The Nanoelectronic Modeling Tool (NEMO) and its Expansion to High Performance Parallel Computing

Gerhard Klimeck [a], Chris Bowen [b], Timothy B. Boykin [c], Fabiano Oyafuso [d], Carlos H. Salazar-Lazaro [a], Adrian Stoica [a], and Thomas Cwik [a]

[a] Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109
[b] Texas Instruments Inc., Dallas, TX 75265
[c] University of Alabama in Huntsville, Huntsville, AL 35899
[d] University of Illinois, Beckmann Institute, Urbana, IL 61801

Abstract

Material variations on an atomic scale enable the quantum mechanical functionality of devices such as resonant tunneling diodes (RTDs), quantum well infrared photodetectors (QWIPs), quantum well lasers, and heterostructure field effect transistors (HFETs). The design and optimization of such heterostructure devices requires a detailed understanding of quantum mechanical electron transport. NEMO is a general-purpose quantum device design and analysis tool that addresses this problem. This paper highlights two recent NEMO developments: 1) genetic algorithm based parameter optimization and 2) full band hole transport simulation.

Introduction

The NASA/JPL goal to reduce payload in future space missions while increasing mission capability demands miniaturization of measurement, analytical and communication systems. Currently, typical system requirements include the detection of particular spectral lines, associated data processing, and communication of the acquired data to other subsystems. While silicon device technology dominates the commercial microprocessor and memory market, semiconductor heterostructure devices maintain their niche for light detection, light emission, and high-speed data transmission. The production of these heterostructure devices is enabled by the advancement of material growth techniques, which opened up a vast design space. The full experimental exploration of this design space is unfeasible and a reliable design tool is needed.

Military applications have similar system requirements to those listed above. Such requirements prompted a device modeling project at the Central Research Laboratory of Texas Instruments (which transferred to Raytheon Systems in 1997). NEMO was developed as a general-purpose quantum mechanics-based 1-D device design and analysis tool from 1993-97. The tool is available to US researchers by request on the NEMO web site[1]. NEMO is based on the non-equilibrium Green function approach, which allows a fundamentally sound inclusion of the required physics: bandstructure, scattering, and charge self-consistency. The theoretical approach is documented in references [2, 3] while some of the major simulation results are documented in references [4-6]. This paper highlights two recent NEMO developments at JPL: 1) genetic algorithm based device parameter optimization and 2) full band hole transport simulation.

Genetically Engineered Quantum Devices

Heterostructure device designs involve the choice of material compositions, layer thicknesses, and doping profiles. In this work the RTD is used as a vehicle to study the effects of structural and doping variations on the electron transport. A parallelized genetic algorithm package (PGAPACK)[7] is combined with NEMO to optimize structural and doping parameters. The electron transport simulations are based on a single band model, which incorporates[3] effects of non-parabolic bands in the longitudinal and transverse directions relative to electron transport. The model parameters are derived from a tight binding sp3s* multiband model. The computation of the non-parabolic single band model executes about 60 times faster than the computation of the full band sp3s* model. This dramatic increase in speed allows inclusion of Hartree charge self-

consistency with non-parabolicity in the transverse direction. The double integral in total energy and transverse momentum (see Eq. (1) below) is carried out explicitly[2] in the inner loop of the charge self-consistency.

In the numerical experiment described in Figure 1, five parameters (2 doping concentrations, N_1, N_2, and 3 thicknesses, T_1, T_2, T_3) are varied within the genetic algorithm in order to achieve the best fit to an experimental current-voltage (I-V) curve. The simulation is started from a random population of 200 parameter sets. The doping population is logarithmically distributed around the nominal values by factors of 10 ($N_1 \in [1x10^{17}, 1x10^{19}]$, $N_2 \in [1x10^{14}, 1x10^{16}]$). The layer thickness population is uniformly distributed around the nominal value by 10 monolayers ($T_1 \in [1,17]$, $T_2, T_3 \in [6,26]$). In each generation 63 of the worst genes are dropped out of the population and new genes are generated[7] from the rest by mutation and crossover. Mutation allows the parameters to leave the original parameter range. The genetic algorithm converges to the nominal structure values, well within the experimental uncertainty as shown in Figure 1. Again it is found that the barrier widths must be increased in the simulation by 1 or 2 monolayers versus the nominal values to achieve the best agreement with experimental data[5].

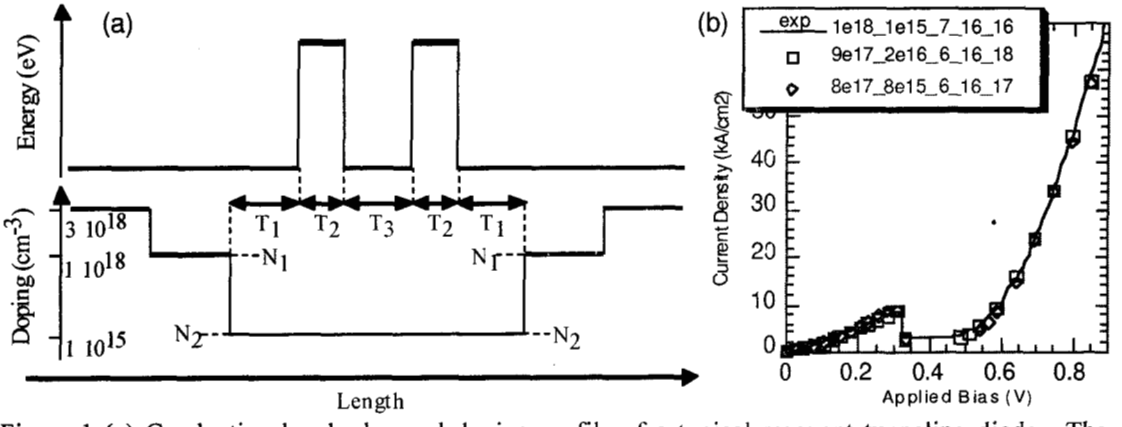

Figure 1 (a) Conduction band edge and doping profile of a typical resonant tunneling diode. The central device region is typically undoped. The low doped spacer thickness, the barrier thicknesses and the well thickness are labeled T_1, T_2, and T_3, respectively. The low spacer doping and the central device doping are labeled N_1 and N_2, respectively. These five parameters were varied in the simulations reported here. (b) I-V characteristics of an InGaAs/InAlAs resonant tunneling diode. The nominal structure is described by $T_1=T_2=16$ml (monolayers), T3=7ml, $N_1=1x10^{18}$ cm^{-3}, and $N_2=1x10^{15}$ cm^{-3}. The solid line shows experimental data[5], without the noise in the valley current region. The dots indicate two different "optimal" parameter sets. The curves are labeled by $N_1\_N_2\_T_1\_T_2\_T_3$.

The design of a fitness function, which evaluates the performance of a particular parameter set, is important for convergence of the optimization algorithm. The experimental curve is used as a reference in this work. The fitness of each particular simulated I-V characteristic is evaluated against this reference. The fitness function includes six contributions: relative and absolute error of all data points, peak/valley current and voltage and the slope at the peak/valley. These six errors are added with different relative weights to make up the overall fitness. Different relative weights will result in different "optimal" structures as shown in Figure 1b.

This work is the first step to integrate NEMO within a high performance parallel computational environment. A desired curve can now be entered as the target of the simulation and the genetic algorithm is expected obtain the optimal parameter set. Future work will utilize this method to analyze the vast material and structure parameter space. It is planned to evaluate other optimization techniques such as simulated annealing and directive approaches as well.

Full band Simulation of Hole Transport

In most high-speed quantum devices an attempt is made to utilize the high electron mobility in III-V materials. Carrier transport research has, therefore, focussed on pure electron transport. Optical devices, however, typically involve quantum states in the valence bands, i.e. hole states. To begin the study of electron and hole transport in laser structures the pure hole transport in a hole-doped RTD is examined. The model RTD considered here consists of 10 ml AlAs barriers with a 20 ml GaAs well. To avoid complications due to notch states outside the RTD a linear potential drop is applied.

The heavy- and light-hole bands in a typical III-V semiconductor are close enough in energy that states will mix due to heterostructure confinement and applied biases. The treatment of the mixing of the states is complicated by the anisotropy of the valence bands in different crystal directions such as [100], [110], and [111]. A second nearest-neighbor tight-binding model has been implemented to fully model the anisotropy in these three directions. The spin-orbit interaction and an explicit spin basis are included. The major disadvantage of this model is the increased numerical complexity as described in the next paragraph.

The nearest neighbor model sp3s* basis is typically represented as a 10x10 matrix. Since each anion (cation) only couples to the nearest cation (anion) the tight binding basis can be split into a 5x5 matrix with twice as many sites. Since the recursive Green function algorithm[2] scales linearly with the site number N and the matrix inversion on each site scales with the cube of the basis B a speed-up of $\Delta B^3/\Delta N=2^3/2=4$ can be achieved by splitting the basis and increasing the site number. This basis split is not possible in the second-nearest neighbor model anymore. The inclusion of an explicit spin basis within a second-nearest-neighbor model increases the basis matrix size on each site to 20x20. In a device of 50 spatial nodes the recursive Green function algorithm[2] now requires 50 matrix inversions of the size of 20x20, instead of 2x50=100 matrix inversions of the size 5x5. The number of operations needed per energy increases roughly by a factor of $4^3/2=32$. Another added difficulty associated with hole transport is the increased number of hole resonances compared to typical electron devices, thereby increasing the number of energy nodes in the simulation. Overall, the simulation of hole transport is numerically approximately 100-200 times more expensive to compute than electron transport. Such increased computing intensity warrants the parallelization of the NEMO code at the level of the energy integration, where each energy is independent of the other (since we do not include scattering here).

The central part of Figure 2 depicts the transverse subband dispersion of the hole RTD states at zero bias. The non-parabolicity, strong mixing, and anti-crossing of heavy- and light-hole states is evident. In addition two transmission coefficients as a function of energy for transverse momenta k=0 and k=0.05 are shown. The sharpness of the transmission coefficient curves with variations over 6 orders of magnitude highlights the need for a sophisticated gridding technique.

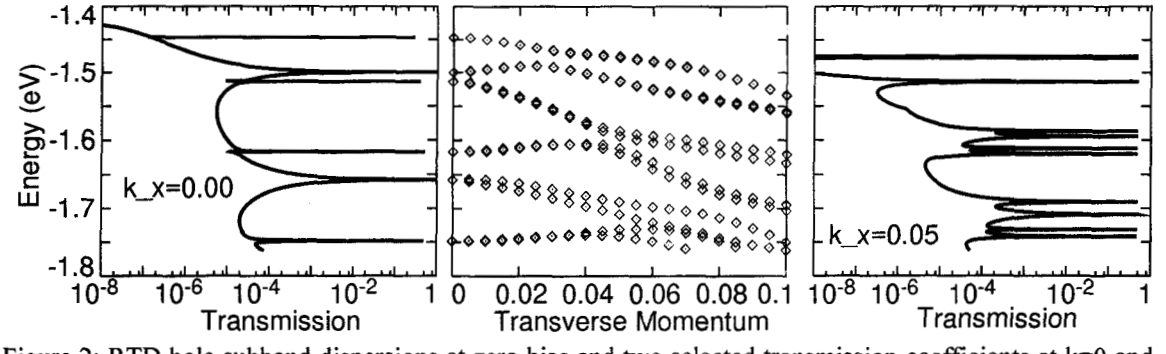

Figure 2: RTD hole subband dispersions at zero bias and two selected transmission coefficients at k=0 and k=0.05. The subband dispersion shows strong band mixing and anti-crossing. The transmission coefficients show a correspondingly strong dependence on transverse momentum k.

The transmission coefficient is a strong function of transverse momentum, which highlights the need for full band integration including an explicit treatment of the transverse momentum:

$$I \propto 2\pi \int kdk \int T(k,E)\left(f_L(E) - f_R(E)\right)dE \qquad (1)$$

Assuming parabolic transverse dispersion which results in

$$T(k,E) \approx T\left(k = 0, E - \frac{\hbar^2 k^2}{2m_0 m^*}\right) \qquad (2)$$

where m* is the parabolic transverse effective mass, equation (1) is often approximated as

$$I \propto \rho_{2D} \int T(k = 0, E_z)\left(F_L(E_z) - F_R(E_z)\right)dE_z \qquad (3)$$

where ρ_{2D} is the two-dimension density of states. This approximation, named after Tsu-Esaki, has been shown to be inaccurate for conduction band[5,6,8] and interband[9] RTDs. Since the Tsu-Esaki assumption relies on momentum independent transmission coefficients, which is shown not to be the case here (Figure 2), it can be expected that the Tsu-Esaki approximation breaks down for pure hole RTD transport as well.

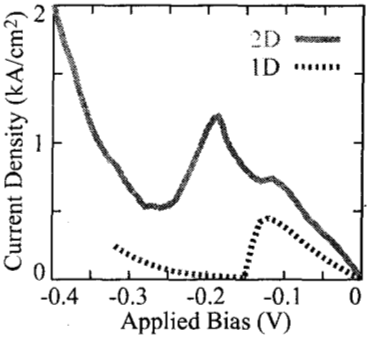

Figure 3: simulated hole RTD I-V characteristics. 2D simulation according to Eq. (1). 1D simulation according to Eq. (3) significantly underestimates the current.

Figure 3 compares two simulated I-V characteristics: one which includes the explicit transverse momentum integration (2D-integral, Eq. (1)) in the [100] direction and a simulation which assumes parabolic bands according to the Tsu-Esaki formula (1D-integral, Eq. (3)). Two qualitatively different I-V curves are obtained. The 1D integral approach does not account for all the transport channels throughout the device. In fact it can be shown that the current flow may occur off the Brillouin zone center in this device resulting in a significantly increased valley current.

Summary

We present the first NEMO simulations driven by a genetic algorithm to optimize parameters such as layer thicknesses and doping profiles. The convergence of the initially random population of devices to experimental specified device parameters is demonstrated. For hole transport devices a second-nearest neighbor sp3s* has been implemented. This model demonstrates the breakdown of the Tsu-Esaki approximation and the need for full band (2D) integration to achieve qualitatively correct answers.

Bibliography

1. NEMO, Nanoelectronic Modeling, in http://www.raytheon.com/rtis/nemo/.
2. R. Lake et al., J. Appl. Phys., **81**(12), 7845 (1997).
3. R. Lake et al., phys. stat. sol. (b), **204**, 354 (1997).
4. G. Klimeck et al., Appl. Phys. Lett., **67**(17), 2539 (1995).
5. G. Klimeck et al., IEEE DRC, 1997: p. 92.
6. R. C. Bowen et al., J. Appl. Phys., **81**, 3207 (1997).
7. D. Levine, http://www-unix.mcs.anl.gov/~levine/PGAPACK/index.html, Parallel Genetic Algorithm Library.
8. T. B. Boykin, R. E. Carnahan, and R. J. Higgins, Phys. Rev. B, **48**, 14232 (1993). T. B. Boykin, R. E. Carnahan, and K. P. Martin, Phys. Rev. B, **51**, 2273 (1995). T. B. Boykin, Phys. Rev. B, **51**, 4289 (1995).
9. M. S. Kiledjian, J. N. Schulman, K. L. Wang, and K. V. Russeau, Phys. Rev. B, **46**, 16012 (1992).