Simulation of a 2D pn junction in silicon thin film incorporating quantum transport for carriers

Shahriar Al Imam

A Thesis
in
The Department
of
Electrical and Computer Engineering

Presented in partial fulfillment of the requirements
for the degree of Master of Applied Science at
Concordia University
Montreal, Quebec, Canada

September 2006

Your file Votre référence
ISBN: 978-0-494-20739-0
Our file Notre référence
ISBN: 978-0-494-20739-0

ABSTRACT

Simulation of a 2D pn junction in silicon thin film incorporating quantum transport for carriers

Shahriar Al Imam

Silicon nanostructures have recently been a subject of interest demonstrating optimistic optical properties like luminescence. The scientific community predicts quantum effects to be the predominant cause for such optical properties of silicon nanostructures, hence it becomes prudent to pursue the roots of such reduced dimensional devices. With this view as a motive, a simulation model for a 2D thin film quantum confined 2D pn junction in silicon is developed in this work.

A thin film silicon layer is considered in the regime of strong confinement. A pn junction in such a film is considered so that the carriers are confined in thickness dimension while they are quantum mechanically transported along the device length. The transverse dimension in considered infinitely wide for plane wave approximation. For device simulation, after a careful study of various schemes to incorporate quantum effects (Van Dort model, Density Gradient Method, etc.) it was decided to use the more rigorous self-consistent Schrödinger-Poisson method. Keeping in mind the computational resource constraints, for problem formulation, decoupled 1D set of equations for carrier transport is deployed. For electrons, the well known single-band effective mass Hamiltonian is used while for holes, multi-band effective mass Hamiltonian with light and heavy holes is applied (though a full 6 band k.p Hamiltonian and spin orbit interaction is required to account for a full featured valence band, no effective work has been done to use such a formulation for a reduced dimensional device). Overall discretization is done using the finite element method with matrix representation of equations. The ohmic contacts in longitudinal direction are simulated with semi infinite open boundary contacts through self energy matrix, and broadening of energy states is incorporated. The simulation is done in Matlab as it gives the highest flexibility (in comparison to Silvaco, Femlab and C++ with the latter being unrealistically involved in numerical solution algorithm). For solution, instead of Wigner function or Green's function, a more direct wave-function perspective is taken. First the equilibrium condition was simulated, and then extension under externally applied voltage was carried out.

As the results show, confinement of carrier in lateral dimension results in energy quantization, and consequently subbands. As the material is degenerately doped, the number of carrier is comparable to existing states and carriers exist in excited states also. The occupancy of only three subbands upholds the earlier assumption that only a few subband are occupied. And along with the retention of subband shapes along device length, validates decoupling of the dimensions. The depletion region width is found to be more than that predicted by 3D junction equations. This may be due to the fact that the thin film cannot fully screen the electric field. Current voltage characteristics also do not show any significant tunneling current.

ACKNOWLEDGEMENT

I would like to take this opportunity to thank some of the people without whose support this report would not have been achieved.

I would first like to express my gratitude towards my supervisor Dr. Mojtaba Kahrizi, who has guided me towards this achievement. I had started research in this topic with Dr. Victor Rossokhaty, and I was stuck in the middle point when he passed away. It was Dr. Kahrizi who took the time and efforts to guide me successfully finish this project. As my supervisor, he always had time for me and my queries and his knowhow in the subject matter has refined me from time to time. His encouragement towards this topic and his support, both academic and otherwise has kept me aiming to excel. I remain ever grateful to him for giving me the opportunity to continue work in this field. If not for him, I doubt I would have come this far.

I would also like to thank the faculty of the department, specially the teachers with whom I have taken course. Besides Dr. Kahrizi, I would like to thank Dr. Katarzyna Radecka for being so supportive during my course with her and afterwards.

I would like to take a moment in remembrance of Dr. Victor Rossokhaty, with whom I had started this work. I would only state that whatever I have achieved in this research was instigated by him, and I pursue his dream in this work.

At this point I have to thank my family and friends, whose love and affection kept me going. I am grateful to my parents for believing in me and being there always for me, sheltering from whatever and whenever, and mention specially my wife, for being so supportive and understanding during this period. She had been my support and the source of my energy for pursuing this research, and beyond. She has been weathered in the process and still had the energy to absorb all the hardships and have a smile for me. Thanks to my daughter, who was born during this time, for providing me with the greatest joy of my life. Having her in my arms, and a look at her face has been worth living for.

Thank you all.

Shahriar Al Imam

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

2-DEG	: Two Dimensional Electron Gas
SOI	: Silicon On Insulator
IC	: Integrated Circuit
3D	: Three Dimension
2D	: Two Dimension
1D	: One Dimension
PL	: Photoluminescence
EL	: Electroluminescence
PS	: Porous Silicon
eV	: Electron Volt
meV	: Milli-Electron Volt
QW	: Quantum Well
DG	: Density Gradient
QHD	: Quantum Hydro Dynamic
NEGF	: Non-Equilibrium Green's Function
SP	: Schrödinger – Poisson
SCF	: Self Consistent Field
DOS	: Density Of States
FEM	: Finite Element Method
MOSFET	: Metal Oxide Semiconductor Field Effect Transistor
MOS	: Metal Oxide Semiconductor

LIST OF SYMBOLS

Symbol	Parameter
J	Current density
Q	Electronic charge
μ	Mobility
k_B	Boltzamnn constant
n	Electron density
ρ	Charge density
m	Mass of particle
U, V	System potential
π	Pi
γ	Source term
δ	Dirac delta
ε	Dielectric constant
λ	Wave length
ν	Frequency
ψ	Wave function
Γ	Broadening matrix
Σ	Self energy matrix
Φ	Basis function
φ	Expansion coefficient
T	Temperature
E_g	Band gap energy
E_f	Fermi energy
E_c	Conduction band edge
E_v	Valence band edge
N_d, N_a	Doping concentration
f_{FD}	Fermi-Dirac distribution
$F_{-1/2}$	Fermi integral
g(E)	Density of states
E_m	Subband energy
$\hbar$	Reduced Plank's constant
Λ	Bohm potential
G	Green's function
I	Identity matrix
H	Hamiltonian matrix
k	Wave vector

Chapter 1: Introduction

Optical and optoelectronic properties of silicon nanostructures have recently been a subject of interest. Though bulk silicon is an indirect semiconductor and generally unsuitable for optoelectronic devices, structures such as porous silicon or quantum dots have been reported to emit light. Due to the lack of sustainability and other problems like carrier injection, controlled direct recombination, etc. an effort is being given to understand the origin of such light generation and to explore the scope of using silicon nanowires or quantum wells for light emitting purposes. It is suggested that lack of long range order in nanostructures might result in quenching of the energy dispersion characteristics [1], and along with quantum confinement in one or more dimensions could result in direct recombination of carriers to generate light [7]. Other possible origins could be surface states, as surface defects may initiate localized direct recombination and high surface to volume ratio in reduced dimensional structures emphasizes such light generation. Some research has been done with lateral junctions in quantum wells (mainly directed in constructing approximate analytical models for the device) which demonstrates high-frequency operation and is suggestive of the possibility of light generation. Hence, instead of approximate analytical model, constructing a simulation model and exploring the possibility of direct recombination for such devices in silicon is in order.

1.1. Objective

The primary objective of this thesis work is to investigate a thin film silicon p-n junction incorporating quantum effects. As the thickness of the thin film is reduced, it essentially becomes a quantum device with 2D electron gas (2-DEG) having 1D confinement, and a quantum mechanical approach becomes necessary to model such a device. Models with approximate analytic analysis are only suggestive of features of such devices, so a numerical simulation model becomes necessary for a more comprehensive insight. Henceforth, a simulation model is constructed for such a device. Due to the requirement of unfeasibly huge computational resource required in direct descretized approach for a 2D device, it becomes prudent to search for an alternative method to develop a simulation model. So in the process of simulating such a device, in this work, decoupled set of 1D equation is applied to the 2D bipolar system.

1.2. Motivation

Recently conducted studies have resulted in the anticipation that a silicon nanostructure might be a sustainable source of light [8, 11, 17-18]. But with simple nanostructures it is yet hard to achieve reproducibility and the search for a nanodevice in silicon for light emission is on the way. In this regard, a thorough analysis of such reduced dimensional device is sought for. Thin films of silicon are now possible on silicon on insulator (SOI) structure, and if thickness is reduced

below 5nm then quantum confinement occurs in that direction and this may pave the way towards silicon optoelectronics.

Reporting of simulation results of a lateral p-n junction in quantum well in silicon is rare. Analytic studies on other material do indicate high-frequency operation of such a device. Other studies suggest that quantum effects could be the origin behind electroluminescence and light emission from silicon nanostructures. So a comprehensive simulation of a lateral p-n junction in silicon quantum well could help find a sustainable silicon light emitting structure.

1.3. Application

Most of the SOI structures are now being used for unipolar devices. But for photonics it becomes important to use a bipolar device for recombination of holes and electrons. In that regard, a thin film silicon p-n junction could offer a new possibility. Silicon light emitting device could find two very important uses:

1. Optical and electronic circuit integration to achieve integrated optoelectronics, which now is impossible as compound semiconductor alloys are used for optoelectronics.
2. Optical interconnect could replace slow electric interconnects, which remain a bottleneck towards high-speed integrated chips, which (the ICs) for a long time to come will continue to be fabricated in silicon.

1.4. Scope of thesis

In this thesis, a lateral p-n junction in a quantum confined thin layer of silicon, i.e. a quantum well is simulated. Studies of various schemes to incorporate quantum effects (Van Dort model, Density Gradient Method, etc.) has been made and finally decision was made to use the more rigorous self-consistent Schrödinger-Poisson method. Keeping in mind the computational resource constraints, for problem formulation, a decoupled 1D set of equations for carrier transport has been adopted. For electrons the well known single-band effective mass Hamiltonian was used to determine the transport behavior while for holes, after a careful analysis an appropriate effective mass Hamiltonian instead of the more involved multi-band approach without incorporating split orbit, but accounting for heavy & light hole effects was used. The usage of a 6-band or 4-band Hamiltonian to reduced dimensional system is still a topic of advanced research, and the implications are not yet assertive.

Overall, finite element method and matrix representation of equations with semi infinite open boundary contacts through self energy matrix was used. For programming, comparing Silvaco, Femlab, C++ and Matlab, Matlab was chosen as it gives the highest flexibility, without the unnecessary involvement into building libraries and algorithm to solve numerical formulation. For solution, instead of using Wigner function or Green's function method, wave-function perspective was chosen. Quantum confinement along the thickness was

incorporated, plane wave approximation in transverse direction and potential dependant state functions in the longitudinal direction is assumed. First simulation included equilibrium condition, and then it was extended for externally applied voltage. So the scope of the project remains to simulate p-n junction in silicon quantum well using a 2D carrier transport model.

1.5. Organization of thesis report

The thesis report is organized chapter wise and is outlined below:

Chapter 1: Introduction – in this chapter a prologue of the work done, the intensions, objectives, motivation, possible implications and the scope is briefed.

Chapter 2: Literature Review –this chapter presents the trend of research community toward silicon optoelectronics. A theoretical finding for the root of silicon luminescence and work done so far for achieving silicon electroluminescence is highlighted. Then work done in quantum well junctions and silicon nanostructures is also presented.

Chapter 3: Quantum Mechanics for Nanodevices – as nanodevices are reduced in dimension, quantum confinement transpires and a quantum mechanical approach becomes imperative for device analysis. This chapter presents some basics of quantum mechanics as used in nanodevices.

Chapter 4: Developing the model: Methodology – in this chapter explanation of the physical considerations made to develop a model for such a

quantum device is presented. Here for the first time (to the author's knowledge) decoupled set of 1D equation for simulating a 2D bipolar device is introduced.

Chapter 5: Formulating the model: Simulation –this chapter illustrates the intrigue details of the method of formulating the model for simulation.

Chapter 6: Simulation results – in this chapter, simulation findings are summarized.

Chapter 7: Conclusion – this concludes the report and highlights a directive for future work.

Chapter 2: Literature review

A considerable amount of work has been done in the field of silicon luminescence starting from theoretical investigation of the origin of such light emission to search for a sustainable light source in silicon. Also quite a number of silicon nanostructures and nanodevices have been studied. A few relevant with this research topic is presented herewith.

2.1 Light properties of silicon

2.1.1 Indirect to direct band gap material

Wherever there exists a system where electron and hole wave functions overlap in direct and reciprocal space, there exist a finite possibility of direct recombination and photon generation. But the luminescence strength and quantum efficiency of emission depends upon the extent of the overlap and transition probability. Using this viewpoint, scientists have engineered different approaches to convert an indirect material to a direct one [1].

1. In impurity assisted radiation, an impurity that has an energy level in the band gap of an indirect material is used as an intermediate state where the electrons and holes become localized and eventually result in radiative recombination. For this impurity level must be dispersed in k-space to be an efficient emission center. In silicon, impurities

usually act as indirect recombination center, but isoelectronic centers [2] and rare-earth elements [3] act as radiative center.

2. In alloys, the band structure itself is engineered to transform an otherwise indirect material to a direct material. Usually two or more groups IV elements (e.g. Si and Ge) are alloyed to shift the energy bands a little so that a direct transition becomes allowable.
3. Quantum confinement can also increase the probability of a direct recombination. Two such possibilities exist: in zone folding introduced by Gnutzmann and Clausecher [4], an ultra short period super lattice with periodicity comparable with lattice constant induces greater overlap of conduction band states at zone edge with valence band states at zone center, and hence increases the possibility of a direct recombination. On the other hand, for quantum confinement in one or more dimensions such overlap becomes innate. This is thought to be the case in porous silicon and its light emission phenomena. Qualitatively, the confinement of the carriers in real space causes their wave functions to spread out in momentum space, which increases the likelihood of strongly radiative transitions. In addition, scattering at the wire or dot boundaries, which in fact is a very likely occurrence, can supply the needed momentum more readily in a coned structure.

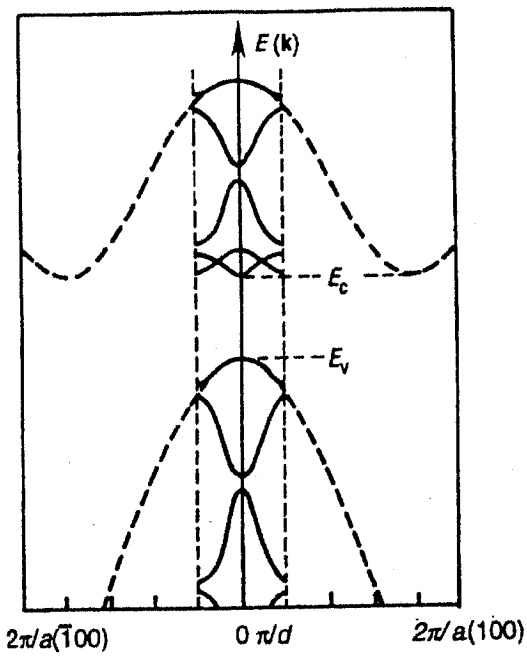


Figure 2.1: Zone folding resulting in overlapping of bands, converting an otherwise indirect bandgap material to a direct one [1].

4. Hybrid approaches, where a direct band gap material is grown on or joined with silicon.

2.1.2 Photo and electroluminescence in silicon

Vigorous research regarding light from silicon began with the discovery of efficient luminescence from porous silicon [5, 6]. Since then a number of studies have investigated the luminescence in silicon [6, 7]. Though most of the studies involved photoluminescence, a few considered electroluminescence also.

If an electron is excited form its original state to a higher state, it gains and stores potential energy. When the system relaxes, i.e. the electron comes back to its original state, the extra energy is emitted as photon and we get light coming out

of the material. This phenomenon is called luminescence. If the excitement of electron in the first place occurs due to an incident photon then the luminescence is called photoluminescence (PL). The phenomenon is interesting in itself as to finding the origin of photoemission from the material. But for an electronic light emitting device, this becomes somewhat irrelevant as the excitation is provided by light, not electrons. If an electron is excited by means of electric stimulants, then the resulting photoemission is called electroluminescence and this is the phenomena that one should be looking for obtaining a light emission devices.

Silicon nanocrystals have recently attracted attention as being possibly a strong candidate to realize efficient silicon based light source operating at room temperature [8]. The band gap in this case is enlarged with respect to bulk silicon due to quantum confinement effects and intense visible PL at room temperature is obtained [9, 10]. In crystalline silicon nanoparticles, the wave function of electrons and holes have been found to be delocalized over the corresponding energy range, which is dependant on the nanoparticle size [11]. The photoluminescence is studied and the PL spectra have a blue shift suggestive of quantum confinement of excitons, which proves that quantum confinement in silicon is a source of possible light emission. Blue shift of PL energy in amorphous silicon nanowells have been observed to occur with a decrease in well thickness [12, 13]. Silicon quantum dots and wells have been studied for visible luminescence [14]. It is found that the PL spectra strongly depend on well thickness.

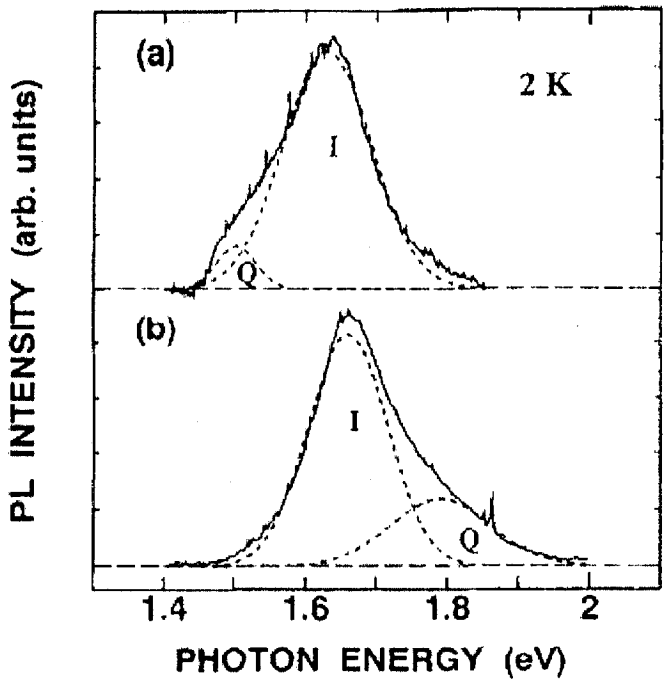


Figure 2.2: photoluminescence phenomena as found in silicon QW [14].

Theoretical calculations of thickness-dependence on bandgap energy for 2D silicon wells [15, 16] have been carried out. One of the band-peak energy is roughly consistent with theoretical calculations based on the quantum confinement model. The results imply that emission may have been caused by radiative recombination in silicon well [14].

Due to problems such as electrical pumping of nanostructures embedded in an insulating matrix, very little work has been done on silicon electroluminescence, but reports have been found for the existence of such occurrence [8, 17, 18] in silicon. This obviously makes the case for silicon as a prospective candidate for light emission device.

Chen et al. fabricated pn homojunctions between heavily doped n-type porous silicon (PS) and p type PS and formed n^+p^+ junction with phosphorous and boron dopants [19]. A standard mesa etch procedure was used to confirm a mesa etched structure. Steiner et al. also fabricated PS pn homojunction [20] but this time without etching mesa structures. In both cases EL was demonstrated.

2.1.3 Relaxation of dispersion relation in silicon

Crystalline silicon is not a source of efficient light emission at room temperature, as its band structure has an indirect band gap of ~1.12eV and it shows a small exciton binding energy of about ~15 meV. In spite that a promising approach to overcome the indirect nature of the material is the relaxation of the k-selection rule due to the spatial confinement in low dimensional nanostructures [21]. As the particles get confined in space, due to uncertainty principle, their momentum gets delocalized and the dispersion relation is relaxed, thereby creating a possibility of direct recombination.

There are several theoretical studies on size dependent decay times carriers in silicon nanocrystals [22, 23], but no or very few studies on the decay times of silicon quantum well is comprehensively reported.

In a study [24], the authors have investigated theoretically the optical properties of free standing silicon quantum wires, using a realistic empirical tight binding model. The excitonic effects are included within the effective mass approximation which is found to be appropriate. It was seen that the exciton

oscillator strengths for quantum wires with small sizes can be as large as that for a direct semiconductor such as GaAs.

The excitonic effects in silicon quantum wires using simple two-band effective mass model with parabolic electron and hole bands has been investigated [25]. The electron and hole masses were obtained by fitting parabolic bands to the valence and conduction band extrema. This gives an indication that using effective mass approximation rather than Lűttinger-Kohn Hamiltonian for a reduced dimensional structure is more appropriate, or at least safer.

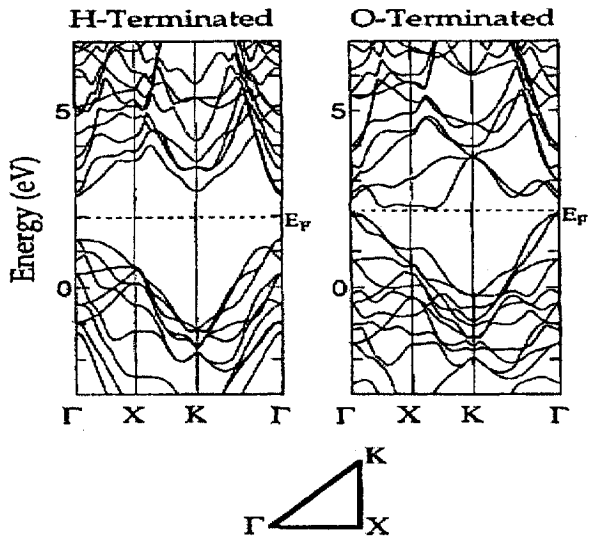


Figure 1.3: Band dispersion relations of the H-terminated and the O-terminated silicon slabs (quantum well) showing direct band gap nature at Γ point [15]

2.2 Lateral junction in Quantum well (QW):

Silicon based inter-subband lasers have been studied [26]. It has been found that phonon confinement reduces the scattering rate by a factor of two to four as compared to the case of no confinement. Thus an enhancement of a factor of two to four would be expected in the inter-subband lifetimes and their differences. This

feature, in addition with the inherently weaker nonpolar optical scattering makes silicon based QW structures favorable candidate for inter-subband lasers [27, 28].

A high frequency lateral junction photodiode is reported in GaAs quantum well [29]. An analytic model is developed using the sheet concentrations of electron and holes. The well width is 25 nm so strong quantum confinement is absent, and a non quantum approach has been taken for the analytic model.

The idea of a two dimensional p-n junction formed as a contact between two regions of quantum dimensional film with different types of conductivity is proposed [30]. In this approach, a conformal mapping technique has been adopted to propose an approximate analytic model of such a quantum well junction. In this study also, no quantum transport mechanism is considered due to the width of the well which is well beyond strong confinement regime. These two approaches do not specifically deal with a particular material, so the same procedure could be extended for any suitable material, in this case, to silicon thin films.

2.3 Models of quantum devices

Incorporating quantum effects has become more significant in recent years, as fabrication technology reaches nano regime. But existing computational tools mostly lack this incorporation, and various methods have been devised to achieve this incorporation. The ones that do incorporate this effect is mostly for dealing with the transitional devices which experiences quantum effects as parasitic distortions to otherwise classical formulation. So race for a suitable industry grade

quantum device simulator is on. Due to huge computational overhead required for a full quantum mechanical approach, methods have been developed to bypass such extensive approaches and use some fitting techniques or use the semi-classical models for incorporating these quantum effects. The trick remains to achieve better accuracy to the actual device, with reduced computational load. Comparison of the methods has been reported [31] and a few major methods are briefed here.

2.3.1 The density gradient (DG) method

The density gradient method is an approach compatible with the drift-diffusion treatment used in device simulators [32]. Different methods have been proposed, one of them is presented here. It applies a quantum potential correction in the density current expression:

$$\vec{J}_n = qD_n\vec{\nabla}n - qn\mu_n\vec{\nabla}(\Psi - \Lambda) - \mu_n n k_n T\vec{\nabla}\ln n_{ic} \qquad (2.1)$$

$$\Lambda = -\frac{\gamma\hbar^2}{6m}\frac{\nabla^2\sqrt{n}}{\sqrt{n}} \qquad D_n = \frac{k_n T}{q}\mu_n$$

The factor γ has been introduced to adjust the quantum correction which has been obtained after a few simplifications. In this way it accounts only one mass in DG model. It could also be adjusted depending on operating temperature and device structure (bulk, SOI, double gate). Concerning the boundary conditions, they are the same as in a semi-classical scheme, only that at contacts, the quantum correction is zero.

2.3.2 Quantum Hydro-Dynamic model (QHD)

Like DG method, many formulation of this model is also available [33]. One simple illustrative presentation looks as follows: for a spatially independent effective mass in 1D, equations are solved for carrier density ρ, angular momentum p and average energy W. A full mathematical QHD description including three moments for electron and holes, and requires seven equations (including Poisson's equation). The quantum correction comes from Wigner function-corrected BTE with incorporating the Bohm potential [34] Q.

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}\left(\frac{\rho p}{m}\right) = 0 \tag{2.2a}$$

$$\frac{\partial (\rho p)}{\partial t} + \frac{\partial}{\partial x}\left(\frac{\rho p^2}{m}\right) + \frac{\partial}{\partial x}\left(U + \frac{Q}{3}\right) + \frac{\partial}{\partial x}(\rho kT) = \left(\frac{\partial (\rho p)}{\partial t}\right)_C \tag{2.2b}$$

$$\frac{\partial W}{\partial t} + \frac{\partial}{\partial x}\left(\frac{pW}{m}\right) + \frac{\partial}{\partial x}\left(\frac{\rho pkT}{m}\right) + \frac{\rho p}{m}\frac{\partial}{\partial x}\left(U + \frac{Q}{3}\right) - \left(\frac{\rho \hbar^2}{12\,m}\right)\frac{\partial}{\partial x}\left(\frac{1}{\rho}\frac{\partial \rho}{\partial x}\right)\frac{\partial}{\partial x}\left(\frac{p}{m}\right) = \left(\frac{\partial W}{\partial t}\right)_C \tag{2.2c}$$

$$W = \frac{3\rho kT}{2} + \frac{\rho p^2}{2m} - \left(\frac{\rho \hbar^2}{24\,m}\right)\frac{\partial}{\partial x}\left(\frac{1}{\rho}\frac{\partial \rho}{\partial x}\right) \tag{2.2d}$$

2.3.3 Van Dort model:

This model was developed by Van Dort et al. and has been quite successful in describing devices which include triangular potential well [35]. The key feature is to concede that quantum confinement changes bandgap and hence the surface wave function of electrons, and eventually the energy and carrier concentration.

$$\Psi_S^{QM} = \Psi_S^{CONV} + \Delta\varepsilon / q + E_n \Delta z \tag{2.3a}$$

$$E_g^{QM} = E_g^{CONV} + \tfrac{13}{9}\Delta\varepsilon \tag{2.3b}$$

$$\Delta\varepsilon \approx \beta\left(\frac{\varepsilon_{Si}}{4\,qkT}\right)^{1/3} \max\left(E_n(0),0\right)^{2/3} \tag{2.3c}$$

$$n_i^{QM} = n_i^{CONV} \exp\left[\left(E_g^{QM} - E_g^{CONV}\right)/2\,kT\right] \tag{2.3d}$$

2.3.4 Self-consistent Schrödinger-Poisson model:

This model is a more accurate model of the actual quantum mechanical phenomena as it does not approximate any correction potential, rather self-consistently solves the basic Schrödinger's equation with Poisson's equation. The only semi-classical point is the effective mass and the extension of a single electron wave function to many-body theorem. One particular formulation of the method adopted in SILVACO simulator is given below.

$$n(r)=\frac{2k_BT}{\pi\hbar^2}\{\sqrt{m_lm_{t1}}\sum_i|\Psi_{li}|^2\ln\left[1+\exp\frac{E_F-E_{li}}{k_BT}\right]+\sqrt{m_lm_{t2}}\sum_i|\Psi_{t1i}|^2\ln\left[1+\exp\frac{E_F-E_{t1i}}{k_BT}\right] \quad \textbf{(2.4a)}$$
$$+\sqrt{m_{t1}m_{t2}}\sum_i|\Psi_{t2i}|^2\ln\left[1+\exp\frac{E_F-E_{t2i}}{k_BT}\right]\}$$

$$\nabla\cdot(\varepsilon\nabla U)=q^2[N_D-N_A-n(x)] \quad \textbf{(2.4b)} \qquad n(r)=\sum_{\infty}|\Psi_\alpha(r)|^2\cdot f_0(\varepsilon_\alpha-E_F) \quad \textbf{(2.4c)}$$

Self-consistent model is vastly used to analyze the behavior of nanoscale devices. It has long been adopted for n-type inversion layer in a silicon MOS structure for energy levels, population and charge distribution analysis [36]. The same self-consistent approach has been also assumed [37] to describe a cylindrical quantum wire for considering electronic confinement.

2.3.5 Non Equilibrium Green's Function (NEGF) formalism

For quantum mechanical modeling of nanodevices, two formalisms are often used to avoid the huge computational burden of direct solution of the Schrödinger-Poisson approach. One is the Wigner Equation, commonly solved by finite difference method. A recent study [38] of this equation with Monte Carlo

method has shown to resolve both quantum interference and dissipation effect due to scattering. This method has been compared with NEGF formalism based simulator [39], and found to be equally effective for solving resonant tunneling diodes. This indicates that NEGF is an alternative choice to solve S-P method in quantum devices.

Chapter 3: Quantum Mechanics for Nanodevices

The purely mathematical postulations that make quantum mechanics the tool for actually describing nature is too intricate and exceed the scope of this report, hence only a brief ideological perception of some of quantum mechanical concepts as applied to nanodevice is presented here. The more used Schrödinger's version of the mechanics is adopted as it gives a better insight to the subject, and provides for a practical means of modeling a nanodevice.

3.1 Energy, density and current operators

The operator associated with the system energy is called Hamiltonian. Hamiltonian contains operations associated with kinetic and potential energies and for a particle in one dimension can be written as:

$$H_{operator} = \frac{-\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x) \tag{3.1}$$

Operating on wave function with Hamiltonian produces the Schrödinger's equation. In time independent Schrödinger's equation, the operation may produce specific values for energy called energy eigenvalues. This situation can be shown in the form

$$H_{op} \cdot \psi_i = E_i \psi_i \tag{3.2}$$

Where the specific values of energy are called energy eigenvalues and the functions are called eigenfunctions.

Two other operators are of immense importance to electrical engineers, the electron density operator and the current operator. It is regarded that wave function of a particle is associated with probability of finding that particle in a particular place at a particular instance of time. In fact, the absolute magnitude of squared wave function gives the probability density. Since the wave function is complex in general, the probability of a particle is given by

$$probability \propto |\psi|^2 = \psi\psi^* \tag{3.3}$$

From the requirement that wave function has to be normalized, one gets,

$$\int \psi\psi^* d^3r = 1 \tag{3.4}$$

For many body system, if a certain particle follows a certain distribution (such as electrons follow Fermi-Dirac distribution) then the number of that particle in a particular state is given by

$$n(E_m) = |\psi_m|^2 \cdot f(E_m) \tag{3.5}$$

So the overall particle concentration is

$$n = \int_{-\infty}^{+\infty} n(E_m) dE = \sum_{m=1}^{\infty} |\psi|^2 \cdot f(E_m) \tag{3.6}$$

For the current operator, one way is to start from consideration of current conservation, more popularly known as the continuity equation postulating that the change in charge density ρ is related to the divergence of current density J through

$$\frac{\partial}{\partial t}\rho(r,t) = -\nabla \cdot J(r,t) \tag{3.7}$$

Here it is seen that time dependence of charge density (temporal dependence of a scalar field) is related to net current into or out of a region of space (spatial dependence of a vector field). For a particle of charge '*e*' one can identify that $\rho = e|\psi|^2$, so that

$$\frac{\partial \rho}{\partial t} = e\frac{\partial |\psi|^2}{\partial t} = e\frac{\partial (\psi\psi^*)}{\partial t} = e\left(\psi^* \frac{\partial \psi}{\partial t} - \psi \frac{\partial \psi^*}{\partial t}\right) \tag{3.8}$$

Now from Schrödinger's equation,

$$i\hbar \frac{\partial}{\partial t}\psi(r,t) = \left(-\frac{\hbar^2}{2m}\nabla^2 + V(r)\right)\psi(r,t) \tag{3.9}$$

Multiplying both sides by $\psi^*(r,t).e/i\hbar$ from left,

$$e\psi^*(r,t)\frac{\partial}{\partial t}\psi(r,t) = ei\frac{\hbar}{2m}\psi^*(r,t)\nabla^2\psi(r,t) - e\frac{i}{\hbar}\psi^*(r,t)V(r)\psi(r,t) \tag{3.10}$$

Thus the first term of charge change rate is found. Taking the complex conjugate of the Schrödinger's equation and multiplying both sides by ψ(r,t) from left results in the second term, finally arriving at:

$$\begin{aligned}\frac{\partial \rho}{\partial t} &= \psi^*(r,t)ei\frac{\hbar}{2m}\nabla^2\psi(r,t) - \psi(r,t)ei\frac{\hbar}{2m}\nabla^2\psi^*(r,t) + \frac{e\psi\psi^*}{i\hbar}(V(r) - V(r)) \\ &= ei\frac{\hbar}{2m}\nabla\cdot\left(\psi^*\nabla\psi - \psi\nabla\psi^*\right) \\ &= -\nabla\cdot J\end{aligned} \tag{3.11}$$

So finally one finds,

$$J = -ei\frac{\hbar}{2m}\left(\psi^*\nabla\psi - \psi\nabla\psi^*\right) \tag{3.12}$$

3.2. 3D Schrödinger's equation

Usually solving the 3D Schrödinger's equation is quite engaging and requires large calculations as a general system is coupled within itself in different dimensions. But if the coordinates are separable then separating the 3D equation into three sets of decoupled 1D equation is possible. Solving these three 1D equations independently one could arrive at the final solution. This for example is possible if potential can be separated into an x-, y- and z-dependant part such that:

$$U(\vec{r}) = U_x(x) + U_y(y) + U_z(z) \tag{3.13}$$

Then the wave function can also be written in product form:

$$\psi(\vec{r}) = X(x) + Y(y) + Z(z) \tag{3.14}$$

Where each of the right hand functions is obtained by solving a separate one dimensional Schrödinger's equation:

$$E_x X(x) = \left(-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + U_x(x) \right) X(x) \tag{3.15a}$$

$$E_y Y(y) = \left(-\frac{\hbar^2}{2m}\frac{d^2}{dy^2} + U_y(y) \right) Y(y) \tag{3.15b}$$

$$E_z Z(z) = \left(-\frac{\hbar^2}{2m}\frac{d^2}{dz^2} + U_z(z) \right) Z(z) \tag{3.15c}$$

The total energy E is equal to the sum of energies associated with each of the three dimensions: $E = E_x + E_y + E_z$.

In some problems it is seen that the quantities may not be separable in Cartesian coordinates but could be separable in cylindrical or spherical

coordinates. For example, the potential in a hydrogen atom $U(\vec{r}) = -q^2/4\pi\varepsilon_0 r$ cannot be separable in *(x,y,z)*, but it is separable in (r,θ,Φ) and the wave function may be written in the form:

$$\psi(r,\theta,\phi) = [f(r)/r]Y_l^m(\theta,\phi) \tag{3.16}$$

Here the radial wave function *f(r)* is obtained by solving the radial Schrödinger's equation:

$$Ef(r) = \left(-\frac{\hbar^2}{2m}\frac{d^2}{dr^2} + U(r) + \frac{l(l+1)\hbar^2}{2mr^2}\right)f(r) \tag{3.17}$$

So it is evident that symmetry plays an important role in solving 3D Schrödinger's equation, when it implies the prospect of separating the coordinates.

3.3 Multi-electron picture

The Schrödinger's equation that is given above is for a one particle system. But most of the real life devices have a large number of electrons and so one has to incorporate one electron's interaction to rest of the particles, which makes a direct implementation of an accurate system of equations virtually impossible to implement. Considering a system of only 2 electrons, a helium atom, calculation of eigenstates requires solving the two-electron Schrödinger's equation:

$$E\Psi(\vec{r}_1,\vec{r}_2) = \left(-\frac{\hbar}{2m}\nabla^2 + U(\vec{r}_1) + U(\vec{r}_2) + U_{ee}(\vec{r}_1,\vec{r}_2)\right)\Psi(\vec{r}_1,\vec{r}_2) \tag{3.18}$$

Where r_1 and r_2 are the coordinates of the two electrons and U_{ee} is the potential energy due to their mutual repulsion, $U_{ee}(\vec{r}_1,\vec{r}_2) = e^2/4\pi\varepsilon_0|\vec{r}_1 - \vec{r}_2|$. This is

more difficult to solve than the one-electron Schrödinger's equation, but not impossible. However, this approach quickly becomes inconceivable as size is increased to mesoscopic system with numerous electrons of multilevel energy.

Two general approaches are there to account for such systems. One is to consider one-electron system at a time, and for numerous such considerations, arrive at the multi-particle picture. This single particle approach is taken in Monte Carlo simulations. But the drawback is that it cannot account for any inter-particle interaction. The other approach is to analyze a given system from the collective viewpoint. The use of multi particle statistics and transport equations is adopted to arrive at the final solution. This approach approximates the interaction among particles better, but becomes an exceedingly involved.

One particular approach, the 'Self Consistent Field' Method, which is widely used in this field employs the idea of considering only one electron instead of calculating all the interactions among different electrons and the effect of rest of the system is considered in the so called Self Consistent Field (SCF), and one adopts a proper SCF method to solve the system.

Much work has been done in this SCF method and many sophisticated versions have been developed over the years. It is an amazingly well picture which provides a reasonably accurate description of the multi-particle system. There is no mathematically convincing way to prove the accuracy of this SCF method, but it describes every atom of the periodic table to such extent that people have come to rely heavily on this method.

3.4. Self Consistent Field Method: Mesoscopic Systems

In SCF, one calculates the wave function of a particle, from an equivalent Schrödinger's equation for single electron, and incorporates the effect of rest of the system to a semi-classical parameter called the effective mass, and uses a self consistent field, which is so called because it is calculated self-consistently from electrostatic viewpoint of the system. For submicron systems, the SCF method translates to the following:

1. Guessing an initial potential for the system, and calculating the state functions of the relevant particle by solving the single particle effective mass Schrödinger's equation.
2. From the state functions, calculating the particle density using appropriate multi-particle statistics (e.g. Fermi-Dirac distribution or Fermi Integral for electrons and holes).
3. Using this particle density, solving Poisson's equation for finding the potential of the system.
4. Computing steps 1 to 3 self-consistently until convergence is achieved.
5. Calculating other required parameters, once self-consistency has been achieved.

As the system dimensions get smaller and smaller, other quantum effects like confinement and hot electron effect starts to emerge as significant deviations

and single electron charging effect may also become important. The dimensional mismatch between bulk contacts and device could also play a part. Level broadening within the device due to contacts becomes predominant. For the regime where single electron charging energy is comparable to level broadening and the thermal energy of the system particles, SCF method is a good approximation. In fact it provides one of the most accurate analyses in this regime.

3.5. Hole dispersion relation

The approximation that has been used so far for finding the dispersion relation of electron is effective mass approximation and used as single-band effective mass Hamiltonian to describe the particle. This approximation holds good for bulk semiconductor conduction electrons, where indeed the effect of the whole crystal could be coupled into a semi-classical concept of effective mass and the dispersion relation can be approximated as a parabola in the range of interest.

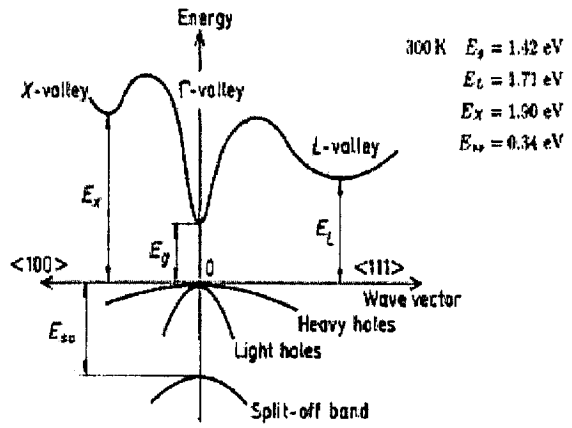


Figure 3.1: Band diagram of a direct band gap material showing the valence band dispersion relation.

The dispersion relation of a hole is not so straight forward. The Schrödinger's equation only deals with the non-relativistic physics of a system. For electrons traveling at high velocities, relativistic effects can become significant and requirements demand use of the Dirac equation instead. Typically in solids the velocities are not high enough to require this, but the electric fields are very high near the nuclei of atoms leading to weak relativistic effects that can be accounted for by adding a spin-orbit correction to the Schrödinger's equation. Then again the optical properties of the hole at the top of the valence band gives rise to heavy and light hole effects, making single band effective mass Hamiltonian futile for hole dispersion relation, and a multi-band effective mass equitation, sometimes called the full band approach is called for. This gives a rather complicated dispersion relation for hole in the valence band.

3.6. Density of States

Electrical engineers eventually wish to describe the current-voltage characteristics of semiconductor devices. Since current is due to the flow of charge, an important step in the process is to determine the number of electrons and holes in the semiconductor that will be available for conduction. Since by Pauli's exclusion principle, only one electron can occupy a given quantum state, the number of carriers that can contribute to the conduction process is a function of the number of available energy or quantum states. It should be noted that in forming bands in solid, the band of allowed energies are actually made up of

densely placed discrete energy levels. So it becomes an important parameter how many states are there with a particular energy per unit volume of the solid. This density of electronic quantum energy states are called density of states (DOS). As in 3D there might be degenerate states with the same energy but different wave vector k, it then is only logical to calculate the density of state in k-space and then convert it into energy space for a more accurate result.

3.6.1 Bulk density of states

If one considers a particle trapped in a 3D potential well given by

$$\begin{array}{lll} V(x,y,z)=0 & & 0<x<a,0<y<a,0<z<a \\ V(x,y,z)=\infty & \text{for} & elswhere \end{array}$$

Using the separation of variable technique the Schrödinger's equitation can be solved to give

$$\frac{2mE}{\hbar^2}=k^2=k_x^2+k_y^2+k_z^2=\left(n_x^2+n_y^2+n_z^2\right)\left(\frac{\pi^2}{a^2}\right) \qquad \textbf{(3.19)}$$

Where n_x, n_y and n_z are integers. As negative integers yield the same wave function with a negative value but the same probability, as long as quantum states are concerned, only positive integers are needed for a full description of all possible k values. Now if a 3D k-space is defined, then only one quadrant of positive k_x, k_y and k_z components make up possible k states. As the quantum states are discrete, distance between two states in k-space in a particular dimension is:

$$k_{x+1}-k_x=(n_x+1)\left(\frac{\pi}{a}\right)-(n_x)\left(\frac{\pi}{a}\right)=\left(\frac{\pi}{a}\right) \qquad \textbf{(3.20)}$$

Generalizing for 3D, the volume V_k for a particular k state is $V_k=(\pi/a)^3$. The density of quantum states in k-space can now be determined. A differential volume in k-space is given by $4\pi k^2 dk$. So the differential density of quantum states in k-space is given by:

$$g_T(k)dk = 2\left(\tfrac{1}{8}\right)\frac{4\pi k^2 dk}{\left(\pi/a\right)^3} = \frac{k^2 dk}{\pi^2}a^3 \qquad (3.21)$$

Converting the states into energy space,

$$k^2 = \frac{2mE}{\hbar^2} \Rightarrow k = \frac{\sqrt{2mE}}{\hbar} \quad \text{And the differential} \quad dk = \frac{1}{\hbar}\sqrt{\frac{m}{2E}}dE$$

Putting these values into the density of state equation in k-space,

$$g_T(E)dE = \frac{a^3}{\pi^2}\left(\frac{2mE}{\hbar^2}\right)\cdot\frac{1}{\hbar}\sqrt{\frac{m}{2E}}dE = \frac{4\pi a^3}{h^3}(2m)^{3/2}\cdot\sqrt{E}dE \qquad (3.22)$$

So the density of states in energy space becomes:

$$g(E) = \frac{4\pi(2m)^{3/2}}{h^3}\sqrt{E} \qquad (3.23)$$

3.6.2 Density of states in thin films: 2D structures

In very thin films, the quantum confinement is in one dimension, in the direction of film thickness and one effectively deals with a case of quantum well. In the confinement dimension, the energy becomes quantized to form discrete subbands far apart in energy from one another, so it can no longer be assumed to be a band in that dimension. Suppose this dimension is x direction, giving effectively a 2D Electron Gas which is free to move in any direction in the y-z

plane but are confined in the x dimension so that it has discrete k_x values and hence E_x values. For the other dimensions' contribution to the total energy, a similar strategy as before can be deployed, only that this time dealing in in-plane energy would result in:

$$E = E_x + E_p \rightarrow E_p = E_y + E_z \text{ such that } \frac{2mE}{\hbar^2} = k^2 = k_y^2 + k_z^2 = \left(n_y^2 + n_z^2\right)\left(\frac{\pi^2}{a^2}\right)$$

This is now effectively a 2D system, giving a 2D plane of k values with a differential area of $S_k=(\pi/a)^2$ associated with each quantum state:

$$g_T(k)dk = 2\left(\tfrac{1}{4}\right)\frac{2\pi k dk}{\left(\pi/a\right)^2} = \frac{kdk}{\pi}a^2 \tag{3.24}$$

Again converting the k-space into energy space,

$$g_T(E)dE = \frac{a^2}{\pi}\sqrt{\frac{2mE}{\hbar^2}}\cdot\frac{1}{\hbar}\sqrt{\frac{m}{2E}}dE = \frac{4\pi m a^2}{h^2}dE \tag{3.25}$$

So the resultant density of states is given by

$$g(E) = \frac{4\pi m}{h^2} \tag{3.26}$$

3.6.3 Density of state in nanowire: 1D structure

A nanowire is a one dimensional structure such that the two transverse dimensions confine the electrons in such a way that the only dimension the electron has a band nature left is the longitudinal direction. In the other two dimensions, its energy is quantized and has discrete subband values, giving rise to

2D subbands, while in the longitudinal direction, it has still the dispersion relation given by

$$\frac{2mE}{\hbar^2} = k^2 = k_z^2 = \left(n_z^2\right)\left(\frac{\pi^2}{a^2}\right) \tag{3.27}$$

So for this 1D system which ultimately results in a length association with every quantum state $L_k=\pi/a$:

$$g_T(k)dk = 2\left(\tfrac{1}{2}\right)\frac{dk}{\left(\pi/a\right)} = \frac{dk}{\pi}a \tag{3.28}$$

Converting the k-space to energy space,

$$g_T(E)dE = \frac{a}{\pi}\cdot\frac{1}{\hbar}\sqrt{\frac{m}{2E}}dE = \frac{a}{h}\sqrt{\frac{2m}{E}}dE \tag{3.29}$$

So finally the density of states can be written as

$$g(E) = \frac{1}{h}\sqrt{\frac{2m}{E}} \tag{3.30}$$

3.6.4 Density of states in Quantum Dots: 0D system

A quantum dot represents a quantum mechanical potential well in three dimensions for carriers (electrons and/or holes). It means that the electronic states are discrete, like that of an atom. This leads to a number of properties that are significantly different from quantum wells. These differences arise from the discrete density of states, as opposed to the continuous density of states in quantum wells. For example, state blocking effects (due to Pauli's exclusion principle) play an important role in carrier dynamics in quantum dots. Also

temperature dependence is very different since carriers can only be thermally excited to a limited number of well-separated excited states. Thermal excitation of carriers cannot therefore homogeneously broaden the states. The homogeneous broadening, which is then due mainly to interactions with phonons, is negligible especially at low temperatures. So the density of states results in discrete values of energy.

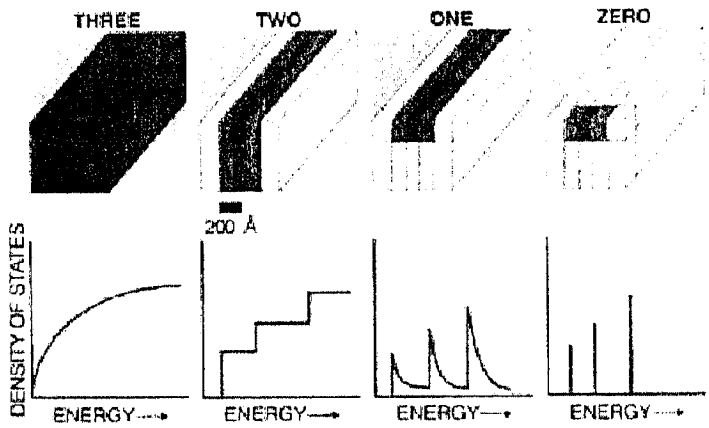


Figure 3.2: Density of states of different dimensional systems.

Chapter 4: Developing the model: Methodology

The problem of developing a theoretical model for a 2D pn junction device requires solution of some inherently intrigued problems. In order to address all the relevant issues in a proper acceptable manner, the problem at hand is first divided into smaller problems, and then each of the problems are dealt with individually. Consolidating all the solution in an integral manner provides the final model of the device. In that process, the methodology adopted is illustrated in this chapter. First it was decided better to go for a numerical analysis rather than an analytic one, as in real devices, the numerical solution tend to give more precise results and better describe the involved physics. Among many formulations at hand for numerical analysis, the very well known Finite Element Method is deployed as this gives the most straight forward physical description of the modeling, then the transport of carriers are quantum mechanically analyzed to give a comprehensive numerical model for such abstraction. Afterwards, the solution procedure is discussed and Green's Function method is adopted only for proper incorporation of the effects of boundary.

4.1. Numerical analysis

Numerical analysis is the study of algorithms for the problems of continuous mathematics (as distinguished from discrete mathematics) using basic arithmetical operations like addition. Some problems it deals with arise directly

from the study of calculus; other areas of interest are real variable or complex variable questions, numerical linear algebra over the real or complex fields, the solution of differential equations, and other related problems arising in the physical sciences and engineering.

Many problems in continuous mathematics do not possess a closed-form solution. In these situations, one has two options left: either one tries to find an approximate solution using asymptotic analysis or one seeks a numerical solution. Some problems can be solved exactly by direct methods. However, no direct methods exist for most problems. In such cases it is sometimes possible to use an iterative method. Such a method starts from a guess and finds successive approximations that hopefully converge to the solution. Even when a direct method exists, an iterative method may be preferable because it is more efficient or more stable.

Sometimes, continuous problems must be replaced by a discrete problem whose solution is known to approximate that of the continuous problem; this process is called discretization. For example, the solution of a differential equation is a function. This function must be represented by a finite amount of data, for instance by its value at a finite number of points in its domain, even though this domain is a continuum. So discretization becomes key point for resource management.

The algorithms of numerical analysis are routinely applied to solve many problems in science and engineering. In fact, almost all supercomputers are continually running numerical analysis algorithms.

As a consequence, efficiency plays an important role and a heuristic method may be preferred above a method with a solid theoretic foundation because it is more efficient. Generally, numerical analysis uses empirical results of computation runs to probe new methods and analyze problems, though it of course also employs mathematical axioms, theorems and proofs.

4.2. FEM formalism

The most important factor in numerical solution to Schrödinger's equation is the formation of Hamiltonian matrix. Once done that, the rest part becomes rather trivial to solve. It so happens that a quantum system about always constitutes of a number of different energy eigenstates, which in the end gives rise to its own set of system equations. This results effectively in a system of equations , often linear ones, hence linear algebra comes into play. The Hamiltonian in quantum mechanics becomes a matrix for the whole state space and finding the diagonal representation of Hamiltonian remains the major problem. An example would better explain the situation.

When dealing with a 1D infinite quantum well where an electron is trapped inside, the electron energy will be quantized and it will have a finite probability distribution of occupancy in different eigenstates. But to describe the whole

system, all the eigenstates has to be considered. Each eigenstate gives rise to its own wave function which is a solution of the Schrödinger's equation.

For eigenstate 1: $$\left\{-\frac{\hbar^2}{2m_x^*}\frac{\partial^2}{\partial x^2}+U(x)\right\}\Psi_1(x)=E_1\Psi_1(x) \quad (4.1a)$$

For eigenstate 2: $$\left\{-\frac{\hbar^2}{2m_x^*}\frac{\partial^2}{\partial x^2}+U(x)\right\}\Psi_2(x)=E_2\Psi_2(x) \quad (4.1b)$$

.....

For eigenstate n: $$\left\{-\frac{\hbar^2}{2m_x^*}\frac{\partial^2}{\partial x^2}+U(x)\right\}\Psi_n(x)=E_n\Psi_n(x) \quad (4.1c)$$

If all such equations are written in the following form, then with the coefficients of the equations can be obtained for a matrix, and eventually the whole system could be expressed in matrix formation.

Eigenstate 1: $$\left\{\frac{\hbar^2}{2m_x^*}\frac{\partial^2}{\partial x^2}+U(x)\right\}\Psi_1(x)+0\cdot\Psi_2(x)+\ldots+0\cdot\Psi_n(x)=E_1\Psi_1(x)+0\cdot\Psi_2(x)+\ldots.+0\cdot\Psi_n(x) \quad (4.2a)$$

Eigenstate 2: $$0\cdot\Psi_1(x)+\left\{\frac{\hbar^2}{2m_x^*}\frac{\partial^2}{\partial x^2}+U(x)\right\}\Psi_2(x)+\ldots+0\cdot\Psi_n(x)=0\cdot\Psi_1(x)+E_2\Psi_2(x)+\ldots.+0\cdot\Psi_n(x) \quad (4.2b)$$

.....

Eigenstate n: $$0\cdot\Psi_1(x)+0\cdot\Psi_2(x)+\ldots+\left\{\frac{\hbar^2}{2m_x^*}\frac{\partial^2}{\partial x^2}+U(x)\right\}\Psi_n(x)=0\cdot\Psi_1(x)+0\cdot\Psi_2(x)+\ldots.+E_n\Psi_n(x) \quad (4.2c)$$

Thus the Hamiltonian that describes the whole system is formed.

$$\begin{bmatrix} \left\{\frac{\hbar^2}{2m_x^*}\frac{\partial^2}{\partial x^2}+U(x)\right\} & 0 & .... & 0 \\ 0 & \left\{\frac{\hbar^2}{2m_x^*}\frac{\partial^2}{\partial x^2}+U(x)\right\} & .... & 0 \\ .... & .... & .... & .... \\ 0 & 0 & .... & \left\{\frac{\hbar^2}{2m_x^*}\frac{\partial^2}{\partial x^2}+U(x)\right\} \end{bmatrix} \begin{bmatrix} \Psi_1(x) \\ \Psi_2(x) \\ .. \\ \Psi_n(x) \end{bmatrix} = \begin{bmatrix} E_1 & 0 & .... & 0 \\ 0 & E_2 & .... & 0 \\ .... & .... & .... & .... \\ 0 & 0 & .... & E_n \end{bmatrix} \begin{bmatrix} \Psi_1(x) \\ \Psi_2(x) \\ .. \\ \Psi_n(x) \end{bmatrix} \tag{4.3}$$

The situation is not resolved. Still a way to discretize the 2nd order differential operator into finite space is needed. A number of ways remain, but the most widely accepted method is the Finite Element Method.

Consider a function which is already discretized for numerical solution as in the figure below.

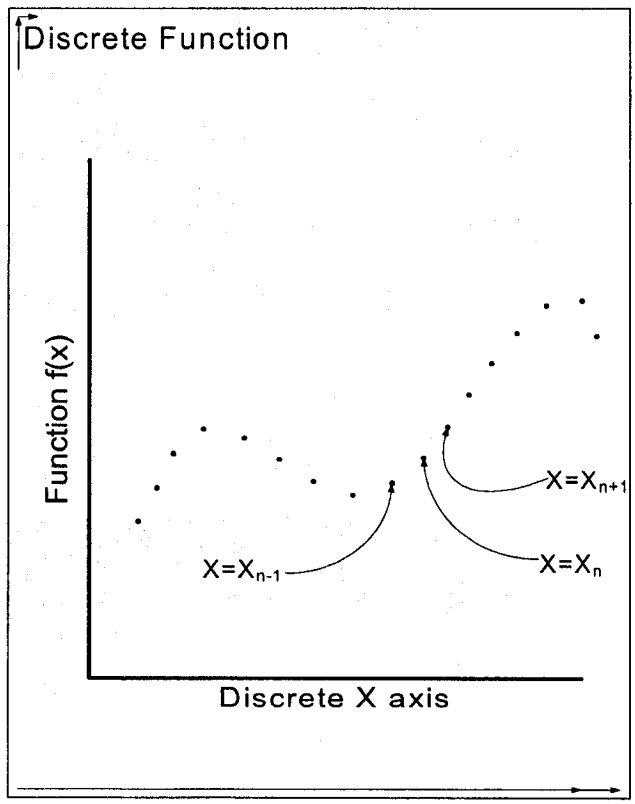


Figure 4.1: Arbitrary discrete function.

As it is considered to be a 1D case only, x is chosen to be the space dimension. In this case, the function can be expressed as a vector given by:

$$f(x) \Rightarrow \{f(x_1), f(x_2), f(x_3) \ldots\ldots\ldots f(x_{n-1}), f(x_n)\} \tag{4.4}$$

Then the FEM method finds the 2nd derivative as following:

$$\frac{\partial^2}{\partial x^2} f(x)\Big|_{x=x_n} = -f(x_n - 1) + 2f(x_n) - f(x_n + 1) \tag{4.4}$$

For a 2D case, first a 2D *(n+1) by (n+1)* grid is created, where *h=1/(n+1)* is the grid spacing. So the function *U* becomes discrete and the elements *U(i,j)* becomes approximate solutions at *x=i*h* and *y=j*h*. This is shown below for *n=7*:

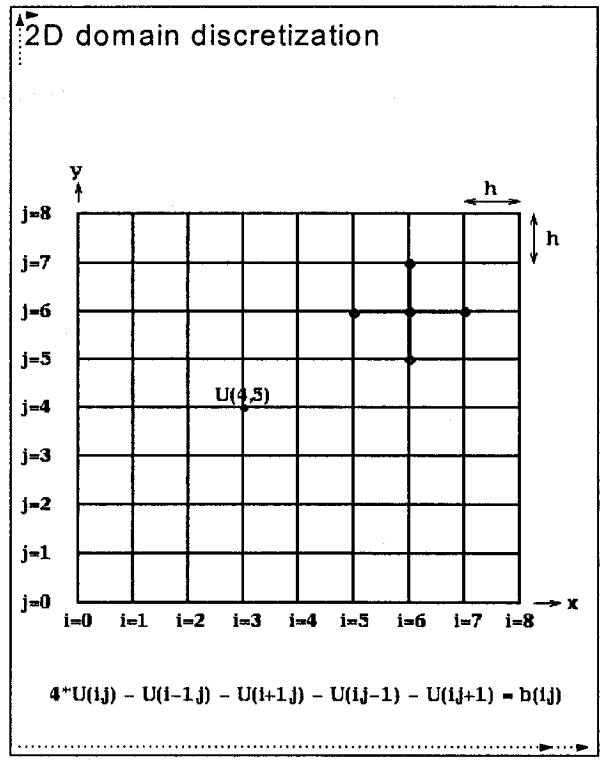


Figure 4.2: The 5-1 point domain discretization for a 2D system.

Then for a particular point (center of the thick stencil), in the 5-1 formalism, FEM suggests that the neighboring 4 points (as given in the thick stencil) have to be known to find the formulation of the 2nd derivative of the point in the middle and the formula is:

$$\nabla^2 U(x)\Big|_{x=x_n} = 4U(x_n, y_n) - U(x_n - 1, y_n) - U(x_n + 1, y_n) - U(x_n, y_n - 1) - U(x_n, y_n + 1) \quad \textbf{(4.5)}$$

The above linear equation relating *U(i,j)* and the value at its neighbors (indicated by the thick stencil) must hold for $1 \le i, j \le n$, giving $N=n^2$ equations in N unknowns. When *(i,j)* is adjacent to a boundary *(i=1 or j=1 or i=n or j=n)*, one or more of the *U(i±1, j±1)* values is on the boundary and therefore has a value.

4.3. Single-band effective mass Hamiltonian

When talking about electrons in semiconductor, what is actually referred to is the so called quasi particle, in this case a free to move electron in the conduction band. The nomenclature 'free' is used for the fact that the electron is not bound to its parent atom, and can move within the material with some freedom, and in the process, it emulates somewhat the motion of a free electron. This observation then becomes the basis of use of single-band effective mass equation.

$$i\hbar \frac{\partial}{\partial t} \Psi(r,t) = E_{band}(-i\nabla)\Psi(r,t) + U(r,t)\Psi(r,t) \quad \textbf{(4.6)}$$

In room temperature, and in normal conditions, the electrons in the conduction band tend to occupy the lowest possible energy. Looking at the dispersion relation of silicon, it is seen that the electrons of interest always lies in

the bottom part of the conduction band, which can be well approximated by a parabolic nature, like that of a free electron. So electrons in that range have energy:

$$E = E_{cond} + \frac{\hbar^2 k_x^2}{2m_x^*} + \frac{\hbar^2 k_y^2}{2m_y^*} + \frac{\hbar^2 k_z^2}{2m_z^*} \tag{4.7}$$

In a crystal, electrons face at least three distinct potentials: the periodic (or near periodic in case of extremely small nanostructures) potential due to lattice atoms, the local electric field potential due to externally applied excitations, and the scattering potential. So instead of solving the Schrödinger's equation using all three potential, a semi-classical approach adopts only the potential due to scattering and external force, and incorporates the effect of crystal potential in the parameter: effective mass. This is called the single-band effective mass equation. The conduction band near local minima consists only of a single dispersion relation, so the term 'single-band' enters.

4.4. Full band Hamiltonian for hole

The one-band effective mass model works very well for having an isotropic parabolic band that is well separated from the other bands. This is usually true for conduction band in wide bandgap semiconductors. But the valence band involves multiple closely spaced bands that are strongly anisotropic and non-parabolic. Close to the Γ-point the energy dispersion can usually be expressed in the form (*A, B, C* are constants)

$$E(\vec{k}) = E_v - Ak^2 \pm \sqrt{B^2k^2 + C^2\left(k_x^2k_y^2 + k_y^2k_z^2 + k_z^2k_x^2\right)} \tag{4.8}$$

This dispersion relation can be described by a 4X4 matrix: $\left|h(\vec{k})\right| = -PI - T$, where

$$[T] = \begin{bmatrix} Q & 0 & -S & R \\ 0 & Q & R^+ & S^+0 \\ -S^+ & R & -Q & -Q \\ R^+ & S & 0 & \end{bmatrix} \tag{4.9a}$$

$$P \equiv E_v + \frac{\hbar^2\gamma_1}{2m}\left(k_x^2 + k_y^2 + k_z^2\right) \tag{4.9b}$$

$$Q \equiv \frac{\hbar^2\gamma_2}{2m}\left(k_x^2 + k_y^2 - 2k_z^2\right) \tag{4.9c}$$

$$R \equiv \frac{\hbar^2}{2m}\left[-\sqrt{3}\gamma_2\left(k_x^2 - k_y^2\right) + i2\sqrt{3}\gamma_3k_xk_y\right] \tag{4.9d}$$

$$S \equiv \frac{\hbar^2\gamma_3}{2m}\left[2\sqrt{3}\left(k_x - ik_y\right)k_z\right] \tag{4.9e}$$

Usually this gives quite accurate eigenvalues of *[h(k)]* near the two highest valence bands very close to the Γ-point. But to get better agreement over a wider range of the *k*-values and to include the split-off band, one often uses a three band *[h(k)]* of the form:

$$[T] = \begin{bmatrix} P+Q & 0 & -S & R & -S/\sqrt{2} & \sqrt{2}R \\ 0 & P+Q & R^+ & S^+ & -\sqrt{2}R & -S^+/\sqrt{2} \\ R^+ & 0 & P-Q & 0 & -\sqrt{2}Q & \sqrt{3/2}S \\ 0 & R^+ & 0 & P-Q & -\sqrt{3/2}S^+ & \sqrt{2}Q^+ \\ -S^+/\sqrt{2} & -\sqrt{2}R^+ & -\sqrt{2}Q^+ & \sqrt{3/2}S & P+Q & 0 \\ \sqrt{2}R & -S/\sqrt{2} & -\sqrt{3/2}S^+ & \sqrt{2}Q & 0 & P+Q \end{bmatrix} \tag{4.10}$$

4.5. Effective mass Hamiltonian for hole

The full-band approach, resulting in multi-band Hamiltonian for holes in the valence band work well for a bulk device. But as it is very much involved with different dimensions, the consequence of using this approach for a nanodevice is still somewhat unclear and a matter of current research. So for a coarser approximation, one uses the already given effective mass of holes in silicon and constructs an effective mass Hamiltonian for holes to get:

$$i\hbar\frac{\partial}{\partial t}\Psi(r,t) = -\frac{\hbar^2}{2m^*_{l,h}}\Psi(r,t) + U_v(r,t)\Psi(r,t) \qquad (4.11)$$

Where U_v is the valence band energy and depending upon the particular band, m^* can be for light or heavy hole. The spin interaction band split is ignored for the present purpose.

4.6 Validity of effective mass in nanoscale

In bulk materials, most of the atoms reside well inside the surface, and a general consideration can be made that the surface is a far away ending of an otherwise periodic array of atoms. With this view, the crystal can be thought to be effectively made of only periodic atoms, giving rise to periodic potential. Consequently, Bloch's theorem can be applied and use of Krönig-Penny model can be made to treat the macroscopic potential due to applied bias or macroscopic space charge, ultimately resulting in band diagrams of the material. A deeper investigation into the matter reveals that in the lowest parts of the conduction

band, the energy dispersion relation is approximated to be parabolic. Now instead of explicitly determining the effect of the perfectly periodic crystal on conduction electron, one incorporates the idea of effective mass, which accounts for the crystal interaction at least near the band minima.

Effective mass is defined by analogy with Newton's second law $F=ma$. Using quantum mechanics it can be shown that for an electron in an external electric field E:

$$a = \frac{1}{\hbar^2}\frac{d^2\varepsilon}{dk^2}qE \tag{4.12}$$

where a is acceleration, $\hbar=h/2\pi$ is reduced Planck's constant, k is the wave number (often loosely called momentum since $k = \rho/\hbar$), $\varepsilon(k)$ is the energy as a function of k, or the dispersion relation as it is often called. From the external electric field alone, the electron would experience a force of $\boldsymbol{F}=q\boldsymbol{E}$, where q is the charge. Hence under the model that only the external electric field acts, effective mass m^* becomes:

$$m^* = \hbar^2\left[\frac{d^2\varepsilon}{dk^2}\right]^{-1} \tag{4.13}$$

For a free particle, the dispersion relation is quadratic, and so the effective mass would be constant (and equal to the real mass). In a crystal, the situation is far more complex. The dispersion relation is not even approximately quadratic, in the large scale. However, wherever a minimum occurs in the dispersion relation, the minimum can be approximated by a quadratic curve in the small region around

that minimum. Hence, for electrons which have energy close to a minimum, effective mass is a useful concept.

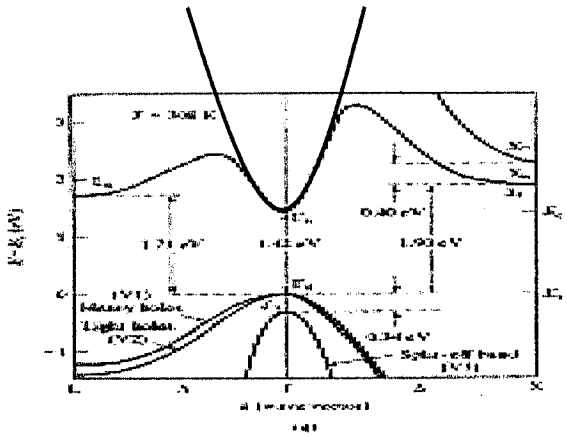


Figure 4.3: Parabolic band approximation of the conduction band [figure reconstructed by author]

In energy regions far away from a minimum, effective mass can be negative or even approach infinity. Effective mass, being generally dependent on direction (with respect to the crystal axes), is a tensor. However, for most calculations the various directions can be averaged out.

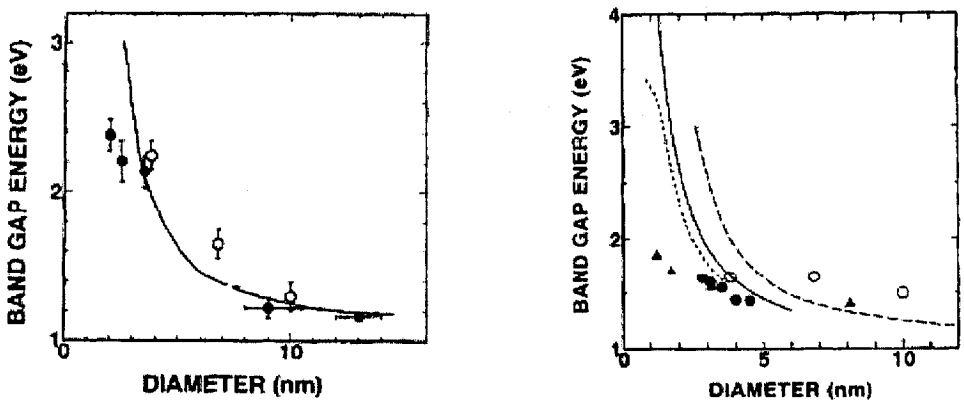


Figure 4.4: Band gap energy as determined by effective mass approximation in different studies, showing the validity range of effective mass approximation in nanoscale [21].

This concept of effective mass is valid for bulk materials where crystal can be assumed to be periodic, but in nanostructures the periodicity is no longer there

as dimensions get smaller. Hence the band diagram also fails for nanostructure thereby invalidating the concept of effective mass. Experiments have shown that well below 5nm, one can no longer use the effective mass approximation.

4.7 2D direct discretized Hamiltonian

In direct discretization, the 2D Schrödinger's equation is considered and using FEM the 2D Hamiltonian matrix is directly created. In this case underlying 2D equation, for example in the case of 2D electron gas, becomes:

$$-\frac{\hbar^2}{2m_x^*}\frac{\partial^2\Phi(x,z)}{\partial x^2}-\frac{\hbar^2}{2m_z^*}\frac{\partial^2\Phi(x,z)}{\partial z^2}+E_c(x,z)\Phi(x,z)=E_l\Phi(x,z) \qquad \textbf{(4.14)}$$

This time it should be noted that two different effective mass for different directions is used for generality and the well potential is also considered to be non zero (more precisely, the minimum of conduction band in a 2D electron gas in solid). The obvious effect is that the wave function now becomes 2D and the potential function also in general becomes 2D. The problem with such solutions is that direct discretization of the system requires huge computational resource.

Using the given algorithm for 2D FEM, one can discretize the above partial differential equation. And the general Hamiltonian looks like below:

$$h(x,z)=\begin{bmatrix} \alpha_1 & \beta & .. & 0 & 0 \\ \beta & \alpha_2 & & 0 & 0 \\ .. & .. & .. & .. & .. \\ 0 & 0 & .. & \alpha_{N_{x-1}} & \beta \\ 0 & 0 & .. & \beta & \alpha_{N_x} \end{bmatrix} \qquad \textbf{(4.15)}$$

The gripping factor about the Hamiltonian is that unlike the 1D case, the elements in the matrix are not single values; instead they themselves are complete matrices, such as,

$$\alpha_N(x)=\begin{bmatrix} 2t_x+2t_z-qE_{c_1}(x) & -t_z & .. & 0 & 0 \\ -t_z & 2t_x+2t_z-qE_{c_2}(x) & .. & 0 & 0 \\ .. & .. & .. & .. & .. \\ 0 & 0 & .. & 2t_x+2t_z-qE_{c_{Nx}}(x) & -t_z \\ 0 & 0 & .. & -t_z & 2t_x+2t_z-qE_{c_{Nz}}(x) \end{bmatrix} \tag{4.16}$$

is the one dimensional system of equations, while

$$\beta=\begin{bmatrix} -t_x & 0 & .. & 0 \\ 0 & -t_x & .. & 0 \\ .. & .. & .. & .. \\ 0 & 0 & .. & -t_x \end{bmatrix} \tag{4.17}$$

is the other dimensional coupling matrix. So together, the rather simple looking Hamiltonian is now in fact an n^2 by n^2 matrix.

4.8 Decoupled set of 1D equations

In this approach, the basic 2D equation remains the same as before,

$$-\frac{\hbar^2}{2m_x^*}\frac{\partial^2\Phi(x,z)}{\partial x^2}-\frac{\hbar^2}{2m_z^*}\frac{\partial^2\Phi(x,z)}{\partial z^2}+E_c(x,z)\Phi(x,z)=E_l\Phi(x,z) \tag{4.14}$$

But this time instead of trying to directly expand the differential operator according to 2D FEM, the method involves decoupling of dimension variance in an attempt to separate the dimensional dependence.

First a dimension is chosen to start with, say x. Then starting from one end, say $i=1$ (as shown in figure below) mesh points are gradually considered from

U(1,1) to *U(1,7)* and this 1D system is solved as an eigenvalue problem. This is repeated for every value of *i* till the other end. Then the position dependant eigenvalue *(E_1, E_2..... etc.)* is used as the system potential for 1D systems starting from say *j*=1(in the figure below) and is solved for points *U(1,1)* to *U(7,1)* and go up to *j*=7. This gives the total 2D solution.

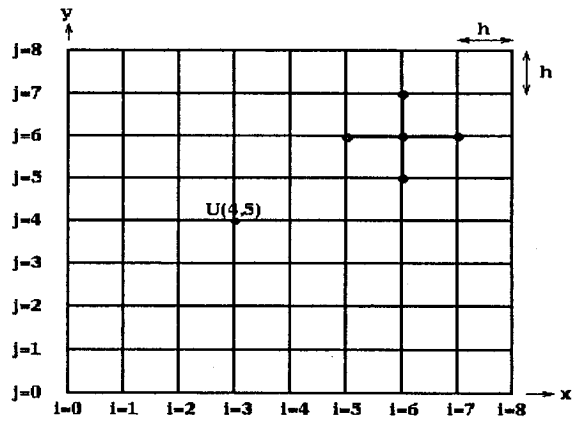


Figure 4.5: Using the discrete 2D mesh for the decoupled 1D system of equations.

The formal approach is as below: As before the starting point is the 2D Schrödinger's equation (the two dimensions being *x* and *z*):

$$-\frac{\hbar^2}{2m_x^*}\frac{\partial^2\Phi(x,z)}{\partial x^2}-\frac{\hbar^2}{2m_z^*}\frac{\partial^2\Phi(x,z)}{\partial z^2}+E_c(x,z)\Phi(x,z)=E_l\Phi(x,z) \qquad \textbf{(4.14)}$$

Then the differential equation is expanded in orthonormal basis function such as:

$$\psi(x,y)=\sum_m \delta(x-x')\phi_m(x',y)\varphi_m(x') \qquad \textbf{(4.18)}$$

This actually is quite similar to Fourier analysis. In Fourier analysis, there already exists the set of orthnormal functions: *sinθ, sin2θ, sin3θ, .., cosθ, cos2θ, cos3θ, ..*, and the objective becomes just to find the numeric coefficient of these

basis functions. But in a more general case, any orthonormal basis function can be chosen, as in this case the individual vertical slice dependence of wave function. Instead of a given function, these basis functions have to be found from another underlying equation. Then the coefficients could also be functions instead of numerals, but the basic mathematical concept remains the same.

For thin body, which is about always a good assumption in nano regime, quantum confinement in a particular direction introduces subbands, and only a few subbands are usually effectively occupied. Accordingly, wavefunction is expanded in the already mentioned orthonormal basis. The eigenfunctions and the associated eigenenergies are obtained by solving a one-dimensional wave equation in the *z*-direction within each *x'* valued slice.

$$-\frac{\hbar^2}{2m_z^*}\frac{\partial^2}{\partial z^2}\phi_m(x,z)+V(x,z)\phi_m(x,z)=E_m\phi_m(x,z) \qquad \textbf{(4.19)}$$

It should be noted here that the position dependant eigenvalue is not the eigenstate of the system, as eigenstate does not differ with position in the system. This energy could be interpreted as to represent the bottom of a particular subband, which varies with position along the channel. The envelope wave functions are zero at the boundary for infinite potential well. By using the orthonormal basis, the basic 2D equation can be transformed to a mode-space basis. By retaining only a few occupied modes the computational burden can be significantly reduced. Here it could be stated that the confinement dimension of the device in question here is less than the bulk excitonic Bohr radius, so carriers

are strongly confined along device thickness. This results in energy quantization in that dimension, and eventually subbands are formed as a result.

Table 1: Bulk Excitonic Bohr radius of different semiconductor materials [tabulated by author from various sources]

Materials	Excitonic Bohr radius in angstrom
Cadmium Sulfide	315
Zinc Sulfide	50
Lead Sulfide	204
Lead Selenide	460
Cadmium Selenide	61
Silicon	55
Zinc oxide	18
Copper Chloride	10
Indium Arsenide	340
Indium Antimonide	540
Cadmium Telluminide	100

The geometry offers yet another simplification. If it is assumed that the shape of the confined mode does not change along the length, which is same as assuming the potential variation within the system is slow with respect to electron wavelength, the result is

$$-\frac{\hbar^2}{2m_x^*}\frac{\partial^2}{\partial x^2}\varphi_m(x,z)+V(x,z)\varphi_m(x,z)=E_n\varphi_m(x,z) \tag{4.20}$$

So in a nutshell the algorithm could be described stepwise as follows:

1. Discretize the 2D space coordinates according to the system.
2. Choose a particular direction as the basis function coordinate and expand the wavefunction as below:

$$\psi(x,y)=\sum_m \delta(x-x')\phi_m(x',y)\varphi_m(x') \tag{4.18}$$

3. Solve for each slice of the chosen coordinate as individual 1D system and the problem reduces to 1D eigenvalue problem:

$$-\frac{\hbar^2}{2m_z^*}\frac{\partial^2}{\partial z^2}\phi_m(x,z)+V(x,z)\phi_m(x,z)=E_m\phi_m(x,z) \quad (4.19)$$

4. Using the position dependant eigenvalue from the previous equation as the potential profile for the other dimension, solve the following equation as another eigenvalue problem:

$$-\frac{\hbar^2}{2m_x^*}\frac{\partial^2}{\partial x^2}\varphi_m(x,z)+V(x,z)\varphi_m(x,z)=E_n\varphi_m(x,z) \quad (4.20)$$

5. Finally find the overall wave function from the orthonormal basis function expansion given in step 2.

4.9 Validity of decoupled set approach

A closer look into the approach reveals some significant assumptions in the mathematics. First it has been assumed that the domain of operation is thin. This is usually true for practical quantum devices such as a thin body MOSFET or nanowire or even a 2D pn junction. So this assumption actually does not cause a serious problem to its applicability in this case.

On the other hand, another important assumption is that the potential variation with respect to the electronic wavelength is slow. This implies that the shape of the confined subband in the transverse direction effectively maintains same shape, and its variation in the longitudinal direction remains insignificant. In fact the only fast varying potential that an electron faces in a practical device is the atomic potential due to the lattice points, which ultimately is ignored by always considering the smeared out average electronic concentration in device model.

Though the assumptions seem quite reasonable, the effect it has may become significant in some cases. The two inherent implications remain prudent: subbands with different energies do not couple and some coupling information of a subband with itself is also lost. But as it is seen, for a thin body MOSFET this method works well [42] and also results well for a thin 2D pn junction.

4.10 Choice of representation

The choice of representation may become important for particular systems. In formulating a theory of quantum transport there remains a choice of what representation to use, the optimum one depends on the problem at hand. A representation based on eigenstates is often convenient for analytical calculations since the Hamiltonian is diagonal. On the other hand, a real space representation is intuitively more appealing. In dealing with decoupled 1D device, it remains convenient to use the eigenstate representation for the transverse dimensions but a discrete real space lattice for the longitudinal direction gives better applicability.

For devices with a large (effectively infinite) cross-section, it is common to ignore the transverse confining potential and use periodic boundary conditions in that direction since the real boundary conditions are believed to have minimal effect on the observed properties. The transverse eigenstates are then given by plane waves.

4.11 Direct discretization: Computationally expensive

As it is seen that direct discretization imposes serious memory problem, there has been rigorous studies how to efficiently solve the direct discretized 2D system, more generally referred to a as 2D Poisson problem, as the operator is Laplacian. It is not the intension of this project to give a comprehensive comparison of different algorithms to solve, but to make the case strong for the proposed method, only a brief discussion is introduced.

Table 2: A comparison between different types of solution to Laplace's equation

Algorithm	Type	Serial time	PRAM	Storage time	Process time
Dense LU	D	N^3	N	N^2	N^2
Band LU	D	N^2	N	$N^{3/2}$	N
Inv(P)*b	D	N^2	Log N	N^2	N^2
Jacobi	I	N^2	N	N	N
Sparse LU	D	$N^{3/2}$	$N^{1/}2$	N*log N	N
CG	I	$N^{3/2}$	$N^{1/2}$*log N	N	N
SOR	I	$N^{3/2}$	$N^{1/2}$	N	N
FFT	D	N*log N	log N	N	N
Multigrid	I	N	$(\log N)^2$	N	N
Lowerbound		N	log N	N	

Key to abbreviations:

Dense LU : Gaussian elimination, treating *P* as dense

Band LU : Gaussian elimination, treating P as zero outside a band of half width $(n-1)$ near diagonal

Sparse LU : Gaussian elimination, exploiting entire zero-structure of P

Inv(P)*b : Pre-compute and store inverse of P, multiply it by source $b(i,j)$

CG : Conjugate Gradient method

SOR : Successive Over Relaxation

FFT : Fast Fourier Transform based method

The first column in the table identifies the algorithm, except the last entry, which gives a simple Lower Bound on the running time for any algorithm. The Lower Bound is obtained as follows. For the serial time, the time required simply to print each of the N solution components is N. For the *PRAM* time, which assumes as many processors as we like and assumes communication is free, it is to be noted that the inverse $inv(P)$ of the discrete Poisson matrix P is dense, so that each component of the solution $U = inv(P)*b$ is a nontrivial function of each of the N components of b. The time required to compute any nontrivial function of N values in parallel is $log(N)$.

The second column says whether the algorithm is of Type *D=Direct*, which means that after a finite number of steps it produces the exact answer, or of Type *I=Indirect*, which means that one step of the algorithm decreases the error by a constant factor $\rho<1$, where ρ depends on the algorithm and N. This means that if one wants the final error to be ε times smaller than the initial error, one must take m steps where $\rho^m \leq \varepsilon$. To compute the complexities in the table, m has to be chosen

so that ρ^m is about as small as the discretization error, i.e. the difference between the true solution *u(i*h,j*h)* and the exact discrete solution *U(i,j)*. There is no point in making the error in the computed *U(i,j)* any smaller than this, since this could only decrease the more significant error measure, the difference between the true solution *u(i*h,j*h)* and the computed *U(i,j)*, by a factor of 2.

The second and third columns give the running time for the algorithm on a serial machine and a *PRAM*, respectively. A *PRAM* can have as many processors as needed (shown in the last column), and communication is free. Thus, the *PRAM* time is a lower bound for any implementation on a real parallel machine. Finally, the fifth column indicated how much storage is needed. *LU* decomposition requires significantly more storage than the other methods, which requires just a constant amount of storage per grid point.

This includes methods like *Dense LU* (Gaussian Elimination) even though these are much slower than the fastest methods, because these slower methods solve much more general linear systems than the much faster but more specialized algorithms like Multigrid. This table illustrates that there is often an enormous payoff for using an algorithm specially tuned for the system at hand. *Band LU* is Gaussian Elimination specialized to take advantage of the fact that none of the zero entries of P more than n entries away from the diagonal ever "fill-in" during Gaussian Elimination, so we can avoid both storing them and doing arithmetic with them. *Sparse LU* exploits the fact that many other entries in the L and U factors of P are zero, and avoids both computing with or storing any zero entries.

*Inv(P)*b* means pre-computing and storing the exact *N-by-N* inverse of *P* (the cost of this pre-computation is not counted in the complexity), and then doing a matrix-vector multiplication. On a serial machine, this is no faster than any of the subsequent methods, and uses much more storage.

Jacobi, SOR (Successive Over Relaxation) and CG (Conjugate Gradients) can be thought of as performing most nearest neighbor operations on the grid (CG requires summing across the grid too). In other words, information at one grid point can only propagate to an adjacent grid point in 1 iteration. Since the solution U(i,j) at each grid point depends on the value of $b(l,m)$ at every other grid point, it take n steps for information to propagate all the way across the grid. This would not lead to a good solution in fewer than $n=\sqrt{(N)}$ steps, which is what these methods require.

4.12 Green's function formalism

A significant advancement with the decoupled approach could be attained if the solution method is instead of direct, is the Green's function method. But its advantages are only apparent in complex systems where non-ballistic transport is considered with incoherent scattering effects and for a more rigorous contact analysis. It also comes into help if we are trying to simulate a 3D system. But for simpler situation like a 2D infinite potential well, it becomes unnecessarily overburdened.

4.13 Self energy matrices

For the device in question, the carriers are confined in the lateral direction so that the boundary condition assumes that the carrier concentrations, i.e. their wave functions become zero in the silicon- oxide interface. But in the longitudinal direction, there are two contacts which in fact inject electrons and holes to the system and have to maintain equilibrium condition for any supplied current. So to incorporate these contacts into the actual device, some modeling method has to be adopted. A very suitable candidate in this case is the self energy matrix modeling the semi-infinite leads. Here it is assumed that the contacts are so abundant in carrier concentration that the thermodynamic equilibrium is maintained within the contacts (considering them as reservoirs of carriers) and the Fermi level inside remains constant.

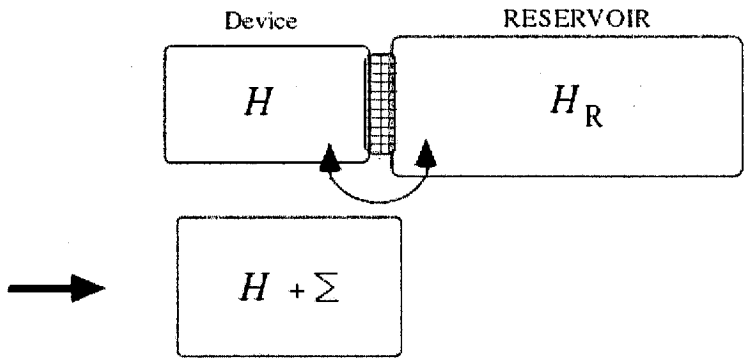


Figure 4.6: Coupling between the contacts (seen as reservoirs of carriers) and the actual device [41].

The concept of self-energy is used in many-body physics to describe electron–electron and electron–phonon interactions. In the present context,

however, it can be used to describe the effect of a semi-infinite contact. In general, a 'device' connected to a large reservoir and has an overall Hamiltonian matrix of

$$\begin{bmatrix} H & \tau \\ \tau^{+} & H_R \end{bmatrix} \tag{4.22}$$

where the dimension of Hamiltonian for the contact, H_R is huge compared to that of the device H. The overall Green's function has the form

$$\begin{bmatrix} G & G_{DR} \\ G_{RD} & G_R \end{bmatrix} = \begin{bmatrix} (E+i0^{+})I-H & -\tau \\ -\tau^{+} & (E+i0^{+})I-H_R \end{bmatrix}^{-1} \tag{4.23}$$

Here G is related to the device, while G_R is related to the reservoir. The other two are interactions among the two. Because the concern is only about the details inside the device and not inside the reservoir for analysis purposes, only G is of interest (and not in G_R or G_{DR} or G_{RD}). It then becomes is straightforward to show that [40]

$$G = \left[(E+i0^{+})I-H-\Sigma\right]^{-1} \approx \left[EI-H-\Sigma\right]^{-1} \tag{4.24a}$$

where $\Sigma = \tau g_R \tau^{+}$ and $g_R = \left[(E+i0^{+})I-H_R\right]^{-1}$ (4.24b)

This shows that the effect of the coupling to the reservoir can be accounted for by adding a self-energy matrix Σ to the Hamiltonian H. This is a very general concept that allows eliminating the huge reservoir and working solely within the device whose dimensions are much smaller. It is to be noted that unlike 0^{+}, Σ is not necessarily an infinitesimal quantity and it can be finite with a value defined by the coupling to the reservoir.

Arbitrary reservoir coupling can be calculated from (4.24b) in general, resulting in coupling matrices. So to find the self energy matrix Σ, a huge inversion is required:

$$\sum(m,n)=\sum_{\mu,\upsilon}\tau(m,\mu)g_R(\mu,\upsilon)\tau^+(\upsilon,n) \tag{4.25}$$

The indices *m, n* refer to points within the device while μ,υ refer to points inside the reservoir. However, the coupling matrix τ couples the points within the device to a small number of points on the surface of the reservoir, so that only g_R (μ,υ) for points (μ,υ) that are on the surface is needed. This surface Green's function can often be calculated analytically assuming a given reservoir model.

For the case in hand, the self-energy can be obtained from fairly elementary arguments without worrying about surface Green's functions. The self-energy matrix *Σ(E)* that accounts for the semi-infinite lead on the left is given by (*t* being the onsite coupling energy $t=\hbar^2/2m^*a^2$)

$$\Sigma_1(E)=\begin{array}{c} \\ |1\rangle \\ |2\rangle \\ \vdots \\ |N-1\rangle \\ |N\rangle \end{array}\begin{array}{c} \begin{array}{cccccc} |1\rangle & |2\rangle & \cdots & |N-1\rangle & |N\rangle \end{array} \\ \begin{bmatrix} -te^{ik_1a} & 0 & & 0 & 0 \\ 0 & 0 & & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & & 0 & 0 \\ 0 & 0 & & 0 & 0 \end{bmatrix} \end{array} \tag{4.26}$$

where $E=E_c+U_1+2t(1-\cos k_1a)$

In other words all that is needed is to add a term te^{ik_1a} to *H(1,1)* and the semi-infinite lead is accounted for exactly, as far as calculating the Green's function is concerned. In the self-energy method it is assumed that we only have outgoing

(not incoming) waves at the ends. The fact that an actual device has incoming waves as well from the contacts is irrelevant when calculating G, as it is the retarded Green's function representing the response of the system to an impulse excitation within the device: $(EI - H - \Sigma)G = I$, and hence the appropriate boundary condition for G is that it only has outgoing waves at the ends. This means that when calculating G it can be assumed that just outside the device (and hence inside the contacts)

$$\psi_0 = \psi_1 e^{ik_1 a} \tag{4.26}$$

Hence only this term (with a negative sign to incorporate outgoing wave) is added to point 1 of the device Hamiltonian to take care of the open boundary condition. Similarly the self-energy matrix that accounts for the semi-infinite lead on the right has only one non-zero term at point N which is given by

$$\Sigma_2(N,N) = -te^{ik_2 a} \tag{4.27}$$

The Green's function is obtained from

$$G = [EI - H - \Sigma_1 - \Sigma_2]^{-1} \tag{4.28}$$

where the self-energy functions account for the open boundary conditions exactly. It has been demonstrated that the results agree quite well with those obtained directly using periodic boundary conditions [41].

The self-energy method is computationally more intensive, since it requires integration over energy. However, the periodic boundary conditions in an attempt to model the contacts, merely get rid of end effects through the artifact of wrapping the device into a ring. An open system has a continuous energy

spectrum, while a ring has a discrete energy spectrum. The electron density is obtained by integrating over energy and is relatively unaffected by the discretization of the spectrum at least at room temperature. But the difference would be apparent, when the device comes out of thermal equilibrium. The full power of the self-energy method becomes apparent when we model a device under bias—a problem that cannot be handled with periodic boundary conditions.

4.14 Broadening effect

If a nanodevice is so small that the carriers get confined, then there exist quantized energy states. In itself, this energy distribution can be calculated and accounted for in that device. But whenever a device is mentioned, it is always implied that corresponding contacts are there, as without the contact of the device, which remain the connection of the device to outer world, the device is of no use. But in the confined nanodevice, there is an inherent effect of the contacts on that device. It broadens whatever energy levels the device might have had and must be correctly accounted for device modeling. Depending upon the physical nature of the contact, the coupling becomes strong or weak, which determines how the levels inside the device will be broadened. It so happens that self-energy method correctly addresses this issue. There are two factors that distinguish Σ_1 and Σ_2 from ordinary Hamiltonians. Firstly, they are energy dependent. Secondly, they are not Hermitian. It can be written [41]

$$H+\Sigma_1+\Sigma_2=\left(H+\frac{\Sigma_1+\Sigma_1^+}{2}+\frac{\Sigma_2+\Sigma_2^+}{2}\right)+\left(\frac{\Sigma_1-\Sigma_1^+}{2}+\frac{\Sigma_2-\Sigma_2^+}{2}\right)=\hat{H}-i\frac{\Gamma_1}{2}-i\frac{\Gamma_2}{2} \quad \textbf{(4.29)}$$

where $\hat{H}=H+\frac{\Sigma_1+\Sigma_1^+}{2}+\frac{\Sigma_2+\Sigma_2^+}{2}$ and $\Gamma_1=i\left[\Sigma_1-\Sigma_1^+\right]$, $\Gamma_2=i\left[\Sigma_2-\Sigma_2^+\right]$

These self-energy terms have two effects. It changes Hamiltonian which changes the eigenstates and their energies. More importantly, it introduces an imaginary part to the energy determined by the 'broadening' functions Γ_1 and Γ_2. The former represents a quantitative change; but the latter represents a qualitative change as regards to level broadening.

A particular representation might diagonalize H. This representation will not necessarily diagonalize Γ_1 and Γ_2. Interesting quantum interference effects can arise from non-diagonal elements of Γ_1 and Γ_2. But if Γ_1 and Γ_2 are also simultaneously diagonalized then the eigenenergies of $(H+\Sigma_1+\Sigma_2)$ will be given by

$$\varepsilon-i(\gamma_1-\gamma_2)/2 \quad \textbf{(4.30)}$$

where ε, γ_1 and γ_2 are corresponding diagonal elements of $\hat{H}$, Γ_1 and Γ_2 respectively. This could be viewed as a broadening of the energy level from a delta function $\delta(E-\varepsilon)$ into a line of the form

$$\frac{\gamma_1+\gamma_2}{(E-\varepsilon)^2+((\gamma_1-\gamma_2)/2)^2} \quad \textbf{(4.31)}$$

As γ_1 and γ_2 are generally energy dependant coupling coefficients of contacts to the device, the expression (4.31) may have a non-Lorentzian shape.

The imaginary part of the energy implies that the wave function and the associated probability decays with time which can be written in the form

$$\Psi \sim e^{(-i\varepsilon t/\hbar)} e^{(-\gamma_1 t/2\hbar)} e^{(-\gamma_2 t/2\hbar)} \tag{4.32}$$

$$|\Psi|^2 \sim e^{(-\gamma_1 t/\hbar)} e^{(-\gamma_2 t/2\hbar)} \tag{4.33}$$

Electrons in this state will eventually discharge into the left and right leads with time constants $\hbar/\gamma_1$ and $\hbar/\gamma_2$ respectively. Thus $\hbar/\gamma_1$ and $\hbar/\gamma_2$ represent the rates at which an electron initially in a particular state will escape into the left and right states respectively, as seen from the self-energy matrix element

$$\Sigma(1,1) = -te^{ika} \rightarrow \Gamma(1,1) = 2t\sin(ka) = \hbar\upsilon/a \tag{4.33}$$

Which is quite reasonable since the rate of escape from a lattice site of size a should be v/a.

Chapter 5: Simulation: Formulating the model

As the theoretical basis of the device model is gradually constructed, it then remains to implement the model in a simulator to evaluate the performance of the model. The underlying considerations include from choosing of the software to detailed implementation method for different parts. The details for such a simulation are presented here.

5.1. Choice of software

Whenever the question of numerically solving some system of equations arises, the choices become obvious. Either a versatile programming language, like C++ is used, where basic codes for solving every bit of mathematical operation is generated from scratch, or any of the mathematical programs can be used, which already have the subroutines to do the intricate arithmetic and numeric operations. Choosing a programming language for this project would have been over ambitious as unnecessary time and effort would go into developing codes for mathematical operations, which is not the main problem in hand. So the effective choice was in choosing one of the mathematical software. The practical choice was MATLAB, FEMLAB, MATHEMATICA and MAPLE. Among these, the first two actually are adept to handling floating point operation of complex and huge matrices, while the later two only pose good option for symbolic mathematics. So essentially the choice was limited to two.

Among the two, FEMLAB do better job in physical situation involving different systems, but for this case, as only the Schrödinger's equitation has to be solved, MATLAB is as good. Moreover it offered flexibility of script programming, and considering all the aspects, MATLAB was chosen.

One other important consideration came into play. Usually in more practical cases, a device problem should be solved self-consistently, that is iteratively and simultaneously solving Schrödinger's equation with Poisson's equation, which at the moment is not possible with FEMLAB.

5.2. Device structure

The device in question is essentially a 2D device. The longitudinal direction is considered to be the x-direction, while the lateral dimension is in the z-direction. y-dimension is the transverse direction and the device is so wide in this direction that the carriers in this direction are free to move and characterized by plane wave. So in essence, the device is distributed in the x-z plane. In the longitudinal direction, the device ends in two contacts, the Anode and the Cathode from where electrons and holes are injected into the device. In the lateral dimension, silicon device is sandwiched between two layers of oxide, ensuring confinement in both sides. Half of the device is p type and rest half is n type, so the metallurgical junction is at the middle. Doping is considered to be step type so the junction is abrupt, so that the basic physical effects become clear. As for device is fabricated in silicon, for confinement, the z-dimension is 5nm wide, while in the longitudinal

direction, it has a length of 100nm. So there is no confinement, but still a quantum mechanical state function gives a good description. In y-direction, the device is several micrometers wide so effectively it becomes large.

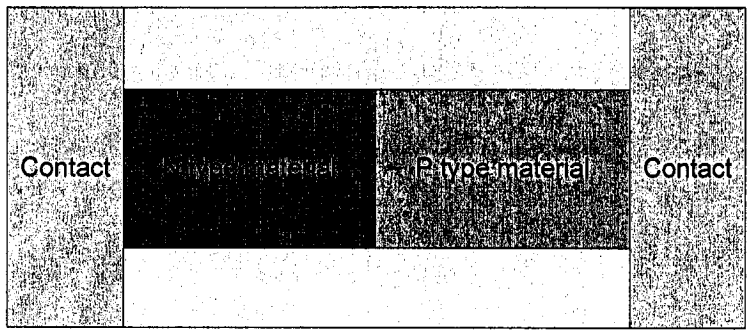


Figure 5.1: Schematic diagram of the device structure in 2D. The transverse direction is not shown.

5.3. Meshing

Two meshing schemes were considered, the triangular meshing and the rectangular meshing. For triangular meshing, the formulation of the 2D laplacian and the 1D Hamiltonian becomes involved and the basic physical aspects are curtained, while the rectangular meshing in general has a resolution problem. But the fact that the device is strictly rectangular in shape, made the case for rectangular meshing.

In rectangular meshing, the grid points were set so that every feature is well accounted for, such as the silicon-silicon oxide junction, semiconductor-metal contact, metallurgical junction between p and n type etc. The spacing is same in the z and x directions, and the spacing a was determined by the upper limit

$$\Delta E \leq 0.8t \qquad (5.1)$$ where t depends upon a.

This is for the fact that under this limit, the actual dispersion relation between energy and discretized momentum of carrier given by

$$E = 2(1 - \cos ka) \tag{5.2}$$

is well approximated by the continuous dispersion relation

$$E = \frac{n\hbar^2 k^2}{2m^*} \tag{5.3}$$

5.4. Lateral dimensional Hamiltonian

In MATLAB, with rectangular meshing for a 1D case in the *z*-direction, an *n by n* Hamiltonian matrix is in order, where *n* is the number of discrete points in that direction. So if the lateral width is *z* nm and the grid spacing is *a* nm, then *n=z/a*. As *n* is increased, the size of the matrix also increases and requires more and more computational resource. The coupling constant for the equation is

$$t_z = \frac{\hbar^2}{2m_z^*(\Delta z)^2} \tag{5.4}$$

Where Δz is the differential distance between mesh points and m_z^* is the z-directed effective mass of concerned carrier, (electron or hole). This is the onsite energy that every particle feels due to lattice, accordingly discritized into FEM formalism.

The device is first sliced into vertical slices of width a so that we have L_x/a vertical slices. It is assumed that all quantities are constant in x-direction within each slice, and a lateral Hamiltonian is constructed for each slice. So for a particular slice, the actual Hamiltonian both for electron and hole becomes:

$$H_z(x_n)=\begin{array}{c|cccccc} & |z_1\rangle & |z_2\rangle & \cdots & |z_{n-1}\rangle & |z_n\rangle \\ |z_1\rangle & 2t_z+U(1,x_n) & -t_z & \cdots & 0 & 0 \\ |z_2\rangle & -t_z & 2t_z+U(2,x_n) & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ |z_{n-1}\rangle & 0 & 0 & \cdots & 2t_z+U(n-1,x_n) & -t_z \\ |z_n\rangle & 0 & 0 & \cdots & -t_z & 2t_z+U(n,x_n) \end{array} \quad \textbf{(5.5)}$$

For example, with n=5 at the 10th slice, the total Schrödinger's equation in matrix form becomes:

$$\begin{bmatrix} 2t_z+U(1,10) & -t_z & 0 & 0 & 0 \\ -t_z & 2t_z+U(2,10) & -t_z & 0 & 0 \\ 0 & -t_z & 2t_z+U(3,10) & -t_z & 0 \\ 0 & 0 & -t_z & 2t_z+U(4,10) & -t_z \\ 0 & 0 & 0 & -t_z & 2t_z+U(5,10) \end{bmatrix} \cdot \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \\ \psi_5 \end{bmatrix} = \begin{bmatrix} E_1(10) & 0 & 0 & 0 & 0 \\ 0 & E_1(10) & 0 & 0 & 0 \\ 0 & 0 & E_1(10) & 0 & 0 \\ 0 & 0 & 0 & E_1(10) & 0 \\ 0 & 0 & 0 & 0 & E_1(10) \end{bmatrix} \cdot \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \\ \psi_4 \\ \psi_5 \end{bmatrix} \quad \textbf{(5.6)}$$

For the practical device in consideration, the size is n=10.

5.5. Boundary condition

It is assumed that the carrier wave functions go to zero at the semiconductor-oxide interface; hence the Si-SiO_2 barrier is assumed to be infinitely high. This may seem contrary to the practical device, but actually accounts well for confinement. The energy difference between the materials is quite high to confine the carrier within the device (about 8eV in SiO_2 and 1.12eV in Si), and only a small evanescence tail exist into the oxide layer, even that can be accounted for with a slight modification of the effective mass. The confinement is inherent to the 1D Hamiltonian as seen here as follows.

For a particular grid point z_n, the differential operator results in FEM formalism:

$$\left.\frac{d^2\psi(z)}{dz^2}\right|_{z=z_n} = \frac{2\psi(z_n)-\psi(z_{n-1})-\psi(z_{n+1})}{(\Delta z)^2} \tag{5.7}$$

So if the boundary point is considered where $n=1$ so $z_n=1$ then

$$\frac{d^2\psi(1)}{dz^2} = \frac{2\psi(1)-\psi(0)-\psi(2)}{(\Delta z)^2} \tag{5.8}$$

Now as it can be seen that the point $z=0$ for which the wave function $\psi(0)=0$ is within the oxide layer and this essentially means that there exists no carrier within the oxide layer, and the boundary is effectively that of an infinite well. Similar argument can be explained for the last point $z=n$, giving rise to same infinite potential well boundary condition.

5.6. Energy grid

In the longitudinal x-direction, the device is coupled with two contacts, each injecting a particular type of carrier (anode injecting hole and cathode injecting electron). So the Schrödinger's equation is no longer an eigenvalue problem in this dimension, rather the wave functions for a given eigenenergy has to be found for a corresponding longitudinal energy. The longitudinal energy is the energy of a carrier in relevant contact that can be injected to the device. Since the device is degenerately doped, the thermal equilibrium Fermi energy lies well within the corresponding band, and particles with any energy between Fermi energy and the band edge can be injected with high probability. So to account for this the longitudinal energy has to be discretized and solution has to be found for

all such energy. Hence, for longitudinal Hamiltonian, a grid in the energy has to be defined.

For the device in hand, the reference energy is the conduction band energy E_{cn} in the n side. Taking $E_{cn}=0$ then the energy range for electrons remain from 0 to E_{Fn}. Similarly the energy range for holes is from E_{Fp} to E_{vp}. The energy grid spacing is taken to be 0.5meV, which is quite typical for such calculations. In equilibrium, $E_{Fn}=E_{Fp}$ while for a bias of V, it is given by $E_{Fn}=E_{Fp}+qV$.

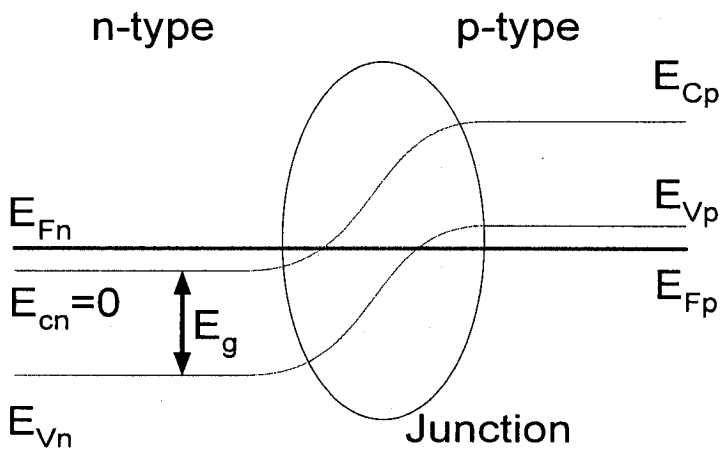


Figure 5.2: The schematic diagram of the band diagram of a pn junction under equilibrium condition with no bias. The system Fermi potential is the same in every point of the system.

5.7. Longitudinal Hamiltonian

In decoupled mode, the longitudinal x direction is treated separately to z direction, and a 1D Hamiltonian is used for x directed Schrödinger's equation. As the device is vertically sliced for decoupling, the lateral dimensional Hamiltonian gives the eigenvalue for each vertical slice for a particular subband, that energy

acts as the on site effective potential for the longitudinal direction. So for a particular subband, the Hamiltonian looks like

$$H_x = \begin{array}{c} \\ |x_1\rangle \\ |x_2\rangle \\ \vdots \\ |x_{n-1}\rangle \\ |x_n\rangle \end{array} \begin{array}{c} \begin{array}{cccccc} |x_1\rangle & |x_2\rangle & \cdots & |x_{n-1}\rangle & |x_n\rangle \end{array} \\ \begin{bmatrix} 2t_x+E_1 & -t_z & \cdots & 0 & 0 \\ -t_z & 2t_x+E_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 2t_x+E_{n-1} & -t_z \\ 0 & 0 & \cdots & -t_z & 2t_x+E_n \end{bmatrix} \end{array} \tag{5.9}$$

This Hamiltonian is used for both electrons and holes, but for different carriers, it has different effective mass. More over as electrons and holes are injected from opposite contacts, the numbering of the grid points is also reversed, i.e. an arbitrary node x_m for electron becomes x_{n-m} for hole.

5.8. Open boundary condition and carrier injection

Semi-infinite contacts are attached to the device as anode and cathode. Because the potential in the contacts is assumed to be uniform, the solutions in the semi-infinite contacts are plane waves. If a unit amplitude wave is injected from the cathode (electron), then some portion reflects from the device and some transmits across and exits the perfectly absorbing anode contact

$$\phi_m(x) = 1e^{ik_x x} + r_m e^{-ik_x x} \ldots\ldots\ldots x < 0 \tag{5.10a}$$

and

$$\phi_m(x) = t_m e^{ik_x x} \ldots\ldots\ldots x > L_x \tag{5.10b}$$

Where r_m and t_m are reflection and transmission coefficients for cathode injection into mode m and L_x is the length of the active device.

By solving (4.20) subject to the boundary conditions, (5.10a) and (5.10b), the wave function due to the injection of a unit amplitude wave from the cathode is found. This translated to the matrix formation gives the self energy matrices of dimension *n by n* with only one non-zero term

$$\Sigma_1(i,j) = -t_x e^{ik_x x}\delta_{1,j}\delta_{1,i} \tag{5.11a}$$

$$\Sigma_2(i,j) = -t_x e^{ik_x x}\delta_{n,j}\delta_{n,i} \tag{5.11a}$$

For carrier injection a source vector term γ_1 is required of *n by 1* dimension. It has only one non-zero term accounting for carrier injection, and also the level broadening.

$$\gamma_1(1) = i\left[\Sigma_1(1,1) - \Sigma_1^*(1,1)\right] = 2t_x \sin k_x a = \hbar\upsilon(k_x)/a \tag{5.12}$$

So for longitudinal direction the equation that is solved is

$$\left[EI - H - \Sigma_1 - \Sigma_2\right]\phi = -i\gamma_1 \tag{5.13}$$

Similar arguments also account for holes that are injected from anode.

5.9. Steady state carrier density

Electron density for a confined mode m and with injected wave vector k_x is obtained from summing all transverse (y-directed) mode with wave vector k_y:

$$n_e^m(k_x,x) = \frac{1}{W}\sum_{k_y}\left|\phi(k_x,x)\right|^2 f_{FD}\left(E - E_{F,C}\right) \tag{5.14}$$

Here k_x refers to the x-component of the wave vector of an electron with total energy E in the cathode contact and the probability that the state with energy

E is occupied within the contact is given by the Fermi-Dirac distribution f_{FD}. This is valid because the contacts are always under thermal equilibrium.

As only positive (or negative, depending upon the position of cathode in the left or right) k_x values are injected, so summing (5.14) over all positive k_x gives the 2D (x-y plane) charge density. As the transverse direction is virtually infinite, the summation in that dimension can be converted to integral over transverse energy. The longitudinal summation can also be changed into integral and the final result can be expressed as [43]:

$$n_e^m(x) = \int_{-\infty}^{\infty} n_e^m(x, E_l) dE_l \tag{5.15}$$

Considering the conduction band (valence band while considering hole calculations) as infinitely wide it can be shown that (including spin degeneracy) [43]

$$n_e^m(x, E_l) = \frac{1}{\hbar a}\sqrt{\frac{2m_y^* k_B T}{\pi}} F_{-1/2}(E_{F,c} - E_l)\frac{A_e^m(x, E_l)}{2\pi} \tag{5.16}$$

Here $F_{-1/2}$ is the Fermi integral of order -1/2 and A_e^m is the local density of state:

$$A_e^m(x, E_l) = a\frac{dk_x}{dE_l}|\phi_m(x, k_x)|^2 \tag{5.17}$$

Similar calculations are assumed for hole also.

5.10. Current calculation

For the current calculation, instead of directly using the current operator, the transmission viewpoint is adopted for its simplicity. The current due to

injection of an electron into a particular mode m with a particular longitudinal energy E_l is found from

$$I_m(E_l) = \frac{q}{W}\sum_{k_y} T_m^c(E_l)v_x(E_l)f(E - E_{F,c}) \tag{5.18}$$

here T_m^c is the current transmission coefficient for cathode contact, and v_x is the velocity of electron in x direction. The net current due to electrons in mode m can be obtained by summing all transverse energy and longitudinal energy to obtain

$$I_m = \int_0^\infty I_m(E_l)dE_l \tag{5.19}$$

where

$$I_m(E_l) = \frac{q}{h}T_m^c(E_l)\frac{1}{\hbar a}\sqrt{\frac{2m_y^* k_B T}{\pi}}F_{1/2}(E_{F,c} - E_l) \tag{5.20}$$

The transmission coefficient for cathode in mode m is given by

$$T_m^c(E_l) = \frac{I_{transmitted}}{I_{incident}} = 1 - |r_m|^2 \tag{5.21}$$

From the boundary condition (5.10a) it is seen that the transmission coefficient can be expressed as

$$T_m^c(E_l) = 1 - |\phi_m(x=0) - 1|^2 \tag{5.22}$$

Similarly the current due to holes can be calculated and the total device current is found by adding electron and hole current.

5.11. Poisson's equation formulation

The formulation of the 2D matrix is not only conceptually difficult but in effect also quite difficult to formulate in MATLAB also, as the source term in the system of 2D equations are not as easy to comprehend. In fact, quite some manipulation is in order to bring the system of equations in a more standard array orientation.

To write this as a linear system in the more standard form $\Delta U=\rho$, where U and ρ are column vectors, a linear ordering of the unknowns $U(i,j)$ has to be chosen. For example, the natural row ordering shown below

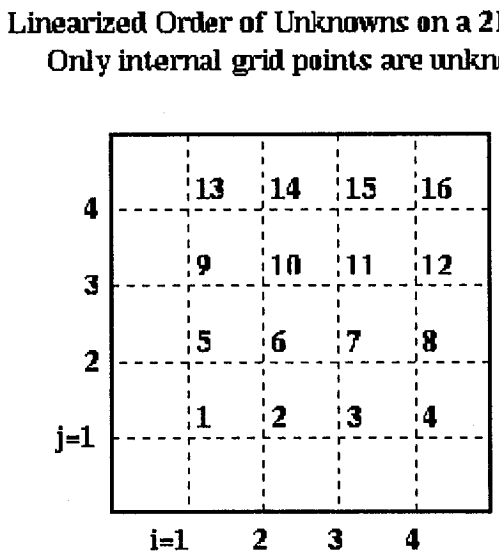


Figure 5.3: The ordering of unknown points in a 4X4 2D discrete space.

leads to the linear system $\Delta U=\rho$:

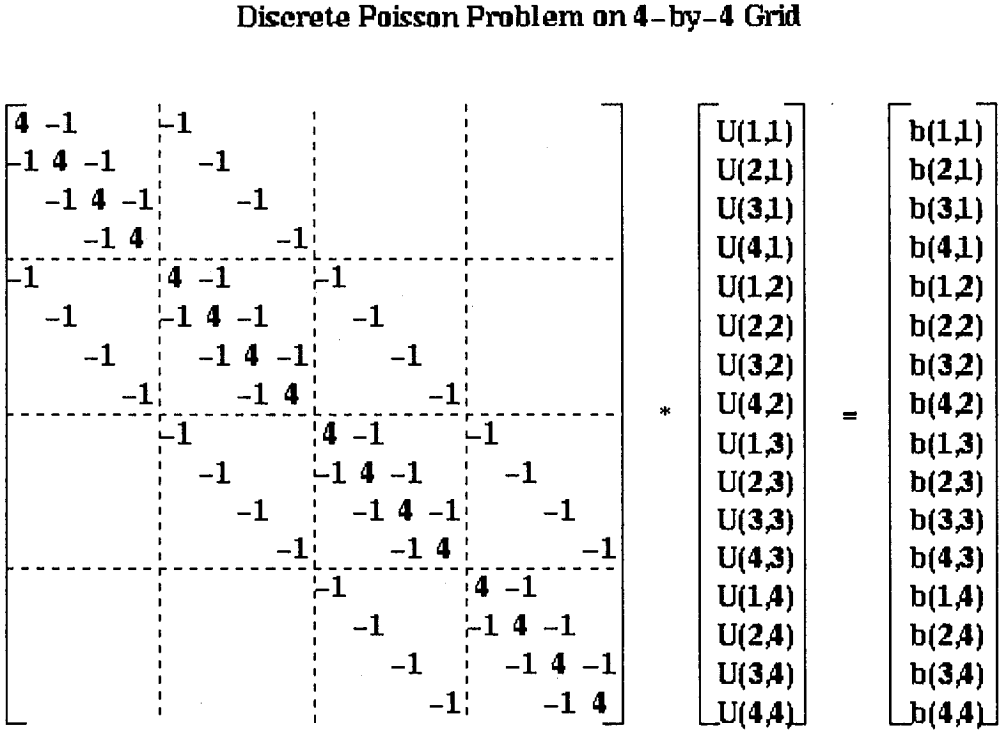


Figure 5.4: The source term reshaping and formation of the 2D matrix equation for the grid above.

It should be noted that in above mentioned example, the outer points are boundary which we consider to be zero for wave function as our well is infinite, so effectively the unknown points reduce to 16 instead of 25.

5.12. Biased system

Once the device under equilibrium has been simulated then the same procedure was adopted to simulate the device under forward bias only. In this case the reference was still chosen as $E_{Cn}=0$ and for an applied voltage V, the difference in Fermi levels were $E_{Fn}=qV+E_{Fp}$. The increment of voltage was done in steps of 0.01 volts and to construct the current voltage characteristics.

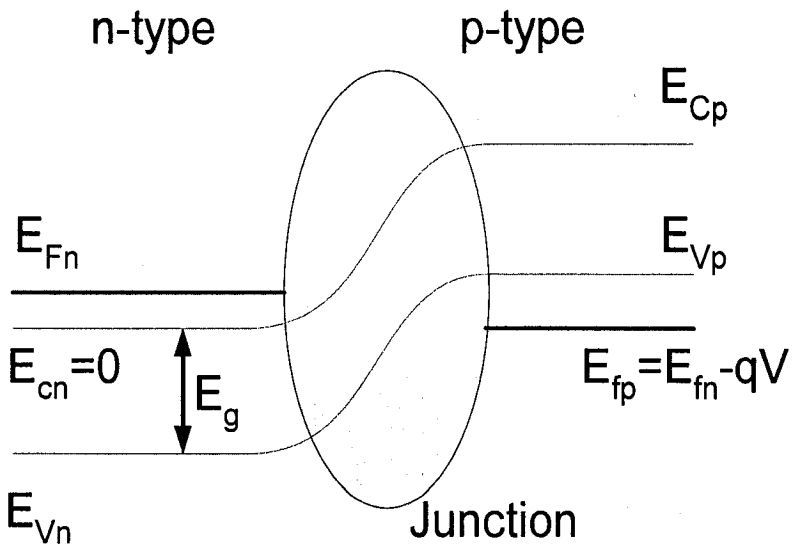


Figure 5.5: The schematic band diagram of a pn junction under forward bias condition. The applied voltage creates a difference in the electrode thermo chemical potential.

5.13. Self-consistent analysis

Self consistency is achieved as the successive calculated potential profile difference becomes lower than the preset allowed difference. As the energy grid has been set with a resolution of 0.5meV, this in fact sets the limit to the maximum achievable accuracy and any error resulting in lesser value is deemed acceptable. This then sets the convergence criteria.

First an initial guess was made for the system potential. The initial guess only plays a part in achieving convergence faster, and for a well formulated problem does not significantly pose a problem. With the guess potential, the quantum transport equations are solved to get the overall carrier density. Then along with the doping profile, this is inputted to the Poisson's equation to get the potential of the system. Then the new potential is compared with the previous potential, and if the difference is larger than the allowed error, the new potential

replaces the old one and quantum transport is solved again with this new potential. The process goes on until self-consistency is achieved.

5.14. Flowchart

The overall algorithm can be described as follows:

1. Guess a potential profile for the whole system.
2. With this potential profile, calculate the vertical slice confined 1D system equation to get the confined wave function and position dependant eigenenergy.
3. Using the position dependant eigenenergy as the effective local potential, solve the longitudinal system equation with Green's function formulation to incorporate contacts and carrier injection.
4. Calculate the carrier density in the device.
5. Using the carrier density and the doping profile, solve Poisson's equation in 2D to find the system potential.
6. Check whether the new calculated potential differs within the given range with the initial potential.
7. If self-consistency is not achieved, repeat steps 2-6.
8. Once self-consistency is achieved, calculate current and other relevant properties from carrier density and potential profile.

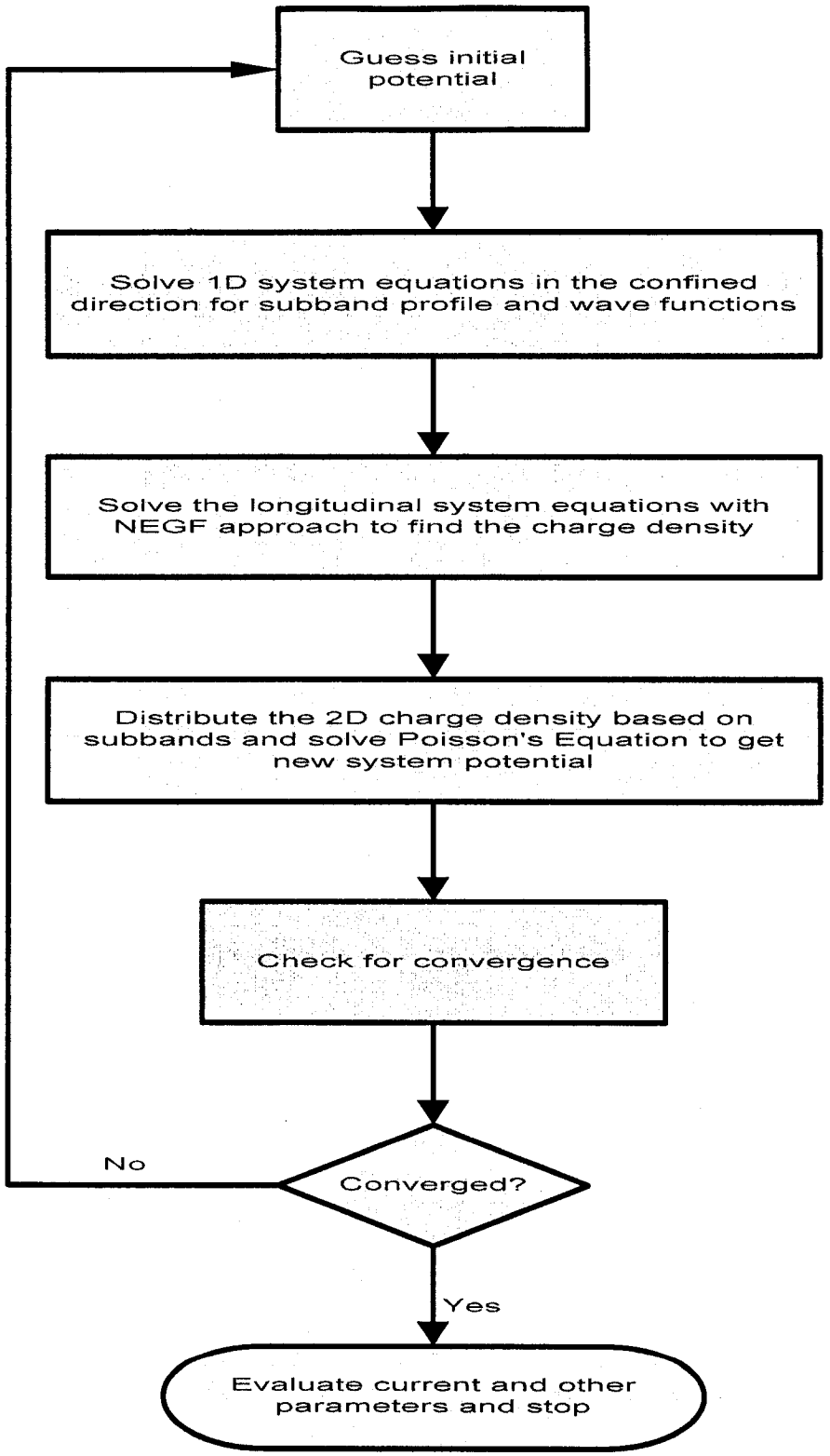


Figure 5.6: Flow chart of the decoupled system of 1D equations adopted in this work.

Chapter 6: Simulation results

In this chapter the key simulation results are presented which demonstrate the effect of quantum coupling. First the equilibrium characteristics are presented, for no external bias application, and then biased characteristics is presented.

6.1 Equilibrium potential profile

6.1.1. Built-in potential

A longitudinal section is selected along the device length, through the midpoint of the thickness, and the self-consistent potential is plotted. It can be seen that the difference in potential in the n-side and the p-side is the built-in potential, V_{bi}, and as for any degenerately doped junction, the magnitude is larger than the band gap energy.

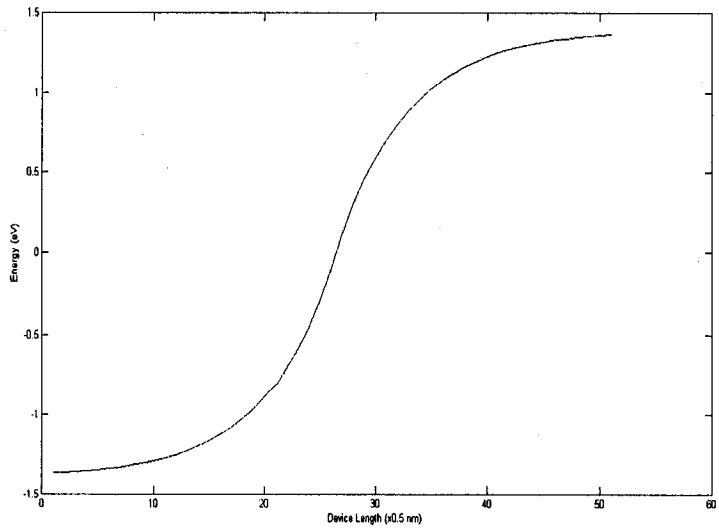


Figure 6.1: Built-in Potential along the device length axis

In equilibrium, a degenerately doped n-type material will have an equilibrium Fermi level within the conduction band. So the Fermi level is situated above the conduction band edge. On the other side, a degenerately doped p-type material will have equilibrium Fermi level within the valence band, hence it will be situated below the valence band edge. Now as in equilibrium, it is required that throughout the device, the Fermi energy be the same, so Fermi energy in the n-type material is exactly equal to the Fermi energy in the p-type material. So the p-type valence band edge will have higher energy than the n-type conduction band edge. Hence, the p-type conduction band being above the p-type valence band by band gap energy, E_g, the difference in conduction band edge between the two sides is given by

$$\Delta E = E_{cn} + E_{Fn} + E_{Fp} + E_g \tag{6.1}$$

So the built in potential is given by

$$V_{bi} = E_{Fn} + E_{Fp} + E_g \tag{6.2}$$

6.1.2 Depletion region

As seen in the above figure, there remain two quasi-neutral regions where the potential has no slope, while in the middle there exists a rise in potential from n-side to the p-side. This rise in potential, which is equal to the built in potential, is in fact the depletion region, where the material is depleted of free carriers. The length of the depletion region is calculated from the point where the potential starts to rise till the point where the potential becomes constant again. It can be

seen that the value of the depletion region width is about one order more than predicted by 3D equations (which is about 2-3nm),

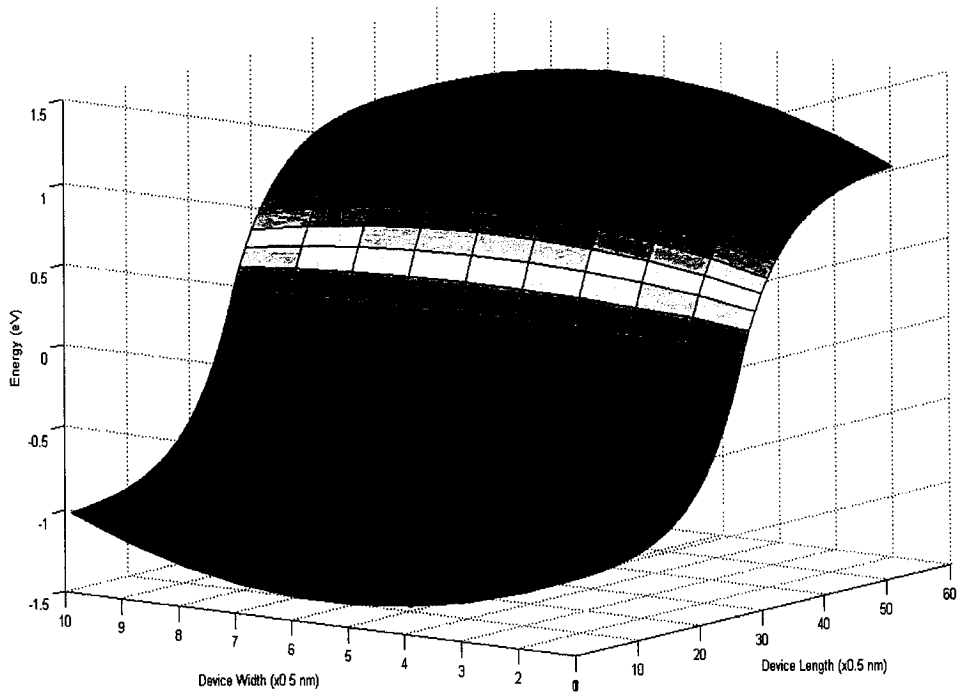


Figure 6.2: Conduction Band of the 2D pn junction in silicon

Most probably the reason is lack of screening of the electric field by the thin film device. Because the active device and the depleted dopant charges confined within are very thin in width, it only screens a small part of the potential field that surround the 3D space around the charges. For this reason the depletion region is extended beyond the prediction of otherwise classical calculation.

6.1.3 Lateral dependence of potential

Unlike the bulk counterpart, the results demonstrate a lateral dependence of device potential along the confinement direction. If a section along the confinement direction z is taken, and the potential profile is observed, this lateral dependence is observed.

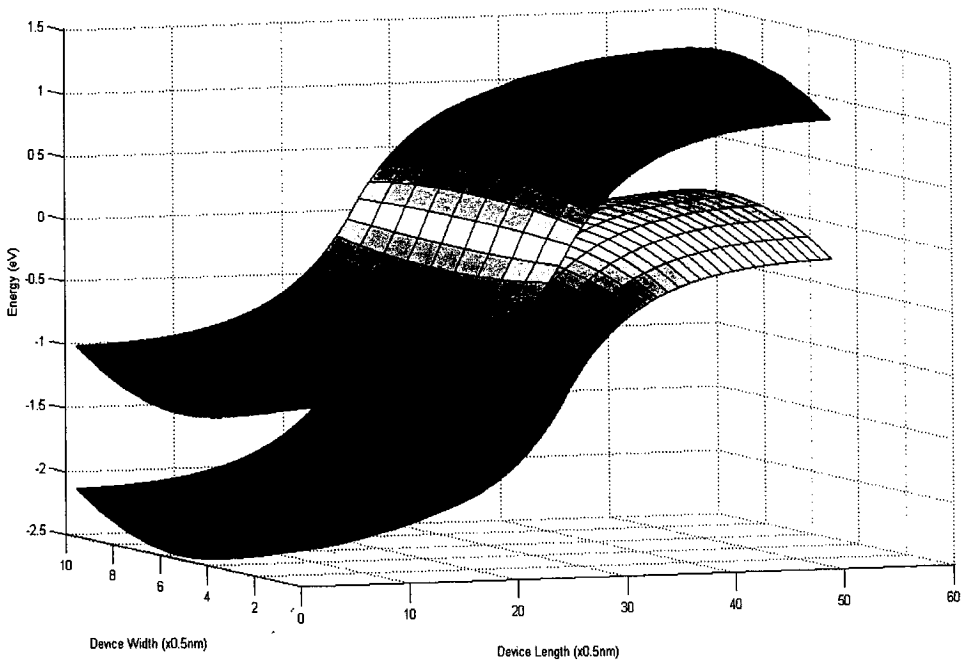


Figure 6.3: 2D Band Diagram of the 2D pn junction

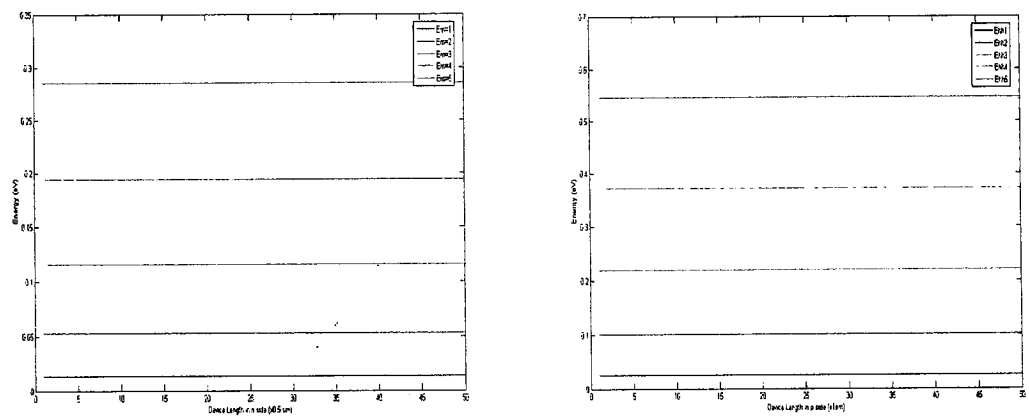
Most probably this lateral potential variation occurs for the quantum confinement of carriers. As the carriers are confined within the device width, the carriers tend to accumulate in the middle part of the thickness, and the device edges are left with fewer carriers. But the doping remains constant in the whole

device thickness. This gives a net charge even in the quasi neutral regions of the device along the lateral dimension and the potential is changed in accordance.

6.2 Equilibrium carrier concentration

6.2.1 Subband

As the device dimension in the lateral direction (z-axis) is only 5nm, which is smaller than the excitonic Bohr radius in silicon, the device operates in strong confinement regime. The results also uphold this fact as the formation of subbands is clearly seen. The subband energies are given below.



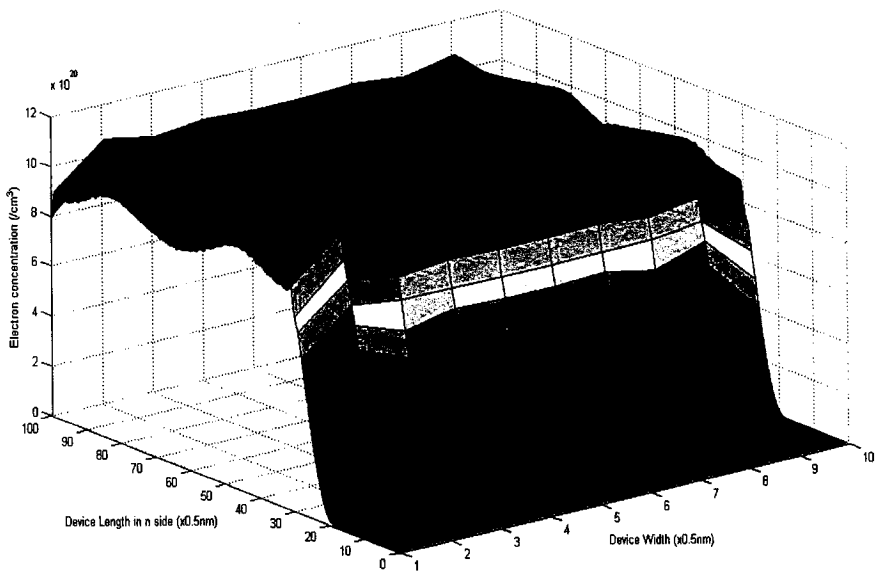
(a) (b)

Figure 6.4: Subband energy of (a) electrons in n-side, (b) holes in p-side.

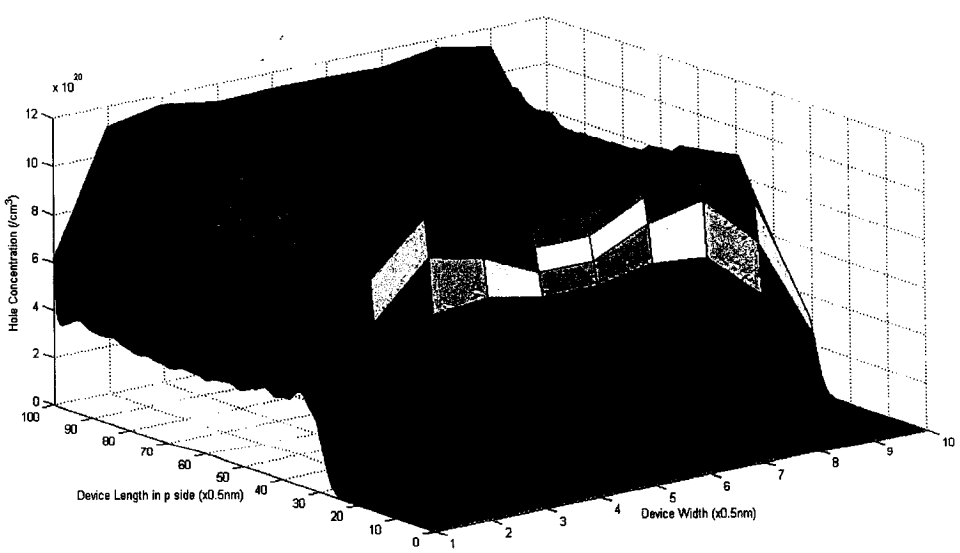
As the device is thin, the earlier assumption of the application of the decoupled mode was that only a few subbands are occupied. In the result it becomes evident that indeed only the first 3 subbands are occupied, verifying the earlier assumption and thereby validates the use of the decoupled system of equations.

6.2.2 Electron and hole concentration

The total electron and hole concentration is given below. It is clearly shown there exists a region around the metallurgical junction where no free carriers exist. This is the depletion region.



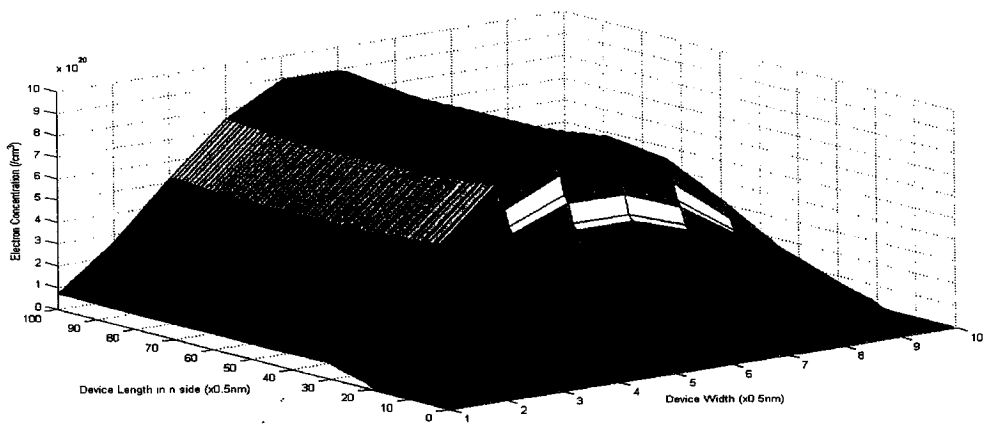
(a)



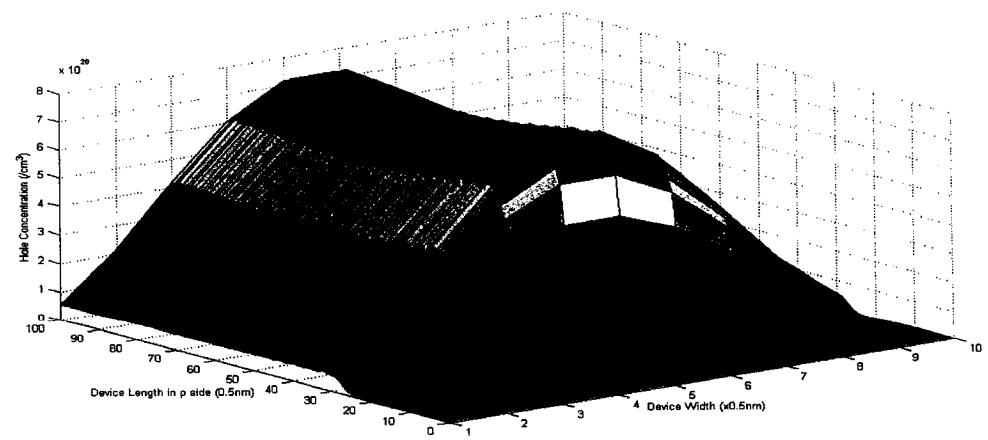
(b)

Figure 6.5: Total carrier concentration with no external bias (a) electron in n-side (b) hole in p-side.

Subband wise, the electron and hole concentration is given below. As the device is degenerately doped, the number of available carriers is comparable to the existing states, so more that one subband is occupied. In fact, in this simulation up to five subbands were considered, but it is found that only the first three are effectively occupied. Even if the higher subbands are injected with electrons and hole from their respective contacts, they immediately occupy the available lower subband to attain lower energy.

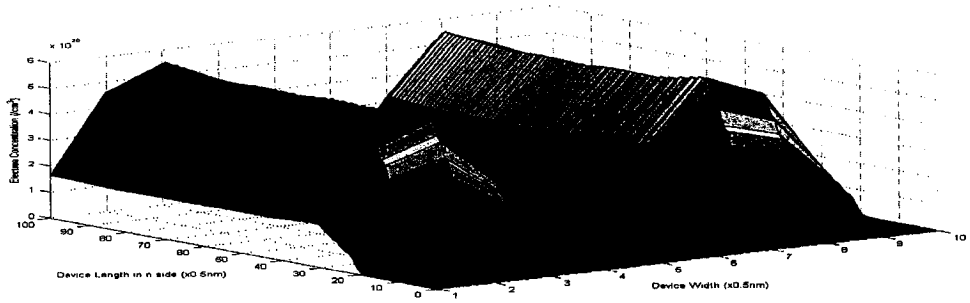


(a)

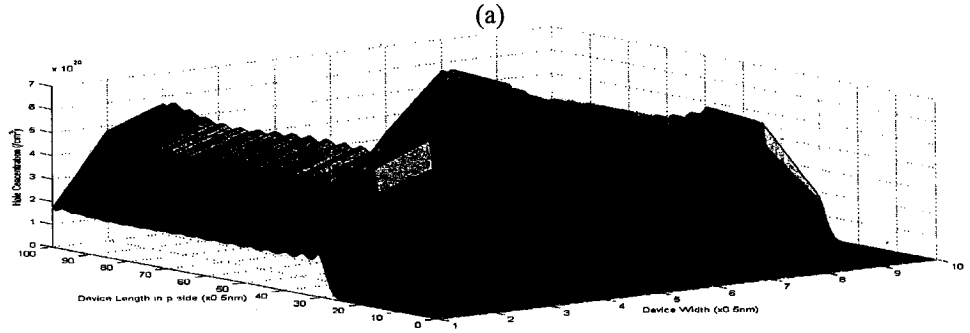


(b)

Figure 6.6: Carrier concentration in the first subband (a) electron in n-side, (b) hole in p-side.

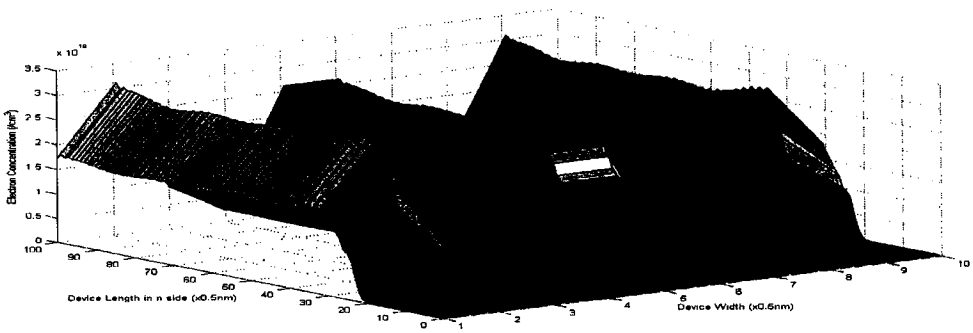


(a)

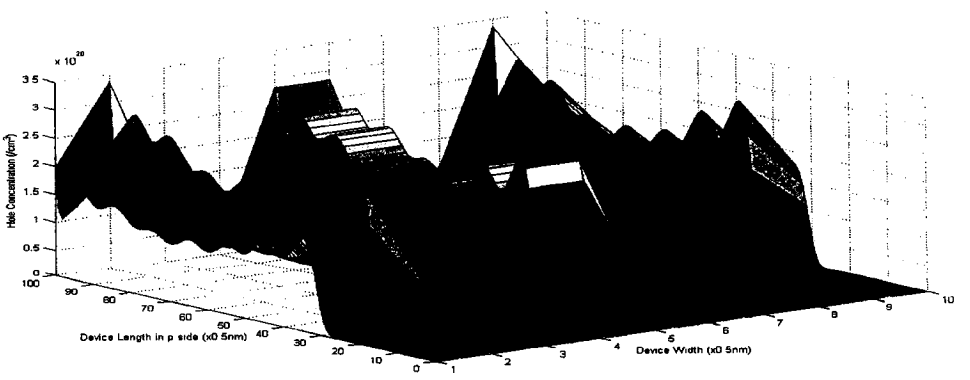


(b)

Figure 6.7: Carrier concentration in 2nd subband (a) electron in n-side, (b) hole in p-side.

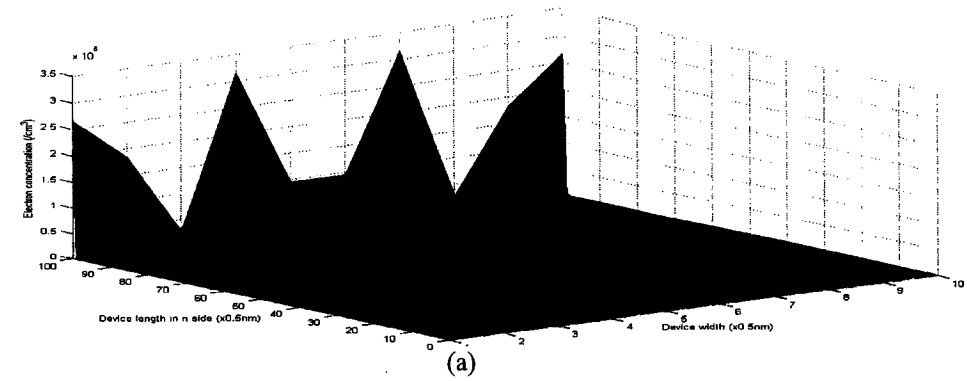


(a)

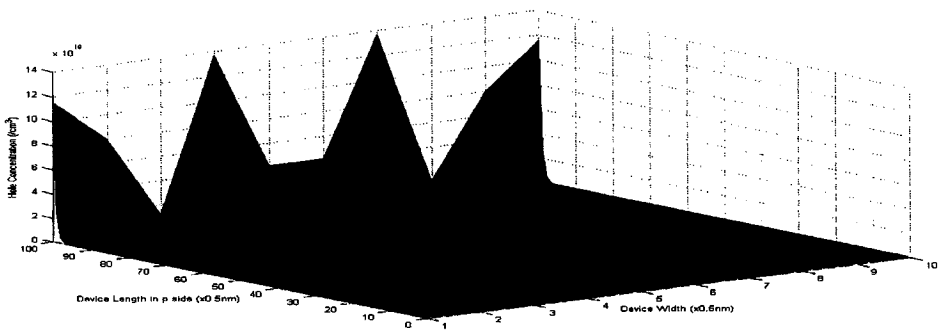


(b)

Figure 6.8: Carrier concentration in 3rd subband (a) electron in n-side, (b) hole in p-side.

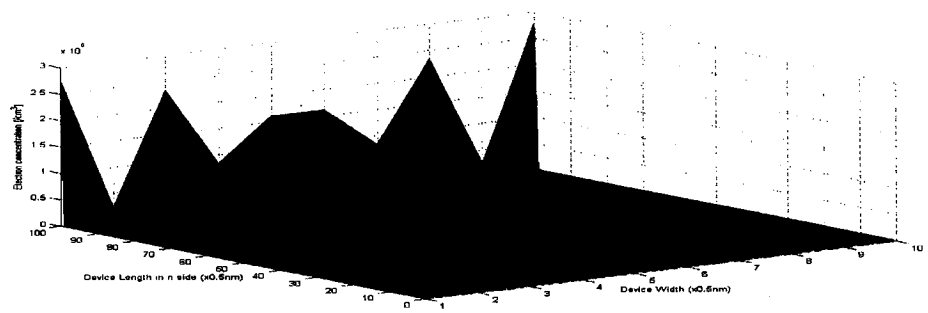


(a)

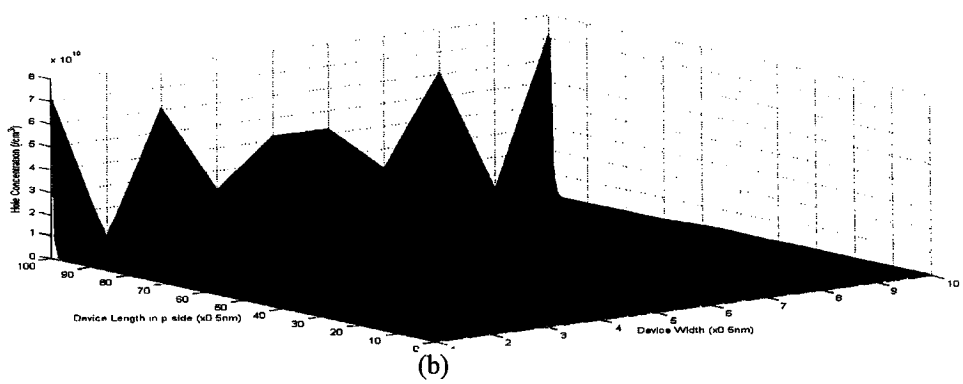


(b)

Figure 6.9: Carrier concentration in 4th subband (a) electron in n-side, (b) hole in p-**side.**



(a)



(b)

Figure 6.10: Carrier concentration in 5th subband (a) electron in n-side, (b) hole in p-side.

6.3 I vs. V curve

The current versus voltage curve is given below. A bias of 1.4 volts is applied and the current is plotted for the normalized voltage.

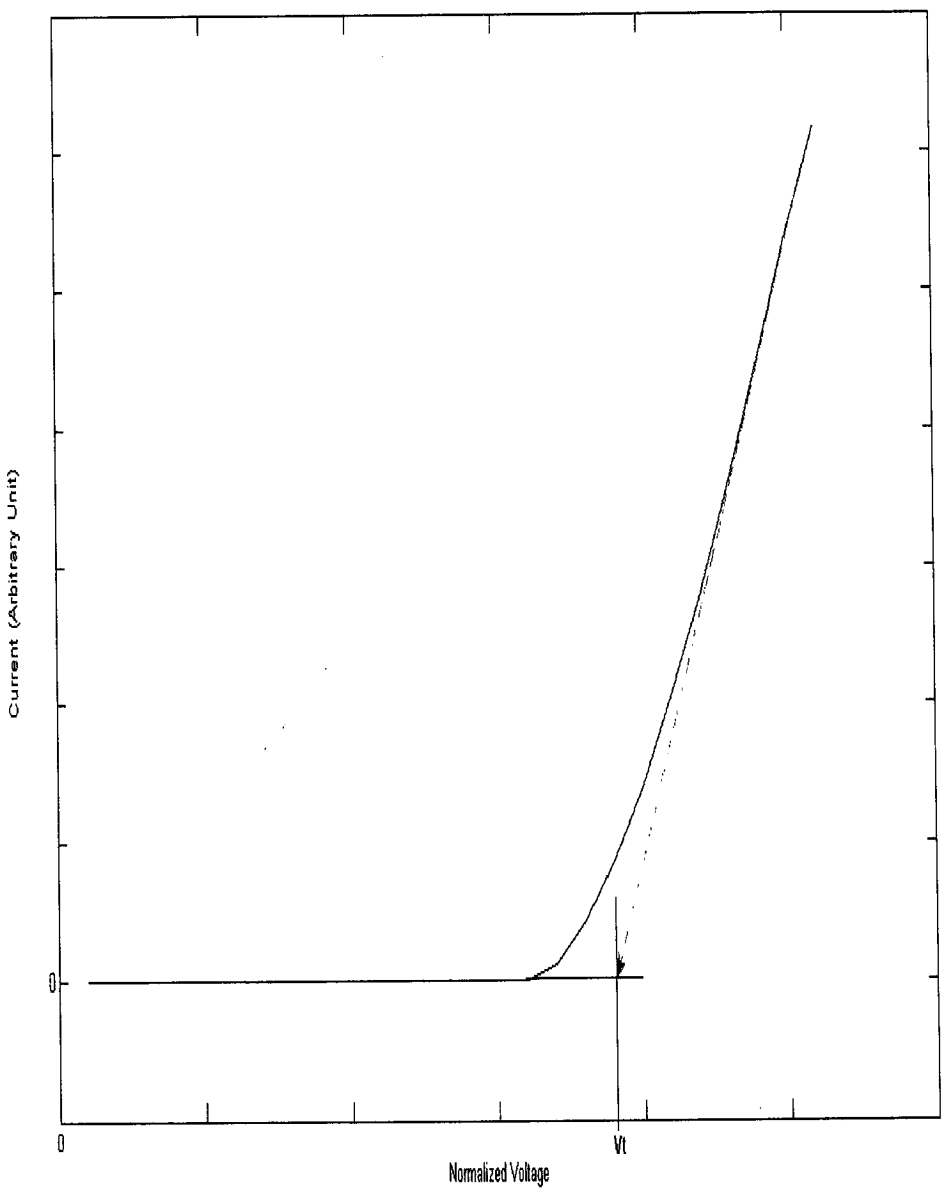


Figure 6.11: Current vs. Voltage curve of the 2D pn junction.

Chapter 7: Conclusion and Contribution

In this work an attempt is made to formulate and investigate a 2D quantum transport based simulation model for a thin film pn junction fabricated in silicon. The possibility of silicon nanostructures as optically viable semiconductor remains the main motivation, other solely electronic properties in such devices also need to be explored in these early stages. In that respect this work can be seen as an attempt which could pave the way for further probing.

There has been considerable work on developing quantum transport model for unipolar device structures like MOS, but not on bipolar devices. Most of these works concentrate on electron as carrier and at best, sometimes consider holes as the sole carrier. But in pn junction, both electrons and holes exist and create a junction with associated depleted region, which then creates the basic device. In order to have light generation, it is only obvious that a device must be injected with both electrons and holes to sustain carrier recombination. In that regard, pn junctions in reduced dimensional system may become important. In this work, a workable simulation model is developed for the pn junction in silicon thin film, employing a methodology which has had success in dealing with unipolar devices only. To the author's knowledge, no such work (simulation model of silicon based quantum well pn junction) has been done before. So it could be pointed out that the contribution of this work could be listed as follows:

- The 2D Schrödinger's equation is decoupled for a bipolar system into a set of 1D equation and applied simultaneously to both electrons and holes.
- A simulation model for 2D pn junction in silicon is developed.
- Simulation is carried for both equilibrium and biased condition.
- Future work is directed towards modeling light emission analysis.

7.1 Discussion

In this work a 2D pn junction is investigated. First a review of work done so far suggested that silicon can overcome its trademark of indirect bandgap material and contribute to light emission in nanoscale, and even a 1D confinement of a thin film could produce light. This worked as the motivation to develop a quantum transport model for such a device. The bulk pn junction usually deploys the drift diffusion model, incorporating the Boltzmann transport equation for transport modeling and Poisson's equation for electrostatics, but this approach fails to account for the quantum nature of carriers which becomes dominant in confined structures. To incorporate the quantum effects, a fuller quantum mechanical approach is taken here to generate a quantum transport model for the carriers. In this case as the system is effectively 2D, a full 2D direct discretization becomes computationally burdensome and may become impossible to apply for a fairly large system. So a decoupled set of 1D equations is used and the validity and assumptions are justified for the investigated structure. As device energy states

becomes broadened when they are connected to contacts, and to incorporate the contacts to the device, the open semi-infinite contact method and self energy matrices are used to introduce them in the equations. Wave injection is then considered to treat carrier injections and the decoupled set of 1D Schrödinger's equation is self-consistently solved with Poisson's equation to give a self-consistent result. The result shows that the carriers are largely confined within 3 subbands, which validates the assumption of a few occupied subband for a thin film, and the shape of the confined subband also remains almost the same along the device length, thus validating the decoupled method. The built in potential is found to be more than the band gap energy of the material as it is degenerately doped, and depletion region is found to be extended beyond of that predicted by 3D junction model. This result is perhaps expected as similar extended depletion region is also encountered for a metal-semiconductor junction. The current curve shows no significant tunneling current that is seen in a tunneling diode of the same doping level in 3D, perhaps for the extended depletion region which reduces the tunneling probability. In the end the results show carrier confinement and energy quantization; this could be the origin of direct recombination of the carriers to generate light in silicon.

7.2 Future work

In this work a basic quantum transport model is applied for the carriers in a 2D thin film pn junction. In doing so the single band effective mass equation is

used for the electrons. Since it is not clear what the formation of a full band approach (which is more accurate for the holes in the valence band) would take for a quantum confined structure, the multiband effective mass approach was taken for the holes. In future, this arena needs to be investigated and a better model for holes in confined structures could be presented. Also the effect of spin orbit coupling could to be explored.

Hot electron effects have to be investigated further for this device. The main mechanisms of recombination have to be examined and then the recombination current part of the total current would be determined more accurately.

As fabrication technologies are getting better by the day, nanowires, instead of thin films could be readily available in near future and so a pn junction in nanowire has to be investigated. In that case the model has to be modified for 2D confinement. As the computing resources are also becoming more powerful, a 3D device modeling may become feasible for such devices.

Reference

[1]. S. S. Iyer, Y.-H. Xie, "Light emission from silicon"; Science, Vol.260, pp40-46, 2 April 1993.

[2]. Roger A. Faulkner, "Toward a Theory of Isoelectronic Impurities in Semiconductors"; Physical Review, Vol.175, pp991-1009, 1968.

[3]. H. Ennen, J. Schneider, G. Pomrenke, and A. Axmann, "1.54-μm luminescence of erbium-implanted III-V semiconductors and silicon"; Applied Physics Letters, Vol.43, pp943-945, 1983.

[4]. U. Gnutzmann, K. Clausecker, "Theory of direct optical transitions in an optical indirect semiconductor with a super lattice structure"; Applied Physics A: Materials Science & Processing, Vol 3, No.1, pp9-13, 1974.

[5]. C. Pickering, M. I. J. Beale, D. J. Robbins, P. J. Pearson, R. Greef, "Optical studies of the structure of porous silicon films formed in p-type degenerate and non-degenerate silicon"; Journal of Physics C: Solid State Physics, Vol.17, No.35, pp6535-6552, 20 December 1984.

[6]. L. T. Canham, ""; Applied Physics Letters, Vo.57, Issue10, pp.1046-1048, 3 September 1990.

[7]. V. Lehmann, U. Gosele, "Porous silicon formation: A quantum wire effect"; Applied Physics Letters, Vol.58, Issue8, pp.856-858, 25 February, 1991.

[8]. Irrera, D. Pacifici, M. Miritello, G. Franzo, F. Priolo, F. Iacona, D. Sanfilippo, G. Di Stefano, P. G. Fallica. "Electroluminescence properties of light emitting devices based on silicon nanocrystals"; Physica E: Low-dimensional Systems and Nanostructures. Vol.16, Issues3-4, pp.395-399, March 2003.

[9]. T. Shimizu-Iwayama, K. Fujita, S. Nakao, K. Saitoh, T. Fujita, N. Itoh, "Visible photoluminescence in Si^+ -implanted silica glass"; Journal of Applied Physics, Vol.75, Issue12, pp.7779-7783, 15 June, 1994.

[10]. F. Priolo, G. Franzò, D. Pacifici, V. Vinciguerra, F. Iacona, A. Irrera, "Role of the energy transfer in the optical properties of undoped and Er-doped interacting Si nanocrystals"; Journal of Applied Physics, Vol.89, Issue1, pp.264-272, 1 January, 2001.

[11]. Y. Kanemitsu, "Efficient light emission from crystalline and amorphous silicon nanostructures"; Journal of Luminescence, Vol.100, pp.209-217, 2002.

[12]. Y. Kanemitsu, M. Iiboshi, T. Kushida, "Photoluminescence dynamics of amorphous Si/SiO_2 quantum wells"; Applied Physics Letters. Vol.76, Issue16, pp.2200-2202, April 17, 2000.

[13]. Y. Kanemitsu, T. Kushida, "Size effects on the luminescence spectrum in amorphous Si/SiO_2 multilayer structures", Applied Physics Letters, Vol.77, Issue22, pp.3550-3552, November 27, 2000.

[14]. Y. Kanemitsu, S. Okamoto, "Visible luminescence from silicon quantum dots and wells"; Material Science and Engineering B, Vol. 48, pp.108-115, 1997.

[15]. H. Kageshima, "Effects of oxygen termination on the electronic states of silicon quantum slabs"; Surface Science, Vol.357-358, pp.312-316, 1996.

[16]. S. B. Zhang, A. Zunger, "Prediction of unusual electronic properties of Si quantum films"; Applied Physics Letters, Vol.63, Issue10, pp.1399-1401, September 6, 1993.

[17]. G. Franzò, A. Irrera, E.C. Moreira, M. Miritello, F. Iacona, D. Sanfilippo, G. Di Stefano, P.G. Fallica, F. Priolo, "Electroluminescence of silicon nanocrystals in MOS structures"; Applied Physics A, Vol.74, No.1, January 2002.

[18]. A. Irrera, D. Pacifici, M. Miritello, G. Franzò, F. Priolo, F. Iacona, D. Sanfilippo, G. Di Stefano, P. G. Fallica, "Excitation and de-excitation properties of silicon quantum dots under electrical pumping"; Applied Physics Letters, Vol.81, Issue10, pp.1866-1868, September 2, 2002.

[19]. Z. Chen, G. Bosman, R. Ochoa, "Visible light emission from heavily doped porous silicon homojunction pn diodes"; Applied Physics Letters, Vol.62, Issue7, pp.708-710, February 15, 1993.

[20]. P. Steiner, F. Kozlowski, W. Lang, "Light-emitting porous silicon diode with an increased electroluminescence quantum efficiency"; Applied Physics Letters, Vol.62, Issue21, pp.2700-2702, May 24, 1993.

[21]. Y. Kanemitsu "Light emission from porous silicon and related materials"; Physics Reports, Vol. 263, pp1-91, 1995.

[22]. T. Takagahara, K. Takeda, "Theory of the quantum confinement effect on excitons in quantum dots of indirect-gap materials"; Physical Review B, Vol.46, Issue23, pp.15578–15581, December 15, 1992.

[23]. K. Ito, S. Ohyama, Y. Uehara, S. Ushioda, "Visible light emission spectra of individual microstructures of porous Silicon"; Applied Physics Letters, Vol.67, Issue17, pp.2536-2538, October 23, 1995.

[24]. G. D. Sanders, Y. Chang, "Theory of optical properties of quantum wires in porous silicon"; Physical Review B, Vol.45, No.16, April 15, 1992.

[25]. G. D. Sanders, Y. Chang, "Optical properties of free-standing silicon quantum wires"; Applied Physics Letters, Vol.60, Issue 20, May 18, 1992.

[26]. G. Sun, L. Friedman, R. A. Soref, "Intersubband lasing in silicon-based quantum well structures"; Superlattices and Microstructures, Vol.22, No.1, 1997.

[27]. G. Sun, L. Friedman, R. A. Soref, "Intersubband lasing lifetimes of SiGe/Si and GaAs/AlGaAs multiple quantum well structures"; Applied Physics Letters, Vol.66, Issue 25, pp.3425-3427, June 19, 1995.

[28]. G. Sun , L. Friedman, "Heavy-hole scattering by confined nonpolar optical phonons in a single $Si_{1-x}Ge_x$/Si quantum well"; Physical Review B, Vol.53, Issue7, pp.3966–3974, February 15, 1996.

[29]. N. Tsutsui, V. Ryzhii, I. Khmyrova, P. O. Vaccaro, H. Taniyama, T. Aida, "High-frequency performance of lateral p-n junction photodiodes"; IEEE journal of Quantum Electronics, Vol.37, No.6, June 2001.

[30]. A. Sh. Achoyan, A. Ê. Yesayan, Ê. M. Kazaryan, S. G. Petrosyan, "Two-dimensional p-n junction under equilibrium conditions"; Semiconductors, Vol.36, No.8, pp.903-907, 2002.

[31]. H. Kosina, G. Klimeck, M. Nedjalkov, S. Selberherr, "Comparison of Numerical Quantum Device Models"; International Conference on Simulation of Semiconductor Processes and Devices (SISPAD 2003), Boston, MA, pp.171 -174, September 3-5, 2003.

[32]. M. G. Ancona, "Quantum correction to the equation of state of an electron gas in a semiconductor"; Physical Review B, Vol.39, No.13, pp. 9536-9540, 1989.

[33]. H. L. Grubin, J. P. Kreskovosky, "Quantum moment balance equations and resonant tunneling structures"; Solid State Electronics, Vol.32, No.12, pp.1071-1075, 1989.

[34]. H. L. Grubin, J. P. Kreskovosky, T. R. Gocindan, D. K. Ferry, "Uses of the quantum potential in modeling hot-carrier semiconductor devices"; Semiconductor Science and Technology, Vol.9, pp.855-858, 1994.

[35]. M. J. Van Dort, et al., "A simple Model for Quantization Effects in Heavily-Doped Silicon MOSFETs at Inversion Conditions," Solid-State Electron, 1994.

[36]. F. Stern, "Self-consistent results for n-type Si inversion layers"; Physical Review B, Vol.5, No.12, June 15, 1972.

[37]. J. H. Luscombe, A. M. Bouchard, "Electron confinement in quantum nanostructures: self-consistent Poisson-Schrödinger theory"; Physical Review B, Vol.46, No.16, October 15, 1992.

[38]. L. Shifren, D. Ferry, "Wigner function quantum Monte Carlo"; Physica B, Vol.314, pp.72-75, 2002.

[39]. R. Lake, G. Klimeck, R. Bowen, D. Jovanovic, "Single and multiband modeling of quantum electron transport through layered semiconductor devices"; Journal of Applied Physics, Vol.8, pp.7845-7869, 1997.

[40]. S. Datta, "Electronic Transport in Mesoscopic Systems"; (Cambridge University Press, UK, 1997).

[41]. S. Datta, "Nanoscale device modeling: the Green's function method"; Superlattices and Microstructures, Vol. 28, No. 4, 2000.

[42]. R. Venugopal, Z. Ren, S. Datta, M. S. Lundstrom, "Simulating quantum transport in nanoscale transistors: Real versus mode-space approaches"; Journal of Applied Physics, Vol.92, No.7, October 2002.

[43]. S. Datta, "Quantum Transport: Atom to Transistor"; Cambridge University Press.

Appendix

Sample MATAB code

```
_____Inputs (MKS Units)_____

Length=1e-7;          % Device length in X direction
Thickness=5e-9;       % Device thickness in Z direction
Width=5e-6;           % Device width in Y direction
Oxide=5e-9;           % Oside thickness in Z direction
a=5e-10;              % Slice width
b=5e-10;              % Mesh point distance in each vertical slice
c=5e-10;              % Mesh point distance in oxide
mode=5;               % Number of subbands to be considered
Volt=1.4;             % Applied Voltage

_____Constants (units in MKS except energy in eV)_____

K=8.61735E-5;         % Boltzmann constant
q=1.6e-19;            % Electronic charge
hbar=1.0544e-34;      % Reduced Plank's constant
mex=0.916;            % Effective electron mass in X direction
mey=0.916;            % Effective electron mass in Y direction
```

```
mez=0.916;          % Effective electron mass in Z direction
mpx=0.916;          % Effective hole mass in X direction
mpy=0.916;          % Effective hole mass in Y direction
mpz=0.916;          % Effective hole mass in Z direction
melec=9.1e-31;      % Electron rest mass in kg
T=300;              % Absolute Temperature in Kelvin
epsilon0=8.8542e-12;  % Permittivity of free space
epsilonr=11.8;      % Relative permittivity of silicon
epsilonox=12.8;     % Relative permittivity of oxide

_____Calculated or chosen intermediate variables_____

Nx=round(Length/a);          % Number of vertical slices
Nz=round(Thickness/b);       % Number of nodes in device slice
Nox=round(Oxide/c);          % Number of nodes in oxide
Nthick=Nz+2*Nox;             % Number of nodes in each vertical slice
kt=K*T;                      % Thermal energy
Ae=2*mez*melec/(hbar)^2;     % Coefficient of Electron Schrodinger Eq.
%Ap=2*mpz*melec/(hbar)^2;  % Coefficient of Hole Schrodinger Eq.
tez=(hbar^2)/(2*mez*melec*b^2*q);  % Electron coupling constant
tex=(hbar^2)/(2*mex*melec*a^2*q);  % Electron coupling constant
%tpz=(hbar^2)/(2*mpz*melec*b^2*q);  % Hole coupling constant
%tpx=(hbar^2)/(2*mpx*melec*a^2*q);  % Hole coupling constant
Efn=0.4;                     % Fermi energy of anode (reference value)
```

```
Ecn=0.001;                % Conduction band of cathode side (reference value)
Efp=Efn-Volt;             % Fermi energy of cathode (reference value)
Ecp=0.001;                % Conduction band edge of n side (reference value)
Em=zeros(Nx,mode);        % Eigenenergy in each vertical slice
U=zeros(Nthick,Nx);       % Potential profile
Eleamax=0.125;            % Maximum injection energy for electron from anode
Eleamin=Ecn;              % Minimum injection energy for electron from anode
Elecmax=0.125;            % Maximum injection energy for electron from cathode
Elecmin=Ecp;              % Minimum injection energy for electron from cathode
Eldiff=5e-4;              % Energy grid spacing
Elp_max=0.8*tpx;          % Maximum injection energy for hole
Elp_min=0.001;            % Minimum injection energy for hole
Elp_diff=5e-4;            % Energy grid spacing for hole
sigma1=zeros(Nx);         % Self energy matrix for Anode
sigma2=zeros(Nx);         % Self energy matrix for Cathode
gama1=zeros(Nx,1);        % Source matrix for electron injection
nelec=zeros(Nx,mode);     % Electron concentration in the device
nhole=zeros(Nx,mode);     % Hole concentration in the device
NE_anode=round((El_electron_anode_max-
El_electron_anode_min)/El_electron_diff);% Energy Grid for electron
NE_cathode=round((El_electron_cathode_max-
El_electron_cathode_min)/El_electron_diff);% Energy Grid for electron
NP=round((Elp_max-Elp_min)/Elp_diff);% Energy Grid for hole
```

```
phiz=zeros(Nx,Nz,mode);          % Eigenfunction in each vertical slice
phixa=zeros(round((Eleamax-Eleamin)/Eldiff),Nx,mode);
phixc=zeros(round((Elecmax-Elecmin)/Eldiff),Nx,mode);
nanodemode=zeros(Nx,mode);
nanodem=zeros(Nx,Nz,mode);
nanode=zeros(Nz,Nx)';          % 2 dimensional electron density for anode injection
ncathodemode=zeros(Nx,mode);
ncathodem=zeros(Nx,Nz,mode);
ncathode=zeros(Nz,Nx)';      % 2 dimensional electron density or cathode injection
eprob=zeros(Nz,Nx)';          % 2 dimensional electron probability
nelec=zeros(Nx,Nz);          % 2 dimensional electron density
nhole2D=zeros(Nx,Nz,mode);
nhole3D=zeros(Nz,Nx)';       % 3 dimensional hole density
phole3D=zeros(Nz,Nx)';       % 3 dimensional hole probability
u=U;                         % Calculated potential from Poisson's Equation
epsilon=epsilon0*epsilonr;   % Permittivity of silicon
epsilonx=epsilon0*epsilonox; % Permittivity of oxide

_____2D Laplacian matrix for 2D Poisson's Equation_____
d=zeros(Nthick+2,Nx+2);
ind=1;
for jj=2:Nthick+1;
   for ii=2:Nx+1;
```

```
      d(jj,ii)=ind;
      ind=ind+1;
   end
end
d=-delsq(d);
for ii=(Nx*Nox)+1:Nx:Nx*(Nox+Nz);
   d(ii,ii)=-3;
   d(ii,ii-1)=0;
   d(ii+Nx-1,ii+Nx-1)=-3;
   d(ii+Nx-1,ii+Nx)=0;
end
dinv=inv(d);

_____Energy grid for longitudinal Energy_____

Elea=linspace(Eleamin,Eleamax,round((Eleamax-Eleamin)/Eldiff));
Elec=linspace(Elecmin,Elecmax,round((Elecmax-Elecmin)/Eldiff));

_____Doping profile calculation_____

Ndevice=-(1e21)*ones(Nz,Nx);
for ii=1:Nx/2;
   Ndevice(:,ii)=1e21;
end
```

```
_____Self consistent solution_____

for mm=0:.05:0.4;
   Volt=mm,
itteration=1,
error=inf,
check=1;
while error>Eldiff
        <statements>
end


_____Device Potential_____

   Udevice(:,:)=U(Nox+1:Nox+Nz,:);    % Potential inside the device
   Uave=mean(Udevice);                % For carrier injection in 1D


_____ Lateral Hamiltonian matrix for each vertical slice_____

for ii=1:Nx;                  % For each slice...
     hz=2*tez*diag(ones(1,Nz))-(tez*diag(ones(1,Nz-1),1))-(tez*diag(ones(1,Nz-
1),-1))+diag(Udevice(:,ii));
     [V,D]=eig(hz);
       for jj=1:mode;          % For each mode in a particular slice.....
          phiz(ii,:,jj)=V(:,jj);  % Wavefunctions of modes in different slices
          Em(ii,jj)=D(jj,jj);     % eigenenergy of each mode at different slices
          eprobm=abs(phiz(ii,:,jj)).^2;
```

```
        end
end

_____Hamiltonian matrix in longitudinal direction_____

for ii=1:round((Eleamax-Eleamin)/Eldiff);  % For each energy.....
    coska=1-((Elea(ii)-Uave(1))/(2*tex));
    ka=acos(coska);
    sigma1(1,1)=-tex*exp(i*ka);
    coska=1-((Elea(ii)-Uave(Nx))/(2*tex));
    ka=acos(coska);
    sigma2(Nx,Nx)=-tex*exp(i*ka);
    gama1=diag(i*(sigma1-sigma1'));
    for jj=1:mode;          % For each mode in a particular longitudinal energy
        hx=(2*tex*diag(ones(1,Nx))+diag(Em(:,jj)))-(tex*diag(ones(1,Nx-
1),1))-(tex*diag(ones(1,Nx-1),-1));
        G=Elea(ii)*eye(Nx)-hx-sigma1-sigma2;
        phixa(ii,:,jj)=-i*G\gama1;
        n(:,jj)=((sqrt(2*mey*melec*kt/pi))*(abs(phixa(ii,:,jj)).^2)*(fermi(Efn-
Elea(ii),-0.5)))/(hbar*2*pi*gama1(1,1));
1).^2)).*(sqrt(2*mey*melec*kt/pi)).*(fermi(Ef-Ele(ii),-0.5));
        end
        nanodemode=n+nanodemode;
end
```

```
_____Calculating 2D electron density_____

for ii=1:Nx;
    for jj=1:mode;
        nanodem(ii,:,jj)=nanodemode(ii,jj)*((abs(phiz(ii,:,jj))).^2);
    end
end
for ii=1:mode;
    nanode=nanodem(:,:,ii)+nanode;
end


_____Poisson's Equation _____

rho=[zeros(Nox,Nx);(nanode'-Ndevice);zeros(Nox,Nx)];
rhovect=reshape(rho',[],1);
u=dinv*rhovect;
beta=-q*a^2/epsilon;
u=beta*u;
Upoisson=reshape(u,[],Nthick)'./[epsilonx*ones(Nox,Nx);epsilon*ones(Nz,Nx);ep
silonx*ones(Nox,Nx)];
```