A New Computational Model for Heterojunction Resonant Tunneling Diode

J. Gene Cao
IMA
University of Minnesota
Minneapolis, Minnesota 55455

Patrick Roblin
Department of Electrical Engineering
The Ohio State University
Columbus, Ohio 43210

Abstract

Following the full band structure modeling approach of Roblin & Muller [15], in this paper, we develop a new computational model for heterojunction Resonant Tunneling Diode (RTD), which accounts for full band structure. It is based on the Generalized Wannier Function (GWF) representation [6, 8, 9]. This model uses spatially varying band structure and a matrix solver that is very stable and efficient. It can handle arbitrary number of bands (arbitrary number of off-diagonals). The model gives a systematic procedure to compute transmitted/reflected current of a given RTD devices. As a demonstration we present several calculations for GaAs-$Ga_{.7}Al_{.3}As$-GaAs-$Ga_{.7}Al_{.3}As$-GaAs RTD using this model. In these examples we use the band structures of GaAs and AlAs calculated using a pseudo-potential empirical method and use their arithmetic average for matrix elements at the interface.

1 Introduction

Pioneered by Esaki, Tsu and Chang [3, 5, 19] in early 70's, resonant tunneling devices have generated a lot of interests. This paper is concerned with Resonant Tunneling Diode (RTD) that has potential applications in high-speed devices: oscillators, amplifiers, harmonic multipliers and logic elements [2, 12, 13]. In particular, we consider a RTD that has a structure such as GaAs-AlAs-GaAs-AlAs-GaAs. The two AlAs layers serve as potential barriers and the GaAs piece between them as potential well in the GaAs bulk. (We may use other materials than GaAs and AlAs for bulk and barrier respectively.) When a stream of electrons are incident from the left with an energy E, they may be either reflected or transmitted due to the double barriers and the potential well. This paper aims to establish a mathematical model that can be used to compute the probability of an electron being transmitted/reflected.

RTD is a quantum device. Many quantum mechanical phenomena cannot be explained by classical mechanics. One interesting example is tunneling, that is, an electron can transmit through a barrier under certain condition. What is more fascinating is *resonant* tunneling: at some energy level (*resonant state*) an electron can go clean through a *double* barrier but has a very low probability to getting through a *single* barrier; and at neighboring lower and higher energy levels electrons are partly or completely blocked by the double barrier. Associated with resonant tunneling is the *negative dynamic resistance* phenomenon that the current decreases as the voltage increases. Our goal in this paper is to give a systematic procedure to quantitatively identify these *generic* energy levels for a given RTD device with structure described above, which can be used by an application engineer.

There has been a lot work on modeling RTD (see [11, 18] for an overview). All the existing multiple sequential scattering (MSS) models use only a few bands (most often only one or two bands). We are going to present a model that allow us to use *arbitrary* number of bands. This model is based on Generalized Wannier Functions, introduced by Kohn & Onffrey [10], for spatially varying one dimensional crystal structure. Our approach is based on that (1) the Generalized Wannier Functions (GWF) form an orthonormal basis of a L^2 space that contains all the solutions of the Schröedinger equation for the crystal structure we are interested in; (2) each GWF is exponentially localized, and (3) away

from the GaAs-AlAs interface with this basis, GWF is nearly the same as WF. Therefore, except at interface, the GWF Hamiltonian matrix elements equal the Fourier coefficients of the band structure, as for the WF Hamiltonian matrix elements for perfect crystal. We compute Fourier coefficients of the band structure using the empirical pseudo-potential method of Cohen and Bergstresser [1, 4].

The computational model presented in this paper is a major improvement of the multi-band model Roblin [16] that accounts for only one or two bands and could handle only very small structures. Moreover, only RTDs with very *thin* barriers were reported in [16]. Our model removes all these major limitations and the numerical code is very stable. In addition, Roblin [16] used the same material GaAs through out the RTD and our model can handle heterojunction of two or more different materials. Numerical results for RTD with structure of GaAs-$Ga_{.7}Al_{.3}As$-GaAs-$Ga_{.7}Al_{.3}As$-GaAs are presented in Section 7 below.

The purpose of this paper is to illustrate the numerical algorithm of this model. The detailed comparison of this model with other band model will be presented elsewhere [17].

2 The Mathematical Formulation of the Problem

2.1 The Generalized Wannier Functions

In a perfect 1-D crystal, the eigenfunctions of the Schröedinger operator with a periodic potential are Bloch functions $\phi(k,x)$, where k is the wave number and x is the spatial coordinate. Bloch functions are periodic functions of the wave number k and can be expanded as Fourier series

$$\phi(k,x) = \sqrt{\frac{a}{2\pi}} \sum_{m=-\infty}^{\infty} w(m,x) e^{ikma},$$

where a is the lattice constant, and Fourier coefficients $w(m,x)$ are called Wannier Functions. Although Wannier Functions are not eigenfunctions of the Schröedinger operator, they have several nice properties: they are exponentially localized and invariant under a lattice translation $w(m,x) = w(0,x-ma)$; they consist of an orthonormal basis of solution space, and under this basis the Hamiltonian matrix elements are the Fourier coefficients of the band-structure, which make them very useful in many practical computations as well as theoretical discussions.

In the RTD we are considering, there are two AlAs layers inserted into the GaAs bulk. Consequently the long-range perfect periodicity is lost and Bloch functions no longer exist. Since Wannier Functions are exponentially localized, however, they are near-sighted: a WF at a given lattice site almost does not see or is almost not affected at all by changes a few lattice sites away. Therefore, away from the interfaces, WFs "survive" with very little modification. Only at the interfaces, WFs need significant changes to reflect the interface coupling, which leads to the generalization of Wannier Function to imperfect crystal, called Generalized Wannier Function (GFW) [10, 11]. It can be shown that GWFs exponentially converge to WFs away from the interface.

In our full band model and numerical computations, we use GWF for RTD as a composite of modified WFs for perfect GaAs and AlAs. Intuitively, our GWFs may be constructed in the following way. We start with a GaAs bulk that is assumed to be perfect one dimensional crystal. (Although the GaAs bulk is finite, it does not matter much because WFs are "near-sighted".) We have WFs for this GaAs bulk. We then cut off two slices from GaAs bulk and fill in with AlAs in these two slots to form a double barrier. At the space locations within these two slots, we replace the WFs for GaAs with WFs for AlAs. This is equivalent to replace the Fourier coefficients of the band structure of GaAs with that of AlAs in the two blocks on the diagonal of the Hamiltonian matrix. This idea bears a striking resemblance to the concept that chemical bonds can be transfered from one compound to another, behind which there may be interesting connections worth further study.

Corresponding to the GaAs-AlAs-GaAs-AlAs-GaAs RTD structure, there are five blocks on the diagonal of the Hamiltonian matrix. The first, third and fifth represent GaAs, and the second and fourth represent AlAs. The off-diagonal entries between these blocks and within the band (from $-N_B$-th diagonal to N_B-th diagonal) are determined by the coupling of GaAs and AlAs interfaces. At the interface lattice locations, the GWF Fourier coefficients are obviously very different, which we approximate using the average of the Fourier coefficient of the GaAs and that of AlAs for simplicity. More accurate evaluation of the Fourier coefficients at the interfaces, without constructing GFWs, remains for future study.

Generalized Wannier Functions inherit most of the nice properties of Wannier Func-

tions, except translation invariance of course [8]. In this paper, we are not concerned with the proof of these properties. In stead, we simply assume that each wave function can be expanded in Generalized Wannier Functions, and the Hamiltonian matrix can be constructed in the way stated in previous paragraph.

2.2 The Problem in GWF Formulation

The behavior of a crystal electron is described by its wave-function. In stead of solving directly the Schröedinger equation for wave-function ψ, we first formally expand ψ as a superposition of Generalized Wannier functions $|n\rangle$ [15],

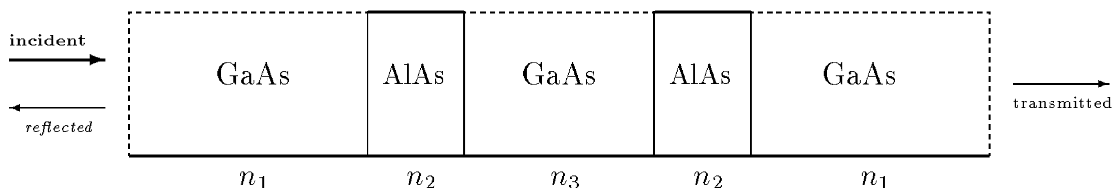
$$|\psi\rangle = \sum_{p=-\infty}^{\infty} f(n)|n\rangle,$$

where $f(n)$ is the Fourier coefficient of $|\psi\rangle$ and $|f(n)|^2$ is the probability of the presence of the electron at lattice site n. Consequently, wave-function ψ is uniquely determined by wave vector $\mathbf{f} \equiv \{f(n)\}_{n=-\infty}^{\infty}$. In the Generalized Wannier Function representation, therefore, the Schröedinger equation for wave function ψ is equivalent to an algebra equation for wave vector $\mathbf{f}$,

$$\mathcal{H}\mathbf{f} = E\mathbf{f},$$

where $\mathcal{H}$ is the Hamilton matrix and E is the conduction or valence band structure.

Consider a RTD with structure depicted below, where n_1, n_2 and n_3 are positive integers indicating numbers of lattice sites. We assume that the device axis (from left to right) is parallel to the lattice vector.



Suppose that there is one incident wave from the left (emitter) with $2N_B$ nearest neighboring waves, half propagating to the right (collector) and half being reflected to the

left. (N_B is any given positive integer. Therefore we can use arbitrary number of bands in our simulations.) Then, the energy **band structure** may be written as

$$E = E(k) = \sum_{p=-N_B}^{N_B} H_p \exp(p\ kaj), \tag{1}$$

where a is the lattice constant ($a = 5.43\text{Å}$ for GaAs), k is the wave number (to be determined), and $\{H_p\}_{p=0}^{N_B}$ are the Fourier coefficients of the band structure of GaAs or AlAs depending on the location of the lattice site along the device axis, satisfying $H_{-p} = H_p$.

Then the Hamilton matrix $\mathcal{H}$ is an infinite banded Hermitian matrix with $2N_B + 1$ bands. The equation for wave vector $\mathbf{f}$ can be written explicitly as an infinite system of difference equations:

$$\sum_{p=-N_B}^{N_B} H_{n+p} f(n+p) = Ef(n), \qquad n = 0, \pm 1, \pm 2, \cdots. \tag{2}$$

A simple solution of which is $f(n) = r^n$ with $r = \exp(kaj)$.

The group velocity of electrons is then

$$v(k) = \frac{a}{\hbar}\frac{\partial E(k)}{\partial k} = \frac{-2a^2}{\hbar}\sum_{p=1}^{N_B} pH_p \sin(p\ ka). \tag{3}$$

The probability for an electron to be reflected to the left is [15]

$$R = \frac{1}{v(k_1)}\sum_{q=1}^{l} v(k_q)|b_q|^2, \tag{4}$$

where $\{k_q\}_{q=1}^{N_B}$ are wave numbers at the left in proper order, $l \le N_B$ is the number of *real* wave numbers, and $\{b_q\}_{q=0}^{N_B}$ are the "weights" of the reflected wave corresponding to different wave numbers (to be determined later). Similarly, the probability for an electron to transmit to the right is

$$T = \frac{1}{v(k_1)}\sum_{q=1}^{l'} v(k'_q)|c_q|^2, \tag{5}$$

where $\{k'_q\}_{q=0}^{N_B}$, $l' \le N_B$ and $\{c_q\}_{q=0}^{N_B}$ are defined analogously.

In order to find R and T, therefore, we need to find

1. Fourier coefficients $\{H_q\}_{q=0}^{N_B}$ for both GaAs and AlAs, and the Hamiltonian matrix,

2. wave numbers $\{k_q\}_{q=1}^{N_B}$ and $\{k'_q\}_{q=1}^{N_B}$,

3. wave vector $\mathbf{f}$,

4. "weights" $\{b_q\}_{q=0}^{N_B}$ and $\{c_q\}_{q=0}^{N_B}$.

3 The Hamiltonian Matrix

Since GaAs and AlAs are crystals with zincblende symmetry, we may compute the band structures for GaAs and AlAs using the empirical pseudo-potential method of Cohen and Bergstresser [1, 4].

Recall that the band structure may be written as Fourier series as in (1). Suppose that q is the period of band structure in k-space, $E(k+q) = E(k)$. Using the empirical pseudo-potential method, we may first compute the values of $E(k)$ at $N_B + 1$ sampling points, $k = 0,\ T_s,\ \cdots,\ N_B T_s$, where $T_s = q/(2N_B)$ is the sampling period. Using the symmetry of the band structure $E(k)$ at the point of $k = N_B T_s$, we can obtain the values of $E(k)$ at N_B *more* points, $k = (N_B+1)T_s,\ (N_B+2)T_s,\ \cdots,\ (2N_B)T_s$. Since $E(2N_B T_s) = E(0)$, we need only $2N_B$ sampling points, $k = 0,\ T_s,\ \cdots,\ (2N_B - 1)T_s$.

With the values of $E(k)$ at the $2N_B$ sampling points, according to the sampling theorem [14], we can compute Fourier coefficient of the band structure,

$$H_p = \frac{1}{2N_B} rect(\frac{p}{2N_B}) \sum_{n=0}^{2N_B-1} E(nT_s) e^{-n(\pi p/N_B)j}, \qquad (p = 0, 1, \cdots, N_B) \qquad (6)$$

where

$$rect(x) = \begin{cases} 1, & |x| < 1/2, \\ 1/2 & |x| = 1/2, \\ 0, & |x| > 1/2. \end{cases}$$

We need to carry out above calculation for both GaAs and AlAs for their Fourier coefficients. (We may need to shift the band profiles "vertically" such that $E_G(0) = 0$ and $E_A(0) = 1.1$, where the subscripts G and A representing GaAs and AlAs respectively.)

For simplicity, we assume that the band is flat at the contact of two ends. We choose proper index and integers n_1, n_2 and n_3 such that

1. n_2 is the width of the AlAs monolayers, and n_3 is the width of the GaAs well;

2. the band is flat for $n \leq N_B$ and for $n \geq m - N_B$, where $m = 2n_1 + 2n_2 + n_3$;

3. $m \geq 3N_B$.

Let $\{H_p^G\}_{p=0}^{N_B}$ and $\{H_p^A\}_{p=0}^{N_B}$ be the Fourier coefficients for GaAs and AlAs respectively. We use them to assemble the Hamiltonian matrix $\mathcal{H}$ for the GaAs-AlAs structure as follows.

$$\mathcal{H}_{i,j} = \begin{cases} \frac{W_{i,j}^G H_{|j-i|}^G + W_{i,j}^A H_{|j-i|}^A}{W_{i,j}^G + W_{i,j}^A}, & \text{if} \quad |j-i| \leq N_B, \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

where

$$\begin{array}{ll} W_{i,j}^G = 1 - p_i p_j, & i,j = 1, 2, \cdots, m, \\ W_{i,j}^A = 1 - (1-p_i)(1-p_j), & i,j = 1, 2, \cdots, m; \\ p_i = 0, & \text{if the } i\text{th node is in a GaAs region,} \\ p_i = 1, & \text{if the } i\text{th node is in a AlAs region.} \end{array} \tag{8}$$

Note that the two AlAs regions are given by $(n_1+1) \leq i \leq (n_1+n_2)$ and $(n_1+n_2+n_3+1) \leq i \leq (n_1 + 2n_2 + n_3)$, the complement of which is the three GaAs region. In formula (7), we use the arithmetic average of H_p^G and H_p^A to represent the coupling of GaAs and AlAs near the interface. The precise evaluation of the Hamiltonian matrix entries at the interfaces, to our knowledge, requires the actual construction of the GWFs. It remains as a challenge for future research to find an easy way of computing Hamiltonian matrix to any degree of accuracy when GWFs are used as a basis.

4 The wave numbers

To find k_qs, we need first to set up the polynomial equation of $r = \exp(k_q a j)$,

$$\sum_{p=-N_B}^{N_B} H_p r^{N_B+p} = E r^{N_B}. \tag{9}$$

That is,

$$\sum_{p=1}^{N_B} H_p \left[r^{N_B-p} + r^{N_B+p} \right] + (H_0 - E) r^{N_B} = 0. \tag{10}$$

Since both ends are made of GaAs, $\{H_p\}_{p=0}^{N_B} = \{H_p^G\}_{p=0}^{N_B}$. H_0 should be replaced with $H_0 + HV(1)$ at the left end or $H_0 + HV(m)$ at the right end, however, if an addition

potential profile HV is applied to the RTD. In that case, transmitted and reflected waves will have different wave numbers.

For a good band model, N_B should be around 10 or larger. Therefore, in general, equation (10) can only be solved numerically. We compute the roots of polynomial equation (10), $r = r_1, r_2, \cdots, r_{2N_B}$ by calculating the eigenvalues of the corresponding companion matrix (see [20] for example). The matrix method appears to be very stable and efficient.

The next question is how to select and determine the order (or indices) of $k_q = -j(\ln r_q)/a$, $q = 1, 2, \cdots, m$.

The incident and reflected wave numbers $\{k_q\}_{q=1}^{N_B}$ are chosen in following way.

We select the real wave numbers $k_1, k_2, \cdots, k_l$ from the $2N_B$ roots of equation (10) such that the group velocity $\{v(k_q)\}_{q=1}^{l}$ are positive and in decreasing order, and the imaginary wave numbers $k_{l+1}, k_{l+2}, \cdots, k_{N_B}$ are chosen to have positive imaginary parts. In other words, we select $k_1, k_2, \cdots, k_{N_B}$ from the $2N_B$ roots of equation (10) in the following way

1. $k_1, k_2, \cdots, k_l$ are all the real with $v(k_q) > 0$ $(1 \leq q \leq l)$ such that $v(k_1) > v(k_2) > \cdots > v(k_l) > 0$;

2. $k_{l+1}, k_{l+2}, \cdots, k_{N_B}$ are all the complex roots with *positive imaginary* part.

Selection Criterion 1 is for two purposes: (1) the $\{k_q\}_{q=1}^{l}$ are the wave numbers of *reflected* electrons, and (2) the reflected electrons travel at a speed same or slower than incident electrons. (A mathematical formulation of this statement is given in next section in formula (11) for the wave at the left end.) Selection Criterion 2 guarantees all complex waves are damped down.

5 The Wave Vector

The key to determine the wave vector is to use the boundary conditions at two ends to reduce the *infinite* system (2) to a *finite* one.

Since the band is flat at two ends (an assumption supported by the exponential locality of GFWs), *infinite* system (2) can be reduced to a *finite* system for a *finite* wave vector $\mathbf{f}_{1.m} \equiv \{f(n)\}_{n=1}^{m}$, which is our first goal. Here the indices are chosen such that the band is flat for $n \leq N_B$ and for $n \geq m - N_B$, as stated earlier.

Consider a m-equation sub-system of system (2) with $n = 1, 2, \cdots, m$. It involves $m + 2N_B$ variables: N_B at the left contact $(f(-N_B+1), \cdots, f(0))$, m in the middle $(f(1), \cdots, f(m))$, and N_B at the right contact $(f(m+1), \cdots, f(m+N_B))$. Our idea is to use the boundary conditions to express the N_B variables at the left and the N_B variables at the right using the m variables in the middle. In this way, we derive a closed system of m equations for m variables.

5.1 The Boundary Condition at the Left

Since there are N_B reflected waves and one incident wave, the general solution on the **left** may be written as (The *incident wave* plus the sum of N_B *reflected waves.*):

$$f(n) = \exp(ik_1 na) + \sum_{q=1}^{N_B} b_q \exp(-ik_q na), \qquad n = N_B, N_B - 1, \cdots. \tag{11}$$

(The negative sign in (11) indicates the waves are traveling to left, that is, in the negative direction of the device axis.) Left boundary condition (11) gives us the relations between the following three vectors

$$\mathbf{b} = \begin{bmatrix} b_1 \\ \cdots \\ b_{N_B} \end{bmatrix}, \qquad \mathbf{f}_{1.N_B} = \begin{bmatrix} f(1) \\ f(2) \\ \cdots \\ f(N_B) \end{bmatrix}, \qquad \text{and} \qquad \mathbf{f}_{left} = \begin{bmatrix} f(1-N_B) \\ f(2-N_B) \\ \cdots \\ f(0) \end{bmatrix}.$$

We need the first $2N_B$ equations from (11) to derive formulas for $\mathbf{b}$ and $\mathbf{f}_{left}$. With $r_q = \exp(ik_q a)$, the $2N_B$ equations may be rewritten as

$$\sum_{q=1}^{N_B} r_q^{-n} b_q = f(n) - r_1^n, \qquad \begin{cases} n = 1 - N_B, 2 - N_B, \cdots, -1, 0, \\ n = 1, 2, \cdots, N_B. \end{cases} \tag{12}$$

In matrix form, systems (12) become

$$\begin{aligned} \mathbf{R}_1 \mathbf{b} &= \mathbf{f}_{left} - \mathbf{r}_{(1-N_B).0}, \\ \mathbf{R}\mathbf{b} &= \mathbf{f}_{1.N_B} - \mathbf{r}_{1.N_B}, \end{aligned}$$

where

$$\mathbf{R}_1 = (r_q^{N_B - p})_{N_B \times N_B}, \qquad \mathbf{R} = (r_q^{-p})_{N_B \times N_B},$$

(p and q are row and column indices respectively.), and

$$\mathbf{r}_{(1-N_B).0} = \begin{bmatrix} r_1^{1-N_B} \\ \cdots \\ r_1^0 \end{bmatrix}, \qquad \mathbf{r}_{1.N_B} = \begin{bmatrix} r_1 \\ \cdots \\ r_1^{N_B} \end{bmatrix}.$$

Therefore,

$$\mathbf{f}_{left} = \mathbf{R}_1\Big[\mathbf{R}^{-1}\mathbf{f}_{1.N_B} - \mathbf{R}^{-1}\mathbf{r}_{1.N_B}\Big] + \mathbf{r}_{(1-N_B).0}. \tag{13}$$

We further observe that

$$\mathbf{R} = (r_q^{N_B-p})_{N_B\times N_B}\cdot \mathrm{diag}[r_1^{-N_B}, r_2^{-N_B}, \cdots, r_{N_B}^{-N_B}] = \mathbf{R}_1\cdot \mathrm{diag}[r_1^{-N_B}, r_2^{-N_B}, \cdots, r_{N_B}^{-N_B}],$$

$$\mathbf{R}^{-1} = \mathrm{diag}[r_1^{N_B}, r_2^{N_B}, \cdots, r_{N_B}^{N_B}]\cdot \mathbf{R}_1^{-1}.$$

Hence,

$$\mathbf{f}_{left} = \mathbf{R}_1\mathbf{D}_r\mathbf{R}_1^{-1}\mathbf{f}_{1.N_B} + \mathbf{r}_{(1-N_B).0} - \mathbf{R}_1\mathbf{D}_r\mathbf{R}_1^{-1}\mathbf{r}_{1.N_B}, \tag{14}$$

where

$$\mathbf{D}_r = \mathrm{diag}[r_1^{N_B}, r_2^{N_B}, \cdots, r_{N_B}^{N_B}].$$

We note that $\mathbf{R_1}$ is a Vandermonde matrix whose determinant equals $\prod_{i\neq j}(r_i - r_j)$. So $r_1, \cdots, r_{N_B}$ must be distinct and computed accurately, which is needed for $\mathbf{f}_{left}$ given in (14) and for the left "weight" vector (for reflected waves)

$$\mathbf{b} = (\mathbf{D}_r\mathbf{R}_1)^{-1}(\mathbf{f}_{1.N_B} - \mathbf{r}_{1.N_B}). \tag{15}$$

The same comment applies to Vandermonde matrix $\mathbf{S}_1$ below.

Equations (14) and (15) express the N_B variables at the left and the left "weight" vector in terms of the first N_B variables in the middle, which closes up at the left end of the m-equation sub-system mentioned at the beginning of this section.

5.2 The Boundary Condition at the Right

Since there are N_B transmitted waves, the general solution on the **right** may be written as:

$$f(n) = \sum_{q=1}^{N_B} \tilde{c}_q \exp(ik_q' na), \qquad n = m - N_B + 1, m - N_B + 2, \cdots. \tag{16}$$

Let $s_q = \exp(ik'_q a)$ and $c_q = \tilde{c}_q \exp(ik'_q(m+1)a)$. Then, the first $2N_B$ equations in (16) can be written as

$$\sum_{q=1}^{N_B} s_q^n c_q = f(m+n+1), \qquad \begin{cases} n = 0, \cdots, N_B - 2, N_B - 1; \\ n = -N_B, 1 - N_B, \cdots, -1. \end{cases} \tag{17}$$

In matrix form

$$\begin{aligned} \mathbf{S}_1 \mathbf{c} &= \mathbf{f}_{right}, \\ \mathbf{S}\mathbf{c} &= \mathbf{f}_{(m-N_B+1).m}, \end{aligned}$$

where

$$\mathbf{S}_1 = (s_q^{p-1})_{N_B \times N_B}, \quad \mathbf{S} = (s_q^{p-1-N_B})_{N_B \times N_B}, \quad \mathbf{c} = [c_1, \ \cdots, \ c_{N_B}]^T.$$

$$\mathbf{f}_{(m-N_B+1).m} = \begin{bmatrix} f(m-N_B+1) \\ f(m-N_B+2) \\ \cdots \\ f(m) \end{bmatrix}, \quad \mathbf{f}_{right} = \begin{bmatrix} f(m+1) \\ f(m+2) \\ \cdots \\ f(m+N_B) \end{bmatrix}.$$

Therefore,

$$\mathbf{f}_{right} = \mathbf{S}_1 \mathbf{D}_s \mathbf{S}_1^{-1} \mathbf{f}_{(m-N_B+1).m}, \tag{18}$$

where

$$\mathbf{D}_s = \mathrm{diag}[s_1^{N_B}, s_2^{N_B}, \cdots, s_{N_B}^{N_B}].$$

The right "weight" vector can be computed as

$$\mathbf{c} = (\mathbf{D}_s \mathbf{S}_1)^{-1} \mathbf{f}_{(m-N_B+1).m}. \tag{19}$$

Equations (18) and (19) express $\mathbf{f}_{right}$, the N_B variables at the right, and $\mathbf{c}$, the right "weight" vector in terms of $\mathbf{f}_{(m-N_B+1).m}$, the last N_B variables in the middle, which closes up at the right end of the m-equation sub-system.

5.3 The Reduced System for Finite Wave Vector $\mathbf{f}_{1.m}$

With formulas (14) and (18), we can derive the finite system from infinite system (2). The m-equation sub-system

$$\sum_{p=-N_B}^{N_B} H_{n+p} f(n+p) = Ef(n), \qquad \begin{cases} n = 1, 2, \cdots, N_B, \\ n = N_B + 1, N_B + 2, \cdots, m - N_B, \\ n = m - N_B + 1, m - N_B + 2, \cdots, m, \end{cases} \tag{20}$$

can be rewritten in vector form as

$$(\hat{\mathbf{H}} - E)\mathbf{f}_{1.m} = \begin{bmatrix} -\mathbf{U}\mathbf{f}_{left} \\ 0 \\ \cdots \\ 0 \\ -\mathbf{U}^T\mathbf{f}_{right} \end{bmatrix}, \tag{21}$$

where

$$\hat{\mathbf{H}} = \begin{bmatrix} \tilde{H}_0(1) & H_1 & H_2 & \cdots & H_{N_B} & & & \\ H_1 & \tilde{H}_0(2) & H_1 & H_2 & \cdots & H_{N_B} & & \\ & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \\ & & & H_{N_B} & H_{N_B-1} & \cdots & H_1 & \tilde{H}_0(m) \end{bmatrix}_{m\times m},$$

$$\mathbf{f}_{1.m} = \begin{bmatrix} f(1) \\ f(2) \\ \cdots \\ f(m) \end{bmatrix}, \quad \mathbf{U} = \begin{bmatrix} H_{N_B} & H_{N_B-1} & \cdots & H_1 \\ 0 & H_{N_B} & \cdots & H_2 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & H_{N_B} \end{bmatrix}_{N_B\times N_B}. \tag{22}$$

It then follows from (21), (14) and (18) that the reduced linear system for finite wave vector $\mathbf{f}_{1.m}$ is

$$\mathbf{M}\mathbf{f}_{1.m} = \mathbf{y}, \tag{23}$$

where

$$\mathbf{M} = \hat{\mathbf{H}} - E + \begin{bmatrix} \mathbf{U}\mathbf{R}_1\mathbf{D}_r\mathbf{R}_1^{-1} & & & & \\ & 0 & & & \\ & & \ddots & & \\ & & & 0 & \\ & & & & \mathbf{U}^T\mathbf{S}_1\mathbf{D}_s\mathbf{S}_1^{-1} \end{bmatrix}_{m\times m},$$

and

$$\mathbf{y} = \begin{bmatrix} \mathbf{U}(\mathbf{R}_1\mathbf{D}_r\mathbf{R}_1^{-1}\mathbf{r}_{1.N_B} - \mathbf{r}_{(1-N_B).0}) \\ 0 \\ \cdots \\ 0 \end{bmatrix}_{m\times 1}.$$

Note that $\hat{\mathbf{H}}$, $\mathbf{U}$ is given directly by the Hamiltonian matrix. Nevertheless, we need to find $\mathbf{R}_1$, $\mathbf{D}_r$, $\mathbf{S}_1$ and $\mathbf{D}_s$, which depend on $r_q = \exp(ik_q a)$ and $s_q = \exp(ik'_q a)$, $q = 1, 2, \cdots, m$. In order for $\mathbf{R}_1$ and $\mathbf{S}_1$ to be invertible, $r_1, r_2, \cdots, r_{N_B}$ should be m **distinct** roots, so are $s_1, s_2, \cdots, s_{N_B}$.

Also note that, $\mathbf{y}$ depends on r_1 that represents the incidental wave. If there is no incidental wave, but there is an **impulse** at position k, then $\mathbf{y} = [\delta_{nk}]_{n=1}^m$. The following algorithms need only this simple modification to compute the current in the case of an impulse.

6 The Procedure for Computing Transmitted/Reflected Current

In summary, we have derived the following procedure to compute the transmitted and reflected current in a RTD device.

1. Compute the Fourier coefficients for both GaAs and AlAs using formula (6) and sampling points obtained by the empirical pseudo-potential method of Cohen and Bergstresser [1, 4];

2. Set up the Hamiltonian matrix for the given structure using (7);

3. Find wave numbers at the left $\{k_q\}_{q=1}^{N_B}$ by solving polynomial equation (10) using matrix method at both ends; ($\{k_q\}_{q=1}^{N_B}$ are to be selected such that (1) $\{k_q\}_{q=1}^{l}$ are all the real roots and $v(k_1) > v(k_2) > \cdots > v(k_l) > 0$, and (2) $\{k_q\}_{q=l+1}^{N_B}$ all the complex roots with positive imaginary part.)

4. Set up $N_B \times N_B$ matrices $\mathbf{U}$, $\mathbf{R}_1$ and $\mathbf{D}_r$, and compute matrix $\mathbf{P} = \mathbf{U}\mathbf{R}_1\mathbf{D}_r\mathbf{R}_1^{-1}$.

5. Set up the right-hand-side vector $\mathbf{y} = [\mathbf{P}\mathbf{r}_{1.N_B} - \mathbf{U}\mathbf{r}_{(1-N_B).0}, 0, \cdots, 0]^T$.

6. Find wave numbers at the right $k'_1,\ k'_2,\ \cdots,\ k'_{N_B}$ in the same way as finding $\{k_q\}_{q=1}^{N_B}$.

7. Set up $N_B \times N_B$ matrices $\mathbf{S}_1$ and $\mathbf{D}_s$, and compute matrix $\mathbf{Q} = \mathbf{U}^T\mathbf{S}_1\mathbf{D}_s\mathbf{S}_1^{-1}$.

8. Set up $m \times m$ coefficient matrix $\mathbf{M} = \hat{\mathbf{H}} - E + \mathrm{diag}[\mathbf{P}, 0, \cdots, 0, \mathbf{Q}]$.

9. Solve the reduced finite system $\mathbf{Mf} = \mathbf{y}$ using a band-matrix solver.

10. Compute the transmitted current T and the reflected current R using formulas (4, 5), and check the result using current conservation law.

This procedure applies to other semiconductor materials than GaAs and AlAs. It has been implemented in both Fortran and Matlab. It has been very efficient and stable in all of our numerical computations with these codes.

7 Some Numerical Results

To demonstrate the effectiveness and efficiency of our model, we use it to compute the transmitted/reflected current of RTD with structure GaAs-$Ga_{.7}Al_{.3}$As-GaAs-$Ga_{.7}Al_{.3}$As-GaAs .

The band structures of GaAs and AlAs in 100 direction, computed with the empirical pseudo-potential method, are depicted in Figures 1 and 2. The reason that we use $Ga_{.7}Al_{.3}$As as the mono-layers is that it is fairly close to GaAs and the average in (7) is a good approximation at the interface. The band structure of $Ga_{.7}Al_{.3}$As is depicted in Figure 3. In Figures 4 and 5, we show transmitted current as a function of energy level for two RTD with different barrier widths but the same well width. We notice that, from Figure 4, there almost no electrons are transmitted when their energy level is lower than 0.1 eV; electrons can go through clean the double barrier when their energy level is about 0.175 eV or 0.570 eV (resonant state). It is interesting that electrons have very low probability of passing through at a higher energy level around 0.275 eV. This behavior becomes more pronounced as the mono-layers become thiner, as shown in Figure 5.

In Figure 6, we present a calculation for a RTD device with barrier width same as well width, $n_2 = n_3 = 6$. Notice the sharp change of the transmitted current in small region near energy level 0.161 eV. This feature suggests RTD as a potential impulse generator.

Also plotted in Figures 4, 5 and 6 are the transmitted current in the Γ valley, T_1, and the transmitted current in the X valley, T_2. They are the first and second term in formula (5) for total transmitted current T. That is,

$$T_1 = \frac{v(k_1')}{v(k_1)}|c_1|^2, \qquad T_2 = \frac{v(k_2')}{v(k_1)}|c_2|^2. \tag{24}$$

8 Concluding Remarks

Using Generalized Wannier Function representation, we have developed a full band algorithm for RTD. Our numerical experiments have shown that this algorithm is very efficient and stable. It allows us to use an arbitrary number of nearest neighboring bands (we used 20 in the calculations presented in §7). It is therefore the natural band model for any future *non-local* multiple sequential scattering algorithm. This full band model accounts automatically for non-effective mass effects and Γ to X band coupling and therefore be useful for calibrating the coupling constant of other multi-band Wannier models.

For convenience, our theory is stated in this paper using GaAs and AlAs as sample materials with *two* junctions. It should be pointed out that, however, our theory can handle the junction of other materials as well, and our approach applies to super-lattice structure with any number of junctions and any number of different materials. More junctions simply mean more blocks in matrix $\hat{\mathbf{H}}$, and different material (in the same block location) simply means different matrix entries (on each diagonals of that block).

It can be seen from our derivation in this paper that we do not actually compute these Generalized Wannier Functions. At least for our purpose in this paper, we even do not need to know what they really are as long as they exist as a basis with the property that allow us to compute the Hamiltonian matrix *easily*. These simplifications come with a price that the approximation of the Fourier coefficients at the interfaces is relatively crude. The same approximation might lead to large errors when the mono-layers are 100% AlAs. We must further study the properties of GWF and derive more accurate formulas for the Fourier coefficients at the interfaces, without explicitly constructing all the GWFs. GWFs themselves are very interesting. They could be used in other applications such as calculating the ground state energies of many-atom systems to reduce the computing time dramatically [8]. These are interesting and challenging problems mathematically as well as "practically" (in engineering applications). We are looking forward to finding the answers and further improve the numerical implementation of our model in specific

applications.

ACKNOWLEDGMENT
This work was supported by Texas Instruments. Dr. Paul Vonallmen of The University of Illinois at Urbana-Champaign kindly provides us his Fortran code that implements the empirical pseudo-potential method.

References

1. W. Andreoni & R. Car, *Phys. Rev.* B 21(8), 3334 (1980).

2. E. R. Brown, J. R. Söderström, C. D. Parker, L. J. Mahoney, K. M. Molvar & T. C. McGill, *Appl. Phys. Lett.* 58, 2291 (1991).

3. L. L. Chang, L. Esaki & R. Tsu, *Appl. Phys. Lett., 24*, 593 (1974).

4. M.L. Cohen and T.K. Bergstresser, *Phys. Rev.* 141(2), 789 (1963).

5. L. Esaki & R. Tsu, *IBM J. Res. Dev. 14*, 61 (1970).

6. J. G. Gay & J. R. Smith, *Phys. Rev.* B 11(12), 4906 (1975).

7. W. Kohn, *Phys. Rev.* B 7(10), 4388 (1973).

8. W. Kohn, *Chem. Phys. Lett.* 208, 167 (1993).

9. W. Kohn, *Density Functional Theory*, ed. by E.K.U. Gross & R. M. Dreizler, Penum Press, New York (1995).

10. W. Kohn & J. R. Onffroy, *Phys. Rev.* B 8(6), 2485 (1973).

11. R. Lake & S. Datta, *Phys. Rev.* B 45(12), 6670 (1992).

12. H. C. Liu & T. C. L. G. Sollner, in *Semicondutors and Semimetals*, edited by R. K. Willardson, A. C. Beer & E. R. Weber, Vol. 41 (1994).

13. A. Rydberg & H. Grñqvist, *Electron Lett.*, 25, 348 (1989).

14. P. Roblin, D. Sc. Dissertation, Washington University, (1984).

15. P. Roblin & W. R. Miller, *Phys. Rev.* B 32, 5222 (1985).

16. P. Roblin, Superlattice and Microstructure, Vol 4, No. 3, 363 (1988).

17. P. Roblin, J. G. Cao & P. Sotirelis, in preparation.

18. P. Sotirelis & P. Roblin, *Phys. Rev.* B, to appear ???.

19. R. Tsu & L. Esaki, *Appl. Phys. Lett.*, *22*, 562 (1973).

20. D. M. Young & R. T. Gregory, *A Survey of Numerical Analysis*, Vol. 1, 219 (1988).

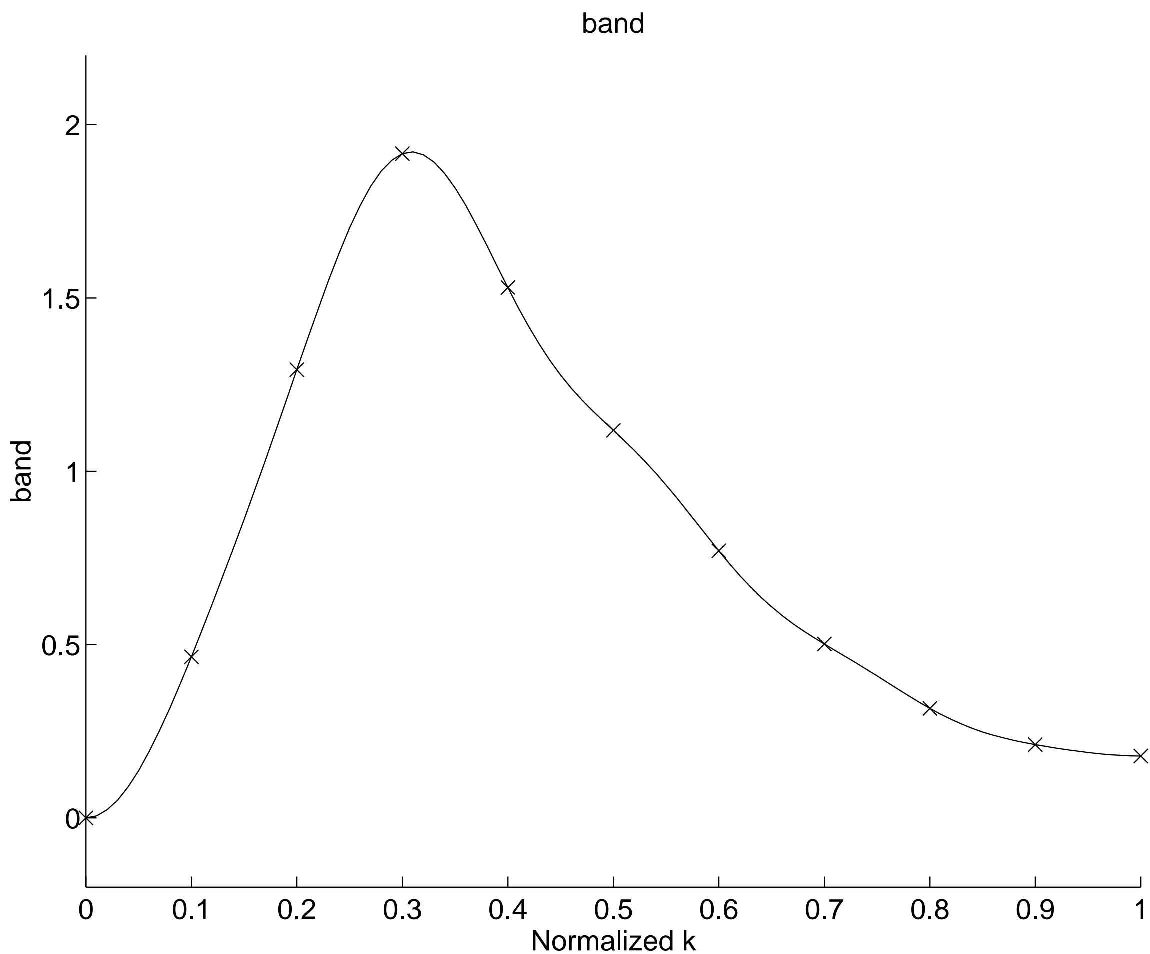


Figure 1: Band structure of **GaAs**, E_{GaAs}, reconstructed using sampling theorem. The 11 sample points in the Brillouin zone (100) are computed using the empirical pseudo-potential method of Cohen and Bergstresser. The Fourier coefficients are computed using inversion formula (6). Note that we have shifted band profiles "vertically" such that $E_{GaAs}(0) = 0$. In addition, $E_{GaAs}(0.3) = 1.9163$ is the maximum value, and $E_{GaAs}(1) = 0.17865$.

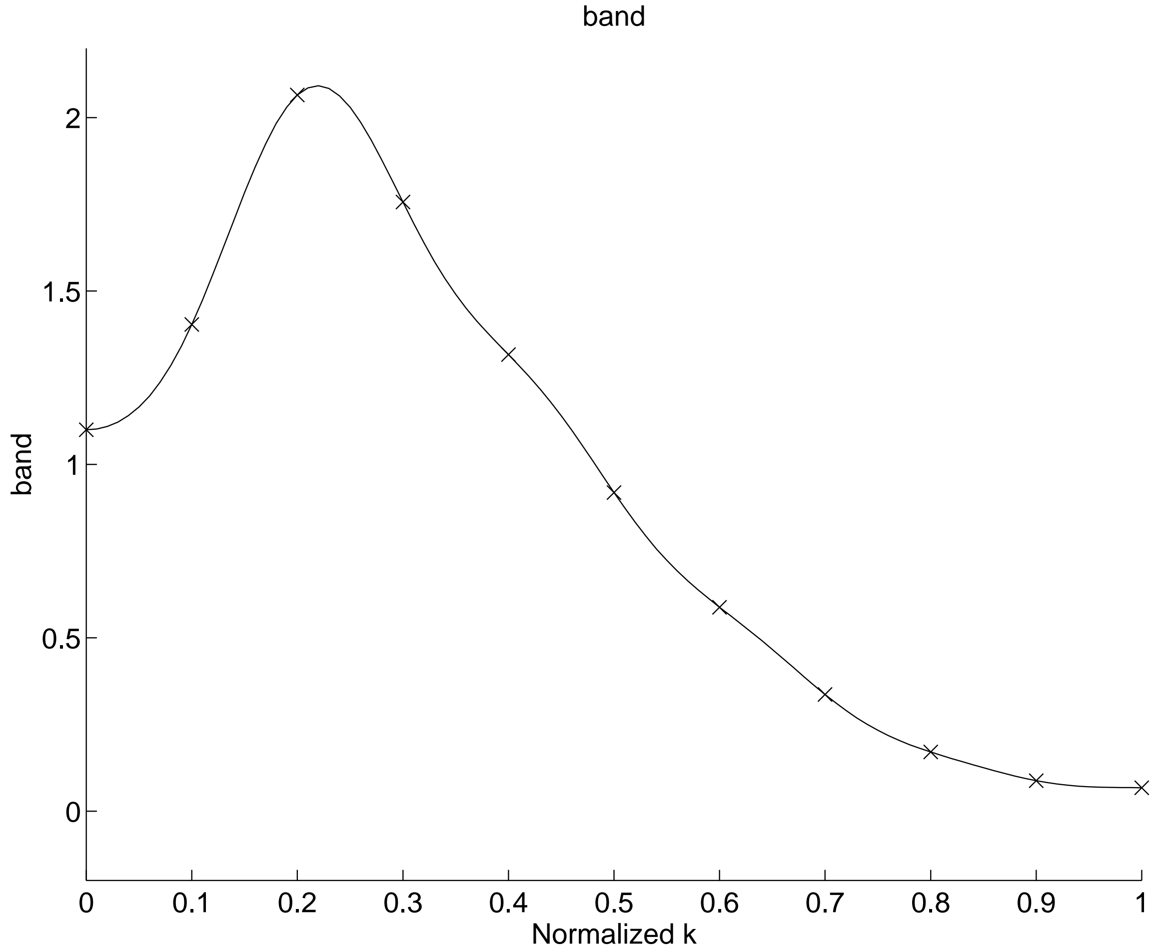


Figure 2: Band structure of **AlAs**, E_{AlAs}. The 11 sample points are computed using the empirical pseudo-potential method of Cohen and Bergstresser. The Fourier coefficients are computed using inversion formula (6). We have shifted band profiles "vertically" such that $E_{GaAs}(0) = 1.1$. In addition, $E_{AlAs}(0.3) = 2.0652$ is the maximum value, and $E_{AlAs}(1) = 0.067590$.

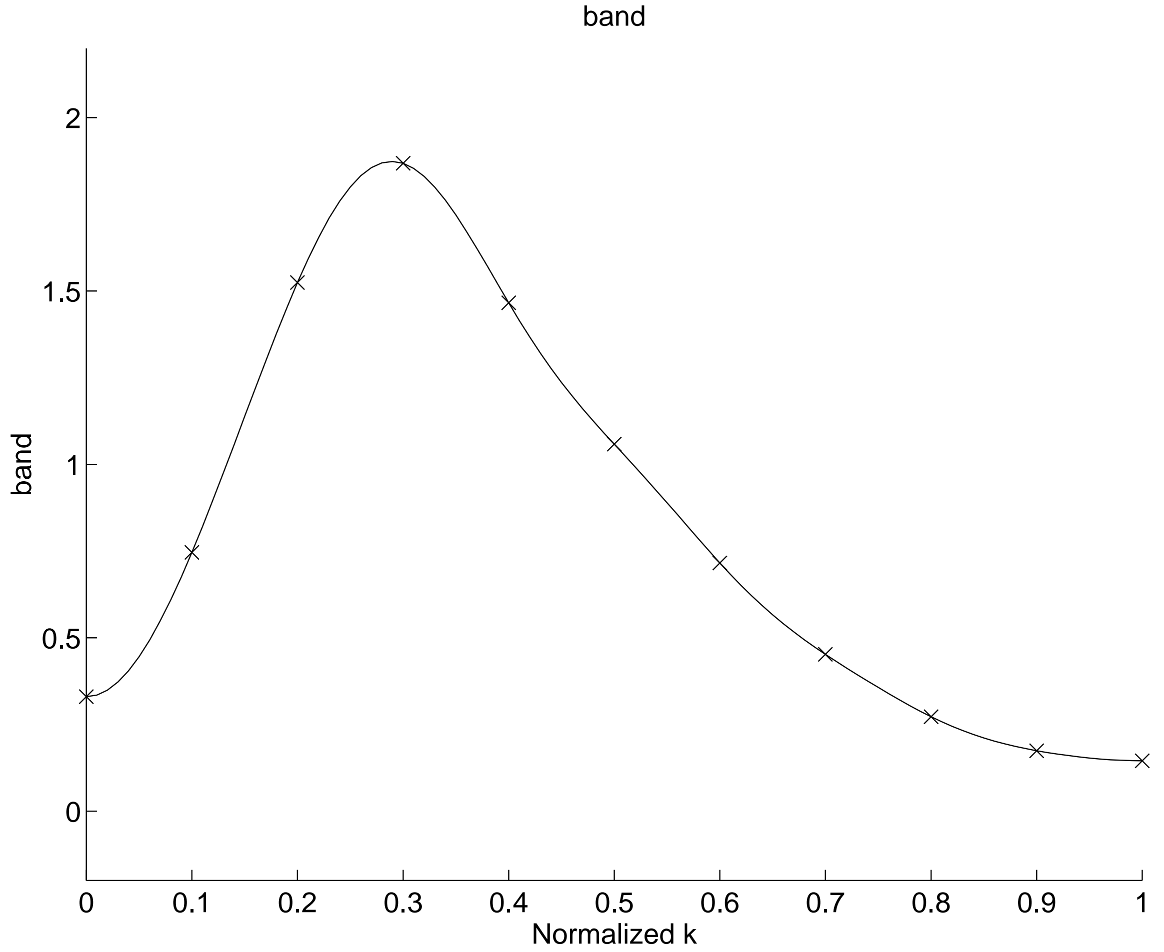


Figure 3: Band structure $(1-x)E_{GaAs} + xE_{AlAs}$ for $x = 0.3$. The 11 sample points are computed by applying the empirical pseudo-potential method of Cohen and Bergstresser to GaAs and AlAs and then using formula $E = (1-x)E_{GaAs} + xE_{AlAs}$ with $x = 0.3$. The Fourier coefficients are computed using inversion formula (6). Note that $E(0) = 0.33$, $E(0.3) = 1.8684$ is the maximum value, and $E(1) = 0.14533$.

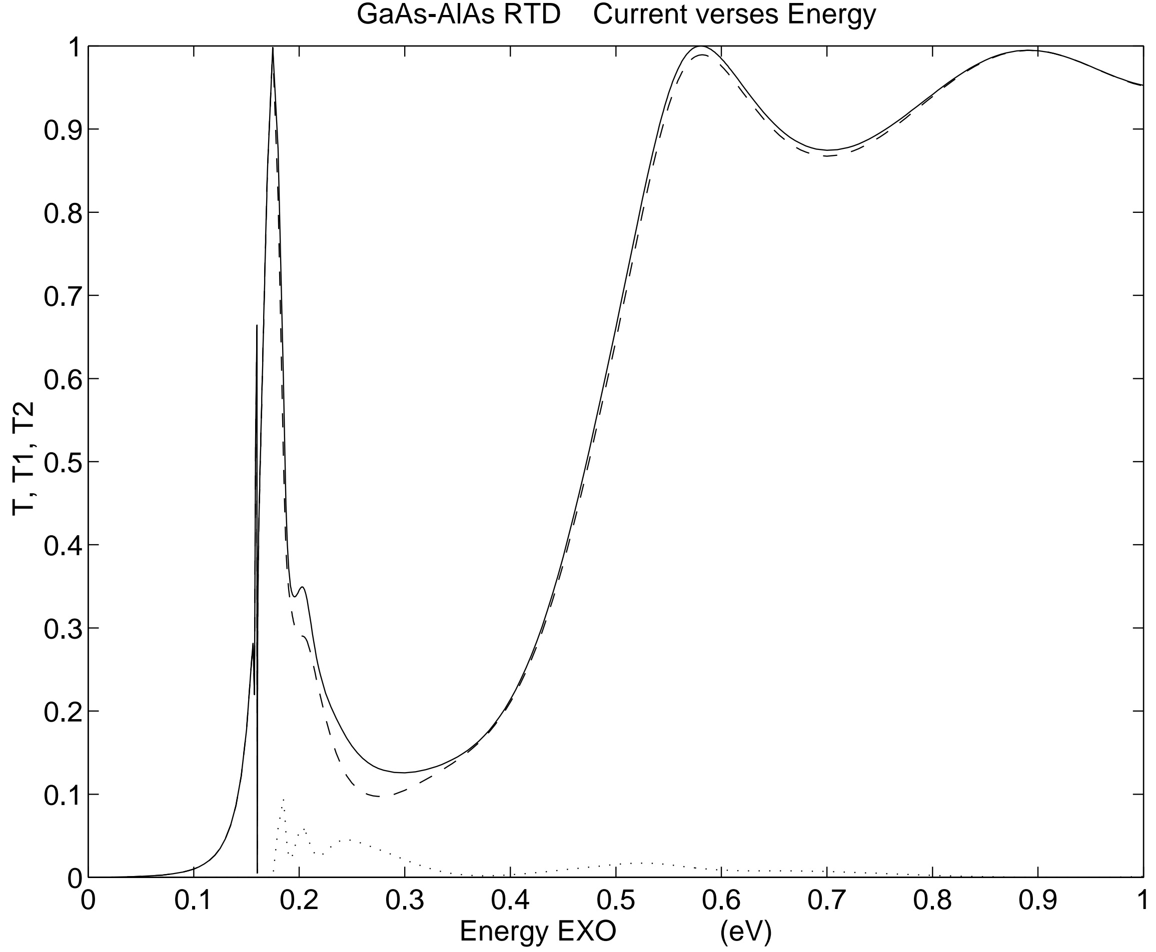


Figure 4: Current corresponding to the band structure with the barriers specified by $(1-x)E_{GaAs} + xE_{AlAs}$ for $x = .3$. No voltage is applied ($V_D = 0.0eV$). $n_1 = 11$, and the barrier width and the well width are equal, $n_2 = n_3 = 8$. Solid line: T, the transmitted current; dash line: T_1, the transmitted current in the Γ band; dot line: T_2, the transmitted current in the X band. Two resonant states are at 0.175 eV and 0.570 eV.

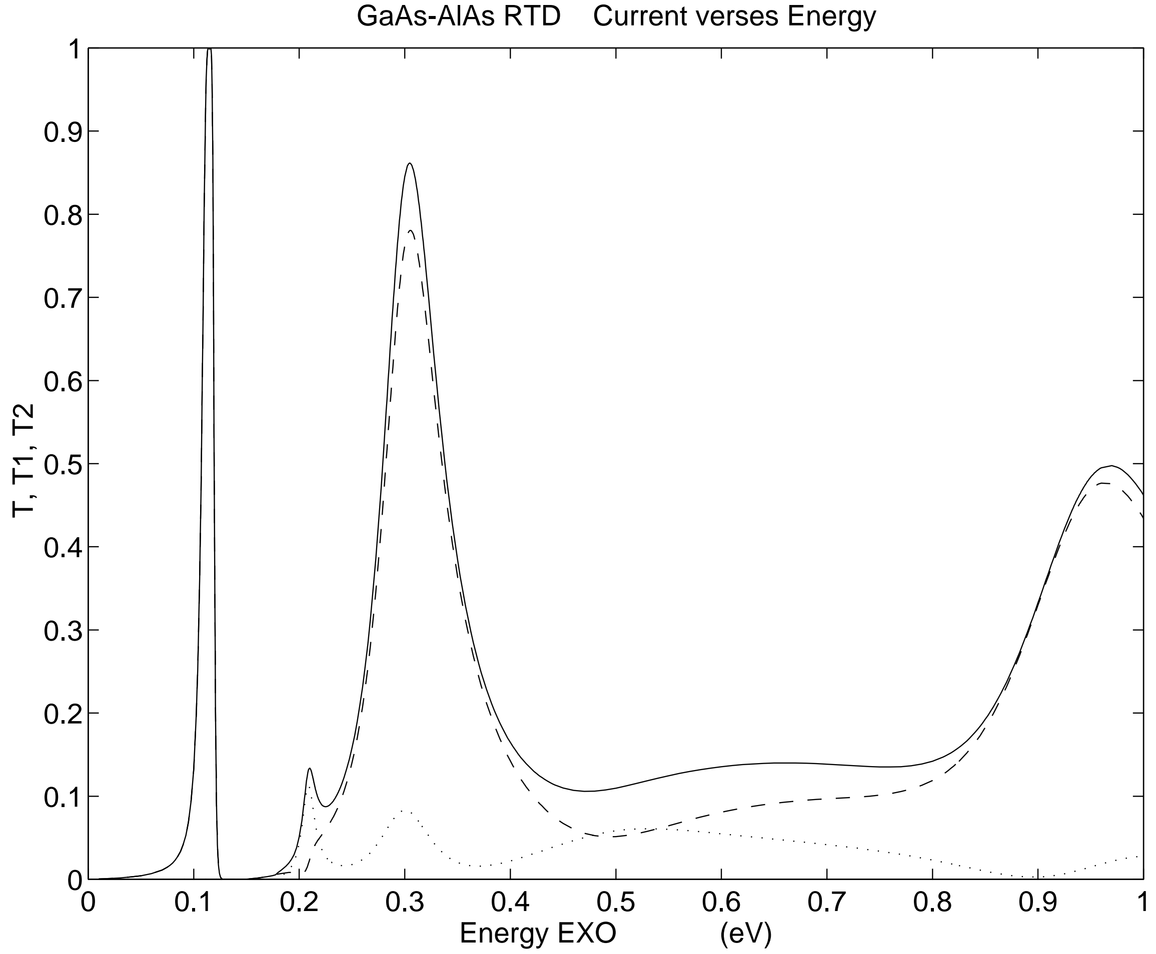


Figure 5: Current corresponding to the band structure with $AlAs$ barriers. No voltage is applied ($V_D = 0.0eV$). $n_1 = 11$, the barrier width $n_2 = 3$, and the well width $n_3 = 8$. Solid line: T, the transmitted current; dash line: T_1, the transmitted current in the Γ band; dot line: T_2, the transmitted current in the X band. A resonant state occurs at 0.116 eV.

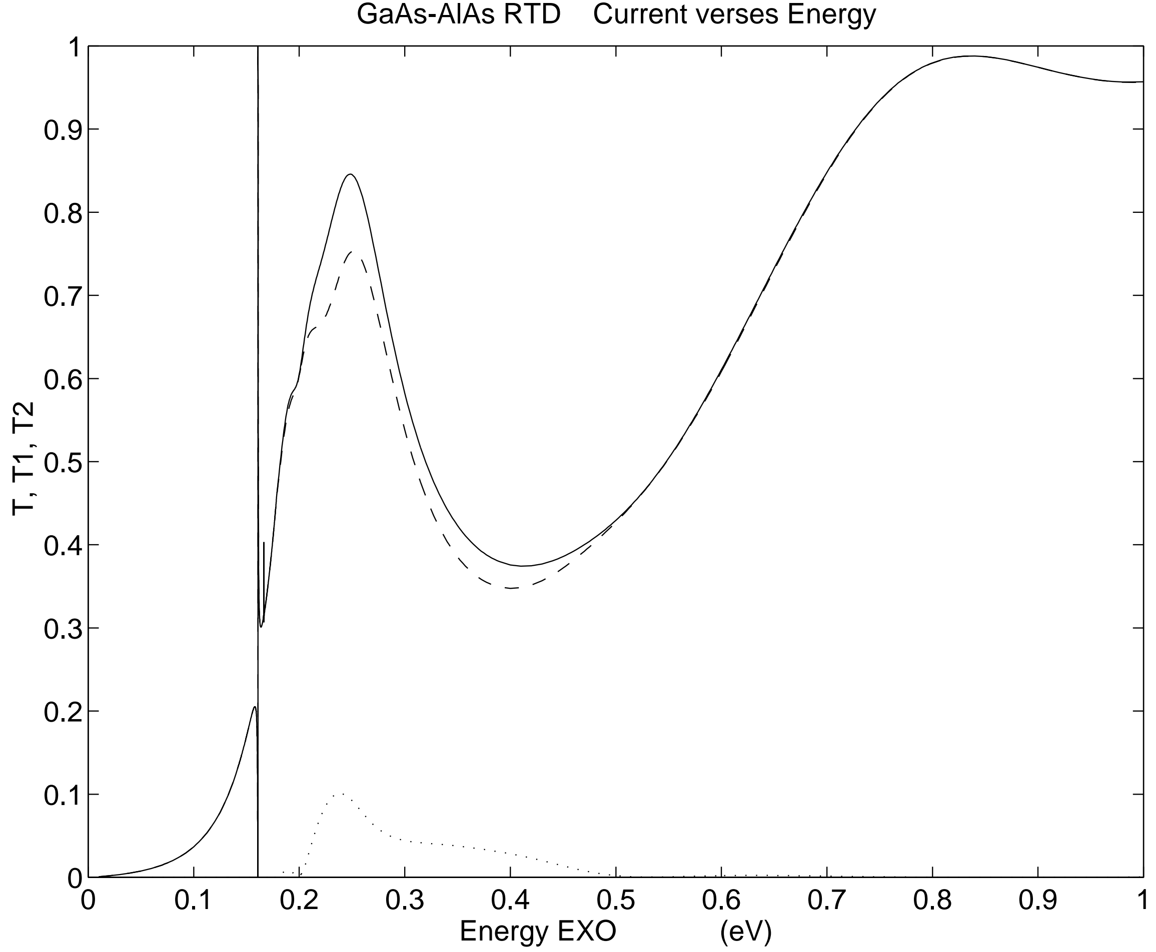


Figure 6: Current corresponding to the band structure with the barriers specified by $(1-x)E_{GaAs} + xE_{AlAs}$ for $x = 0.3$, and no voltage is applied ($V_D = 0.0eV$). $n_1 = 11$, and the barrier width and the well width are equal, $n_2 = n_3 = 6$. Solid line: T, the transmitted current; dash line: T_1, the transmitted current in the Γ band; dot line: T_2, the transmitted current in the X band. A resonant state occurs at 0.161 eV.