of strain $\epsilon_{xy} - \epsilon_{xz}$. For example, the magnitude of the average valley shift amounts to about 0.1 eV in Ge-matched strained silicon, with a maximum strain of $\epsilon \sim 0.04$. The L-minima in Ge are not affected by [001] strain. It should be pointed out [9] that, to a linear approximation, one should expect that the band discontinuity in the Si$_{1-x}$Ge$_x$/Si system will have a small $\sim 10^{-2}$ eV negative shift (type I lineup) up to $x < 0.8$ and a small positive shift (type II lineup) at $x > 0.8$ if the Si(001)-matched condition is used, which corresponds to the low fractional thickness $l_1/l \sim 0$. The conduction carriers in the case of Si(001)-matched $a_{|i|$ preserve their $\Delta$-character, i.e. no change of the constituent effective mass throughout the heterostructure has to be expected. The case of Ge(001)-matched layers which corresponds to the high fractional thickness $l_1/l \sim 1$ is different: there is a positive shift of up to a few $10^{-1}$ eV as the concentration of Ge goes up. Besides, at $x > 0.8$ there is a change of the carrier symmetry from $\Delta$ to L, which implies a considerable change of the effective mass tensor, with respect to both its longitudinal and lateral components. The particular stoichiometry of Ge/Si has been used here for demonstrative purposes, since in this case the L-states are relatively well separated from the $\Delta$ states and the entire consideration can be carried out statistically in terms of a simple single-band model. To this end, the band discontinuity scales linearly with the fractional thickness $\lambda \equiv l_1/l$, i.e. $\Delta\nu(\lambda) = \Delta\nu_1\lambda + \Delta\nu_2(1 + \lambda)$, with $\Delta\nu_1 \approx 0.01$ Ha in the extreme case of $\lambda \rightarrow 1$, when the Ge slab is relaxed and the Si is strained, whereas in the opposite case of $\lambda \rightarrow 0$, i.e. when the Ge is strained and the Si is relaxed, the gap between the lowest $\Delta$-state in Si and the L-state in Ge is about $\Delta\nu_2 \approx 0.02$ Ha. Generalization to a multiband situation is rather straightforward.

To stay within a reasonable range of relevance, it was assumed that the band discontinuity in the Si/Ge heterostructure is 0.01 Ha (type II lineup), which is weighted more toward structures with smaller fractional thicknesses $l_1/l$ (thicker Si slabs) and obviously is less accurate for larger $l_1/l \sim 1$ (more like a Ge-matched structure). This choice is also due to the fact that experimentally it is more challenging to grow Ge-rich Si$_{1-x}$Ge$_x$ slabs thicker than just a few layers, since in that case the system usually develops thread dislocations because of the high tensile stress. The results of numerical simulations for the effective heteromass $M_1$, as a function of the fractional thickness and lateral momentum $k_1$ in a prototype Si$_{1-x}$Ge$_x$/Si stratified system are shown in figure 4. The effects of the constituent mass anisotropy and the band discontinuities are obviously adding to each other, illustrating the case of $\Delta\mu\Delta\nu > 0$.

The current density given by expressions (8) and (11) was evaluated using results for $M_1(l_1/l, k_1)$ numerically obtained from equations (13)–(20) and illustrated in figure 5. Cases with intermediate values for the band lineup $\Delta\nu_{bo}$ (not shown) are comprised between the two curves, one with $\Delta\nu_{bo} = 0$ and the other with $\Delta\nu_{bo} = 0.15$ eV, which show a broad minimum in the tunneling current.

Another extreme situation, this time of the destructive 1D and 3D effects on the tunneling current, is realized in the AlGaAs/GaAs quantum well heterostructure, which is the subject for the remaining part of this section. According to $ab$ initio theoretical studies, as well as cyclotron and optical measurements, see [1] and references therein, there are three major valleys in the electronic structure of GaAs, which are located at the $\Gamma$, L and X points, in the order of their alignment on the energy scale, the first is the lowest. This order is known to be opposite to the other end compound, AlAs, and scales on the phase diagram of Al$_x$Ga$_{1-x}$As almost linearly [1]. This gives a good opportunity to manipulate the effective mass components by just changing the stoichiometry in the ternary alloy, whereas our particular choice was made for Al$_{0.5}$Ga$_{0.5}$As, since in that case the lowest occupied valley is located at the X point. Ab initio calculations for Al$_{0.5}$Ga$_{0.5}$As and GaAs as the operating gate shows that for $M_1(l_1/l, k_1)$ varying between 0 and 0.097 a.u., with the band offset between AlGaAs and GaAs set to zero in order to emphasize the effect of the mass anisotropy. The results of similar calculations but with a band offset of $\Delta\nu = 0.2$ eV are shown in the middle panel of figure 6. Similarly to the case of the SiGe/Si system, the tunneling current density $J_{th}$ in a field-effect transistor set using the AlGaAs/GaAs heterostructure as a channel, Al$_2$O$_3$ as the insulating layer and Au/Cr as the operating gate shows a considerable reduction due to the resonance mass effect, as shown in figure 7.

There are different trends with respect to the heterostructure dimensional parameters related to the two mechanisms
Figure 4. The effective longitudinal mass $M_z$ in the prototype Si$_{1-x}$Ge$_x$/Si heterostructure as a function of the Si$_{1-x}$Ge$_x$ fractional thickness $l_1/l$, at various $k_\parallel$. Left panel: zero band offset $\Delta\nu=0$; right panel: non-zero band offset $\Delta\nu=0.15$ eV. The sizes of the characters and the line thicknesses correlate with the magnitude of $k_\parallel$. The two effects add to each other at all $k_\parallel$. The dashed line goes through the maxima of $M_z$ and serves as a guide to the eye. The graph at the bottom illustrates Max($M_z$) due to mass anisotropy (3D effect) and Max($M_z$) due to band offset (1D effect) as functions of the fractional thickness $l_1/l$ at various total heterostructure periods $l$: the thin line stands for the 1D effect of the band offset and the bold line describes the 3D effect of the effective mass anisotropy. The constituent effective mass tensor components used in equation (17) are taken from table 1.

Figure 5. The tunneling current density $j_a$ in the metal/insulator/Si$_{1-x}$Ge$_x$/Si heterostructure system as a function of the fractional thickness $l_1/l$ of the constituent layers. There is a clear drop of current around $l_1/l$ where $M_z(k_\parallel)$ has a maximum, cf. figure 4. The electronic parameters of SiO$_2$ are used in the insulating region. The energy of the carriers is fixed at $E = 0.4$ eV, a reasonable approximation justified by the energy distribution at lower applied fields.

of the tunneling channel. Anisotropy of the band dispersion demonstrates a strong impact of the structural dimensions on the heteromass $M_z$: for systems with smaller period $l$, tunneling carriers with high enough lateral momentum $k_\parallel$ become heavier in a heterostructure with a larger fraction of GaAs, i.e. when the carriers have an opportunity to spend more time in the space domain with less localization. Also, the maximum enhancement of $M_z$ is achieved if the actual thickness $l_1$ of the ternary compound-based slab does not change for various $l$. This is in contrast to the effect of the variable work function which is essentially one dimensional: $M_z$ at various global periods $l$ is mostly enhanced at a fixed ratio between the band-offset-forming constituent slabs, which is at around $l_1/l_2 \sim 3$ for AlGaAs/GaAs. These trends are depicted at the bottom of figure 6, where Max($M_z$) is plotted as a function of the fractional thickness $l_1/l$ at various total heterostructure periods $l$: the thin line stands for the 1D effect of the band offset and the bold line describes the 3D effect of the effective mass anisotropy.

Among the other distinct features easily seen in figure 6 is the opposite trends of the two mechanisms with respect to the period of the heterostructure: at smaller $l$, the 3D effect of the anisotropic localization tensor is dominant, whereas the multi-barrier effect prevails at larger $l$. Therefore, there is a whole list of parameters accessible by using a heterostructure such as the anisotropy of the bonding covalency, distinctive work functions, structure dimensions and the observation that the two effects can in general add or subtract, which can be used in practice to design a system with predetermined tunneling features. As such, reduction of the gate leakage in field-effect transport devices seems to be one of the potential areas for application.

5. Conclusions

The tunneling of carriers in stratified superlattice systems is analyzed in terms of the constituent effective mass tensor. WKB analysis shows that the tunneling current scales...
Figure 6. Similar to figure 4, but for the prototype AlGaAs/GaAs heterostructure. The effects due to the band offset and mass anisotropy have anti-resonance character and interfere destructively at around $k_1 \sim 0.033$ au. Note the two ranges of clustering for $M_z$ maxima: the one due to the mass anisotropy occurs at $l_1 \sim 0.45l$, whereas the one due to the band offset is mostly around $l_1 = 0.85l$. The graph at the bottom illustrates $\text{Max}(M_z)$ as a function of the fractional thickness $l_1/l$ at various total heterostructure periods $l$: the thin line stands for the 1D effect of the band offset and the bold line describes the 3D effect of the effective mass anisotropy. The constituent effective mass tensor components used in equation (17) are taken from table 2.

Figure 7. The same as figure 5, but for the metal/insulator/AlGaAs/GaAs prototype multilayer structure. The insulating region is represented by Al$_2$O$_3$.

 inversely with the square root of the mass along the normal to the interfaces and is proportional to the mass parallel to the interface. This observation alone hints toward an alternative method to control the tunneling current through a given barrier by tweaking the kinematic properties of the carriers in its side regions. Modulation of the side regions through adding extra heterojunctions is suggested as one such recipe. It is shown in detail how the tunneling probability in stratified systems is controlled not only by the differential effective mass effect in the direction of the tunneling current, but also by a kinematic coupling between the longitudinal and transverse momenta of the carriers scaled with the differential effective mass effect transverse to the current. The effect is shown to be present even if the work functions of the constituent slabs are identical, i.e. even if there is no potential barrier between the slabs. Analysis is assisted by the derivation of a formula for an effective superlattice representing the side regions of an intervening insulator, which demonstrates clear minima of the tunneling current at certain configurations of the superlattice constituent dimensions. The findings show how the relatively simple concept of the anisotropic effective mass tensor can be used in designing tunneling devices with desired features, particularly with reduced leakage current. The method is applied to a binary heterostructure by evolving an effective equation for the effective mass and band structure for the entire system around the dispersion minimum; this is in fact a recursive method and can easily be applied to systems with any number of chemically different constituent slabs. The purpose of the method is to offer a universal recipe for the design of the side regions of an intervening barrier in order to control the tunneling current through the kinematic coupling effect. As an example, Si$_{1-x}$Ge$_x$/Si strained layer heterostructures and Al$_x$Ga$_{1-x}$As/GaAs quantum wells are demonstrated to reduce the tunneling current at certain fractional thicknesses and stoichiometries of the constituent slabs without affecting the lateral mobility. Interference between the well-known 1D effect of the band offset between stoichiometrically different layers in a multilayer structure and the 3D effect of the variable effective mass anisotropy is shown to lead to either destructive or constructive effect on the current. This is due to the different origins of these two mechanisms: the 1D effect is related to the work function modulation and is determined in general by the surface potential of the constituent materials,
whereas the 3D effect is based on the modulation of the covalent bonding, which manifests itself in the kinematics through the anisotropic mass tensor. This makes the suggested methods of tunneling control more volatile in practice. Based on these observations, it is concluded that the anisotropy of the carrier mass serves as an extra tool for control of the tunneling current, which is particularly important in overall reduction of the power consumption in pertinent technologies. The method is also considered as complementary to the existing high-\(k\) methods in gate-leakage reduction technologies.

Acknowledgments

One of the authors (SH) is grateful to A Bratkovsky for helpful discussions on tunneling and related problems.

References