The Role of the Collisional Broadening of the States on the Low-Field Mobility in Silicon Inversion Layers

by

Gokula Kannan Jayaram Thulasingam

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved November 2017 by the
Graduate Supervisory Committee:

Dragica Vasileska, Chair
David Ferry
Stephen Goodnick
David Allee

ARIZONA STATE UNIVERSITY

December 2017

ABSTRACT

Scaling of the Metal-Oxide-Semiconductor Field Effect Transistor (MOSFET) towards shorter channel lengths, has lead to an increasing importance of quantum effects on the device performance. Until now, a semi-classical model based on Monte Carlo method for instance, has been sufficient to address these issues in silicon, and arrive at a reasonably good fit to experimental mobility data. But as the semiconductor world moves towards 10nm technology, many of the basic assumptions in this method, namely the very fundamental Fermi's golden rule come into question. The derivation of the Fermi's golden rule assumes that the scattering is infrequent (therefore the long time limit) and the collision duration time is zero. This thesis overcomes some of the limitations of the above approach by successfully developing a quantum mechanical simulator that can model the low-field inversion layer mobility in silicon MOS capacitors and other inversion layers as well. It solves for the scattering induced collisional broadening of the states by accounting for the various scattering mechanisms present in silicon through the non-equilibrium based near-equilibrium Green's Functions approach, which shall be referred to as near-equilibrium Green's Function (nEGF) in this work. It adopts a two-loop approach, where the outer loop solves for the self-consistency between the potential and the subband sheet charge density by solving the Poisson and the Schrödinger equations self-consistently. The inner loop solves for the nEGF (renormalization of the spectrum and the broadening of the states), self-consistently using the self-consistent Born approximation, which is then used to compute the mobility using the Green-Kubo Formalism.

To my Amma and Appa

ACKNOWLEDGMENTS

This work has been wholly motivated by my mentor and guide, Dr. Dragica Vasileska without whose support it would not have taken this form. Her patience in helping me understand new concepts, and more importantly, the trust she placed in my capability to learn such a difficult field has played a significant role in shaping this work.

I am grateful to Dr. David Ferry and Dr. Stephen Goodnick for being a part of my Graduate Advisory Committee. I would also like to extend my sincere thanks to the kindness shown to me by Dr. David Allee. Without his support, I would have found it extremely hard to secure a full Teaching Assistantship throughout the majority of this work. I would also like to thank the Late Prof. D.K.Schroder for his absolutely stunning lectures, and discussion that inspired and motivated me to learn Device Physics. I would also like to thank the High Performance Computing Cluster at Arizona State University, and the Extreme Science and Engineering Discovery Environment (XSEDE), for their support, without which this computationally intensive work would not have been feasible. I would like to extend my appreciation to the School of Electrical, Computer and Energy Engineering at Arizona State University for providing me this opportunity to pursue my PhD degree. I also take this opportunity to thank Darleen Mandt , Lynn Pratte and Esther Korner for helping me with all the official documents.

I would like to thank all my colleagues in the Computational Electronics group for their help. I am totally indebted to my family for the love and support I received in my quest for higher education. I would also like to thank all my friends, especially for their support without which I would not have seen through the tough times.

TABLE OF CONTENTS

Page

LIST OF TABLES vii

LIST OF FIGURES viii

CHAPTER

1 INTRODUCTION 1

1.1. Transistor Scaling 1

1.2. Transport in Semiconductors 4

1.2.1. Overview of Semiclassical Transport 4

1.2.2. Failure of the BTE 7

1.2.3. Quantum Transport- an Overview 11

1.3. Summary 21

2 GREEN'S FUNCTION FORMALISM 23

2.1. Review of the Formalism – an Introduction 23

2.2. The Many Body Problem 23

2.3. Green's Function Formalism 25

2.4. Hamiltonian and Second Quantization 27

2.5. Green's Function 29

2.5.1. Time Evolution Pictures 29

2.5.2. Contour-Ordered Green's Functions 30

2.6. Equations of motion for the Green's function 35

2.7. Evaluation for Green's Functions 38

CHAPTER Page

3 REVIEW ON SCHRED AND SCATTERING MECHANISMS IN SILICON 43

3.1. SCHRED - Recap. 43

3.1.1. Models and Features 43

3.1.2. The Poisson Equation..48

3.1.3. Time independant Schrödinger Wave Equation 50

3.1.4. 2D Sheet charge density and total density 51

3.1.5. Flow Chart of the SCHREDV2.0 program 52

3.2. Scattering in Silicon - Diffusive Trasnport..54

3.2.1. Coulomb Scattering..54

3.2.2. Surface-Roughness Scattering 56

3.2.3. Electron-Phonon Interaction 59

3.2.4. Deformation Potential Scattering...60

3.2.5. Nonpolar Optical Phonon Scattering 63

4 MANY BODY EFFECTS AND CONDUCTIVITY 66

4.1. Review on Many-body effects..66

4.1.1. Screening under Random Phase Approximation 67

4.1.2. The Exchange and Correlation effects in Hartree Theory 70

4.2. Review on Conductivity 72

4.2.1. Linear Response Theory 73

5 SCHRED INTEGRATION AND IMPLEMENTATION OF GF CORE 79

5.1. Coupling of the nEGF Solver to the SCHRED V2.0 Solver 79

CHAPTER Page

5.2. MPI – Parallelization of the nEGF Energy Integration 79

5.3. Electron-Phonon Self-Energy Implementation..85

5.4. Flowchart of the Overall Program 89

6 SIMULATION RESULTS 96

6.1. Scattering rates in the First Born approximation 96

6.1.1. Broadening of the States across MOSFET Generations 97

6.2. Results of the SCHRED-nEGF Code 101

6.2.1. The real DOS – Collisional Broadening of the States.....................102

6.2.2. DOS for Self-consistent phonon calculation..................................103

6.2.3. Mobility plot 104

7 CONCLUSION AND FUTURE WORK 108

REFERENCES 110

LIST OF TABLES

Table Page

3.1: Degeneracy Factors for Transition Between Unprimed and Primed Subbands, for Both g- and f-phonons. 65

5.1: Simulation Time 82

LIST OF FIGURES

Figure Page

1.1: Transistor Innovations for the Technology Generations – from Intel 22nm Announcement Presentation…………………………………………..……….…... 3

1.2: Green's Function Based Approaches Usage……………………………………..15

2.1: Feynman's Diagram…………………………………………………………….. 41

3.1: Schematic Description of the Three Orthogonal Coordinate Systems: Device Coordinate System (Dcs), Crystal Coordinate System (Ccs), and Ellipse Coordinate System (Ecs). From Lundstrom and Co-workers, with Permission………………. 44

3.2: Subband Structure in the Inversion Layer of Regular and Surface-channel Strained-si Layer……………………………………………………………………. 47

3.3: Flow Chart of SCHREDV2.0……………………………………………….... 53

3.4: Normalized Magnitudes for Gaussian(Red) and Exponential(Blue) Models for Roughness Parameters $\Delta = 1.5\ nm$ and $L = 0.243\ nm$ ………………………………58

3.5: 3D Surface Roughness Model for the Power Spectrum – (a) Gaussian Spectral Model, (B) Exponential Spectral Model for $\Delta = 1.5\ nm$ and $L = 0.243\ nm$ ………...59

3.6: g-type and f-type Phonons Transitions……………………………………….. 65

4.1: Feynman Diagram for the Effective Screened Interaction (Polarization Diagram). 68

4.2 : Dyson's Equation for the Screened Interaction in Diagrammatical View…………69

4.3: Screening Wavevector q_s Vs q vector. E1 is a Sample Subband Energy Level, E_F Is the Fermi Level………………………………………………………………70

4.4: Dyson's Equation for Retarded/Advanced Green's Function…………………….. 75

Figure Page

4.5: Ladder Diagram for the Interaction Between the Green's Function(Solid Lines) and the (Impurity) Scattering Event (Dashed Lines) that Contribute to Conductivity…75

5.1: Speedup Obtained per Iteration of the Inner Negf Loop……………………….. 81

5.2: Self-consistent Born Approximation for Electron-phonon Interaction………… 87

5.3: Overall Program………………………………………………….……………… 92

5.4: nEGF Loop Solver - with the MPI Core for Calculating the E_F Integration..…… 93

6.1: First Born Approximation Scattering Rate…………………..…………………... 97

6.2: Left Panel N_A= 10^{17}cm^{-3}, Right Panel N_A= 10^{18}cm^{-}3. First Row Panels: Effective (Real) Dos of the Lowest Three Subbands. Secod and Third Row Panels: Simulation Run with and Without Different Scattering Mechanisms. Fourth Row Panel: Comparision of Simulation Run with All Scattering Mechanisms Included for 2 Sheet Charge Densities…….…………………………………….………………… 98

6.3: Real DOS Vs Energy for Different Doping Concentrations………..……………102

6.4: DOS Comparison Between Ideal DOS (Black Solid Lines), DOS Without Phonon Self-consistency (no-ph), and DOS with Self-consistent Inclusion of Phonon Scattering (ph)………………………………..……………………………103

6.5: Mobility Vs Electric Field in Silicon for Different Doping Generations………....105

6.6: Mobility Vs Electric Field Comparison Between Real Dos with Collisional Broadening of States (CBS) and Ideal Dos Without CBS…………………….. 106

CHAPTER 1. INTRODUCTION

1.1. TRANSISTOR SCALING

Everyday life today revolves around electronics - from phones, tablets, laptops to complicated and larger machinery like cars that require ultimate precision for its operation through the use of computers. The need for more computing power with greater accuracy and increased reliability was enabled by decreasing size of the semiconductor chips (Moore's Law's [1]). And this law has been the driving force of the semiconductor industry for the last few decades. With the advent of the present day technology, which is shifting more towards mobile computing for everyday communication, the need for reliable but accurate computing has become inevitable. Added to this, the increasing use of the internet and the internet based software on handheld devices demands a great performing hardware unit packed in dimensions less than the size of your palm.

The history of semiconductors dates back to the end of the 20th century when semiconductors were used as detectors in radios based on the not so reliable Schottky diode. Further research in the field, lead to devices better in reliability and performance such as the PN diode, the PNP point contact transistor and the BJT (Bipolar Junction Transistor). The BJTs ruled the semiconductor market until the 1970's, when the Metal-Oxide-Semiconductor Field-Effect Transistor (MOSFET) technology proved to be more reliable in terms of scaling and power consumption.

One of the major reasons for the shift to MOSFETs, specifically the Silicon (Si) based (due to its low power consumption) Complementary Metal-Oxide Semiconductor (CMOS) in the 70s and 80s was the ease of the fabrication process in scaling devices.

Specifically, in comparison to their then competition of BJTs, whose fabrication became increasingly difficult as the devices scaled down. This enabled the packing of a high density of logic functions on a single chip and made CMOS the most favored technology for VLSI chips for the past four decades.

The process of scaling in the CMOS is becoming increasingly complex in the past decade after we have entered the sub-50 nm regime. Successful scaling of devices necessitates thinner gate oxides and higher doping concentrations to get better drive currents for the device. But managing device performance at these dimensions also requires us to do more in terms of finding new device structures to manage the heavy short channel effects, that includes quantum effects like space quantization, direct tunneling from polysilicon through the gate-oxide, tunneling from drain to body, tunneling from source to drain etc., to name a few.

Alternative device technologies, namely strained Si, high-K based CMOS transistors overcome some of these limitations. However, they still prove to be a challenge not only in terms of device speed and power optimization, but also possess incredible challenges in changing the existing fabrication technology to maintain the maximum yield to enable mass chip production. And, as we get down to the current gate lengths of 22nm and below, the device technology for the predominantly CMOS based processor architecture has changed into an apparently new 3-D Tri-gate technology (FinFETs), in order to have better control over the channel. This also helps to overcome and avoid gate and substrate leakage issues and several other short channel effects mentioned above which manifest now on a much bigger scale.

But in spite of all of these device changes in terms of the device structure, fabrication and the architecture of the processors, the wafer for the processor scaling is still based primarily on the Si substrate (majorly because of the ease of fabrication of Si with higher yield, and its extremely good device characteristics). Mobility characterization in Si henceforth becomes really important in this scenario. And, as technology moves down to low power based devices like phones, and tablets, exploration of alternative materials like Germanium (Ge) and Gallium Arsenide (GaAs) etc., becomes undeniably interesting.

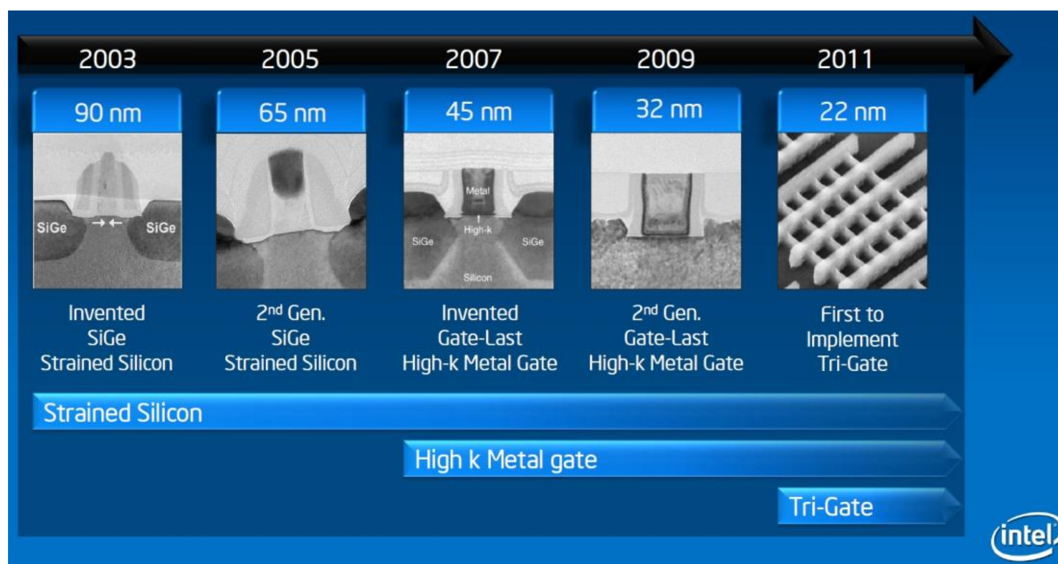

Figure 1.1 : Transistor innovations for the technology generations – *from Intel 22nm Announcement presentation.*[2]

Namely, Ge which lost the battle earlier to Si in the 70s, due to its low power and high leakage issues, is becoming a material of great interest as we go down in power consumption for these devices (where now people are ready to forego and manage with the loss through leakage for the gain through the low power advantage). This leads us to

an important conclusion that it is imperative for us to have a reliable model for characterizing the mobility in these devices (Si and Ge) in both the industrial and scientific community.

1.2. TRANSPORT IN SEMICONDUCTORS

1.2.1. OVERVIEW OF SEMICLASSICAL TRANSPORT

Transport in semiconductor has been traditionally treated using a semiclassical perspective as opposed to using completely quantum viewpoint. Drift-diffusion simulations [3],[4], have been the standard initial choice for modeling devices in the past where the significant length scales were greater than 1 μm [5], and the diffusive regime was valid. This method solves a set of coupled non-linear partial differential equations iteratively using the Gummel's or Newton's method [6]. The model solves the set of the coupled equations iteratively for self-consistency (continuity equations for electrons and holes, the current density equation and the Poisson equation). This model has become increasingly difficult to account for high electric field effects like velocity saturation, velocity overshoot, etc., as the dimensions of the channel kept shrinking. Field-dependent mobility and diffusion coefficients were introduced ad hoc to account for the saturation of the carrier velocity at high electric fields. But this was still more of a correction factor (approximation) to the saturation velocity rather than a complete model that can handle these high-field effects accurately.

Like the drift-diffusion model, the hydrodynamic model treats the propagation of electrons and/or holes in a semiconductor device as the flow of a charged incompressible

fluid. The hydrodynamic model accounts for hot carrier effects that are missing in the standard drift-diffusion model formulation. The classical hydrodynamic model has become a standard industrial simulation tool that incorporates important “hot electron” phenomena in submicron semiconductor devices. Hot electron effects are missing in the simpler drift-diffusion model, which assumes that the electron gas is always at ambient temperature. The hydrodynamic model consists of nonlinear hyperbolic conservation laws for particle number, momentum, and energy (with a heat conduction term), coupled to Poisson's equation for the electrostatic potential. In the momentum and energy conservation equations, charge carrier scattering by phonons is modeled by relaxation time approximations. As such the hydrodynamic model supports “velocity overshoot” in devices with channel lengths less than 200 nm in silicon. The hydrodynamic model can be extended to include quantum/tunneling effects by adding quantum corrections. This model is equivalent to the equation of the electro-gas dynamics. Thus the Homogenous electron gas model (HEG) becomes applicable in modeling electrons as a gas with sound velocity and the flow can be either subsonic or supersonic. The transition from supersonic to subsonic can be modeled as a shock wave. This model is observed to hold well down to 50nm. Below 50 nm and high fields, it fails to give accurate values for the velocity overshoot as the energy relaxation time is calculated from bulk Monte Carlo simulations.

As device dimensions shrink below 50nm, one must solve the conventional semiclassical BTE using particle based techniques like Monte Carlo method or other direct numerical techniques. The Monte Carlo approach till today’s date is the most widely preferred method to solve the BTE (in the long-time limit). The method is popular

mainly because it has the advantage of an easier and intuitive treatment of the carriers (though this comes at the expense of larger computational time) [7][8][9]. The basic idea of the Monte Carlo approach is that the scattering events in devices are random. So, the particle motion is simulated as a free flight halted by a random, instantaneous scattering event. Thus, the computational model consists of stochastic models that generate random free-flight times for each particle, choose a specific scattering mechanism at the end of the free flight, and then compute the energy and momentum at the end of this cycle after scattering. This will then be repeated for the particle for the next free-flight/scattering all over again. The various physical quantities, like drift velocity, current, etc., can now be obtained from sampling particles at various times. Time evolution of these physical variables of the device can therefore be modeled.

A variant to the classical Monte Carlo Technique is the so-called Cellular Automaton (CA) method. The CA method consists of a regular grid of *cells*, each in one of a finite number of *states*. For each cell, a set of cells called its *neighborhood* is defined relative to the specified cell. An initial state (time $t=0$) is selected by assigning a state for each cell. A new *generation* is created (advancing t by Δt), according to some fixed rule (generally, a mathematical function) that determines the new state of each cell in terms of the current state of the cell and the states of the cells in its neighborhood. Typically, the rule for updating the state of cells is the same for each cell and does not change over time, and is applied to the whole grid simultaneously, though exceptions are known, such as the stochastic cellular automaton. This procedure can be applied to solving the BTE by constructing cells made of momentum and real space units and giving state values to each

of the cells, indicating either empty or filled. The state of the cells will be updated in discrete time steps, and repeated according to the above algorithm [10][11].

The scattering matrix approach is also known as the response matrix approach. A scattering matrix of a given semiconductor substrate relates the out-going fluxes to the incident-fluxes, accounting for all of the scattering mechanisms and the electric fields inside the slab [12][13]. The flux here represents the distribution function of the carriers, and is discretized in momentum space. The other advantage of this method is that the solution of the BTE is easily computed by breaking up the device into a series of slabs and then cascading the scattering matrices of the slabs together. However, the disadvantage remains that, each slab needs a library of scattering matrices to be pre-computed, which becomes computationally expensive [14].

1.2.2. FAILURE OF THE BTE

As the device dimensions now keep shrinking to sub-20 nm regime, we can start seeing that a purely semiclassical analysis proves to be inadequate in treating transport in these devices. To illustrate this point let us consider the Boltzmann Transport Equation (BTE) below,

$$\frac{\partial f}{\partial t}+\frac{1}{\hbar}\nabla_k E(\mathrm{k})\cdot\nabla_r f+\frac{d\mathrm{k}}{dt}\cdot\frac{\partial f}{\partial \mathrm{k}}=\sum_{\mathrm{k}'}\left\{\Gamma(\mathrm{k}',\mathrm{k})f(\mathrm{k}')\left[1-f(\mathrm{k})\right]-\Gamma(\mathrm{k},\mathrm{k}')f(\mathrm{k})\left[1-f(\mathrm{k}')\right]\right\} \tag{1.1}$$

The RHS of Eq. (1.1) represents the collision integral, which contains the summation over all scattering mechanisms over all final states **k’**. The function f(**k**,**r**,t) is the carrier distribution function which gives the density of particles with momentum **k** at the point **r**

at time t. Approximate solutions for the distribution function can be found by assuming the drifted-Maxwellian model for low field regime.

In the semiclassical world of device transport, most presently used physical models are based of numerical solution of this BTE. The BTE is made on the assumption that particles obey classical Newtonian laws of motion during the free-flight under the field, thus making approximations like electrons occupy distinct momentum states k, and have almost free particle like behavior with an equivalent effective mass in those states. The stationary theory of electrons as stated above, treats scattering under two important assumptions, one that the scattering events are assumed to be *independent* and second, that they occur *instantaneously in space and time,* thus causing weak and infrequent scattering of the electrons among the momentum states $\{\mathbf{k}\}$. Any applied electric field is now treated only as a weak perturbation (under adiabatic approximation, slowly varying fields) and the field is assumed to be responsible to only accelerate the carriers during free-flight between collisions, and does not interfere with the states themselves, or interfere with the scattering events as such. But this approximation needs further probing when device gate lengths reach sub-20 nm.

To further examine this point on an intuitive level, consider the decreasing device channel dimensions. As device size gets close to the de Broglie wavelength of the carriers under effective mass approximation, their wave nature dominates. Thus, at these low device dimensions, the gate length is on the order of the phase coherence length of the carriers (phase coherence length is the distance over which the carriers retain their phase memory). Thus, carriers take lesser time between collisions, and the scattering events seem to have a shorter time scale. The collision time along with the mean free time between collisions now becomes finite - w.r.t transit timescales in the device. Thus, scattering may no longer be actually independent of the field or instantaneous anymore

and a re-examination of the approximations in the BTE becomes important.

The limits of the Boltzmann transport theory as stated above can be overcome when a fully quantum mechanical viewpoint is used to explain the behavior of these devices. The effect of the above semiclassical approximations can further be examined by looking at Eq. (1.1). In Eq.(1.1), the transition rate probability based of the different scattering mechanisms is given by the Fermi's golden rule,

$$\Gamma_{k\to k'} = \frac{2\pi}{\hbar}\left|V_s^{kk'}\right|^2 \delta\left(\varepsilon_{k'} - \varepsilon_k \pm \hbar\omega\right) \tag{1.2}$$

where, $\left|V_s^{kk'}\right|^2$ represents magnitude of the matrix element squared of the various scattering mechanisms included in the model.

The Fermi's golden rule is derived from first-order time-dependent perturbation theory based on two fundamental assumptions:

- It is assumed that the scattering is infrequent which allows us to impose the long time limit and arrive at the energy conserving delta function.
- No initial state decay, $C_{ns} = 1$ (occupancy of the initial state is approximately 1, which means that the state is not depleted).
- It is assumed that the scattering is instantaneous, i.e., the collision duration time is assumed to be zero.

Questioning the first assumption, that is when the scattering is not infrequent, there is not enough time in between the scattering events. Thus, if the time scale is comparable to the order of $\frac{1}{\omega_{ns}}$, the energy conserving delta function may not be completely formed. Thus the long time limit cannot be directly imposed to arrive at the energy conserving delta-function represented in Fermi's golden rule (instead one might need to account for a modified energy conservation relationship between the initial and final states through a

Lorentzian function). Also, in reality the scattering events can no longer be assumed to be infrequent with certainty, as carriers constantly interact with the surroundings, such as the impurities, surface roughness, or the lattice vibrations (phonons). This now leads us to question the other assumption of no initial state decay. This requires that the initial state occupancy is treated as a function of time. As carriers scatter out of the initial state, state loses its population (which changes its occupation probability) and interacts with the other states leading to its own modification of the energy and population (accounted through the self-energy). This leads to collisional broadening of the States (CBS) and shift in the energy spectrum. Thus, the energy conserving delta function now expands into a Lorentzian function called spectral density function $A_n(k,\omega)$. Examining when this CBS happens in silicon-based devices and what are the different parameters that it depends upon, is one of the major motivations of this work.

Similarly, the scattering can no longer be considered instantaneous for the reasons that were discussed earlier, thus the effect of the field in accelerating the carriers during the collision event has to be considered. This is known as the intra-collisonal field effect (ICFE). Since in this work we are considering near-equilibrium condition, we do not consider intra-collisional field-effect in our model.

To summarize, the state broadening is an important assumption as the device lengths scale to dimensions on the order of sub-20nm and a comprehensive treatment of scattering in Si devices from a QM viewpoint becomes important. The spectral density function can no longer be approximated as a simple energy conserving delta function and the finite momentum state lifetime, due to the state broadening, has to be accounted for when one solves for the transport in these device dimensions.

The above stated reasons further iterate the emphasis on the onset of the failure of the BTE approach and Fermi's golden rule in these device structures. The Monte Carlo

technique, though robust, is still reliant on the BTE (in all its approximations) to handle the scattering and thus suffers from all of the above limitations. Therefore, one has to either radically modify the BTE or start with theory that is more ab-initio. That is, the system (or parts of the system of interest) needs to be modeled from a many-body viewpoint to treat the interactions where some/many of the above semi-classical assumptions do not hold anymore. Namely, the correlations (in space and time) between the particles have to be considered if one wishes to solve for the CBS or the ICFE. Thus, we need to build a more fundamental quantum transport formalism, where we can use suitable semi-classical approximations when needed to reduce complexity of our calculations, thus enabling us to include these above neglected effects.

1.2.3. QUANTUM TRANSPORT - AN OVERVIEW

Quantum transport is the most fundamental of all transport models, as it starts from the Schrödinger wave equations (both time dependent (TDSWE) and time independent (TISWE)) and uses statistical mechanics to model the physics of the device, thus arriving at a set of the so-called quantum kinetic equations. But unlike solving a simple closed system like a 1-D potential barrier, actual semiconductor devices most often represent an open system, with continuous influx/outflow of charges (current) across the terminals that are far from equilibrium. In the case of semiclassical models, this can be accounted with Ohmic or Schottky contact based boundary conditions for the various transport differential equations. But in the case of a quantum open system, one has to go through several un-normalizable scattering states with open boundaries, and a many-body situation where the electron interacts with the system and itself. Hence, one needs to find

the best approach to solve the set of one-body/many-body TDSWE to get the values of the different observables. This has its own challenge in how the boundary conditions are handled in the quantum approach and is treated differently based on each approach within the quantum view.

Many of the quantum effects can be classified into two types, static and dynamic. The static quantum effects namely are tunneling through the gate oxide, and the energy quantization in the inversion layer of a MOSFET. The tunneling happens in case of very thin gate oxides where there is gate leakage via direct tunneling. The size-quantization happens due to inversion charge induced by the gate, thus forming the triangular potential well, leading to the formation of spatially localized subbands. To solve these static effects, one might find it relatively easier as it involves solving only the one-body TISWE.

The real complication in the quantum approach arises when one tries to model the dynamic quantum effects like, collisional broadening of the states due to scattering, intra-collisional field effect, electron-electron scattering, dynamical screening from charged carriers, and other many-body effects. *Density Matrix Method*, *Wigner Function Method*, and the *Green's Functions Formalism* are the methods suitable for addressing some, or all of the above-mentioned issues.

The Density Matrix method [15] and the Wigner function approaches [16] use the full scattering potential (non-local), and compute the interaction of the electron with the scattering event and the field in full form. But the main issue is that the Density Matrix Method is real-space based and the Wigner function is phase-space based. This makes it

increasingly hard when one wants to compute correlations in time, to get the screening wave vector, polarization or calculate the conductivity – which now requires an interaction of two-particles and the system (scattering, external field etc.). Thus, solving for the CB of the states, lifetime of the electrons, or the ICFE is very hard.

This is when the Green's function technique becomes tremendously useful. The technique essentially is based on the integration of the quantum field theory with statistical mechanics to obtain a very solid approach. This approach can be approximated to various degrees of freedom (orders of the perturbations series in the scattering kernel) to get the required observables out of the system. Therein lies one of its biggest advantages of this method over other non-equilibrium quantum transport formalisms.

Green's functions are basically impulse (Linear) responses of the TDSWE system under consideration. Thus, in general, a Green's function for a 1-electron SWE at equilibrium is given by,

$$\left(i\frac{\partial}{\partial \mathrm{t}}-\hat{\mathrm{H}}(\mathbf{r})\right)\mathrm{G_o}(\mathbf{r},\mathbf{r}',\mathrm{t},\mathrm{t}')=\delta\left(\mathbf{r}\ -\mathbf{r}'\right)\delta\left(\mathrm{t}-\mathrm{t}'\right) \tag{1.3}$$

Where the bare one-electron Green's function G_0 is given by,

$$\mathrm{G_o}\left(\mathbf{r},\mathbf{r}',\mathrm{t}-\mathrm{t}'\right)=\theta\left(\mathrm{t}-\mathrm{t}'\right)\mathrm{K_o}\left(\mathbf{r},\mathbf{r}',\mathrm{t}-\mathrm{t}'\right) \tag{1.4}$$

The quantum-field theoretical methods in non-equilibrium statistical mechanics were developed by Martin and Schwinger [17][18], Kadanoff and Baym[19], and were further developed by Keldysh. Later, Ferry and Barker, extended the non-equilibrium Green's function (NEGF) formalism to semiconductor device modeling in 1980 [20]. All these works laid down the foundation of the NEGF theory that we are going to discuss

below briefly, before discussing in more details in Chapter 3.

The transport in the non-equilibrium situation can be broadly divided into 2 cases as below,

- For near-equilibrium situations, one assumes that the distribution function only slightly deviates from the Fermi-Dirac distribution function. Also, low temperatures ensure that the interaction remains weak as the electrons couple weakly especially to the phonons, so that the potential is slowly-varying and the effective mass and single band approximations still holds. For high temperatures, one could solve for higher orders of interaction in the included scattering mechanisms, thus increasing the accuracy of the result and bypassing the initial assumption of independent and instantaneous scattering for the first Born approximation.
- For high electric fields and strongly non-equilibrium transport across the channel, the above approximations fail, and one must solve for the distribution function directly by using the quantum kinetic equations in the NEGF approach.

A diagram that describes the usability of Green's functions approaches for both low-field and high-field conditions, and for bulk systems and devices, is shown in Figure 1.2 approaches in modeling carrier transport in bulk systems, inversion layers and devices under low and high applied electric field. In here, RGFA stands for recursive Green's function approach, and CBR stands for contact block reduction method.

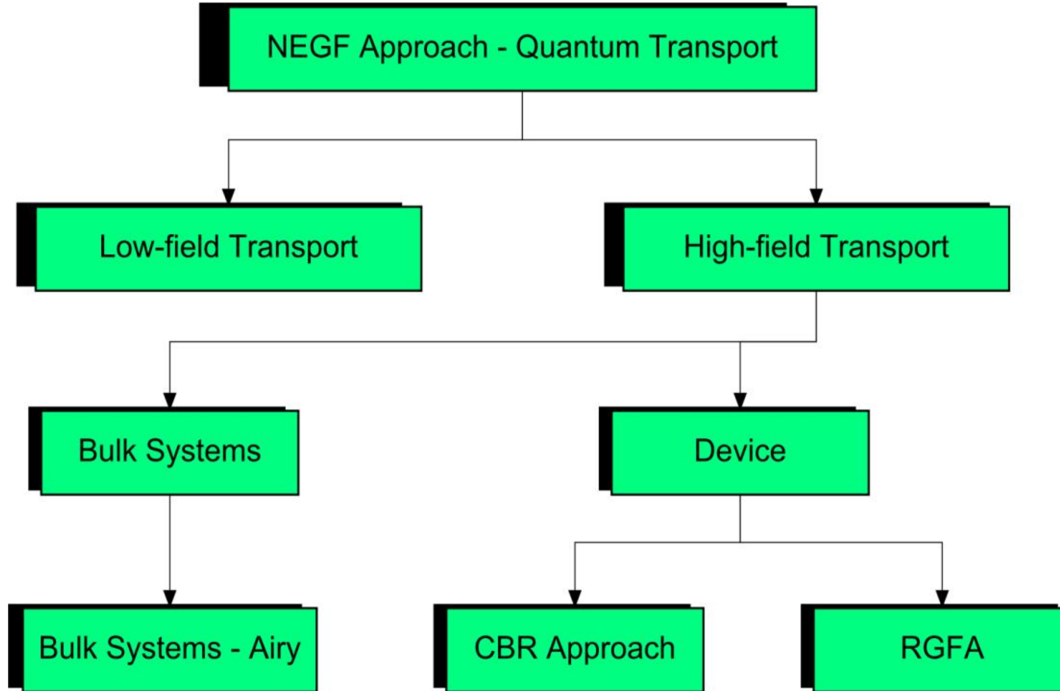

Figure 1.2 : Green's function based approaches

I. LOW FIELD TRANSPORT

Low-field transport in devices is based on the fact that the devices operate in near-equilibrium conditions. Thus, some of the fundamental assumptions of adiabatic variation of the field, effective mass approximation, the single band model of the electrons, and the self-consistent Born approximation, can still be valid in this regime. The following will be a short recapture of the framework developed by [21], which has been extended in this work.

The Green's function technique uses the interaction representation, where both the state vectors and the operators depend on time. The advantage of the interaction picture is that the operators are governed by the "easily" computable unperturbed Hamiltonian of the system, and the state evolution is dependent on the perturbed Hamiltonian (interaction Hamiltonian), which is typically assumed to be small.

The various Keldysh Green's function can be defined as follows. The retarded/advanced Green's function (or correlation function) is defined as:

$$\begin{aligned} G_r(x_1,x_2) &= -\frac{i}{\hbar}\theta(t_1-t_2)\left\langle\left\{\hat{\Psi}(x_1),\hat{\Psi}^+(x_2)\right\}\right\rangle \\ G_a(x_1,x_2) &= \frac{i}{\hbar}\theta(t_2-t_1)\left\langle\left\{\hat{\Psi}(x_1),\hat{\Psi}^+(x_2)\right\}\right\rangle \end{aligned} \tag{1.5}$$

where the brackets denote ensemble average over available states.

For complex Hamiltonians, one applies the perturbation approach of the time-ordered Green's function

$$G(x_1,x_2) = -i\left\langle T\left[\hat{\Psi}(x_1)\hat{\Psi}^+(x_2)\right]\right\rangle \ , \tag{1.6}$$

where T is a time-ordering operator.

The other Green's functions to be defined during non-equilibrium situations are less-than and greater-than Green's functions:

$$\begin{aligned} G^<(x_1,x_2) &= \frac{i}{\hbar}\left\langle\hat{\Psi}^+(x_2)\hat{\Psi}(x_1)\right\rangle \\ G^>(x_1,x_2) &= -\frac{i}{\hbar}\left\langle\hat{\Psi}(x_1)\hat{\Psi}^+(x_2)\right\rangle \end{aligned} \tag{1.7}$$

Looking at the expression for less than Green's function one can "easily" identify that they denote some physical observables; namely at $x_1 = x_2$, it becomes the particle density, and for the same time average becomes the particle density matrix. Similarly the greater than Green's function denotes the hole density and the hole density matrix, respectively, for the same time.

Now, based on Eq.(1.6), we can define time-ordering operators on different branches

of the contour. Then:

$$G_t(x_1,x_2)=\theta(t_1,t_2)G^{>}(x_1,x_2)+\theta(t_2,t_1)G^{<}(x_1,x_2)=G_r+G^{<}=G_a+G^{>} \quad (1.8)$$

and the anti-time ordered Green's function becomes

$$G_{\bar{t}}(x_1,x_2)=\theta(t_2,t_1)G^{>}(x_1,x_2)+\theta(t_1,t_2)G^{<}(x_1,x_2)=G^{>}-G_r=G^{<}-G_a \quad (1.9)$$

The function $\theta(t_1,t_2)$ is defined on the contour with time ordered property. From Eq. (1.8) we now have the following identities:

$$\begin{aligned} &G_r(\omega)-G_a(\omega)=G^{>}(\omega)-G^{<}(\omega) \\ &G_a(\omega)=G_r^{\dagger}(\omega) \end{aligned} \quad (1.10)$$

From the assumption that the fluctuation-dissipation theorem is still valid, in near-equilibrium we have,

$$G^{<}(\omega)=if(\omega)A(\omega), \quad (1.11)$$

where $f(\omega)$ represents the Fermi-Dirac distribution function. Also, from the retarded Green's function one can obtain the spectral density function as,

$$A(\omega)=i\left[G^{r}(\omega)-G^{a}(\omega)\right] \quad (1.12)$$

Thus, from Eq.(1.10), (1.11), (1.12) one needs to have only one independent Green's function, say the retarded Green's function, to compute the rest of the observable physical quantities like collisional broadening of states, renormalization of the spectrum and the charge density.

For the high-field regime the fluctuation dissipation theorem will no longer be valid making the requirement of calculating at least two independent Green's functions

mandatory. For example, following the work of Kadanoff and Baym, and using Feynman rule to expand the perturbation of the contour ordered Green's function according to Wick's theorem, we can arrive at the Dyson equation for the retarded/advanced Green's function as,

$$G_{r,a} = G^{o}_{r,a} + G^{o}_{r,a}\Sigma_{r,a}G_{r,a} \ . \tag{1.13}$$

The equations of motion (Keldysh equation) for the greater-than and less-than Green's functions are,

$$G^{>,<} = \left(1 + G_r\Sigma_r\right)G^{>,<}_{o}\left(1 + \Sigma_a G_a\right) + G_r\Sigma^{>,<}G_a \tag{1.14}$$

These two equations form a coupled set of equations which when solved can give the required transport properties in the non-equilibrium condition.

For the low-field case, under the investigation in this work, it is only needed to solve the Dyson equation for the retarded Green's function which will then be used to calculate the spectral density of states, which after integration over the momentum states gives the density of states. Thus, under the fluctuation-dissipation theorem, the quantum charge density can be computed.

II. HIGH-FIELD TRANSPORT - BULK SYSTEMS (AIRY TRANSFORMS)

In this formalism, the author has adopted a new method to solve the Dyson equation for the retarded Green's function, coupled with the Keldysh equation. The model allows a non-perturbative description of the effects of an external high electric field on electron-

phonon scattering, which is treated under the first-Born approximation within the Kadanoff-Baym-Keldysh nonequilibrium Green's-function approach.

Based on the exact solutions (which happens to be Airy functions) of the Schrodinger equation for a scalar potential along the direction of motion /field, one solves the Dyson's equation for the single-particle retarded Green's function. Recognizing that high fields break the translational symmetry of the system and that momentum is no longer a good quantum number, Airy transforms have been used to handle the position dependence parallel to the applied high field. The spectral density function $A(k,\omega)$ is then computed, and CB and ICFE are investigated.

A weak scattering regime is assumed, with effective mass approximation in the field direction. The nonpolar optical phonon scattering is the only scattering mechanism included in the model, and the field is restricted to a constant electric field in space and time along the direction of the motion (bulk material).

III. STATE-OF-THE-ART IN HIGH FIELD TRANSPORT – BALLISTIC + SCATTERING REGIME – RECURSIVE GREEN'S FUNCTION METHOD (RGFM)

Very brief overview of the model is as follows. The Hamiltonian in the TISWE is changed corresponding to the tight-binding limit. Then the corresponding retarded and less-than Green's function are introduced. To start with, the Dyson equation for the retarded Green's function is solved. To solve this equation, the system is broken down

into an internal region and the external leads. The Dyson equation is used to calculate the Green's function for each of these regions.

Now the Green's function for the total region is calculated using the Dyson's equation, and the effect of contact, the potential, etc., is accounted through the self-energy into the equation for the whole system. Once the Green's function is computed by solving the Dyson equation with the Keldysh equations, one can find the value of the electron density, which can then be used to compute the effective potential self-consistently by solving the Poisson equation.

IV. HIGH FIELD TRANSPORT – CONTACT BLOCK REDUCTION METHOD (CBR) – BALLISTIC REGIME

The CBR method allows one to calculate 2-D or 3-D ballistic transport properties in a device that may have arbitrary shape, potential profile, and most importantly, any number of leads. In this method, quantities like the transmission function and the charge density of an open system can be obtained from the eigenstates of the corresponding closed system defined as $H^0 |\alpha\rangle = \varepsilon_\alpha |\alpha\rangle$ that need to be calculated only once, and the solution of a very small linear algebraic system for every energy step E. The retarded Green's function $\mathbf{G}^R(E)$ can then be calculated via the Dyson equation through a Hermitian Hamiltonian $\mathbf{H^0}$ of a closed system,

$$\mathbf{G}^R\left(E\right) = \mathbf{A}^{-1}\left(E\right)\mathbf{G^0}\left(E\right), \tag{1.15}$$

where $\mathbf{G}^0(E)$ is the retarded Green's function for the closed system (decoupled device). The total numerical cost of the method can be estimated as

$N_{n(r)} = N_{eigen}^2 N_E + N_{eigen}^2 N_{grids}$, where N_E is number of energy steps , N_{eigen} is the number of eigenstates to be used, and N_{grids} is the number of grid points in real-space. Note the absence of large terms like $N_E \times N_{grids}$.

1.3. SUMARY

A low-field mobility solver for quasi-two dimensional electron gas system has been implemented, that involves self-consistent coupling of a 1-D Schrödinger-Poisson solver (SCHRED V2.0) with near-equilibrium Green's function (nEGF) solver. The nEGF part solves self-consistently the Dyson equation for the retarded Green's function using the self-consistent Born approximation for the self-energies. The collisional broadening of the states and the renormalization of the spectrum are then used to calculate the real density of states (DOS). This is done by solving the Dyson's equation for the retarded Green's function [22].

The scattering mechanisms included in the model are Coulomb scattering from the depletion layer and interface/oxide charges, surface-roughness scattering, zero- and first-order intervalley optical phonon scattering processes, and acoustic phonon scattering. Screening of the Coulomb scattering potential is taken into account within the Random Phase Approximation (RPA). For the temperature-dependent RPA polarizability function we use the result given in [23] which is valid for general Fermi-Dirac statistics. For intervalley scattering, which is the dominant scattering mechanism in many-valley semiconductors such as Si and Ge, we follow an approach due to Price [24] and Ridley [25]. We use the same phonon energies and coupling constants in the inter-subband scattering as those in intervalley scattering in bulk Si [26], which is shown to give excellent results for transport in Si devices [27]. The anisotropy of the deformation

potential interaction is treated as described in Refs. [28][29]. Also, the intervalley scattering is solved self-consistently within the self-consistent Born approximation. This is one of the extended features of this work that hasn't been done before in Si systems characterized at near equilibrium.

We divide the solver into 2 parts, an outer loop that consists of the 1-D Schrödinger-Poisson solver, and the inner one that consists of the nEGF solver. The nEGF solver calculates the real DOS, which is then used to calculate the new quantum sheet charge density based on Fermi-Dirac statistics.

The Schrödinger-Poisson outer loop ensures that there is proper treatments of the following short channel effects: space quantization, device transconductance degradation, the finite value of the inversion-layer capacitance (average distance of the carriers in quantum treatment peaks away from the Si-SiO_2 interface, leading to decreased gate capacitance, and decreased inversion charge, thus leading to increased threshold voltage), the impact of polysilicon gate depletion.

In Chapter 2, the treatment of various scattering mechanism is presented, and the overlap integrals are closely examined. In Chapter 3, an overview of the Green's function formalism is given – that will include the general description, followed by the description of the Dyson's equation for the retarded Green's function. Chapter 4 will explain how screening and conductivity is handled in the nEGF approach in this work. This is followed by a short Chapter 5 that explains the design and the implementation of the software. A discussion of the simulation results for the mobility and the DOS is given in Chapter 6. Conclusions from the work accomplished and Future directions of research will be discussed in Chapter 7 and Chapter 8, respectively.

CHAPTER 2. GREEN'S FUNCTION FORMALISM

2.1. REVIEW OF THE FORMALISM – AN INTRODUCTION

In this chapter, the real-time Non-Equilibrium Green's Function (NEGF) technique for modeling transport phenomena in semiconductor devices is briefly outlined. Following which, a brief review of its adoption to near-equilibrium (nEGF) case is discussed. Green's functions [34] are response functions (propagators) that tell us how an excitation propagates through the system.

To study the time evolution of a many-particle quantum system, Kadanoff and Baym [19] formulated the NEGF technique. The Green's functions are defined as the expectation value of certain field operators over the available states of the system. The simplest Green's function of applicability is the one-particle Green's function that contains information on how one particle (an electron or a hole) propagates through the system of all other electrons or holes respectively. In this way it provides us with details on the equilibrium related properties of the system (carrier lifetime, broadening of the states, renormalization of the spectrum, etc.). The solution of the equation of motion for the 1-particle Green's function (that couples it to the two-particle Green's function) is used to calculate the one-particle properties of the system.

2.2. THE MANY BODY PROBLEM

The complete knowledge of an N-body system that has interactions between the particles (pairs of particles or more) requires the computation of N-particle quantities. This is a many-body problem. In such many-body problem, when we derive the equation

of motion of the 1-particle Green's function, we do not get a closed equation and we are led to a higher particle quantity. Seeking the equation of motion of the higher particle quantity will lead to even higher particle quantities. This hierarchy of equations, which ends at the full *N*-particle quantity is called the BBGKY (Bogoliubov-Born Green Kirkwood Yvon) hierarchy.

The purpose of the Green's function hierarchy of equations is to present us with a hierarchy of correlations. Then, assuming that many-particle correlations are weaker, the hierarchy can be truncated. Applying the BBGKY hierarchy to Green's function allows us to make approximations based on correlations.

Thus, the equation of motion for an n-particle Green's function (in the integral form) can be represented in terms of a functional of non-interacting Green's functions and the perturbing potential. One can then solve the above equation iteratively which results in a perturbative series to the lowest order possible to explain the system adequately. However even for the lowest order, this perturbation series becomes complicated to be solved directly due to the higher order derivatives. So under a perturbative approach for near equilibrium conditions, Wick's Theorem [35] is used to reduce this to a combination of time ordered operators. The numerous terms can now be represented in a graphical manner using Feynman diagrams [36]. This enables one to compute the Green's function with relative ease, by summation of the terms that are inferred from the graphs. This, in essence is the Non-Equilibrium Green's function formalism.

The complexity of the diagrams and hence the integrals (summation of terms) increases as ‘n’ increases in the n-particle Green’s function. This forces us to make judicial use of the method, to the lowest possible order required to define the system studied adequately. Thus, after solving for the one-particle function, the two-particle function needs to be solved.

The two-particle Green’s function represents the pairs of excitations that propagate through the system of electrons. For example, in the electron-phonon interaction, the electron interacts with the quantized phonon modes, resulting in an electron-hole bubble that is handled via the creation and annihilation phonon operators. This collective excitation results in a cloud that shields the electrons and gives rise to screening. Similarly, the fluctuations in this collective excitation can be computed via the current correlation function (which is a two-particle Green’s function), and thus calculate the conductivity. The NEGF technique, hence, provides a very powerful and robust technique at the perturbative level for evaluating properties of many-particle systems under various limits. The disadvantage is that it is time and resource consuming.

2.3. GREEN’S FUNCTION FORMALISM

Different formalisms of the Green’s function exist depending on how the averaging of the field operators is done over the states. The zero-temperature formalism averages over the ground state at equilibrium, the finite-temperature formalisms averages over all possible states at equilibrium. The real-time Green’s function (Non-equilibrium Green’s Function at finite temperatures) averages over all the available states when the

system is driven out of equilibrium. Thus, one can compute the Green's function at thermodynamic equilibrium, as well as in non-equilibrium situations. In this work, we limit our study to systems under a near equilibrium condition.

Another exciting distinction in this formalism is that, from these evaluated Green's functions, the expectation value of any observable can be found. This meaning that the Green's functions contain any and all the information about the system under consideration.

As stated earlier, using a perturbative approach, one can deduce the proper Green's functions by approximating the self-energy terms. In this work the quasi-two-dimensional electron gas (Q2DEG) in a MOS capacitor is treated under a near-equilibrium condition. The *second quantization* approach is used to model the interaction between the electrons and the various scattering fields (that are quantized and represented in terms of operators in the occupation number formalism) and are included in the proper self-energy term while solving for the proper Green's function.

As device are getting into the nano-scale regime, the semi-classical limits come into question and there is a growing importance of quantum effects and tunneling at these length-scales. Thus, the basic approach developed in the early 1970s for the NEGF formalism has become increasingly popular during recent years to model transport in mesoscopic devices. The present work adopts this theory which is based on the works of *Vasileska et al*. [23] to near-equilibrium scenario, and adds more features to her model to efficiently model the effect of collisional broadening of the states on density of states, and the mobility of Si based devices under low-field conditions.

2.4. HAMILTONIAN AND SECOND QUANTIZATION

The total Hamiltonian for the complete system is given by,

$$\mathrm{H} = H_e + H_{ph} + H_{e-ph} \tag{2.1}$$

where H_e is the Hamiltonian of non-interacting electrons, H_{ph} is the Hamiltonian of free phonons, and H_{e-ph} is the electron-phonon interaction Hamiltonian.

The H_e is given by,

$$H_e = \int dr \psi^{\dagger}(r)\left[T(r) + U(r)\right]\psi(r) \tag{2.2}$$

where $T(r)$ is the one-electron kinetic energy operator and $U(r)$ is the self-consistent electrostatic potential energy. The kinetic energy operator is obtained from the effective mass approximation, as we consider the transport in the conduction band of silicon where we use parabolic, ellipsoidal energy band structure.

The interaction of the quantum states shown above in H_e are with a classical field $U(r)$, and this is referred to as first quantization. But in the case of the electron-phonon Hamiltonian, the electron quantum states now interact with the quantized harmonic modes of vibrations of phonons, thus leading to a second level of quantization. Hence it becomes convenient to describe the system, in terms of second-quantized operators that operate in occupation number space.

The field operators operating on the space are defined by,

$$\psi(r) = \sum_k u_k(r) c_k$$
$$\psi^\dagger(r) = \sum_k u_k^*(r) c_k^\dagger \tag{2.3}$$

where the operator c_k annihilates the particle in state k , and the operator $c_k^\dagger$ creates the particle in state k. The wavefunctions $u_k(r)$ form a complete set of single-particle eingen-functions with quantum number k. Therefore, the filed operator $\psi(r)$ removes a particle from state at r and the field operator $\psi^\dagger(r)$ creates a particle at state at r. The operators c_k *and* $c_k^\dagger$ satisfy the commutation relations,

$$\left[c_k, c_k^\dagger\right]_\pm = \delta_{kk'} ,$$
$$\left[c_k, c_{k'}\right]_\pm = \left[c_k^\dagger, c_{k'}^\dagger\right]_\pm = 0 , \tag{2.4}$$

where $[A,B]_+ = AB + BA,\ [A,B]_- = AB - BA$, $\delta_{kk'}$ denotes the Kronecker delta, and the plus sign refers to Fermions and the minus sign refers to Bosons. The field operators $\psi(r)$ and $\psi^\dagger(r)$ also satisfy their respective commutation relations.

The Hamiltonian operator for the phonons is given by

$$H_p = \sum_{q\lambda} \hbar\omega_{q\lambda} \left(a_{q\lambda}^\dagger a_{q\lambda} + \frac{1}{2} \right) \tag{2.5}$$

where $\omega_{q\lambda}, a_{q\lambda}^\dagger, a_{q\lambda}$ are the angular frequency, the creation operator and the annihilation operator for mode λ and wavevector q, respectively.

Then, H_{e-ph} is given by

$$
\begin{aligned}
&H_{e-ph} = \int dr \psi^{\dagger}(r)\varphi(r)\psi(r) \\
&where, \qquad (2.6)\\
&\varphi(r) = \frac{1}{\sqrt{V}} \sum_{q\lambda} M_{q\lambda} \left[a_{q\lambda} + a_{q\lambda}^{\dagger} \right] e^{iq.r}
\end{aligned}
$$

where V is the volume of the sample, $M_{q\lambda}$ is the electron-phonon matrix element that depends on the deformation potential tensor.

2.5. GREEN'S FUNCTION

In this section, we briefly introduce the contour-ordered Green function, its perturbation expansion, followed by the Dyson equation for the Green function.

2.5.1. TIME EVOLUTION PICTURES

The Schrödinger, interaction, and Heisenberg pictures in quantum mechanics will be used interchangeably to represent the contour-ordered Green function. Consider a general Hamiltonian represented as,

$$
\mathrm{H} = \mathrm{H}_0 + \mathrm{H}_1 \qquad (2.7)
$$

where $\hat{\mathrm{H}}_0$ is the non-interacting part, $\hat{\mathrm{H}}_1$ is the interacting part such as the electron-phonon interaction, impurity scattering, surface roughness scattering etc.

In the Schrödinger picture, the state vectors are time dependent whereas the operators are time-independent,

$$
i\frac{\partial}{\partial \mathrm{t}}\Psi_{\mathrm{S}}(\mathrm{t}) = \left(\mathrm{H}_{\mathrm{o}} + \mathrm{H}_1\right)\Psi_{\mathrm{S}}(\mathrm{t}) \qquad \mathrm{O}_{\mathrm{S}}(t) = \mathrm{O}_{\mathrm{S}}(t_0) = \mathrm{O}_{\mathrm{S}} \qquad i\frac{\partial}{\partial \mathrm{t}}\mathrm{O}_{\mathrm{S}} = 0 \qquad (2.8)
$$

The interaction picture says that both the state vectors and operators are time-dependant,

$$i\frac{\partial}{\partial t}\Psi_I(t) = H_I(t)\Psi_I(t) \qquad O_I(t) = e^{+iH_o t}O_S e^{-iH_o t} \quad i\frac{\partial}{\partial t}O_I(t) = [O_I, H_o] \qquad (2.9)$$

The Heisenberg picture says that the state vectors are time-independent and the operators are time dependant, meaning

$$i\frac{\partial}{\partial t}\Psi_H(t) = 0;\ O_H(t) = e^{+iHt}\hat{O}_S e^{-iHt};\ i\frac{\partial}{\partial t}O_H(t) = [O_H, H_I] \qquad (2.10)$$

The above different representations are suitably adopted in solitary or in combination, depending on the requirement of the situation to help in obtaining the solution. For example, in treating most of the electron-phonon interactions in this work, interaction representation is best suited to solve for the time evolution of both the operators and the state-vectors while a Heisenberg representation is adopted for the perturbing field. Thus, it becomes easier to create a unitary operator that determines the state vector's time evolution at time t in terms of the state vector at time 0 and solve for it, i.e.

$$\Psi_I(t) = U_I(t,0)\Psi_I(0);\ HU = i\frac{\partial}{\partial t}U \qquad (2.11)$$

2.5.2. CONTOUR-ORDERED GREEN'S FUNCTIONS

Green's functions are thermodynamic averages of the products of field operators $\Psi(r)$ and $\Psi^{\dagger}(r')$, that represent the impulse response of the quantum system under consideration as stated earlier .

The contour ordered Green's function in terms of its field operators is given by,

$$G_{c_k}(\mathrm{r},\mathrm{r}') = -i\left\langle T_{c_k}\left(\Psi(\mathrm{r})\Psi^{+}(\mathrm{r}')\right)\right\rangle \tag{2.12}$$

where c_k is the contour and T_{c_k} is the contour-ordering operator. The above equation can be expanded to a perturbative form using the method as explained in detail by Kadanoff and Baym, or use the equivalent procedure based on the Wick's theorem. One could also use the Feynman rules, which result from the application of the Wick's decomposition to a perturbation expansion. The Feynman rules are essentially a graphical approach to represent the different terms of the Wick's perturbative series, to arrive at the final set of terms for the lowest order required.

If we define the unperturbed electron Green's function as,

$$G_0(\mathrm{r},\mathrm{r}') = -i\left\langle T_c\left(\Psi_I(\mathrm{r})\Psi_I^{+}(\mathrm{r}')\right)\right\rangle \tag{2.13}$$

and phonon Green's function as,

$$\mathrm{D}(\mathrm{r},\mathrm{r}') = \left\langle T_c\left(\varphi_I(\mathrm{r})\varphi_I^{+}(\mathrm{r}')\right)\right\rangle \tag{2.14}$$

Now, Eq. (2.12) can be written as (after perturbative expansion)

$$G(r,r') = G_0(r,r') + \int_C dt_1 \int dx_1 \int_C dt_2 \int dx_2 G_0(r,r_1)\Sigma(r_1,r_2)G_2(r_2,r') \tag{2.15}$$

The above is the integral form of the Dyson equation within the self-consistent Born approximation. $\Sigma(r_1,r_2)$ is the irreducible self-energy for the contour-ordered Green function. Under the self-consistent Born approximation, the corresponding Feynman diagram leads to

$$\Sigma(r_1,r_2) = G(r_1,r_2)\,\mathrm{D}(r_1,r_2) \tag{2.16}$$

One can further define different Green's function based on the different

conditions under which these operators are averaged. The retarded and advanced Green's functions are defined as below,

$$G_r(r_1,r_2) = -\frac{i}{\hbar}\theta(t_1-t_2)\left\langle\left\{\Psi(r_1),\Psi^\dagger(r_2)\right\}\right\rangle \tag{2.17}$$

$$G_a(r_1,r_2) = \frac{i}{\hbar}\theta(t_2-t_1)\left\langle\left\{\Psi(r_1),\Psi^\dagger(r_2)\right\}\right\rangle \tag{2.18}$$

respectively. $\theta(t_2-t_1)$ is the step function. Note that the retarded function can be nonzero only if $t_1 > t_2$ whereas the advanced functions can be nonzero only for $t_1 < t_2$. Also note that the coordinate term r_1, r_2, are four-dimensional quantities for the 3-D space co-ordinates and the time. And the above expression can be averaged over the ground state in equilibrium (zero-temperature), or over all possible states of the system under thermal equilibrium (finite temperature), or averaging over available states of the system (non-equilibrium condition).

In addition to the above retarded and advanced Green's function, in non-equilibrium situations, one also needs additional correlation functions, such as the less-than

$$G^<(r_1,r_2) = \frac{i}{\hbar}\left\langle\Psi^\dagger(r_2)\Psi(r_1)\right\rangle \tag{2.19}$$

and greater-than

$$G^>(r_1,r_2) = -\frac{i}{\hbar}\left\langle\Psi(r_1)\Psi^\dagger(r_2)\right\rangle \tag{2.20}$$

correlation functions. The above four Green's functions are enough to describe most of the required non-equilibrium characteristics of the system, under consideration.

Additionally, one can also define time-ordered functions in terms of the earlier defined Green's functions as,

$$G_t = G_r + G^< = G_a + G^> \tag{2.21}$$

and anti-time-ordered

$$G_{\bar{t}} = G^> - G_r = G^< - G_a \tag{2.22}$$

Green's functions. $\theta(t_1, t_2)$ are the contour functions which are defined according to time order of t_1 and t_2.

By observing the above expressions for the $G^<$ for $r_1 = r_2$ (and at equal times), it can be observed that it resembles the number operator, which essentially is the single particle density. Similarly, one could argue that $G^>$ may be seen as corresponding to the hole particle density, namely the Fourier transformed,

$$\begin{aligned} n(r;E) &= -4 \times \frac{i}{2\pi} G^<\left(r,r;E\right) \\ \mathrm{p}(r;E) &= 4 \times \frac{i}{2\pi} G^>\left(r,r;E\right) \end{aligned} \tag{2.23}$$

Thus, the retarded and advanced Green's functions contain the spectral properties of the system, i.e., the density of states and the renormalized energy spectrum.

In equilibrium situations, where the fluctuation dissipation theorem is valid, one would require only one independent Green's functions (see Eqs. (2.17), (2.18), (2.21)) as the rest of the other three correlation functions can be calculated from this. In non-equilibrium however, the fluctuation dissipation theorem is not valid and we require at least 2 independent Green's functions to proceed with their calculations with respective

equations of motions.

The above equations are assumed for fermions, similar type equations for bosons (namely for phonons – where the commutator operators (boson fields) maintain the sign) exist as follows,

$$D^{>}(\mathrm{r}_1,\mathrm{r}_2) = -i\left\langle H_{e-ph}(r_1)H_{e-ph}(r_2)\right\rangle \tag{2.24}$$

$$D^{<}(r_1,r_2) = -i\left\langle \hat{H}_{e-ph}(r_2)\hat{H}_{e-ph}(r_1)\right\rangle \tag{2.25}$$

$$D_t(r_1,r_2) = \theta(t_1,t_2)D^{>}(r_1,r_2) + \theta(t_2,t_1)D^{<}(r_1,r_2) \tag{2.26}$$

$$D_{\bar{t}}(r_1,r_2) = \theta(t_2,t_1)D^{>}(r_1,r_2) + \theta(t_1,t_2)D^{<}(r_1,r_2) \tag{2.27}$$

$$D_r(r_1,r_2) = D_t - D^{<} = D^{>} - D_{\bar{t}} - \theta(t_1,t_2)\left(D^{>} - D^{<}\right) \tag{2.28}$$

$$D_a(r_1,r_2) = D_t - D^{>} = D^{<} - D_{\bar{t}} - \theta(t_2,t_1)\left(D^{>} - D^{<}\right) \tag{2.29}$$

where $\hat{H}_{e-ph}(x)$ is the second quantized form of the perturbation due to the electron-phonon interaction,

$$\hat{H}_{e-ph}(x) = \sum_{\mathrm{q}} M_{\mathrm{q}\lambda} e^{i\mathrm{q}\cdot\mathrm{r}}\left(\hat{a}_{\mathrm{q}\lambda}e^{-i\omega_{\mathrm{q}\lambda}t} + \hat{a}^{+}_{\text{-}\mathrm{q}\lambda}e^{i\omega_{\mathrm{q}\lambda}t}\right) \tag{2.30}$$

The creation and annihilation operators of the greater than and lesser than Green functions, can be reduced to that of the phonon occupation numbers for the given mode of phonons λ. That is, $\left\langle \hat{a}^{+}_{\mathrm{q}\lambda}\hat{a}_{\mathrm{q}\lambda}\right\rangle = N_{\mathrm{q}\lambda}$.

$$D^{>}(r_1,r_2) = -i\sum_{\mathrm{q}}\left|M_{\mathrm{q}\lambda}\right|^2\left[\left(N_{\mathrm{q}\lambda}+1\right)e^{-i\omega_{\mathrm{q}\lambda}(t_1-t_2)} + N_{\mathrm{q}\lambda}e^{i\omega_{\mathrm{q}\lambda}(t_1-t_2)}\right]e^{i\mathrm{q}\cdot(\mathrm{r}_1-\mathrm{r}_2)} \tag{2.31}$$

$$D^{<}(r_1,r_2) = -i\sum_{\mathrm{q}}\left|M_{\mathrm{q}\lambda}\right|^2\left[\left(N_{\mathrm{q}\lambda}+1\right)e^{i\omega_{\mathrm{q}\lambda}(t_1-t_2)} + N_{\mathrm{q}\lambda}e^{-i\omega_{\mathrm{q}\lambda}(t_1-t_2)}\right]e^{i\mathrm{q}\cdot(\mathrm{r}_1-\mathrm{r}_2)} \tag{2.32}$$

The equilibrium form of the retarded phonon Green's function is,

$$D_r(r_1,r_2) = -2\theta(t_1-t_2)\sum_{\mathrm{q}}\left|M_{\mathrm{q}\lambda}\right|^2 e^{i\mathrm{q}\cdot(\mathrm{r}_1-\mathrm{r}_2)}\sin\left[\omega_{\mathrm{q}\lambda}\left(t_1-t_2\right)\right] \tag{2.33}$$

2.6. EQUATIONS OF MOTION FOR THE GREEN'S FUNCTION

Continuing from the previous section, the expression for the contour ordered Green's function is,

$$\mathrm{G_C}(r_1,r_2) = \begin{bmatrix} G_t & G^< \\ G^> & G_{\bar{t}} \end{bmatrix} \tag{2.34}$$

The Keldysh Green's function is of the form

$$\mathrm{G}_K(r_1,r_2) = \begin{bmatrix} G_r & G_K \\ 0 & G_a \end{bmatrix} \tag{2.35}$$

where $G_K = G^> + G^<$ is the so-called Keldysh Green's function [38]. The Keldysh form of the Green's function is simply the integral form of Dyson's equations written *in real time* upon the application of Langreth's theorem (involving series multiplication).

Under the assumption that these field operators for the above Green's functions are based upon wave-functions that satisfy the Schrödinger equation, one can calculate the equations of motion for the various Green's functions. Thus, the non-interacting (bare) ground state Green's function is given by,

$$\left(i\hbar\frac{\partial}{\partial t_1} - H_o\left(x_1\right) - V\left(x_1\right)\right)\mathrm{G}_K^0(r_1,r_2) = \delta(r_1,r_2)\mathrm{I} \tag{2.36}$$

$$\left(-i\hbar\frac{\partial}{\partial t_1} - H_o\left(x_1\right) - V\left(x_1\right)\right)\mathrm{G}_K^0(r_1,r_2) = \delta(r_1,r_2)\mathrm{I} \tag{2.37}$$

where **I** is the identity matrix. $V(x)$ here is the single-point potential, which corresponds to the potential energy term and not the two-point potential energy operator for particle interactions. These operators represent the plain addition of electrons to the system, without considering any interaction terms.

In this study, as a real device is in a near-equilibrium condition, one needs to consider the driving field and the dissipation processes like Coulomb and phonon scattering into this model. Thus the, equation gets modified on the RHS to include two-point potential terms and higher order Green's functions (two-particle, three-particle Green's function etc) at the ground state. These can further be reduced to a concise form, known as the self-energy term. This is done by using Feynman diagram to expand the perturbative series of the Green's function (in terms of the basic blocks such as the non-interacting Green's function) according to Wick's theorem. What remains as the irreducible part is the self-energy. It can be thought of as the change in the particle's energy due to its interaction with the surrounding system under consideration. In treating transport in our system, the electron's interaction with the ionized impurity, surface roughness and the phonons can be accounted through the self-energy terms. The equations of motion for the full Green's function (*dressed* Green's function) as

$$\left(i\hbar\frac{\partial}{\partial t_1}-H_o\left(x_1\right)-V\left(x_1\right)\right)\mathrm{G}_K(r_1,r_2)=\delta(r_1,r_2)\mathrm{I}+\int dx_3\Sigma_K(r_1,r_3)\mathrm{G}_K(r_3,r_2) \quad (2.38)$$

$$\left(-i\hbar\frac{\partial}{\partial t_1}-H_o\left(x_1\right)-V\left(x_1\right)\right)\mathrm{G}_K(r_1,r_2)=\delta(r_1,r_2)\mathrm{I}+\int dx_3\mathrm{G}_K(r_1,r_3)\Sigma_K(r_3,r_2) \quad (2.39)$$

where the self-energy matrix is given by Keldysh,

$$\Sigma_K = \begin{bmatrix} \Sigma_r & \Sigma_K \\ 0 & \Sigma_a \end{bmatrix} \tag{2.40}$$

The self-energy terms account for both one-point and two-point potentials. The one-point potentials represent scattering from say a fixed localized potential like Coulomb scattering (thus a single time argument), and the two-point potentials represents particle-particle interaction potentials, like in the case of electron-phonon interaction, where electron interacts with a quantized phonon field.

The corresponding Dyson's equations for the Keldysh matrix Green's function, given in (2.40), (2.39) are

$$\mathrm{G}_K = \mathrm{G}_{K0} + \mathrm{G}_{K0}\Sigma_K\mathrm{G}_K \tag{2.41}$$

$$\mathrm{G}_K = \mathrm{G}_{K0} + \mathrm{G}_K\Sigma_K\mathrm{G}_{K0} \tag{2.42}$$

The equations of motion for the less-than and greater-than Green's functions are

$$G^{>,<} = \left(1 + G_r\Sigma_r\right)G_o^{>,<}\left(1 + \Sigma_a G_a\right) + G_r\Sigma^{>,<}G_a \tag{2.43}$$

The retarded and advanced Green's functions satisfy the Dyson equation

$$G_{r,a} = G_{r,a}^o + G_{r,a}^o\Sigma_{r,a}G_{r,a} \tag{2.44}$$

Notice that $G^<$ and the Dyson equations are coupled together, in non-equilibrium one needs to find the solution of $G^<$, so that one can calculate non-equilibrium properties like number operators, current density, etc. In the present work, as we treat near-equilibrium condition, we need to solve only the Dyson equation for the retarded Green's function (to get equilibrium properties like Density of states –DOS) to account for the collisional broadening of the states due to scattering. That is, one calculates the spectral

density function, $A = i(G_r - G_a)$, and then integrates it over the momentum states to get the density of the states (DOS) function.

2.7. EVALUATION OF GREEN'S FUNCTIONS

In the preceding section, the equations of motion for the Green's functions (specifically for the retarded Green's functions) were specified in a generic non-equilibrium picture. In this section a brief recap will be given on how these Green's functions are evaluated based on the work of Vasileska[21][23], who has adopted the NEGF approach to a near-equilibrium condition, which we shall refer to as near-equilibrium Gren's functions (nEGF) from hereinafter. Thus, one can learn how to obtain the one-electron properties, namely the density of states function within this picture.

As stated earlier, all of the one-electron properties to describe the system under consideration can be obtained directly from the one-electron Green's function (by taking expectation values of any observables required). Thus, to solve Eq. (2.44), one needs to evaluate the non-interacting bare retarded Green's function and then calculate the self-energy terms from the scattering matrices for the different mechanisms, and then solve iteratively for the interacting retarded Green's function. Once one obtains the Green's functions, the expectation value of any observable of the system can be calculated.

The equation of motion for the bare Green's function (non-interacting) using Eq. (2.38), is obtained by having both $V(\mathrm{r})$ and Σ equal to zero, i.e.

$$\left(i\hbar\frac{\partial}{\partial t} - H_o(\mathrm{R})\right)G^o(\mathrm{R},\mathrm{R}',t-t') = \delta\left(\mathrm{R}-\mathrm{R}'\right)\delta\left(t-t'\right) \tag{2.45}$$

where $H_o(\mathrm{R})$ is the equilibrium Hamiltonian of the system. Thus, the non-interacting Green's function in momentum space is given by,

$$G^o(z,z',\mathrm{k},\omega)=\sum_n \psi_n^*(z')\psi_n(z)g_{on}(\mathrm{k},\omega)=\sum_n \psi_n^*(z')\psi_n(z)\frac{1}{\hbar\omega-\varepsilon_{\mathrm{k}}-\varepsilon_n} \quad (2.46)$$

Here, $g_{on}(\mathrm{k},\omega)$ is the Fourier transform of the unperturbed subband Green's function $g_{on}(\mathrm{r}-\mathrm{r}',t-t')$. The unperturbed retarded (advanced) Green's functions are then obtained by,

$$G^o_{r/a}(z,z',\mathrm{k},\omega)=G^o(z,z',\mathrm{k},\omega\pm i\eta)\Big|_{\eta\to 0} \quad (2.47)$$

where η is the convergence factor. Once the unperturbed retarded and advanced Green's functions are known, one can calculate the corresponding spectral density function (SDF) from

$$a_{on}(\mathrm{k},\omega)=-2\,\mathrm{Im}\left(g^r_{on}(\mathrm{k},\omega)\right) \quad (2.48)$$

and density of states (DOS) function

$$\rho_o(\omega)=\frac{1}{\pi}\sum_{n,\mathrm{k}} a_{on}(\mathrm{k},\omega)=\sum_n \frac{m^*}{\pi\hbar^2}\theta\left(\hbar\omega-\varepsilon_k-\varepsilon_n\right) \quad (2.49)$$

The full Green's function is now calculated in the *self-consistent Born approximation* (graphically shown in Figure 2.1, for the impurity scattering case) and the broadening of the electronic states is calculated self-consistently. The approximation is just a restatement of the assumption that in a weak scattering regime, where the incident field is much larger than the scattering field, one can assume that the scattering potential does not significantly alter the wavefunction. That is the total field can be replaced with

the incident field. Once this approximation is used, one gets a first-order value for the retarded subband Green's function for the initial guess value. That is when looking at (2.44), one notices that the retarded subband self-energy occurs on both the RHS and LHS of the equation, thus a first-order approximation is required for the self-energy (and the G_R) before one can proceed to solve the equation self-consistently. The above is referred to as *Self-Consistent Born Approximation.*

Extending Eq.(2.46) to full Green's function in the assumption it has the same form under the diagonal approximation,

$$G_r(\mathrm{R},\mathrm{R}',t-t') = \sum_n \psi_n^*(z')\psi_n(z) g_n^r(\mathrm{r},\mathrm{r}',t-t') \tag{2.50}$$

Though the above equation looks like a diagonal one, the self-energy when properly expressed, consists of the various scattering matrix elements, and thus accounts for the different off-diagonal elements corresponding to the different subband indices.

From the above equation, using the diagonal approximation on the Full Green's function (the coupled Dyson equation), and using the value of unperturbed Green's function, the subband retarded Green's function is calculated as,

$$g_n^r(\mathrm{k},\omega) = \frac{1}{\hbar\omega - \varepsilon_{\mathrm{k}} - \varepsilon_n - \Sigma_n^r(\mathrm{k},\omega)} \tag{2.51}$$

and the corresponding retarded subband self-energy is given by,

$$\Sigma_n^r(\mathrm{k},\omega) = \iint dz dz_1 \left[\psi_n^*(z)\Sigma_r(z,z_1,\mathrm{k},\omega)\psi_n(z_1)\right] \tag{2.52}$$

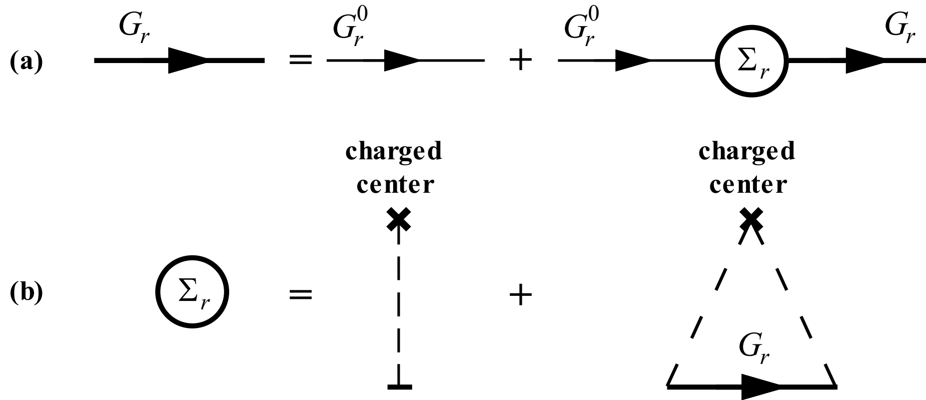

Figure 2.1 : Feynman's diagram:

(a) Eq. (2.44) : Dyson's equation for the retarded Green's function. Thin line represents the non-interacting Green's function and the thick line represents the interacting/full Green's function.

(b) Eq. (2.53) : Self-consistent Born approximation for scattering from impurities.

In silicon, where scattering is considered to be weak (and hence independent), the above integral over the self-energy can be split into separate contributions from each of the scattering mechanisms. Also, applying the self-consistent Born Approximation, the self-energy is calculated in terms of the scattering matrix elements as,

$$\Sigma_n^r(\mathrm{k},\omega) = \sum_m \sum_{\mathrm{q}} \sum_i \left| U_{nm}^i(\mathrm{q}) \right|^2 g_m^r(\mathrm{k}-\mathrm{q},\omega) \tag{2.53}$$

The matrix elements that appear in (2.53) are given in Chapter 3. The summation over i is for the different scattering processes under consideration over m subbbands.

The retarded Green's function is always of the form,

$$g_n^r(\mathrm{k},\omega) = \frac{1}{\hbar\omega - \varepsilon_{\mathrm{k}} - \varepsilon_n - R_n(\mathrm{k},\omega) + i\Gamma_n(\mathrm{k},\omega)} \tag{2.54}$$

where $R_n(\mathrm{k},\omega)$ gives the shift in the subband energies, and $\Gamma_n(\mathrm{k},\omega)$ is proportional to

the inverse of the lifetime of the n-th state (τ_n), where,

$$\Gamma_n(\mathrm{k},\omega)=\frac{1}{2}\sum_m\iint\frac{d^2q}{(2\pi)^2}a_m(\mathrm{q},\omega)\sum_i\left|U^i_{nm}\left(\mathrm{k}-\mathrm{q}\right)\right|^2 \tag{2.55}$$

$$R_n(\mathrm{k},\omega)=\sum_m\iint\frac{d^2q}{(2\pi)^2}\sum_i\left|U^i_{nm}\left(\mathrm{k}-\mathrm{q}\right)\right|^2\frac{\hbar\omega-\varepsilon_\mathrm{q}-\varepsilon_m-R_m(\mathrm{q},\omega)}{\left[\hbar\omega-\varepsilon_\mathrm{q}-\varepsilon_m-R_m(\mathrm{q},\omega)\right]^2+\Gamma^2_m(\mathrm{q},\omega)} \tag{2.56}$$

$\Gamma_n(\mathrm{k},\omega)$ denotes the broadening of the electronic states of the n-th subband and $R_n(\mathrm{k},\omega)$ denotes the energy renormalization term (the shift in the subband energy for the peak of the Lorentzian SDF), where $a_{on}(\mathrm{k},\omega)=-2\,\mathrm{Im}\left(g^r_{on}(\mathrm{k},\omega)\right)$

From Eqn. (2.55), the self-energy terms explicitly link the scattering over the different subbands, thus requiring a self-consistent solution of the Γ function (self-energy's imaginary term). The DOS function then, as explained previously for the non-interacting Green's function, is calculated as a summation over the momentum states of this spectral density function.

The following chapters will give a short review of the approach followed to calculate the screening and conductivity using the Green-Kubo Approach followed by the implementation details of the nEGF solver and its coupling with SCHREDV2.0.

CHAPTER 3. REVIEW ON SCHRED AND SCATTERING MECHANISMS IN SILICON

3.1. SCHRED - RECAP

SCHRED V2.0 is a generalized Schrödinger Poisson Solver, that can treat a multi-valley (user-defined number of conduction bands) semiconductor and specific crystallographic directions (Silicon specific) [39]. The tool has several unique features, namely, the ability to treat strain in silicon, and to treat any other semiconductor capacitor structures made of a material that can be represented using a three CB valley system. This chapter aims to give a brief overview of the SCHRED V2.0 and its capabilities.

3.1.1. MODELS AND FEATURES

In a general case, electrons respond to applied fields with an effective mass that depends on this crystallographic orientation of the field. Hence, in common cubic semiconductors, the dispersion relation in the parabolic band approximation is given by,

$$E_k = \hbar^2 \left[\frac{k_l^2}{2m_l^*} + \frac{k_t^2}{m_t^*} \right] \tag{3.1}$$

Eq. (3.1) describes a band with ellipsoidal constant energy surfaces. The effective mass is a diagonal tensor with different longitudinal and transverse effective masses, m_l and m_t, respectively. Eq. (3.1) is referred to as parabolic band approximation. SCHRED V2.0 employs the above bandstructure method for Silicon. It also employs a coordinate transformation method for a homogenous semiconductor, that uses the given principal effective masses (ellipsoidal effective masses), and a specific given crystallographic

direction (transport, width and wafer directions), to calculate its respective masses in the device coordinate system.

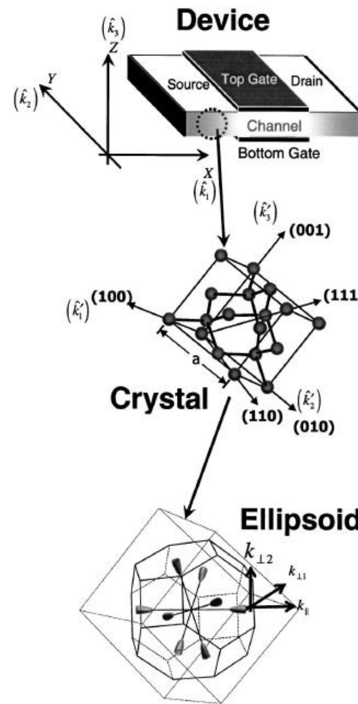

Figure 3.1 : Schematic description of the three orthogonal coordinate systems: device coordinate system (DCS), crystal coordinate system (CCS), and ellipse coordinate system (ECS). *From Lundstrom and co-workers, with permission*.

A homogenous semiconductor can be modeled using SCHRED V2.0 by considering the general structure of the substrate material. In general, the conduction band valley of the material has three valley pairs, which, in turn, have different effective masses along the chosen crystallographic directions. Thus, for a general conduction band ellipsoid (assuming 3 valleys) in the ellipse coordinate system (ECS),

$$E = \frac{\hbar^2 k_{\parallel}^2}{2m_1} + \frac{\hbar^2 k_{\perp 1}^2}{2m_2} + \frac{\hbar^2 k_{\perp 2}^2}{2m_3} \quad . \tag{3.2}$$

In Eq.(3.2), the k-space origin is translated to the conduction-band minima, which serves as the reference for the electronic energy. In compact vector notation, Eq. (3.2) can be written as

$$E = \frac{\hbar^2}{2} k_E^T \left(M_E^{-1} \right) k_E \quad , \tag{3.3}$$

where $k_E = \left(k_{\parallel} k_{\perp 1} k_{\perp 2} \right)^T$ consists of the components of an arbitrary wave vector in the ECS and the inverse M_E^{-1} is a 3×3 diagonal matrix with $m_l^{-1}, m_t^{-1}, m_t^{-1}$ along the diagonal. For a given channel material and for a given conduction band ellipsoid, the directions of the unit basis vectors $k_{\parallel}, k_{\perp 1}, k_{\perp 2}$ relative to the crystal coordinate system (CCS) are known, thus allowing one to write the 3×3 rotation matrix $R_{E \leftarrow C}$, which transforms the components of an arbitrary vector $k_C = (k_1^{'} . k_2^{'} . k_3^{'})^T$ defined in the CCS, to its components in the ECS, i.e.

$$k_E = R_{E \leftarrow C} k_C \quad . \tag{3.4}$$

A similar rotation matrix $R_{C \leftarrow D}$ transforms a wavevector $k_D = (k_1 . k_2 . k_3)^T$ in the device coordinate system (DCS) to k_C in the CCS as

$$k_C = R_{C \leftarrow D} k_D \quad . \tag{3.5}$$

Combining Eq. (3.4) and (3.5) we obtain

$$k_E = R_{E \leftarrow D} k_D \ , \tag{3.6}$$

where the rotation matrix is defined as $R_{E \leftarrow D} = R_{E \leftarrow C} R_{C \leftarrow D}$. Thus

$$E = \frac{\hbar^2}{2} k_D^T \left(M_D^{-1} \right) k_D \ , \tag{3.7}$$

where the inverse effective mass in the DCS is

$$\left(M_D^{-1} \right) = R_{E \leftarrow D}^T \left(M_E^{-1} \right) R_{E \leftarrow D} \quad . \tag{3.8}$$

The different co-ordinate system transformation can be understood from Schematic description of the three orthogonal coordinate systems: device coordinate system (DCS), crystal coordinate system (CCS), and ellipse coordinate system (ECS). *From Lundstrom and co-workers, with permission..* Thus, any homogenous semiconductor, with a specific crystallographic direction (only certain ones) can be represented in the above manner to compute its effective masses along particular high-symmetry crystallographic directions.

Another important feature of the present SCHRED V2.0 code is its ability to solve for conduction band valleys with different offsets. The main effect of strain in tensile strained-Si (that leads to enhanced electron low-field mobility), occurs in the energy band structure. There is a splitting of the two-fold degenerate heavy and light hole bands, which leads to corresponding modifications of the hole effective masses in the valence band. In addition, the six-fold-degenerate conduction-band valleys split into two separate sets of bands: a two-fold degenerate, perpendicular Δ_2-band and a four-fold degenerate, in-plane Δ_4-band. To first order, the ellipsoidal shape of each band in **k**-space is not deformed, so unlike the valence band case, the effective mass of the conduction band remains unchanged. However, the relative energies of each conduction band do shift.

Sometimes, the Δ_2 and Δ_4 bands energy splitting is as large as 0.3 eV, which is one order of magnitude larger than the thermal energy, even at room temperature.

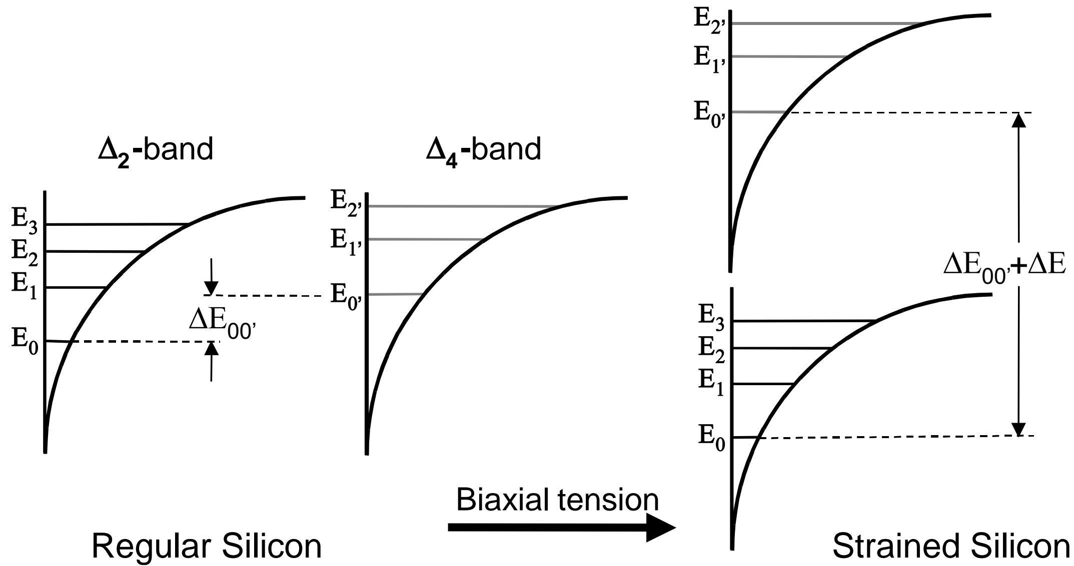

Figure 3.2: Subband structure in the inversion layer of regular and surface-channel strained-Si layer.

This wide splitting suppresses intervalley phonon scattering of electrons from lower valleys to upper valleys, and therefore, reduces the intervalley phonon scattering rate compared with that of the unstrained Si. In the lowered valleys, electrons show the smaller transverse mass in the transport parallel to the interface. These two factors are considered to be the main mechanisms for the observed high mobilities and high transconductances in devices that employ strained-Si layers.

In present day MOS capacitors several modifications to the device structures have been done in order to overcome the high-field effects, specifically the increasing transistor gate leakage current from the ever-thinning oxides. The thinner the oxides,

more quantum mechanical tunneling happens between the gate and the substrate through the oxide layer at high electric fields. Thus, adding the High-K dielectric layer to the existing SiO_2 is a method by which the physical thickness of the gate oxide is increased (by approximately the ratio of the old/new permittivity) without increasing the electrical thickness of the gate (maintaining the electric field at the Si-SiO_2 interface), enabling reduction in the gate leakage current, especially at higher gate fields. SCHREDV2.0 can model this extra layer of high-K dielectric quite effectively by including an extra layer of oxide with a different permittivity.

3.1.2. THE POISSON EQUATION

In order to solve for the potential in the device, one has to solve the 1-Dimensional Poisson's equation,

$$\frac{\partial}{\partial x}\left(\varepsilon(x)\frac{\partial \varphi(x)}{\partial x}\right) = -\rho(x) \ , \qquad (3.9)$$

Where $\varphi(x)$ is the spatially varying potential, $\varepsilon(x)$ is the spatially varying permittivity, ρ is the total charge density.

The Poisson equation is discretized on a finite-difference mesh, for a generalized non-uniform permittivity case. By discretizing the Poisson equation on a general non-uniform mesh without requiring specific boundary conditions, one can effectively solve Poisson equation for any material by varying the permittivity matrix (including the uniform permittivity case, where the matrix reduces to single constant value). Also it is quite easier now to model different layers of materials, as the permittivity takes care of

the boundaries between the layers of different material and does not need any specific boundary conditions for the interface on the Poisson equation mesh.

The normalized form of the Poisson equation is given by.

$$\frac{\varepsilon_{i+1}^{r}+\varepsilon_{i}^{r}}{x_i\left(x_i+x_{i-1}\right)}\varphi_{i+1}^{new}-\left[\frac{\varepsilon_{i+1}^{r}+\varepsilon_{i}^{r}}{x_i\left(x_i+x_{i-1}\right)}+\frac{\varepsilon_{i}^{r}+\varepsilon_{i-1}^{r}}{x_{i-1}\left(x_i+x_{i-1}\right)}+\left(p+n\right)\right]\varphi_{i}^{new}$$
$$+\frac{\varepsilon_{i}^{r}+\varepsilon_{i-1}^{r}}{x_{i-1}\left(x_i+x_{i-1}\right)}\varphi_{i-1}^{new}=-\left(p-n+C\right)-\varphi_{i}^{old}\left(p+n\right) \tag{3.10}$$

where, the new complete coefficients are

$$c_i=-\frac{\varepsilon_{i+1}^{r}+\varepsilon_{i}^{r}}{x_i\left(x_i+x_{i-1}\right)},$$
$$e_i=-\frac{\varepsilon_{i}^{r}+\varepsilon_{i-1}^{r}}{x_{i-1}\left(x_i+x_{i-1}\right)},$$
$$g_i=\left[\frac{\varepsilon_{i+1}^{r}+\varepsilon_{i}^{r}}{x_i\left(x_i+x_{i-1}\right)}+\frac{\varepsilon_{i}^{r}+\varepsilon_{i-1}^{r}}{x_{i-1}\left(x_i+x_{i-1}\right)}+\left(p+n\right)\right],$$
$$R_{0i}=\left(p-n+C\right)+\varphi_{i}^{old}\left(p+n\right) \tag{3.11}$$

Where R_0 is the forcing function.

The finite difference discretization of the 1D Poisson's equation leads to tridiagonal matrix. That is, the Poisson equation can now be represented as,

$$\left[\mathrm{A}\right]\left[\phi\right]=\left[F\right] \tag{3.12}$$

Where F is the forcing function (which is essentially the RHS of Equation (3.12)

The lower-upper triangular matrix (LU) decomposition technique is used to solve the Poisson equation of the above form $\left[\mathrm{A}\right]\left[\phi\right]=\left[F\right]$. In this method the matrix $\left[\mathrm{A}\right]$ is broken

down into upper and lower triangular matrices. $[\phi]$ is then solved by forward and backward substitution.

3.1.3. TIME INDEPENDENT SCHRÖDINGER WAVE EQUATION

The Schrödinger equation is also discretized on a finite difference mesh as well. Thus,

$$-\frac{\hbar^2}{m^* x_i\left(x_i+x_{i-1}\right)}\psi_{i+1}+\left[\frac{\hbar^2}{m^* x_i\left(x_i+x_{i-1}\right)}+\frac{\hbar^2}{m^* x_{i-1}\left(x_i+x_{i-1}\right)}+V_i\right]\psi_i-\frac{\hbar^2}{m^* x_{i-1}\left(x_i+x_{i-1}\right)}\psi_{i-1}$$
$$=E_i\psi_i$$

(3.13)

The above equation is of the form of an eigenvalue problem,

$$\mathrm{A}x=\lambda x \tag{3.14}$$

One way to solve the above problem is to use the online eigenvalue solver libraries like EISPACK (EISPACK is a FORTRAN90 library which calculates the eigenvalues and eigenvectors of a matrix). EISPACK routines, however, require A to be a symmetric matrix. Examining the coefficients of Eq.(3.13), the ψ_{i+1},ψ_{i-1} have x_i, x_{i+1} terms in the denominator that make the matrix asymmetric. Thus, a symmetrization technique has to be employed [40].

The number of eigenvalues and the fact that smallest eigenvalues are to be determined is also specified through an input variable. The output of this program is a new matrix with the required number of eigenvalues in ascending order as specified.

The eigenvalues are then fed as an input to the eigenvector solver. The eigenvector subroutine finds those eigenvectors of a tri-diagonal symmetric matrix corresponding to

specified eigenvalues, using inverse iteration. The calculated eigenvalues represent the subband energy and their corresponding eigenvectors represent the wavefunction of the carriers in the quantum well.

Thus, by solving the above eigenvalue problem, we get the values of the subband energy and the corresponding eigenvectors given the wavefunction in that subband energy level. Once this information is obtained, then the quantum sheet charge density is computed by summing for all the subband contributions.

3.1.4. 2D SHEET CHARGE DENSITY AND TOTAL CHARGE DENSITY

The population of the various subbands is described by the sheet electron density N_n (no of carriers per unit area) by,

$$N_n = \int_0^\infty \rho^{2D}(E) f(E)\, dE \tag{3.15}$$

Where $\rho^{2D}(E)$ is the 2D density of states function (This will get replaced by the scattering based real DOS once we combine the nEGF solver - will output the broadened DOS function), $f(E)$ is the Fermi-Dirac distribution function.

Evaluating integral given by Eq.(3.15) gives,

$$\begin{aligned} N_n &= \int_0^\infty \nu_s \nu_v \frac{m^*}{2\pi\hbar^2} \frac{1}{\left(1+\exp\left(\frac{E_F - E}{K_B T}\right)\right)} dE \\ N_n &= \nu_s \nu_v \frac{m^*}{2\pi\hbar^2} K_B T \log\left(1+\exp\left(\frac{E_F - E_n}{K_B T}\right)\right) \end{aligned} \tag{3.16}$$

Where K_B is the Boltzmann constant, T is the temperature, E_n is the subband energy, E_F is the Fermi energy. The electron density is then calculated over all the subbands to get,

$$n(z) = \sum_n N_n \psi_n^2(z) \ . \tag{3.17}$$

3.1.5. FLOW CHART OF THE SCHREDV2.0 PROGRAM

As shown in the flowchart in Figure 3.3 : Flow Chart of SCHREDV2.0, the Schrödinger equation solver is coupled with the Poisson's equation solver and is iterated until a self-consistent solution is found. The Poisson's equation gives the value of the new potential based on which the Schrödinger equation is solved to obtain new values of subband energies and their wavefunctions. This, in turn, is used to calculate the new sheet charge density and, therefore, the total charge density, which is again used to calculate the new value of the potential by solving the Poisson's equation. The process is repeated until error value of the potential reaches a certain threshold. This process is repeated for the given voltage range. Note that if quantum confinement is not established, the charge is calculated classically for the next iteration step.

This Schrödinger-Poisson solver part forms the outer loop of the self-consistent calculation for the subband structure. The inner loop consists of the nEGF solver that solves self-consistently under the self-consistent Born Approximation of the value of the real DOS is obtained.

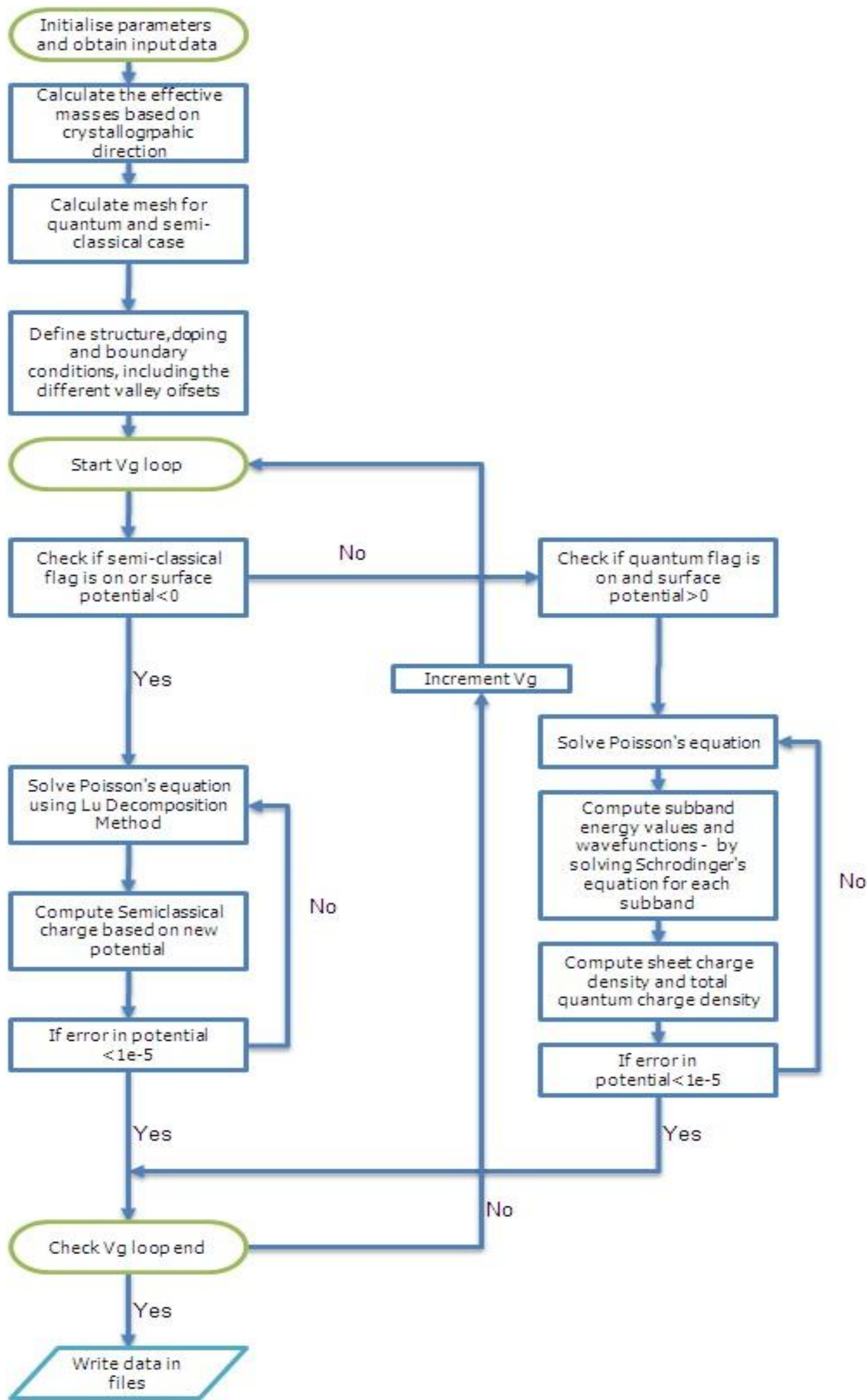

Figure 3.3 : Flow Chart of SCHREDV2.0 .

3.2. SCATTERING IN SILICON – DIFFUSIVE TRANSPORT

In semiconductor transport different scattering mechanisms play a fundamental role in limiting the carrier mobility under various bias conditions. In the following, we review treatment of the relevant scattering mechanisms based on the work of [23], namely the Coulomb scattering (from the depletion and the interface charges), electron-phonon scattering and the surface-roughness scattering, all of which play important roles under low applied fields in limiting the mobility of the carriers in the device. The expressions for the Matrix elements for Coulomb interaction, for scattering between subbands *n* and *m*, are given in section 3.2.1. Surface-roughness scattering is described in section 3.2.2, and in section 3.2.3 we briefly go through the theory of electron-phonon scattering.

3.2.1. COULOMB SCATTERING

Coulomb scattering in 2-Dimensional Electron Gas (2-DEG) can be from the depletion charges, interface charges or the oxide trapped charges. Thus, at low sheet electron densities (low screening) in the inversion layer, we expect Coulomb scattering due to ionized impurities to be important in limiting the mobility.

The matrix element for scattering between n and m subbands for Coulomb scattering is given by [42],

$$\left|\left\langle n\left|U^{depl}(q)\right|m\right\rangle\right|^2 = \left|U_{nm}^{depl}(\mathrm{q})\right|^2 = N_{depl}\left(\frac{e^2}{2\kappa q}\right)^2 A_{nm}^2(q)\int_o^\infty dz_i O_{nm}^2(q,z_i) \qquad (3.18)$$

where N_{depl} is the depletion charge density, z_i is the location of an arbitrary charge

center in the depletion region , **q** is the wavevector and $A_{nm}(q)$ and $O_{nm}(q,z_i)$ are the form factors in the quantized direction, where,

$$A_{nm}(q)=\int_0^{\infty} dz\psi_n(z)e^{-qz}\psi_m(z) \tag{3.19}$$

and,

$$\begin{aligned}O_{nm}(q,z_i)=0.5\left(1+\frac{\varepsilon_{ox}}{\varepsilon_{sc}}\right)e^{qz_i}+0.5\left(1-\frac{\varepsilon_{ox}}{\varepsilon_{sc}}\right)e^{-qz_i}\\ +0.5\left(1+\frac{\varepsilon_{ox}}{\varepsilon_{sc}}\right)\left[e^{-qz_i}\frac{a_{nm}^{(+)}(q,z_i)}{A_{nm}(q)}-e^{qz_i}\frac{a_{nm}^{(-)}(q,z_i)}{A_{nm}(q)}\right]\end{aligned} \tag{3.20}$$

where $a_{nm}^{(\pm)}(q,z_i)=\int_0^{z_i} dz\psi_n(z)e^{\pm qz}\psi_m(z)$.

In (3.20), **q** is a wavevector in the plane parallel to the interface (*xy*-plane in our case).

Similarly, scattering from interface trapped charges is given by,

$$\left|\left\langle n\left|U^{it}(\mathrm{q})\right|m\right\rangle\right|^2=\left|U_{nm}^{it}(\mathrm{q})\right|^2=N_{it}\left(\frac{e^2}{2\kappa}\right)^2\left[\frac{A_{nm}(q)}{q}\right]^2 e^{2qz_i} \tag{3.21}$$

where, N_{it} is the interface charge density located at a distance of z_i.

The matrix element for scattering from the oxide charge, with charge density N_{ox}, is given by,

$$\left|\left\langle n\left|U^{ox}(\mathrm{q})\right|m\right\rangle\right|^2=\left|U_{nm}^{ox}(\mathrm{q})\right|^2=N_{ox}\left(\frac{e^2}{2\kappa}\right)^2\left[\frac{A_{nm}(q)}{q}\right]^2\frac{1-e^{-2qd_{ox}}}{2q} \tag{3.22}$$

where d_{ox} is the oxide thickness.

3.2.2. SURFACE-ROUGHNESS SCATTERING

Surface-roughness (SR) scattering is the elastic scattering of a charged particle from the imperfect Si/SiO_2 interface. The roughness of the interface depends upon wafer processing conditions. Thus, the degree of roughness of the interface depends on the various processing parameters such as the oxidation and annealing temperatures. There are two components to interface-roughness: (1) fluctuation in the subband energy due to the fluctuation of the oxide thickness, which can be interpreted as fluctuating scattering potential leading to a change in the confining potential of the triangular well, and (2) modification of the wavefunction due to the modification of the well thickness and the penetration of the wavefunction in the oxide. The scattering potential causes scattering of the confined carriers, and can be treated perturbatively.

In this work, we use the same approach as used by [23][43], which follows the above idea. The power spectrum (power spectral density), which is a measure of roughness, is generally modeled as a Gaussian function given by,

$$S_G(q) = \pi\Delta^2\zeta^2 \exp\left(-\frac{q^2\zeta^2}{4}\right) \tag{3.23}$$

.

Parameters Δ and ζ characterize the rms height of the bumps on the surface and the roughness correlation length, respectively. The rms height of the bumps gives an indication of how rough the surface is, and the roughness correlation length gives us an idea of how close these imperfections are. Goodnick and co-workers [43] suggested exponential model for the roughness power spectral density after their experimental

measurements proved to be better fit with this model. The power spectral density of the exponential model is given by,

$$S_E(q) = \frac{\pi\Delta^2\zeta^2}{\left(1+q^2\zeta^2/2\right)^{3/2}} \tag{3.24}$$

Extending this exponential to a higher powers one can obtain a more generalized form of the power spectrum, know as self-affine roughness correlation function, that is given by the following expression,

$$S_{SA}(q) = \frac{\pi\Delta^2\zeta^2}{\left(1+q^2\zeta^2/4n\right)^{n+1}} \tag{3.25}$$

where, $n>0$ describes high-q falloff of the distribution. It reduces to exponential correlation for n=0.5 (Figure 3.4, Figure 3.5).

Following [23], the matrix element for scattering between subbands n and m for this scattering mechanism is of the form

$$\left|\left\langle n\middle|U^{sr}(\mathrm{q})\middle|m\right\rangle\right|^2 = S(q)\Gamma_{nm}^2(q) \tag{3.26}$$

Where Γ_{nm} is the matrix elements computed for surface roughness scattering matrix elements computed for surface roughness scattering between subbands n and m.

In order to estimate the value of Γ_{nm}, we use the results of Matsumoto and Uemura, who calculated that in the electronic quantum limit, $\Gamma_{nm}^{(0)} = eE_{av}$, where $E_{av} \propto \left(\frac{1}{2}N_s + N_{depl}\right)$. This is known as the MU [44] definition for SR scattering

Within the MU model, the scattering rate, $\Gamma_{nm}^{(0)}$, is given by,

$$\begin{aligned}\Gamma_{nm}^{(0)} &= \frac{\hbar^2}{2m_z}\frac{d\psi_n}{dz}\frac{d\psi_m}{dz}\bigg|_{z=0} \\ &= \int_0^\infty dz\left\{\psi_n(z)\frac{\partial V(z)}{\partial z}\psi_m(z)-\varepsilon_m\frac{d\psi_n}{dz}\psi_m(z)+\varepsilon_n\psi_n(z)\frac{d\psi_m}{dz}\right\}\end{aligned}. \tag{3.27}$$

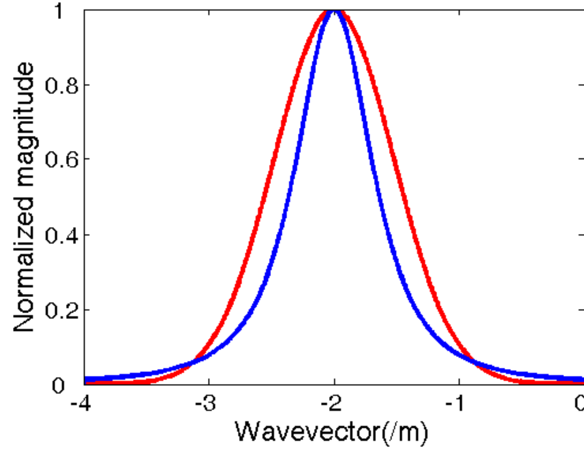

Figure 3.4 : Normalized magnitudes for Gaussian(red) and exponential(blue) models for roughness parameters $\Delta = 1.5\,nm$ and $L = 0.243\,nm$.

This result was further corrected by Ando [45] to account for image charge, to get

$$\Gamma_{nm}(q) = \Gamma_{nm}^{(0)} + \frac{e^2}{\varepsilon_{sc}}\frac{\varepsilon_{sc}-\varepsilon_{ox}}{\varepsilon_{sc}+\varepsilon_{ox}}A_{nm}(q)\left\{\left(N_{depl}+N_s\right)-\frac{1}{2}\sum_i N_i A_{ii}(q)\right\} \tag{3.28}$$

where $\Gamma_{nm}^{(0)}$ is given by (3.27), N_{depl} is the depletion charge density, N_s is the total sheet charge density, N_i is the sheet charge density for subband i and $A_{nm}(q)$ and $A_{ii}(q)$ are overlap integrals. The complete Ando model for the surface-roughness matrix element is employed in this work alongside with the exponential model for the roughness power spectral density.

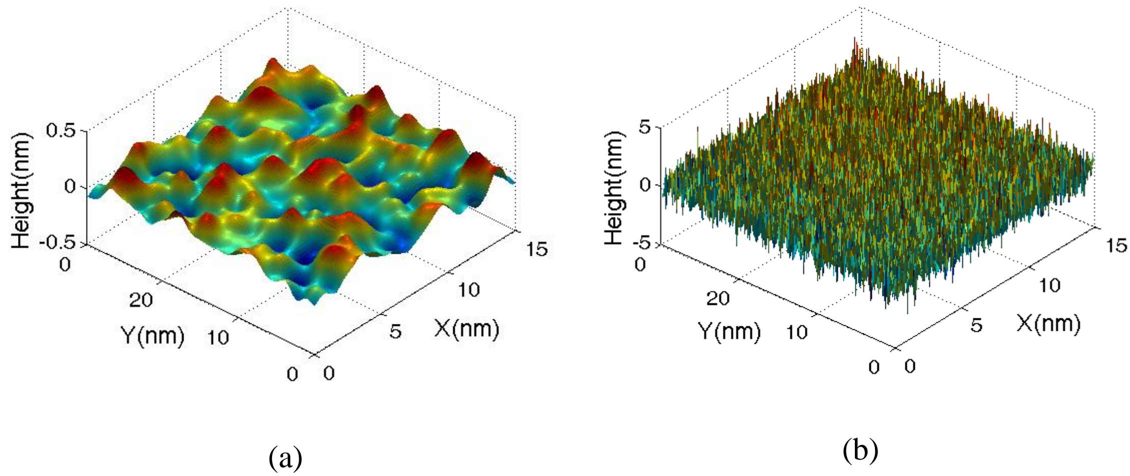

Figure 3.5 : 3D surface roughness model for the power spectrum – (a) Gaussian spectral model, (b) exponential spectral model for $\Delta = 1.5\, nm$ and $L = 0.243\, nm$

3.2.3. ELECTRON-PHONON INTERACTION

Phonons being quantized units of lattice vibrations, interact with the electrons in silicon primarily in three different ways, depending upon their modes of vibration – broadly classified as acoustic or optical. Firstly, an electron interacting with a low energy acoustic phonon results in an elastic acoustic phonon scattering process. Second, the electron interacts with the high energy low momentum phonons, thus transitioning between states in a valley, referred to as intravalley scattering (inelastic process). The third one is where an electron interacts with high-energy high momentum acoustic/optical phonon thus transitioning between states of the same and different valleys, known as intervalley phonon scattering – which is an inelastic process. Thus, as this is a high energy process, they become important when electrons get hot or if the lattice temperature gets high enough to have enough number of optical phonons to cause

significant interaction [28]. A short review of the final equations used in this work will be stated below.

The interaction Hamiltonian is given by,

$$H_{e-ph}(\mathrm{r}) = \sum_{\mathrm{q}} M_{\mathrm{q}\lambda} e^{i\mathrm{q}\cdot\mathrm{r}} \left(\hat{a}_{\mathrm{q}\lambda} + \hat{a}_{\mathrm{q}\lambda}^{+} \right) \tag{3.29}$$

where,

$$M_{\mathrm{q}\lambda} = \left(\frac{\hbar}{2\rho V \omega_{\mathrm{q}\lambda}} \right) \sum_{\mathrm{G}} e^{i\mathrm{G}\cdot\mathrm{r}} \left(\mathrm{q} + \mathrm{G} \right) \cdot \mathrm{e}_{\mathrm{q}\lambda} V_{ea} \left(\mathrm{q} + \mathrm{G} \right) \tag{3.30}$$

where $MN = \rho V$, and ρ is the density of the solid and **G** is the set of all the reciprocal lattice vectors of the solid, $\hat{a}_{q\lambda}\left(\hat{a}_{q\lambda}^{+}\right)$ are phonon annihilation, creation operators corresponding to a wavevector ***q***, and phonon branch λ .

The exact form of the matrix elements for acoustic and nonpolar-optical phonon scattering used in our calculations, are given in sections 2.3.1 and 2.3.2, respectively.

3.2.4. DEFORMATION POTENTIAL SCATTERING

The crystal potential of a semiconductor determines its bandstructure, and this potential is dependent upon the lattice spacing of the constituent atoms. Thus, when a mechanical stress is applied, this induces perturbation in the lattice constant which leads to changes in its crystal potential, which shows up as deformations or gratings in the bands (bandstructure). The potential by which it deforms is dependent on the lattice constant and is referred as deformed potential. These changes in bandstructure potential however are assumed to be small enough so that it does not change the curvature of the

bands, thus the effective mass of the bands remains unchanged.

Thus, lattice vibrations (acoustic phonons) can be thought of a similar strain wave that produces changes in the lattice constant, leading to its respective deformation potential, which are basically the perturbation energy of the bands. These deformation potentials can be deduced from experiments for almost all common semiconductors.

Now this elastic wave can be modeled in a continuum description, which is valid as we consider only long-wavelength phonons for intravalley scattering. Thus, for the isotropic case, we see that deformation potential for acoustic phonons becomes a constant denoted by Ξ (it gives the shift of the band edge per unit elastic strain).

The matrix element, $M_{Q\lambda}$ simplifies to,

$$M_{q'\lambda} = iq' \cdot \mathrm{e}_{q'\lambda} \left(\frac{\hbar \Xi^2}{2\rho V \omega_{q'\lambda}} \right)^{1/2} \tag{3.31}$$

where $\omega_{q'\lambda} = v_{s\lambda} q'$, where $v_{s\lambda}$ is the sound velocity, and $\omega_{q'\lambda}$ is the phonon frequency (both of which correspond to wavevector Q and phonon branch λ). Also, as transverse acoustic phonons (TA) do not contribute to the matrix element the only contribution to deformation potential in the isotropic case, comes from long-wavelength longitudinal acoustic phonons (LA).

For anisotropic semiconductors such as silicon, the deformation potential constant becomes a tensor due to the anisotropy of the bandstructure. So a couple of assumptions can be made in deriving the final expression for the matrix element. Firstly, from experiments it was found that it is a good approximation to use a single energy-dependant

scattering rate for the acoustic and longitudinal phonons. Thus, an effective deformation potential Ξ_{LA}^{eff}, can be used for the LA phonons instead of the tensor.

Also, under the condition that the lattice wavefunctions are essentially harmonic oscillator wavefunctions and the electron wavefunctions are Bloch functions, one can see that the action of the operators $\hat{a}_{q\lambda}^{+}$, $\hat{a}_{q\lambda}$ is to raise and lower energy state of the particular harmonic oscillator mode, thus they reduce to $\sqrt{N_{q\lambda}}$ and $\sqrt{N_{q\lambda}+1}$, respectively. For equilibrium phonons, the $N_{q\lambda}$ is given by the Bose-Einstein distribution.

In the elastic and equipartition approximation, the final expression for matrix elements for acoustic phonon scattering between subbands *n* and *m* is given by,

$$\left|\left\langle n\left|U_{\lambda}^{ac}(\mathrm{q})\right|m\right\rangle\right|^{2}=\frac{k_{B}T}{\rho V v_{s\lambda}^{2}}\left[\Delta_{\lambda,nm}^{eff}(\mathrm{q})\right]^{2}F_{nm} \tag{3.32}$$

The effective deformation potential constant is calculated from,

$$\left[\Delta_{\lambda,nm}^{eff}(\mathrm{q})\right]^{2}=\frac{1}{F_{nm}}\int_{0}^{\infty}dq_{z}\Delta_{\lambda}^{2}(\theta_{q'})\left|\mathbf{F}_{nm}\left(q_{z}\right)\right|^{2} \tag{3.33}$$

where the form factor is defined as,

$$\mathbf{F}_{nm}(q_{z})=\int_{0}^{\infty}dz\psi_{n}(z)e^{iq_{z}z}\psi_{m}(z) \tag{3.34}$$

and the angle $\theta_{q'}$ is between the wavevector $q^{'}$ of the emitted (absorbed) phonon and the longitudinal axis of the valley. Also, the form factor above indicates that the lower subband energy electrons will contribute a greater change in the deformation potential for their change of kinetic energy. This can be attributed to the fact that they will have more

momentum change as they are fixed in a direction.

3.2.5. NONPOLAR OPTICAL PHONON SCATTERING

Silicon being a nonpolar semiconductor has another significant electron-phonon interaction, specifically the nonpolar optical phonon scattering. Unlike the polar compound semiconductors where there is an added electrostatic interaction due to the change of the dipole moment due to the lattice constant perturbation, one can neglect this in nonpolar semiconductors like silicon. The optical phonons (specifically at the zone center) have a high energy and become important in modeling intra-subband and intervalley scattering. For example, in the case of silicon, the treatment of intervalley transitions due to scattering between the minima of the conduction bands becomes important when one considers the dependence of the scattering rate at high fields. Thus, higher order terms of the optical phonon-electron interaction have to be accounted for when deriving the matrix element for the scattering. However, in this work, we treat the device under a near-equilibrium condition, and as the aim of the work is not to solve for the transport parameters like drift velocity at high fields and current, we limit ourselves thus to the zero-order term in the intervalley phonon scattering treatment.

The squared matrix element for nonpolar optical phonon scattering used in this work is given by,

$$\left|\left\langle n\left|U_{\lambda}^{op(0)}\right|m\right\rangle\right|^{2}=\frac{\hbar D_{\lambda}^{2}}{2\rho V\omega_{o\lambda}}F_{nm} \tag{3.35}$$

Where $\omega_{o\lambda}$ is the frequency of the relevant phonon mode independent of the phonon

wave-vector, D_{λ} is the deformation field for that branch of mode λ, F_{nm} is the form factor for the interaction.

In the scattering among the equivalent valleys, there are two types of phonons that might be involved in the process, the g-phonons and the f-phonons (see Figure 3.6:). The g-phonon couples the two valleys along opposite ends of the same axis, i.e. $\langle 100 \rangle$ to $\langle \bar{1}00 \rangle$. The f-phonons couple the $\langle 100 \rangle$ valley with $\langle 010 \rangle$, $\langle 001 \rangle$, etc. Degeneracy factors (g_r) for transition between unprimed (α=1) and primed (α=2) set of subbands, for both g- (*r*=1) and f-phonons (*r*=2) are summarized in Table 3.1 : Degeneracy factors for transition between unprimed and primed subbands, for both g- and f-phonons.

The above scattering matrix elements for the various scattering mechanisms (including the inelastic intervalley scattering) is included into the retarded self-energy expression, and self-consistently solved on an uniform energy mesh to yield the retarded Green's function for the various subbands. From this retarded subband Green's function, one can calculate the broadening of the states, the density of states function and the mobility. More specifically, the intervalley optical phonon scattering has been accounted for in the phonon self-energy, and it is self-consistently solved for within the self-consistent Born Approximation for the first time in this work. This will be explained in detail in the following chapter, where the fundamentals of the Green's function will be first reviewed.

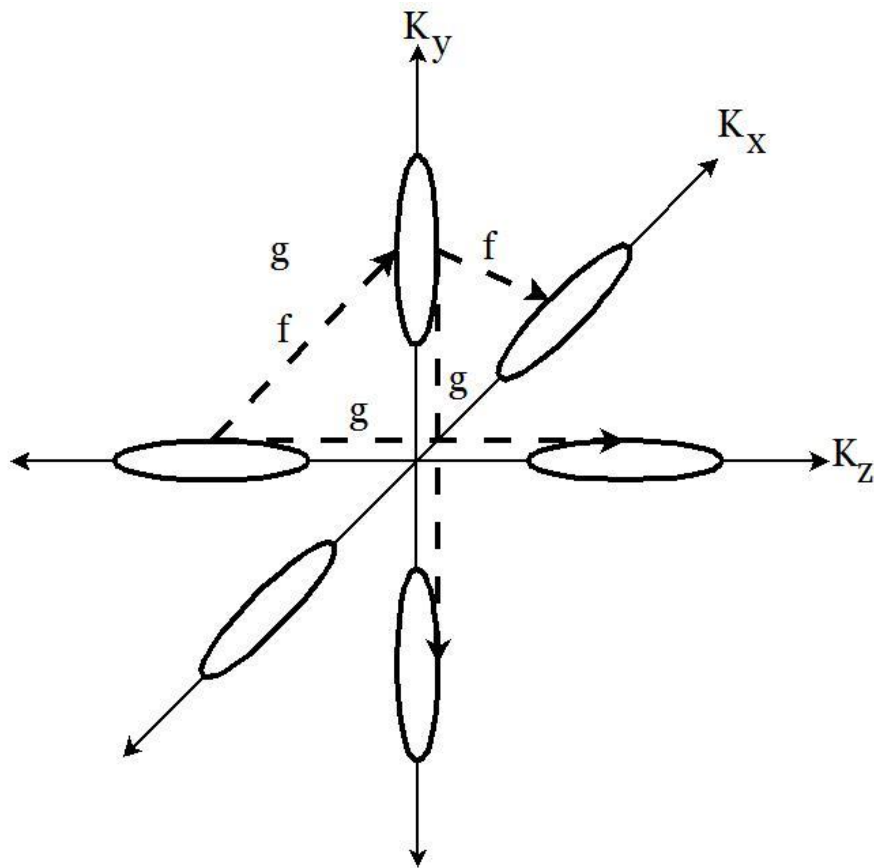

Figure 3.6: g-type and f-type phonons transitions

Table 3.1 : Degeneracy factors for transition between unprimed and primed subbands, for both g- and f-phonons.

Initial Valley/Final Valley	**α=1**	**α=2**
α=1	g_1=1; g_2=0	g_1=0; g_2=4
α=2	g_1=0; g_2=2	g_1=1; g_2=2

CHAPTER 4. MANY-BODY EFFECTS AND CONDUCTIVITY

4.1. REVIEW ON MANY-BODY EFFECTS

Scattering in mesoscopic systems, has traditionally been treated in a semi-classical manner. But the importance of treating screening adequately has always been the priority considering its direct influence on the scattering events (especially in the inversion charge region and temperature variations). Coulomb and surface-roughness scattering have long been known to dominate the low-field mobility at the low and high inversion charge densities, respectively. Thus, in order to treat scattering of these two mechanisms in its entirety, screening from these inversion charges has to be accounted for within a solid framework into our QM model.

The solution of the two-electron Green's function (the density-density correlation function) propagator gives an idea of how the two-particle excitation (plasma) propagates. This "plasma" of charge, now forms a screening potential as seen by the moving electron, and alters the scattering radius. This accounts for the screened scattering matrix elements that can then be included in the Green's function calculation.

On a similar argument the conductivity can also be estimated by calculating the corresponding two particle Green's function propagator for current-current correlations. The above work has been studied extensively by Vasileska and co-workers [23]. In this chapter we will just give a recap of the important results Vasileska and co-workers have derived to account for screening for these two scattering mechanisms and the calculation of the mobility using two particle Green's function propagator.

4.1.1. SCREENING UNDER RANDOM PHASE APPROXIMATION

Scattering of the inversion layer electrons is significantly affected by the screening from the inversion layer charges. In this work, the traditionally well established Random Phase Approximation (RPA) aka the mean-field approximation is used to model the modified dielectric function response. The RPA treats the Hartree potential under a self-consistent field of the external charge and the potential from the electron gas. This accounts for a good approximation of the (Lindhard) dielectric function. One could also view the RPA as a method that yields the dielectric function from a total electron potential whose oscillation averages out at a certain wavevector $\boldsymbol{q}$, hence random phase. It can be noted that the exchange and correlation effects are not treated in the RPA. These are accounted through a density-functional formalism in the one-electron Hamiltonian through the Hartfree-Fock theory. This has been included in the SCREDV2.0 while solving for the self-consistent potential.

The real-time Green's function formalism is used to derive the expression for the screened matrix elements, the polarizability function, and the corresponding susceptibility sensor. The dielectric function (density-density response function) corresponding to the bare polariziability function in Q2D systems is called the Lindhard dielectric function. This function is then obtained through analytic continuation of the dielectric tensor in 2D. This is then included in the retarded subband screened self-energy calculations stated in the previous chapter (through the screened interaction matrix elements).

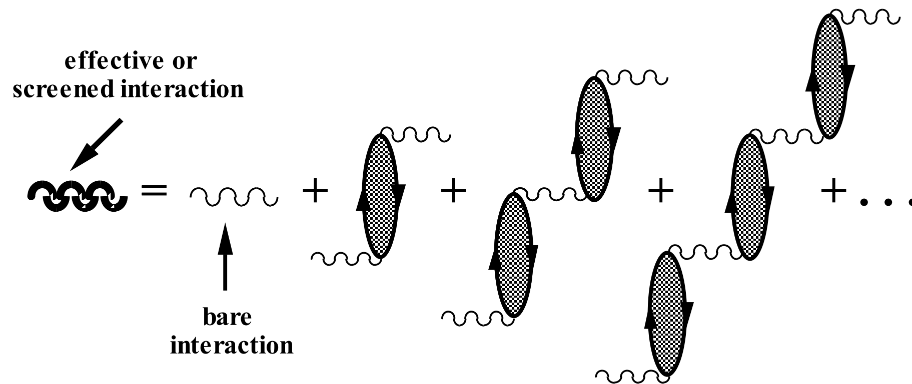

Figure 4.1: Feynman Diagram for the effective screened interaction (polarization diagram)

The above figure represents the perturbative expansion of the screened interaction under the Random Phase Approximation (without the exchange interaction within the particle-hole pair bubble). The shaded loop bubble represents the electron-hole pair interaction. It represents how the particle interacts with itself through the particle-hole bubble. These density fluctuations can be modeled as collective excitations that propagate through the medium using the two-particle Green's function propagator, also known as the density-correlation or polarization propagator.

In this above expansion the bubble-pair represented is the free polarization propagator. This represents the irreducible polarization part which cannot be broken down further as it does not account for any interaction within the bubble. This is the main assumption of RPA, in that the exchange interaction within the bubble is ignored and one proceeds to calculate the screened interaction as a perturbation series based upon the bare bubble-pair. This is valid assumption, since the exchange interaction within the polarization medium is much lesser than the direct interaction. Thus, the direct

interaction, leads to polarization of the medium, which, in turn, shields the main interaction making a weaker effective interaction.

Looking at Figure 4.1, one can see that the perturbation series resembles very closely the Dyson's equations; thus, the screened interaction can be finally written as,

$$W(x,x') = V(x,x') + \int dx_1 \int dx_2 V(x,x_1) P^{(0)}(x_1,x_2) W(x_2,x') \tag{4.1}$$

The diagrammatic representation of the same is given by,

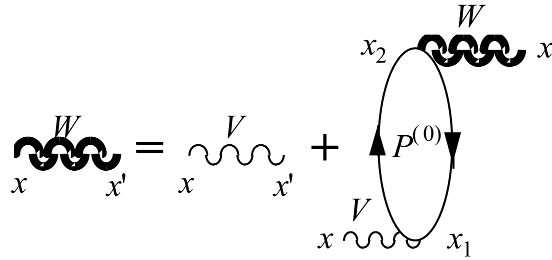

Figure 4.2 : Dyson's equation for the screened interaction in diagrammatical view

For the Coulomb scattering one assumes a locally varying static potential. Then the screened matrix elements for Coulomb interaction are of the form

$$W_{ij}^{eff}(\mathrm{q},\omega) = V_{ij}^{bare}(\mathrm{q}) + \frac{1}{q}\sum_{m,n} F_{ij,nm}(\mathrm{q}) q_{nm}^{s}(\mathrm{q},\omega) W_{nm}^{eff}(\mathrm{q},\omega) \tag{4.2}$$

where $V_{ij}^{bare}(\mathrm{q})$ is the unscreened matrix between subbands *i* and *j*, and $W_{ij}^{eff}(\mathrm{q},\omega)$ represents the effective screened matrix element, q_{nm}^{s} is the screening wavevector. The matrix elements are screened under RPA, in this work, only for the Coulomb scattering. This is due to the fact that surface-roughness is modeled in a method that is closer to a

deformation model of the potential rather than that of an electrostatic interaction based Coulombic model.

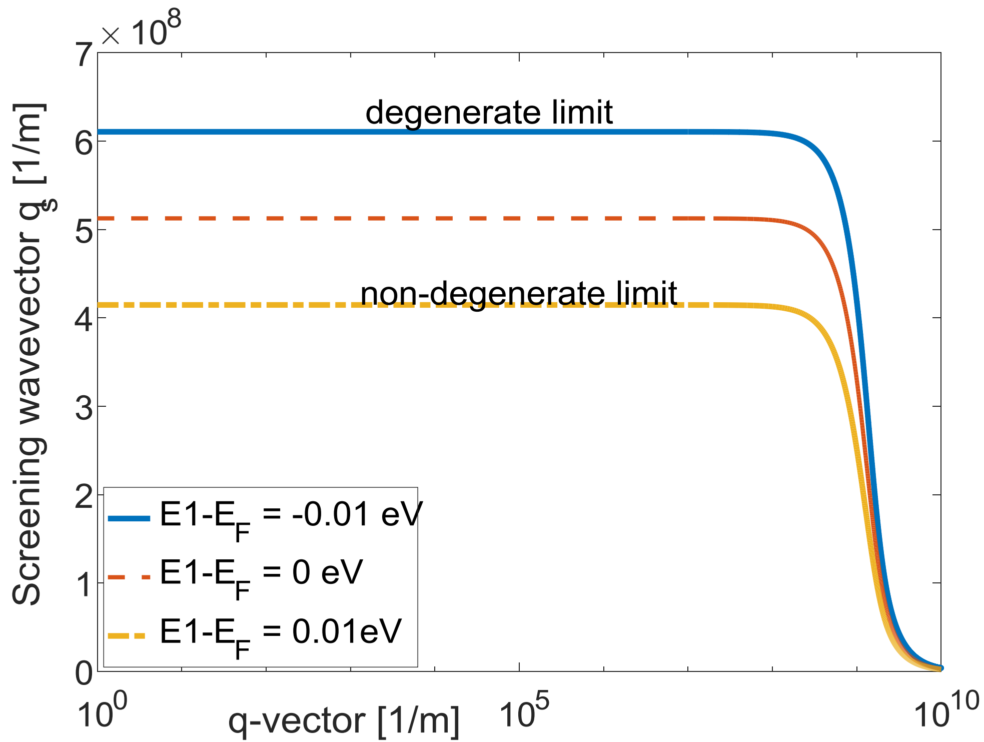

Figure 4.3: Screening wavevector q_s vs q vector. E1 is a sample subband energy level, E_F is the Fermi level.

In Figure 4.3, a plot of the screening wavevector for the degenerate and the non-degenerate conditions is shown. From the plot, it is obvious that the screening wavevector has a larger value, thus resulting in a stronger screening effect for the degenerate limit.

4.1.2. THE EXCHANGE AND CORRELATION EFFECTS IN HARTREE THEORY

The Density Functional Theorem (DFT), states that

$$E[n] = F[n] + \int d^3 r V_{ext}(\mathrm{r}) n(\mathrm{r}) \tag{4.3}$$

i.e. the ground state energy of the system is a function of its charge density $n(r)$, and that the total energy of the system has to be equal to the sum of the minimum ground state energy of the system and the rest of the energy coming off the external interaction with the external potential.

One can derive the exchange correlation potential using the above theoretical assumptions, and under the Local-Density Approximation (LDA), valid for slowly varying potentials one has that,

$$V_{xc}(\mathrm{r}) \approx \frac{d(n\varepsilon_{xc}(n))}{dn} = \varepsilon_{xc}(n) + n\frac{d\varepsilon_{xc}(n)}{dn} \tag{4.4}$$

Thus, the effective Hartree potential now becomes

$$V_{eff}(z) = V_H(z) + V_{xc}(z) \tag{4.5}$$

This potential now modifies the potential profile of the Poisson solver, which goes into the Time –Independent Schrödinger Wave equation. The solution of which then modifies the subband structure and the wavefunctions, which now alter the RPA screened matrix elements. Thus, exchange and correlation are accounted for in the screening implicitly through the Hamiltonian instead of an explicit involvement directly in the screening bubble. This is a valid approximation as long as the exchange-correlation energies are significantly less than the scattering potentials.

4.2. REVIEW ON CONDUCTIVITY

The two-particle Green's function is essentially correlation functions of two sets of field operators, hence two-particle functions. In other words, it represents the propagation of a pair of excitations (excitations here simply refer to holes and electrons: electron – addition of particles and holes - removal of the particles). Thus when these two-particle operators dealt with are the density operators (density-density correlation functions as a special case of the two-particle Green's function) , one calculates the density fluctuations that result in polarizing the medium, creating a cloud of shielding "charge" that screens the external scattering potential. This was explained in the previous section to account for the screening effects.

In this section, similarly, one tries to calculate the related current-current correlation function using the less-than Green's function to calculate the corresponding polarization function to estimate the conductivity (the various moments of the less than Green's function gives the particle density, current density, etc.). This less-than Green's function can later be used either to solve the quantum BTE directly (analogous to the semi-classical BTE but built on the lesser-than Green's function and the lesser-than self-energy) and obtain the conductivity from the solution of that equation, OR one could follow the integrated Green-Kubo approach to calculate these transport coefficients which is also based of ways to find the lesser-than Green's function. The former method is more of a general approach that includes far from equilibrium transport to model high electric field conditions. The equations of motion based on the lesser than Green's

function basically reduce to the equations given by the Green-Kubo approach under near equilibrium conditions.

4.2.1. LINEAR RESPONSE THEORY

In this work, as we deal with near equilibrium conditions on the device, we adopt the use of the results derived based upon the Green-Kubo relations. The Kubo approach is based on the assumption of Linear Response Theory (LRT). The LRT assumes that if the perturbation is small enough, the response of the system should be proportional to the perturbation. It also assumes that the dissipation-fluctuation theorem is valid at near equilibrium conditions. That is, there exists a general relationship between the random fluctuations of an equilibrium system and response of the same system-dissipation- to an external small perturbation, as they share the same origin of the force.

According to the LRT, for a given electric field one can deduce the corresponding current or the current-current correlation function (moments of the lesser-than Green's function). And this correlation function (which is the random fluctuation response) can be related to Spectral density function (which results from a dissipative response to the scattering) using the fluctuation dissipation theorem. This forms the overall basis of the Green-Kubo formalism.

The less-than Green's function (two particle Green's function) is first perturbatively expanded using the S-matrix operator in terms of the field operators. This follows a similar approach as done in the earlier sections. The Wick's theorem is used to express the two-particle Green's functions in terms of the product of one-particle functions,

followed by time ordering of the products of operators using a contour expansion. One then calculates the corresponding expectation value of the current-density operator (which will be in terms of these less-than Green's response functions), from which the corresponding a.c. conductivity (as the conductivity is now a function of frequency and q) is obtained. From this, imposing the limits $\omega \to 0$ (and the differential of the distribution function becomes a delta function at zero temperature at the Fermi surface k_F), we get the d.c. conductivity

$$\begin{aligned} \sigma_{\mathrm{2D}} &= \frac{e^2\hbar^3}{m^{*2}} \sum_n \sum_{\mathrm{k}} \int \frac{d\omega}{2\pi} \left[-\frac{\partial n_F}{\partial \omega} \right] k^2 g_n^r(\mathrm{k},\omega)\, g_n^a(\mathrm{k},\omega) \\ &= \frac{e^2\hbar^3}{m^{*2}} \sum_n \sum_{\mathrm{k}} \int \frac{d\omega}{2\pi} \left[-\frac{\partial n_F}{\partial \omega} \right] k^2 \left| g_n^r(\mathrm{k},\omega) \right|^2 \end{aligned} \tag{4.6}$$

Observing the above equation, it is easy to notice that it represents the two-particle Green's function as a product of the one-particle retarded and advanced Green's function in the ground state. Thus, it corresponds to the lowest order expansion of the two-particle Green's function indicating the "bare" interaction between the single Green's function and the impurity potential. The above equation is also called as the Drude conductivity (also referred to as the d.c form of the a.c conductivity σ_{2D} in its zero-frequency limit) corresponding to the lowest order expansion of the interaction.

The additional contribution comes from the higher order interaction of the impurity scattering event with the exact two-particle Green's function. This can be obtained by using Wick's theorem for the two-particle Green's function, and writing as the time-ordered product of two one-particle Green's function using the S-matrix operator. Thus,

this product of two one-particle Green's functions (which is the polarization function) is averaged over all the states instead of just the ground state. This represents the interaction between the impurity potentials and its connection to both the one-particle Green's functions.

The above interaction can be represented as follows. The one-particle Green's functions can be written as a perturbation series using the Dyson's equation. Thus, represented by diagram as,

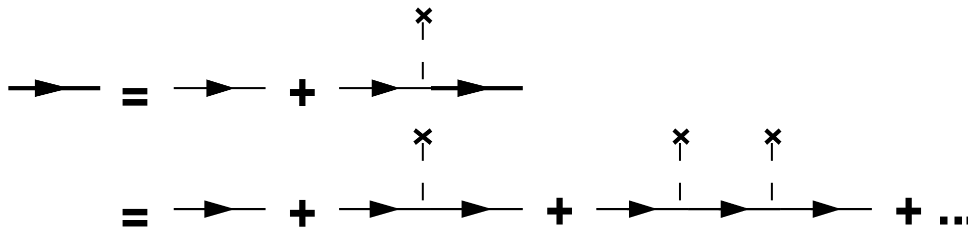

Figure 4.4 : Dyson's equation for retarded/advance Green's function

Now taking the product of two one-particle Green's function would result in a perturbation series that looks like a ladder as shown below in Figure 4.5.

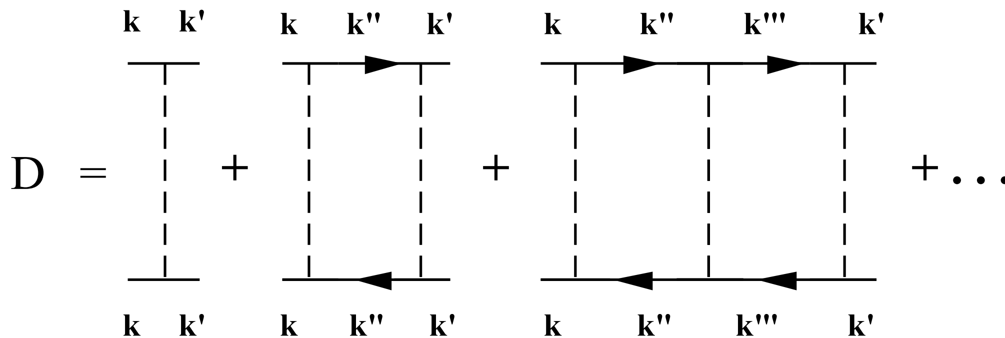

Figure 4.5 : Ladder diagram for the interaction between the Green's function(solid lines) and the (impurity) scattering event (dashed lines) that contribute to conductivity

Hence the two-particle Green's function's full perturbative expansion (that includes all of the interaction within the higher order terms) can now be represented by a *ladder diagram.*

This electron-hole ladder series represents the correction factor to the earlier stated Drude conductivity model, which now includes the higher order interactions. Also, an important assumption here is that the interaction is only due to the impurity scattering, and one takes the Dyson's equation for the impurity scattering for the one-particle Green's function. The dissipative effect of electron-phonon interaction is not considered as a contribution to the conductivity and thus the matrix element in the Dyson's equation, comes only from the impurity and surface-roughness scattering. The polarization function (which is the product $g^r_{om}(\mathrm{q},\omega_+)g^a_{om}(\mathrm{q},\omega_-)$ averaged over the different impurity positions) satisfies the Bethe-Selpeter equation,

$$\begin{aligned}\Pi_n(\mathrm{k},\omega_+,\omega_-) &= g^r_n(\mathrm{k},\omega_+)\,g^a_n(\mathrm{k},\omega_-)\times \\ &\times\left\{k^2+\sum_m\sum_{\mathrm{q}}\frac{\mathrm{k}\cdot\mathrm{q}}{q^2}T_{nm}(\mathrm{k}-\mathrm{q})\Pi_m(\mathrm{q},\omega_+,\omega_-)\right\}\end{aligned} \tag{4.7}$$

The conductivity is now given by,

$$\begin{aligned}\sigma_{2\mathrm{D}}(\Omega) &= \frac{e^2\hbar^3}{m^{*2}}\sum_n\int\frac{d\omega}{2\pi}\left[-\frac{n_F(\omega_+)-n_F(\omega_-)}{\Omega}\right]\sum_{\mathrm{k}}g^r_{on}(\mathrm{k},\omega_+)\,g^a_{on}(\mathrm{k},\omega_-)\times \\ &\times\left\{k^2+\sum_m\sum_{\mathrm{q}}\mathrm{k}\cdot\mathrm{q}\left|U_{nm}(\mathrm{k}-\mathrm{q})\right|^2 g^r_{om}(\mathrm{q},\omega_+)\,g^a_{om}(\mathrm{q},\omega_-)+\dots\right\}\quad .\end{aligned} \tag{4.8}$$

Taking the first order of eq.(4.7) for the polarization function, and neglecting the second part of Eq. (4.8) (as we consider only impurity and surface roughness, for which

the scattering potential becomes a delta function), we get the d.c, conductivity and polarization as,

$$\sigma_{2\mathrm{D}}=2\frac{e^2}{h}\sum_n\int d\omega\left(-\frac{\partial n_F}{\partial\omega}\right)\int_0^\infty\frac{d\varepsilon_{\mathrm{k}}}{2\pi}\Pi_n^{(1)}\left(\varepsilon_{\mathrm{k}},\omega\right)$$

$$\Pi_n^{(1)}\left(\varepsilon_{\mathrm{k}},\omega\right)=\frac{1}{\left(\hbar\omega-\varepsilon_{\mathrm{k}}-\varepsilon_n\right)^2+\Gamma_n^2(\varepsilon_{\mathrm{k}},\omega)}\times \tag{4.9}$$

$$\times\left\{\varepsilon_{\mathrm{k}}+\sum_m\sum_{\mathrm{q}}\frac{\mathrm{k}\cdot\mathrm{q}}{q^2}T_{nm}\left(\mathrm{k}-\mathrm{q}\right)\Pi_m^{(1)}\left(\varepsilon_{\mathrm{q}},\omega\right)\right\}$$

Thus the final conductivity expression becomes,

$$\sigma_{2\mathrm{D}}=2\frac{e^2}{h}\sum_n\int_0^\infty\frac{d\varepsilon_{\mathrm{k}}}{2\pi}\varepsilon_{\mathrm{k}}\Lambda_n\left(\varepsilon_{\mathrm{k}},\varepsilon_F\right)\frac{a_n\left(\varepsilon_{\mathrm{k}},\varepsilon_F\right)}{2\Gamma_n\left(\varepsilon_{\mathrm{k}},\varepsilon_F\right)} \tag{4.10}$$

Where,

$$\Lambda_n\left(\varepsilon_{\mathrm{k}},\omega\right)=1+\sum_m\sum_{\mathrm{q}}\frac{\mathrm{k}\cdot\mathrm{q}}{k^2}T_{nm}\left(\mathrm{k}-\mathrm{q}\right)\frac{a_m\left(\varepsilon_{\mathrm{q}},\omega\right)}{2\Gamma_m\left(\varepsilon_{\mathrm{q}},\omega\right)}\Lambda_m\left(\varepsilon_{\mathrm{q}},\omega\right) \tag{4.11}$$

The above conductivity expression is again for the case of the Fermi's distribution function reducing to a delta function at the zero-temperature limit.

The correction factor to this Drude model is given by,

$$\sigma_{2\mathrm{D}}^{corr}=-2\frac{e^2}{h}\sum_n\int d\omega\left(-\frac{\partial n_F}{\partial\omega}\right)\int_0^\infty\frac{d\varepsilon_{\mathrm{k}}}{2\pi}\varepsilon_{\mathrm{k}}\frac{\left(\hbar\omega-\varepsilon_{\mathrm{k}}-\varepsilon_n\right)^2-\Gamma_n^2(\varepsilon_{\mathrm{k}},\omega)}{\left[\left(\hbar\omega-\varepsilon_{\mathrm{k}}-\varepsilon_n\right)^2+\Gamma_n^2(\varepsilon_{\mathrm{k}},\omega)\right]^2} \tag{4.12}$$

Thus, overall conductivity is given by,

$$
\begin{aligned}
\sigma_{2D} &= \sigma_{\mathrm{2D}}^{Drude} + \sigma_{\mathrm{2D}}^{corr} \\
\sigma_{\mathrm{2D}} &= 2\frac{e^2}{h}\sum_n \int d\omega \left(-\frac{\partial n_F}{\partial \omega}\right) \int_0^{\infty} \frac{d\varepsilon_{\mathrm{k}}}{2\pi}\, \varepsilon_{\mathrm{k}} \frac{a_n(\varepsilon_{\mathrm{k}},\omega)}{2\Gamma_n(\varepsilon_{\mathrm{k}},\omega)} \Big\{ a_n(\varepsilon_{\mathrm{k}},\omega)\,\Gamma_n(\varepsilon_{\mathrm{k}},\omega) + \\
&\quad + \frac{m^*}{\hbar^2}\sum_m \int_0^{\infty} \frac{d\varepsilon_{\mathrm{q}}}{2\pi} \sqrt{\frac{\varepsilon_{\mathrm{q}}}{\varepsilon_{\mathrm{k}}}} \frac{a_m(\varepsilon_{\mathrm{q}},\omega)}{2\Gamma_m(\varepsilon_{\mathrm{q}},\omega)} \Lambda_m(\varepsilon_{\mathrm{q}},\omega) \int_0^{2\pi} \frac{d\varphi}{2\pi} \cos\varphi\, T_{nm}(\mathrm{k}-\mathrm{q}) \Big\}
\end{aligned} \tag{4.13}
$$

which is the expression for D.C conductivity used in this work.

CHAPTER 5. SCHRED INTEGRATION AND IMPLEMENTATION OF PARALLEL GF CORE

5.1. COUPLING OF THE nEGF SOLVER TO THE SCHRED V2.0 SOLVER

The Schrödinger-Poisson solver uses the density of states to compute the sheet density in various subbands, and therefore the total QM charge density in the device. This is where the nEGF solver comes into play, and gives an option of calculating the CBS in the Density of states function. Thus, for every iteration of the outer Schrödinger-Poisson loop, the nEGF solver solves for the retarded subband Green's function self-consistently, and calculates the corresponding self-energy, giving the value of the "real" DOS. This forms the inner loop of the simulator.

While coupling the two, care has to be taken to model the valleys properly. SCHRED considers a three conduction band valley pair model, and thus the nEGF solver has to be updated from a two-valley pair model to that of a three-CB valley pair model. The masses, wavefunctions are passed to the various scattering subroutines, overlap integrals, screening routines. Once the real DOS is solved, the control passes to the outer loop of Schrödinger-Poisson, and the process repeats until the Schrödinger-Poisson threshold value for error is reached for a self-consistent value of the potential.

5.2. MPI – PARALLELIZATION OF THE nEGF ENERGY INTEGRATION

$$\Gamma_n^{elastic}(\mathrm{k},\omega) = \frac{1}{2}\sum_m \iint \frac{d^2q}{(2\pi)^2} a_m(\mathrm{q},\omega) \sum_i \left| U_{nm}^i(\mathrm{k}-\mathrm{q}) \right|^2 \tag{5.1}$$

$$\Gamma_n^{e-ph}(\mathrm{k},\omega)=\sum_m\iint\frac{d^2q}{(2\pi)^2}\sum_i\left|U_{nm}^i\left(\mathrm{k}-\mathrm{q}\right)\right|^2\times$$
$$\left[\begin{array}{l}N_0\left(\dfrac{\Gamma_m^2(\mathrm{q},\omega+\omega_0)}{\left[\hbar\left(\omega+\omega_0\right)-\varepsilon_{\mathrm{q}}-\varepsilon_m-R_m(\mathrm{q},\omega+\omega_0)\right]^2+\Gamma_m^2(\mathrm{q},\omega+\omega_0)}\right)\\+\left(N_0+1\right)\left(\dfrac{\Gamma_m^2(\mathrm{q},\omega-\omega_0)}{\left[\hbar\left(\omega-\omega_0\right)-\varepsilon_{\mathrm{q}}-\varepsilon_m-R_m(\mathrm{q},\omega-\omega_0)\right]^2+\Gamma_m^2(\mathrm{q},\omega-\omega_0)}\right)\end{array}\right] \qquad (5.2)$$

$$\Gamma_n^{tot}(\mathrm{k},\omega)=\Gamma_n^{elastic}(\mathrm{k},\omega)+\Gamma_n^{e-ph}(\mathrm{k},\omega) \qquad (5.3)$$

where k is wavevector, ω is the E_F energy value, n represents the subbands. Also note that in our computations, the k,ω vectors are converted into their respective energy vectors (E_k, E_F) and the integration is done over energy range rather than momentum values.

Looking at the equations for the collisional broadening of the states we see that self-consistent solution is established by starting the iteration with an initial guess value for $\Gamma_n(\mathrm{k},\omega)$, according to the first-Born approximation. Γ_n values have to be computed for all the values of k (no. of points on the E_k energy axis), ω (no. of points on the E_F energy axis), n (no. of subbands), l (no. of valleys) by solving the coupled integral equations. For example, for k =1000, ω =300, n=2, l=3 the problem requires $1000\times300\times2\times3=1.8\times10^6$ times evaluation of the integral equation. This is very computationally expensive. (Note that the 1D Schrödinger-Poisson solver takes negligible amount of time)

But if one looks carefully into the expressions for the Γ-function evaluation, it is independent of ω integration. In other words, if $\hbar\omega_1, \hbar\omega_2, \ldots, \hbar\omega_r$ are *r* points on the E_F energy scale, then evaluation of $\Gamma_n(\mathrm{k}, \omega_2)$ does not depend upon $\Gamma_n(\mathrm{k}, \omega_1)$ *or* $\Gamma_n(\mathrm{k}, \omega_3)$ *or* $\Gamma_n(\mathrm{k}, \omega_4)$,......$\Gamma_n(\mathrm{k}, \omega_n)$. Hence, this part of the integration over E_F points can be parallelized.

Assuming we have 102 points on the E_F scale, and we could run in parallel these 102 E_F integrations on each of the 102 cores available on the local supercomputer, we could technically speed up the computation process significantly.

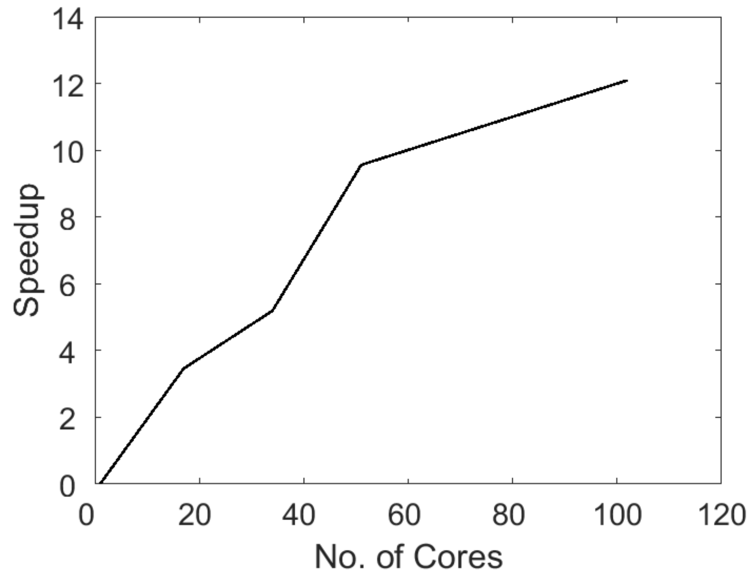

Figure 5.1 : Speedup obtained per iteration of the inner nEGF loop

In Figure 5.1, the speedup is computed based on a reference scale set with the time taken by a single core (serial) for the convergence of the code for a single iteration of the inner GF loop. This process was repeated 3 times in order to get an average value based

on the kernel and cluster variations of the nodes/cores assigned each time. The number of mesh points for each scale is as follows, k=500, ω =102, n=2, l=3. Looking at the figure we observe a maximum speed up of up to 12 times, the time taken by the serial execution of the code on a single core. This apparent speedup can be explained by further looking at the differences within the execution time for different parts of the code on the CPU.

In **Error! Reference source not found.** one notices that the parallelizable part of the code is about 75% of the entire design. Thus, a maximum efficiency of 12 times seems to be consistent with Amdahl's law. This can also be explained by the fact that saturation in speedup is reached beyond a certain point as the transfer of data between the different CPUs/RAM units – memory and IO bandwidth requirements - starts dragging down the speedup increase from increased number of processors.

No. of cores	**Total CPU time (sec)**	**GF Loop total (sec)**	**GF parallel module (sec)**	**Mobility module (sec)**	**GF parallel time/ Total CPU time (%)**
34	612	577	435	16	74.4
1	3041	2887	2282	106	75.05

Table 5.1 : Simulation time

In this work, the Message Passing Interface (MPI) has been used to make the above parallelization possible. MPI is short for message passing protocol – an open source software package developed by a wide variety of people (from industry and academia)

over the last few decades. The motivation is to make it possible to run a set of processes (threads/cores) in parallel on a cluster of nodes. MPI is compatible with most programming languages like C, C++, FORTRAN, Java etc. There is also another variation of the message passing package called as open MP. The difference between MPI and open MP is that, MPI runs the code parallel on different processor cores with a distributed memory - generally many big supercomputing clusters have separate memory for each processor, thus a distributed memory. But open MP on the other hand is mainly for running parallel on a cluster with a shared (cache) memory - an example is the cores on a single processor chip that share the same RAM through a cache memory. The supercomputing cluster at ASU (specifically for large no of cores like >128) uses distributed memory. Thus, the MPI is the preferred package.

So, in MPI, the same version of the code is run in parallel, at the same time on all the number of assigned cores. This is done by modifying the command given to compile and execute the source code, to include the MPI function and other details. That is, information such as the output file name, the number of cores required, the specific processor preferred to run on (different processors may have different number of cores such as 8,16,32 and also their speeds maybe different based on how old they are, etc) can be stated. Once this is stated in the startup script file, then we can edit the source code to actually implement the integration in parallel on the given number of processor cores.

A set of MPI commands can be used to do the following, where we can do several things like letting the core know that the MPI part of the execution has started, make the core hold the execution of the process on that line, and several other commands that help

share data between the cores. Giving the detail for each of this, however, seems unnecessary and only the major details of the parallelization will be explained below.

Let us assume we have "NGRID" points on the E_F scale, and available cores are given my "NUMPROCS", the number of grid points for each core is given by,

$$NDELP = \frac{NGRID}{NUMPROCS} \tag{5.4}$$

The above is important as, as it is not always possible to run say 256 E_F points on 256 processor threads/cores, due to the unavailability of the required cores on the cluster at the given time. Thus, grouping a set of E_F points together on a given core becomes the next best level in getting the speedup we want, which this approach attempts to do.

If "myid" is the ID number of that processor core, the starting point of E_F point of integration for that core is given by,

$$MYDEL_START = NDELP \times myid \tag{5.5}$$

Now inside the Γ loop for E_F integration, we start with the above "mydel_start" and go up to *NDELP*.

Thus the E_F integration is now split among the *NUMPROCS* number of cores. And, when this is evaluated, we issue another MPI command to hold the execution on that core until all other cores reach the same line of execution. After this is done, the calculated Γ values are sent back from the slave cores (say 1,2…numprocs-1) to the master core '0'. Thus the '0' core collects the Γ value for all the E_F points.

This Γ value is now broadcasted using an MPI command back to the rest of the "slave" cores, from where all the slave cores resume their regular execution again. Thus, each core now calculates the error value for itself and checks to see if it is less than the threshold value. If it is not, then the integration loop repeats again with the new $\Gamma_n(\mathrm{k},\omega)$ values. This process repeats until error value gets lesser than the tolerance value of 5×10^{-6} eV. This ensures maximal efficiency as the error is calculated simultaneously by all cores in tandem, and the only time spent idle by the slave cores, is during the collection and re-distribution of the complete set of Γ array values by the master core.

A point to note is that in order to solve for each E_F point on the energy scale we need the previous set of Γ values corresponding to different energies. This is done by the use of the uniform energy mesh on E_F axis. The $\hbar\omega_{q\lambda}$ energy value is divided by the ΔE_F value to find the index by which it is shifted, and thus use that corresponding value of Γ at that ω point. Thus, the solver requires the use of a very fine uniform energy mesh that increases computational complexity leading to the use of parallelization explained in this section

5.3. ELECTRON-PHONON SELF-ENERGY IMPLEMENTATION

Most of the elastic scattering mechanisms included in the nEGF solver require no special treatment, as they are completely accounted for by their scattering matrix elements (as a summation of the square of different scattering matrix elements) included directly into the self-energy matrix for the retarded subband self-energy. But in the case of electron-phonon scattering for the non-polar optical phonons (as is the case with

silicon), calculation of the retarded self-energy for electron-phonon interactions is more complicated as the scattering now becomes inelastic and the self-energy treatment above based on the scattering matrix elements (on a single-point external potential) is not applicable. One needs to consider the quantized phonon field or in other words, the phonon Green's function in conjunction with the electron Green's function for accounting the interaction, and thus the self-energy.

Under the self-consistent Born approximation (as reviewed in chapter 3), the self-energy function due to electron-phonon interaction is given by,

$$\Sigma^{<,>}(x_1,x_2) = iG^{<,>}(x_1,x_2)D^{<,>}(x_1,x_2) \tag{5.6}$$

where $D^{<,>}$ are the phonon Green's functions. And, the retarded/advanced self-energy functions for the electron-phonon interaction are given by,

$$\Sigma_{r/a}(x_1,x_2) = iG_{r/a}(x_1,x_2)D^{>}(x_1,x_2) + iG^{<}(x_1,x_2)D_{r/a}(x_1,x_2) \tag{5.7}$$

and this follows under the assumptions that, the phonon bath that the electron is interacting with, is assumed to be at thermal equilibrium; and secondly, we consider only one-phonon interacting with an electron (one-phonon process) while deriving the self-energy.

The above equation can be further simplified so that the phonon self-energy can be solved self-consistently under the Born approximation. Hence, assuming reasonably small electron densities, the $G^{<}$ term can be neglected, and thus the retarded/advanced self-energy functions simplifies to,

$$\Sigma_{r/a}(x_1,x_2) = iG_{r/a}(x_1,x_2)D^{>}(x_1,x_2) \tag{5.8}$$

The retarded subband self-energy for the phonons is now given by (under diagonal approximation for the retarded Greens function),

$$\Sigma_n^r(\mathrm{k},\omega) = \sum_m \sum_{\mathrm{q}} \sum_i \left| U_{\lambda,nm}^i(\mathrm{q}) \right|^2 \left[\left(N_{\mathrm{Q}\lambda} + 1 \right) g_m^r \left(\mathrm{k}-\mathrm{q}, \omega - \omega_{\mathrm{q}\lambda} \right) + N_{\mathrm{Q}\lambda} g_m^r \left(\mathrm{k}-\mathrm{q}, \omega + \omega_{\mathrm{q}\lambda} \right) \right]$$

(5.9)

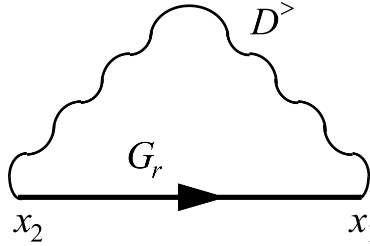

Figure 5.2 : Self-consistent Born approximation for electron-phonon interaction.

Looking at Eq. (5.9), the first term corresponds to phonon emission, and the second term corresponds to phonon absorption. This self-energy thus accounts for the sum of all contributions from the different electron-phonon interactions included through the scattering matrix element.

The final expressions for the broadening of the electronic states $\Gamma_n(\mathrm{k},\omega)$ and the renormalization factor $R_n(\mathrm{k},\omega)$ for the subband energies corresponding the self-energy in Eq.(5.9) now will become

$$\Gamma_n(\mathrm{k},\omega) = \sum_m \iint \frac{d^2q}{(2\pi)^2} \sum_i \left| U^i_{nm}(\mathrm{k}-\mathrm{q}) \right|^2 \times$$

$$\left[N_0 \left(\frac{\Gamma^2_m(\mathrm{q},\omega+\omega_0)}{\left[\hbar(\omega+\omega_0) - \varepsilon_\mathrm{q} - \varepsilon_m - R_m(\mathrm{q},\omega+\omega_0)\right]^2 + \Gamma^2_m(\mathrm{q},\omega+\omega_0)} \right) + (N_0+1) \left(\frac{\Gamma^2_m(\mathrm{q},\omega-\omega_0)}{\left[\hbar(\omega-\omega_0) - \varepsilon_\mathrm{q} - \varepsilon_m - R_m(\mathrm{q},\omega-\omega_0)\right]^2 + \Gamma^2_m(\mathrm{q},\omega-\omega_0)} \right) \right] \quad (5.10)$$

$$R_n(\mathrm{k},\omega) = \sum_m \iint \frac{d^2q}{(2\pi)^2} \sum_i \left| U^i_{nm}(\mathrm{k}-\mathrm{q}) \right|^2 \times$$

$$\left[N_0 \left(\frac{\hbar(\omega+\omega_0) - \varepsilon_\mathrm{q} - \varepsilon_m - R_m(\mathrm{q},\omega+\omega_0)}{\left[\hbar(\omega+\omega_0) - \varepsilon_\mathrm{q} - \varepsilon_m - R_m(\mathrm{q},\omega+\omega_0)\right]^2 + \Gamma^2_m(\mathrm{q},\omega+\omega_0)} \right) + (N_0+1) \left(\frac{\hbar(\omega-\omega_0) - \varepsilon_\mathrm{q} - \varepsilon_m - R_m(\mathrm{q},\omega-\omega_0)}{\left[\hbar(\omega-\omega_0) - \varepsilon_\mathrm{q} - \varepsilon_m - R_m(\mathrm{q},\omega-\omega_0)\right]^2 + \Gamma^2_m(\mathrm{q},\omega-\omega_0)} \right) \right] \quad (5.11)$$

Considering acoustic phonon scattering, being a low energy elastic scattering mechanism, these equations would reduce to the Eqs (3.5),(3.51), similar to the inclusion of sum of square of the matrix elements inside the self-energy term for the Coulomb, surface-roughness and interface-trap scattering mechanisms under *Self-consistent Born approximation*. Thus, the contribution from the acoustic phonon scattering is added to the square of the matrix elements and fed into the same expression as in (3.5), along with the other elastic scattering mechanisms.

But the elastic approximation is not valid for the higher energy optical phonons, thus requires a self-consistent treatment of the phonon self-energy. In order to do this, an uniform energy mesh was taken (with an energy mesh spacing less than ω_0) and the initial value for the proper self-energy for the phonons is obtained from the first Born

approximation of (5.10), (5.11), which under first-order Born approximation becomes,

$$\Gamma_n(\mathrm{k},\omega)=\sum_m \iint \frac{d^2q}{(2\pi)^2}\sum_i \left|U^i_{nm}\left(\mathrm{k}-\mathrm{q}\right)\right|^2 \times \\ \begin{bmatrix} N_0\left[\theta\left(\hbar\left(\omega+\omega_0\right)-\varepsilon_{\mathrm{q}}-\varepsilon_m-R_m(\mathrm{q},\omega+\omega_0)\right)\right] \\ +\left(N_0+1\right)\left[\theta\left(\hbar\left(\omega-\omega_0\right)-\varepsilon_{\mathrm{q}}-\varepsilon_m-R_m(\mathrm{q},\omega+\omega_0)\right)\right] \end{bmatrix} \tag{5.12}$$

And $R_n(\mathrm{k},\omega)$ will be 0 as $\Gamma_n(\mathrm{k},\omega)\to 0$ for first-order Born approximation.

This value of the self-energy will be now used as the initial guess value for the proper self-energy iteration loop in Eq. (5.10) . Thus, the total contribution to the self-energy's imaginary term (broadening of the states Γ_n) is now the sum of the contribution from the above electron-phonon interaction term for Γ_n and the contribution from the rest of the elastic scattering mechanisms to Γ_n included earlier. This now forms the total contribution to the Γ_n function (corresponding to the dressed/full interacting Green's function). This will be used again during the next iteration as the initial guess value for the Γ_n function (instead of the first-order Born approximation value used in the first loop iteration), and iterated until the tolerance value is reached. This accounts for the self-consistent solution of the corresponding full Green's function accounting for the electron-phonon interaction explicitly.

5.4. FLOWCHART OF THE OVERALL PROGRAM

The overall program is detailed in the flowchart below. As shown in Figure 5.3, the Schrödinger-Poisson's solver forms the outer loop of the master iteration, and inside

which runs the nEGF solver – that solves self-consistently, through another loop the value of the real DOS by solving the Dyson's equation for the retarded subband self-energy. Thus, the broadening of the states and the energy renormalization of the spectrum is evaluated.

The initial, first iteration solution of the Poisson equation based on the doping gives a guess potential value based on which the Schrödinger eigenvalue (EISPACK) solver gives the values of the subband wavefunctions. Using the new potential, wavefunction values for the different subbands, and the Ef energy mesh, the inner nEGF loop is iterated until it gives a self-consistent value for broadening of states – Gamma function. Based on this Gamma function, it calculates the value of the real DOS, and thus the sheet charge density.

The Poisson's equation now gives the value of the new potential based on the calculated sheet density (based on the above nEGF solved real DOS). The Schrödinger equation uses the new potential and gives the new values of subband energies and their wavefunctions. This now becomes the input into the nEGF solver. And the process repeats until a self-consistent solution of the Poisson potential is established for the real DOS. In other words, the process is repeated until error value of the potential reaches a certain threshold tolerance for self-consistency.

Also before the start of the solver there is possibility to choose the semi-classical or quantum simulation mode based on either a hardcoded flag value or through the voltage range (by calculating the first iteration value of the Poisson surface position). In the semi-classical mode, the nEGF solver is completely skipped and only the Poisson solver runs

and the semi-classical charge is calculated based on the ideal DOS. In this case a flag can be set to denote semi-classical approximation for the SDF (as the delta function) under first Born approximation, so that one can calculate the ideal DOS based mobility value. Or it could be reset so that one can calculate the mobility for ideal DOS but with self-consistent born approximation instead of just the first Born approximation.

Once the value of the sheet density and potential is calculated, the conductivity is calculated through the Green-Kubo approach. If the ideal DOS flag is set, then an approximation for the value of the SDF is set as a Dirac-delta function while evaluating the polarization equation for the conductivity and the first Born result of the Gamma function is used. If not, the solver runs through the self-consistent calculation for the Gamma function (for the last one iteration of this final, self-consistent potential). And then, the conductivity is calculated as before using the self-consistent solution of the polarizability function (lambda function). Following this, the mobility is evaluated based on this conductivity over the E_F energy range established earlier.

Figure 5.4 explains in detail the nEGF solver's flow. The MPI calls are made at the start of the routine to get the core IDs, calculation of the respective E_F grid points, and establish the array (sub-array slices of E_F grid points) sizes for the Gamma function based on the grid point distribution among the cores.

After this each core runs through the following sequence in parallel. It first calculates the bare overlap factors (matrix elements) for the Coulomb and surface-roughness scattering mechanisms based on the subband wavefunctions, and the Schrödinger-Poisson's potential. Then the acoustic phonons scattering matrix elements are evaluated.

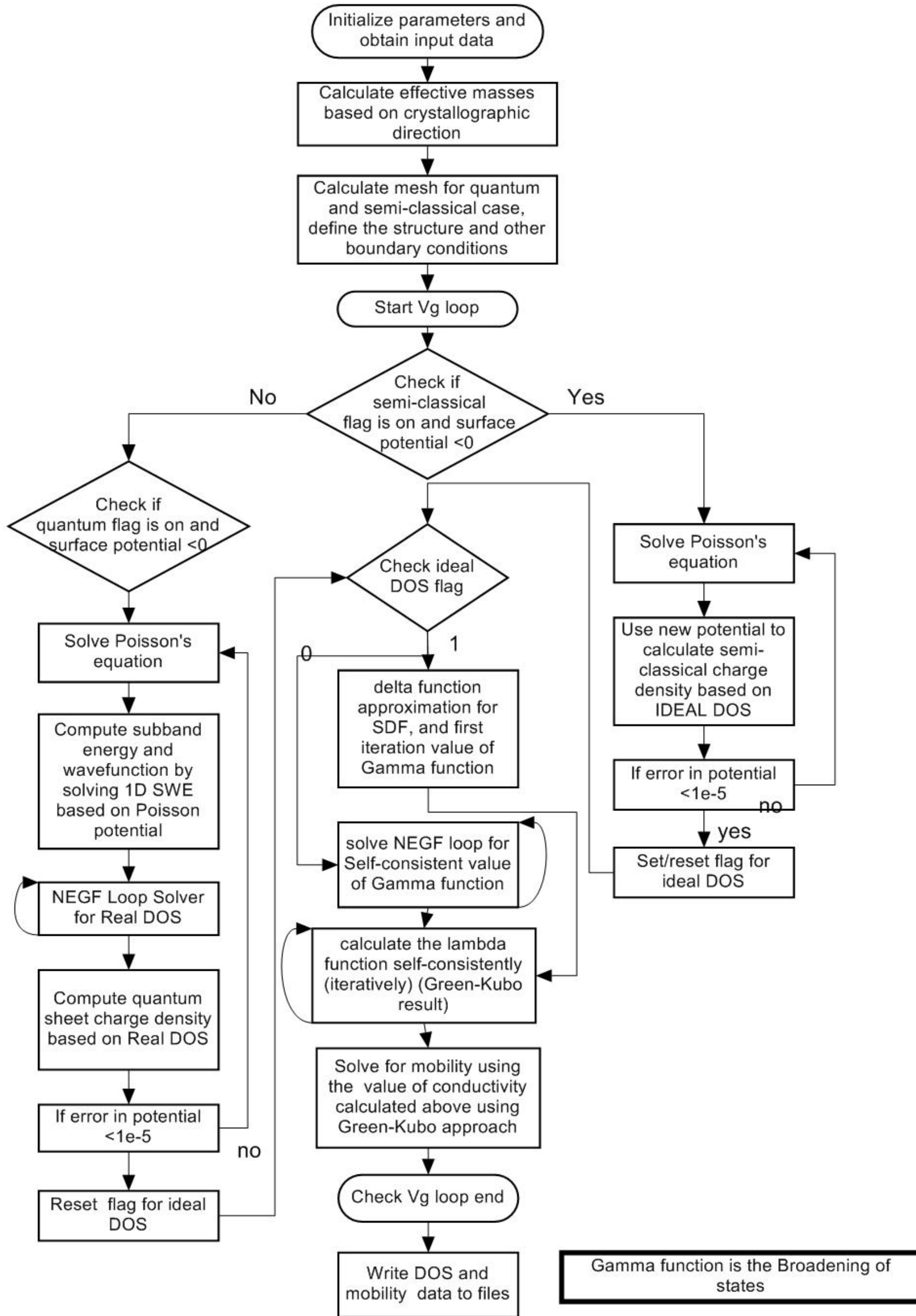

Figure 5.3 : Overall Program

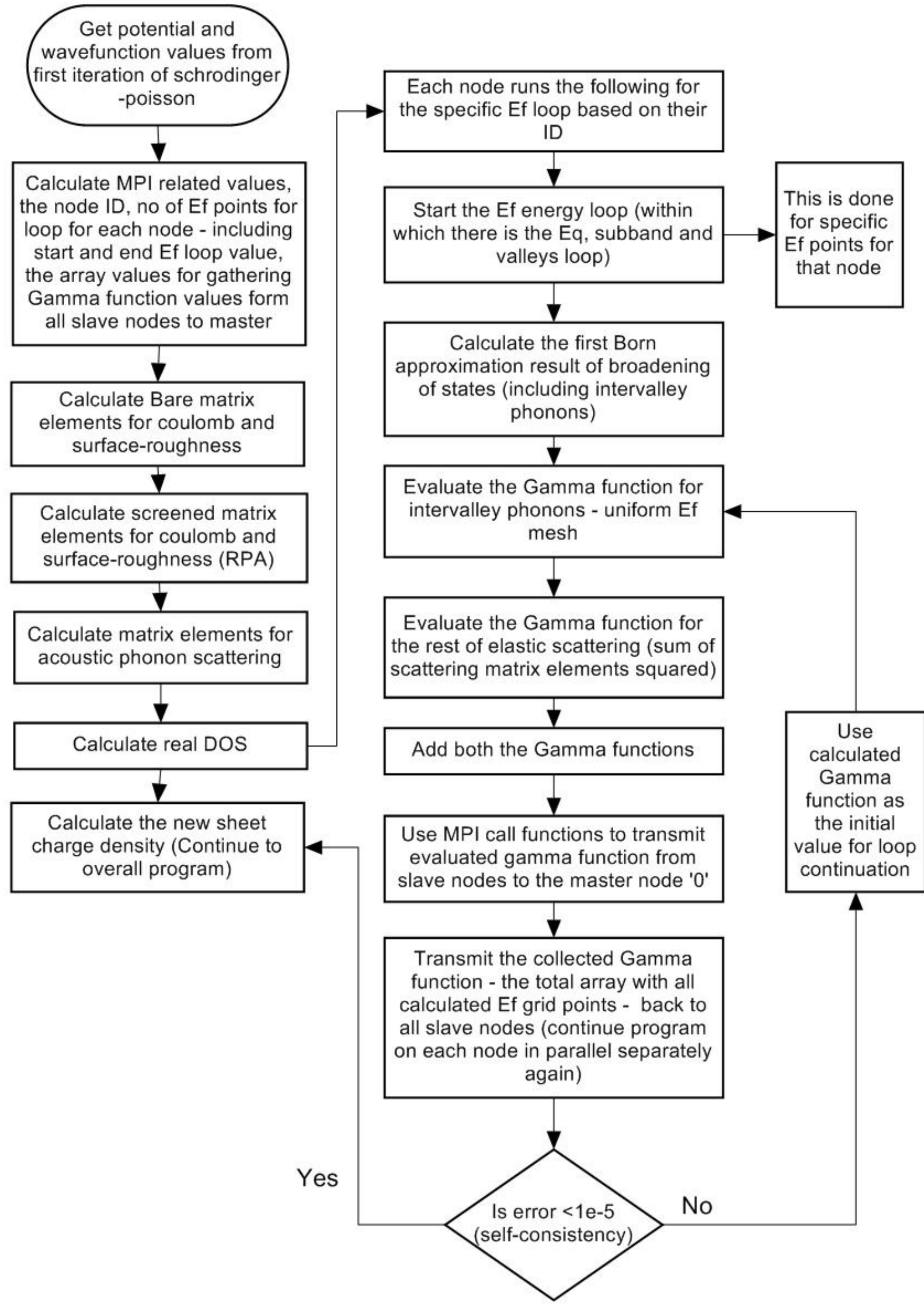

Figure 5.4 : nEGF loop solver - with the MPI core for calculating the Ef integration

The first-Born approximation for the Gamma function is then estimated to provide for the initial guess for the self-consistent Gamma loop that is to follow.

The energy E_F mesh now runs for the given core's grid range, and the corresponding slice of the Gamma function (slice meaning the respective slice of the total Gamma function matrix, corresponding to the E_F range of the Gamma function on that core) is calculated on each core in tandem.

This gamma function has a two contributions, first comes from the intervalley phonons (inelastic, so on a uniform energy mesh the phonon energy difference can be accounted for using a number of E_F grid points). Next part is the elastic contribution through the rest of scattering mechanisms – Coulomb, surface-roughness, acoustic phonons – which go directly into the matrix elements, which are then squared and summed as a net contribution.

At the end of this, an MPI call ensures all cores reach this point, at which the gamma function slices are collected from "slave" cores to the "master" core '0'. The master core now combines the slices of the matrix (E_F sub-array matrices) into the total Gamma function matrix (which now is a function of the 4 variables, the entire E_F range, E_q range, subband and valleys). This complete matrix is then redistributed back to the slave cores, which then calculates the error, and if it is greater than the tolerance value, continues to repeat the loop using the new value of the evaluated gamma function. Thus, one solves for the real DOS, which is then used to calculate the new sheet charge density and therefore the total quantum charge density.

This inner nEGF loop repeats for each iteration of the outer Schrödinger -Poisson's loop, giving the final self-consistent value of the potential based on the broadening of the states and the corresponding quantum charge density. The mobility then is calculated, after which the process is repeated for the next voltage value.

The author is thankful to the following computing facilities for the support they have extended, without which this computationally intensive work would never have been possible.

- This work extensively used the Saguaro High Performance Computing Cluster computer at Arizona State University[46]
- This work extensively used the Ocotillo High Performance Computing Cluster computer at Arizona State University[47]
- This work also used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562.[48]

CHAPTER 6. SIMULATION RESULTS

The simulation results presented in this section were run including the self-consistency of the phonons and parallelizing the code while solving for the broadening of the states.

6.1. SCATTERING RATES IN THE FIRST BORN APPROXIMATION

Figure 6.1 compares the scattering rates of the different scattering mechanisms within the first Born approximation, included in the model individually. This helps us determine qualitatively how the different scattering mechanisms affect the mobility under different transverse electric field conditions. The simulations are carried out at 7.7×10^{17} m^{-3} doping, at a bias voltage of V_G = 5V, at T= 300K, that leads to 1.2×10^{6} V/cm transverse electric field. The roughness correlation length used in these simulations is 15 Å, the RMS height of the bumps is 3 Å. Three sets of simulations were run, each simulation with one of the three scattering mechanism included – surface roughness (SR), Coulomb, acoustic phonon (AP).

Each curve corresponds to individual subband/valleys of that set of simulation. Namely, “SR-11”, indicates surface-roughness scattering only for subband ‘1’ and valley ‘1’. The blue lines indicate the different subband energies for the subband/valleys. The increase (step) in the scattering rate for each curve corresponds to the occurrence of the corresponding subband energy. Looking at the [11] curves for each scattering mechanism, SR is the dominant one, followed by acoustic phonon and the Coulomb scattering. This can be explained by our assumption of the simulation, which is run at

1.2×10^6 V/cm electric field. This field corresponds to high transverse field regime where the surface roughness dominates.

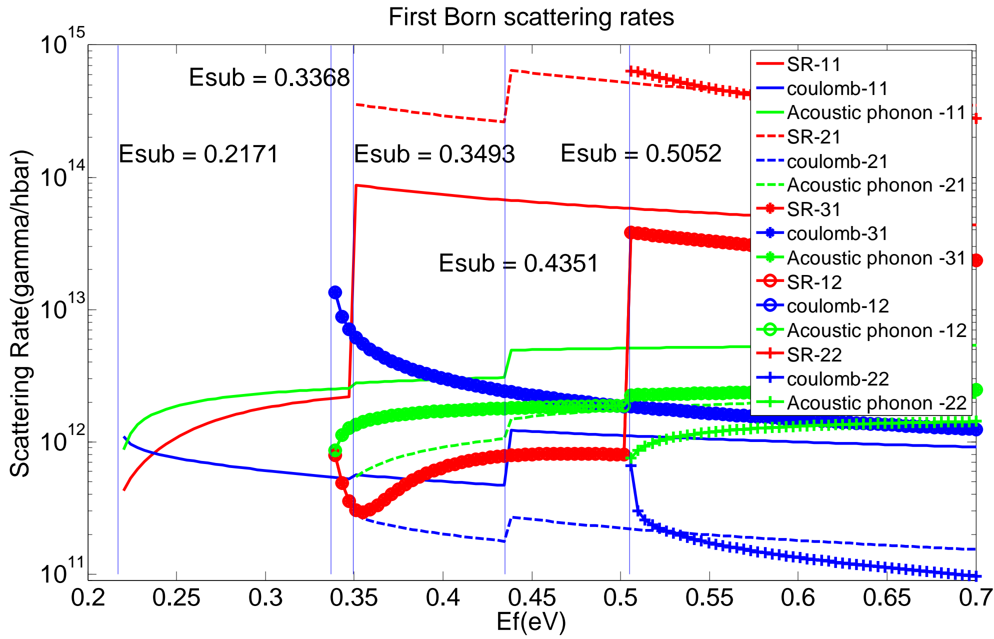

Figure 6.1 : First Born approximation scattering rates

For the subband 2 in valley 1, we notice that the SR curve moves up, while the curves for the AP and Coulomb scattering move down. This increase of SR scattering can also be explained by the high transverse field.

6.1.1. BROADENING OF THE STATES ACROSS MOSFET GENERATIONS

The DOS in a Q2D system is traditionally treated to be independent of the scattering and considered as a step function. This research accounts for the scattering induced broadening in the density of states. The figure below (Figure 6.2) shows the broadening of

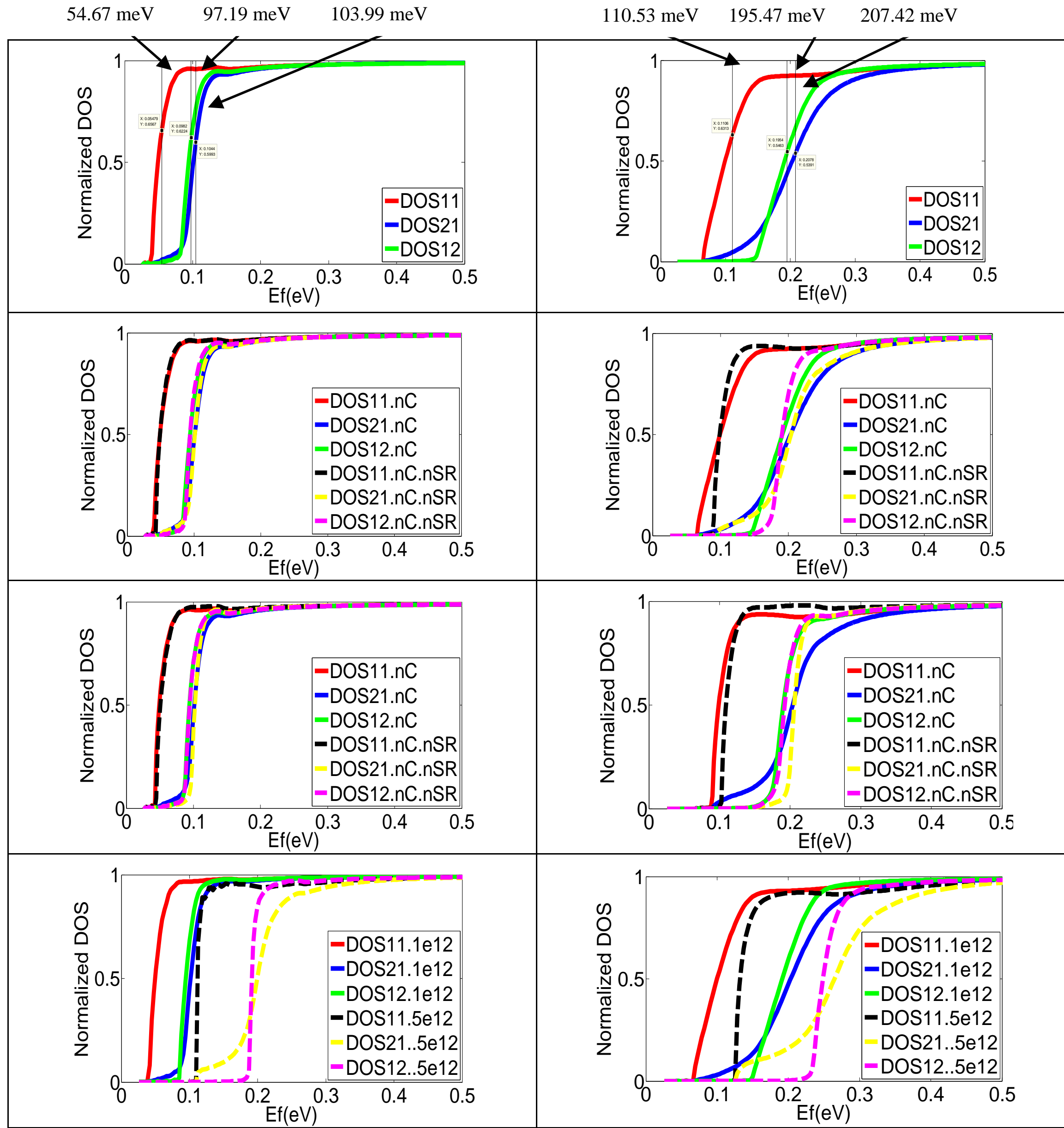

Figure 6.2 : Left panel $N_A = 10^{17}cm^{-3}$, Right panel $N_A = 10^{18}cm^{-}3$. First row panels: Effective(real) DOS of the lowest 3 subbands. The energy values indicate the corresponding subband energies in the order of subbands – 11, 12, 21. Second Row panels: Simulation without Coulomb scattering. "DOS.nC" denotes the real DOS function without Coulomb scattering in comparison to the real DOS with Coulomb scattering denoted by "DOS". Third Row panels: Simulation without Coulomb scattering and without surface roughness scattering. "DOS.nC.nSR" denotes the real DOS function without Coulomb scattering and surface roughness scattering in comparison to the real DOS without Coulomb scattering only denoted by "DOS.nC". Fourth Row panel: Simulation run for 2 sheet charge densities, "DOS.1e12" - $1\times10^{12}cm^{-2}$, "DOS.5e12" - $5\times10^{12}\,cm^{-2}$.

the states across different doping generations, as well as across the different subbands and valleys. The left panels in Figure 6.2 describe the DOS results corresponding to substrate doping density of 10^{17} cm^{-3} and the right panels to substrate doping density of 10^{18} cm^{-3}. The sheet electron density is 10^{12} cm $^{-3}$. The Renormalization of the spectrum is not taken into account in these simulations. The first row corresponds to the case when all scattering mechanisms are incorporated in the model. It describes the effective (real) DOS of the lowest three subbands (2 from unprimed ladder of subbands (11 and 21) and 1 from primed ladder of subbands (12)). The subband energies for the respective subbands are also indicated. The lines denote the ideal DOS step-function. We can observe that the broadening of the states induced by the scattering leads to spreading of the DOS function around this energy value for a given subband.

The second row results correspond to a case when Coulomb scattering is omitted and the bottom panel results correspond to the case when both Coulomb and surface roughness scattering are not included in the theoretical model. One can immediately conclude by the comparison of the results for the DOS function that for the case of substrate doping density of 10^{17} cm^{-3}, Coulomb and surface-roughness scattering are not that significant when compared to phonon scattering as the DOS function shape does not change significantly. This can be reasoned by in the following manner. Although both Coulomb and interface roughness scattering potentials are higher, the subband separation increases and the intersubband scattering reduces due to the smaller overlap of the wavefunctions; thus leading to smaller overlap factors.

The situation is a little different for a substrate doping density of 10^{18} cm^{-3}. Here, although the subband separation is higher, the strength of both Coulomb and surface-roughness potentials increases considerably, thus dominating phonon scattering even at sheet electron density of 10^{12} cm^{-2}. Such doping densities roughly correspond to substrate doping densities of the 32 nm and 22 nm technology nodes.

In the bottom row, we show a plot of the DOS function for two sheet charge densities: 1×10^{12} cm^{-2} and 5×10^{12} cm^{-2}. It can be observed that as the doping density increases, the subband energies and their separation increases which leads to more energy shift between the DOS function corresponding to the different subbands. Also noted is that the second subband of the first valley has the highest broadening. This can be explained by looking at the second and third panel for the 1×10^{18} cm^{-3} doping densities. The DOS function corresponding to this second subband rises much faster to unity (approaching the ideal DOS) when the surface-roughness scattering and Coulomb scattering is absent (third panel), compared to the case when only Coulomb scattering is absent (second panel). This can be attributed to the increased surface-roughness scattering that now dominates the inter-subband scattering, as previously stated.

The Coulomb scattering does not seem as significant in this case. This is also in accordance with our expectation that Coulomb scattering should dominate at lower sheet charge densities. This is further asserted from the second panel data, which shows the increase of the slope in the DOS functions corresponding to the first subband from valley 1 and valley 2 when Coulomb scattering is absent.

6.2. RESULTS OF THE SCHRED-nEGF CODE

In this section we present the results obtained with the integrated code that couples in a self-consistent manner the 1-D Poisson-Schrödinger solver with the nEGF core discussed in the previous section. This was accomplished to make the solver more robust - because of the self-consistent solution of the coupled Poisson-Schrödinger equation, where the Poisson equation is solved on a generic non-uniform mesh using direct LU decomposition method and the Schrödinger equation is solved using the Eigenvalue solver from EISPACK. Also, by doing this we can expand the feature capability of the nEGF solver, as SCHRED V2.0 can model strain in silicon, high-K dielectric, and give us more data like the inversion layer capacitance, subband sheet charge density etc. SCHRED V2.0 is easily adoptable to implement newer devices due to the generic non-uniform mesh adopted while solving for the Poisson and the Schrödinger equation.

When obtaining the simulation results shown in Figure 6.3, we assume transverse electric field of 1.0×10^{6} V/cm. The system is a three conduction band valley pairs system with two subbands per valley. Figure 6.3 shows the DOS simulation for different doping concentrations, at 2.0×10^{16} m^{-3} , 7.7×10^{17} m^{-3} , 2.4×10^{18} m^{-3} doping.

The curve "2e16-subband-11" corresponds to a doping level of 2.0×10^{16} m^{-3}, and '-11' indicates the first subband of the first valley ([nn,kk] – where 'nn' is the subband and 'kk' is the valley). The 11,21,12,22[subband/valley] curves denote the primed subbands for the Δ_4 bands and the 13,23 curves correspond to the unprimed subband ladder of the Δ_2 bands.

6.2.1. The real DOS – Collisional Broadening of the States

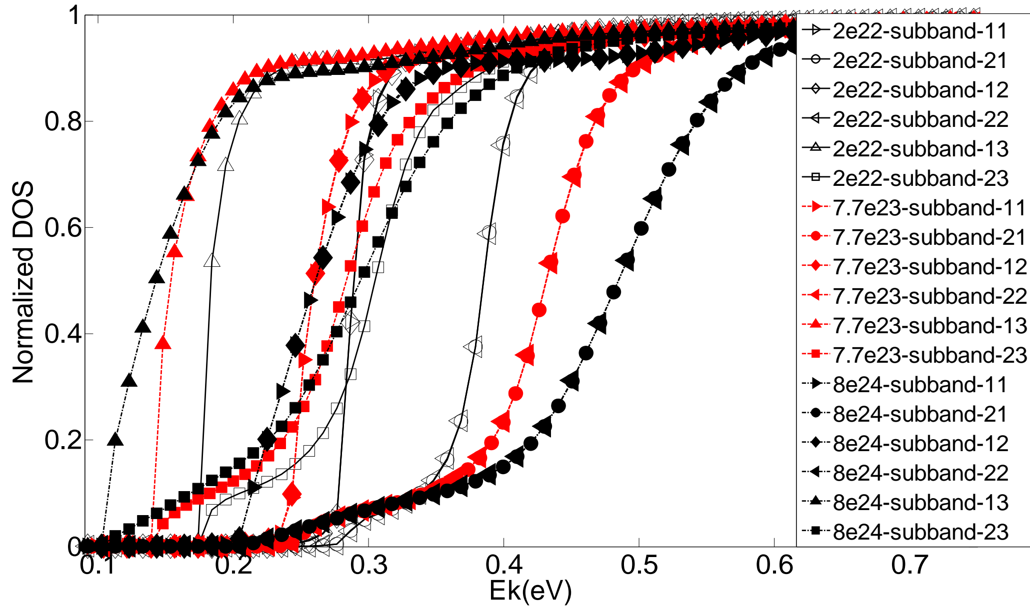

Figure 6.3 : Real DOS vs energy for different doping concentrations

Observing the curves for the three different doping levels, for subband =1, valley = 3, one can easily deduce that the scattering induced collisional broadening of the states is most significantly affecting the 2.4×10^{18} m^{-3} doping concentration. A similar argument can be made for the remaining curves as well. It is also seen that scattering induces a much broader variation specifically in the 21,22 curves more so than others. A probable reason is the one stated in the previous section - The scattering potential of surface roughness is now strong enough (higher transverse field) to overcome the weak overlap of the wavefunctions due to the increased subband separation.

From this and the previous section analysis of the results we might conclude that collisional broadening of the states is more significant in smaller device structures with

very high doping densities due to the Coulomb scattering being increased by two orders of magnitude and interface roughness dominating the intra-subband scattering in this structure at very high doping densities.

6.2.2. DOS for Self-consistent phonon calculation

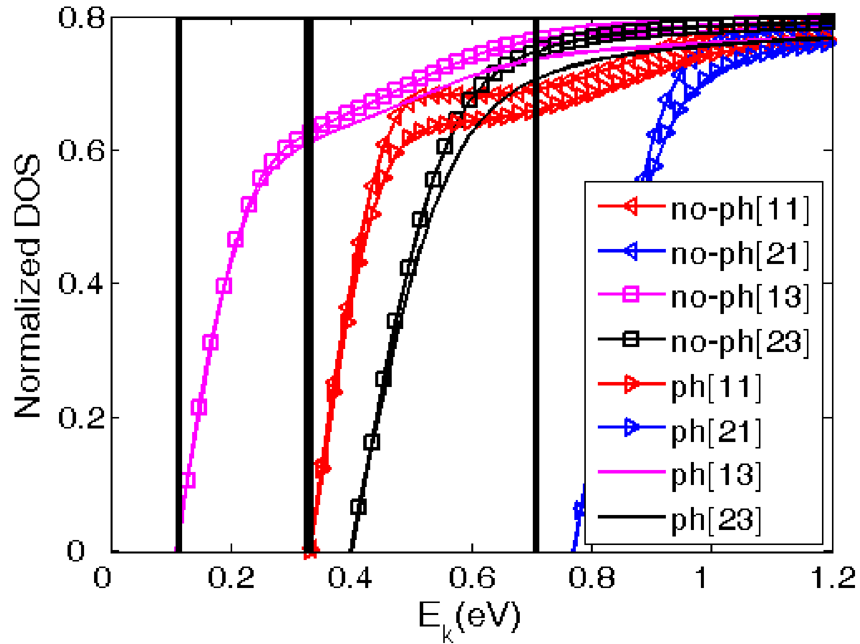

Figure 6.4 : DOS comparison between ideal DOS (black solid lines), DOS without phonon self-consistency (no-ph), and DOS with self-consistent inclusion of phonon scattering (ph).

In Figure 6.4, the DOS - with and without the self-consistent inclusion of intervalley phonon scattering is studied. Also shown is the 2-D ideal DOS step function denoted by the solid black lines. The ideal DOS is a step function and is represented by the solid vertical black lines, thus indicating no broadening of the DOS. Figure 6.4 is plotted for a substrate doping concentration of 2.4×10^{18} cm^{-}3, at an electric field of 1MV/cm. It can be

observed that 'ph' curves for the different subbands have significantly higher broadening of the states, with lowered DOS value as opposed to the 'no-ph' curves. This concurs with our expectation of increased state broadening from the proper inclusion of intervalley phonon scattering within the self-consistent Born approximation. Also, in both of these cases the broadening of the states is quite large as opposed to the case of ideal DOS (black solid line). This can be explained by the higher doping value used in the current simulation dataset.

6.2.3. MOBILITY PLOT

In Figure 6.5, we see that a really good match of the mobility is achieved across all the three doping generations. 'exp 2.0×10^{16} cm^{-3}', 'exp 7.7×10^{17} cm^{-3}', 'exp 2.4×10^{18} cm^{-3}', are the experimental results for 2.0×10^{16} cm^{-3} , 7.7×10^{17} cm^{-3}, 2.4×10^{18} cm^{-3} doping concentrations respectively [49]. Similarly 'sim 2.0×10^{16} cm^{-3}', 'sim $7.7\text{x}10^{17}cm^{-3}$', 'sim $2.4\text{x}10^{18}cm^{-3}$', are our simulation results of this present work , for 2.0×10^{16} cm^{-3} , 7.7×10^{17} cm^{-3}, 2.4×10^{18} cm^{-3} doping concentrations respectively. All the scattering mechanisms are included in the simulation, namely surface roughness (unscreened), Coulomb (screened), Interface trap (screened), acoustic phonon and the intervalley optical phonons scattering. The same set of data for this simulation set is analyzed in the following sections.

A very close agreement between experiment and theory is achieved over the three doping generations for the composite mobility curves. We can see that the universal

mobility curve is especially well justified as the SR dominant high field regime brings the different doping curves to a single one.

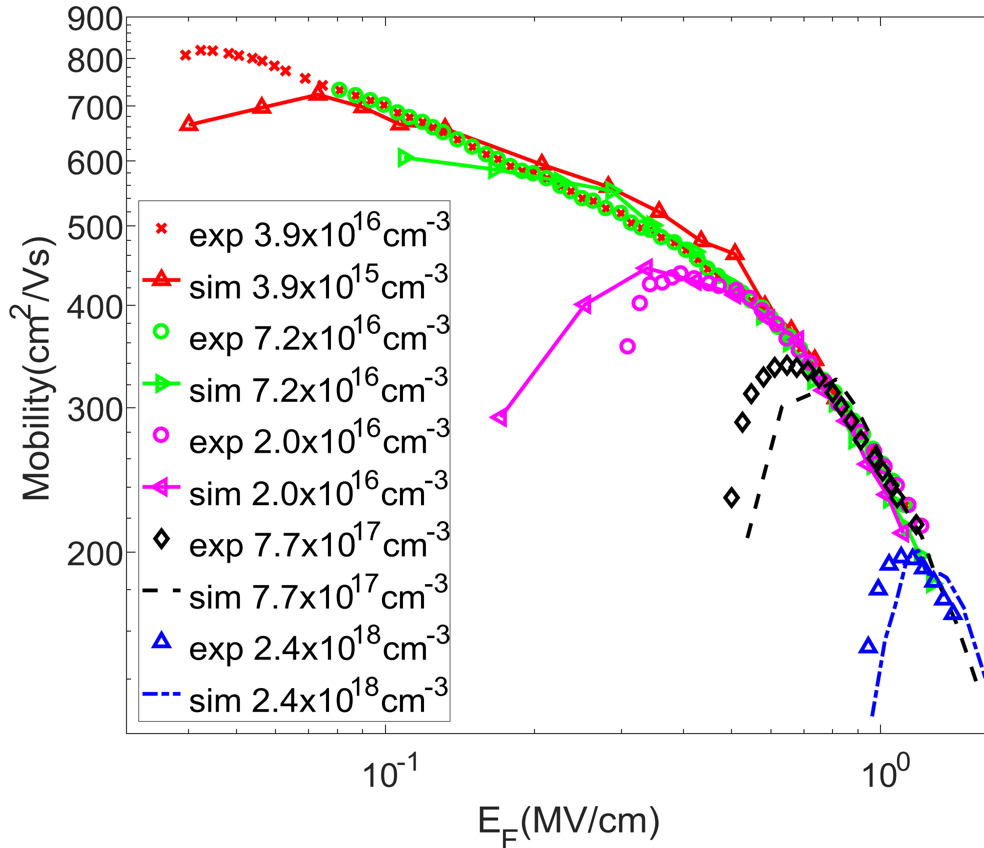

Figure 6.5 : Mobility Vs electric field in silicon for different doping generations. [exp – Experimental [49] ; sim – Simulation (Real DOS)]

A very good fit is observed for both doping concentrations of 7.7×10^{17} cm^{-3}, 2.4×10^{18} cm^{-3} in this region. A very good match is also observed over the relatively lower field values (for instance 5.0×10^{5} V/cm for 2.0×10^{16} cm^{-3}) for all the doping densities as well. In this low field region, as the inversion charges decrease, Coulomb scattering begins to increase and thus brings down the mobility.

A similar match in mobility is observed for the lower doping densities as shown in Figure 6.5, for the cases of 3.5×10^{15} cm^{-3}, 7.2×10^{16} cm^{-3} as well.

In Figure 6.6, one can see the comparison between mobility for real and ideal DOS. The real DOS based mobility is calculated with the self-consistent inclusion of all the relevant scattering mechanisms including the intervalley phonon scattering and thus computing the collisional broadening of the states (CBS).

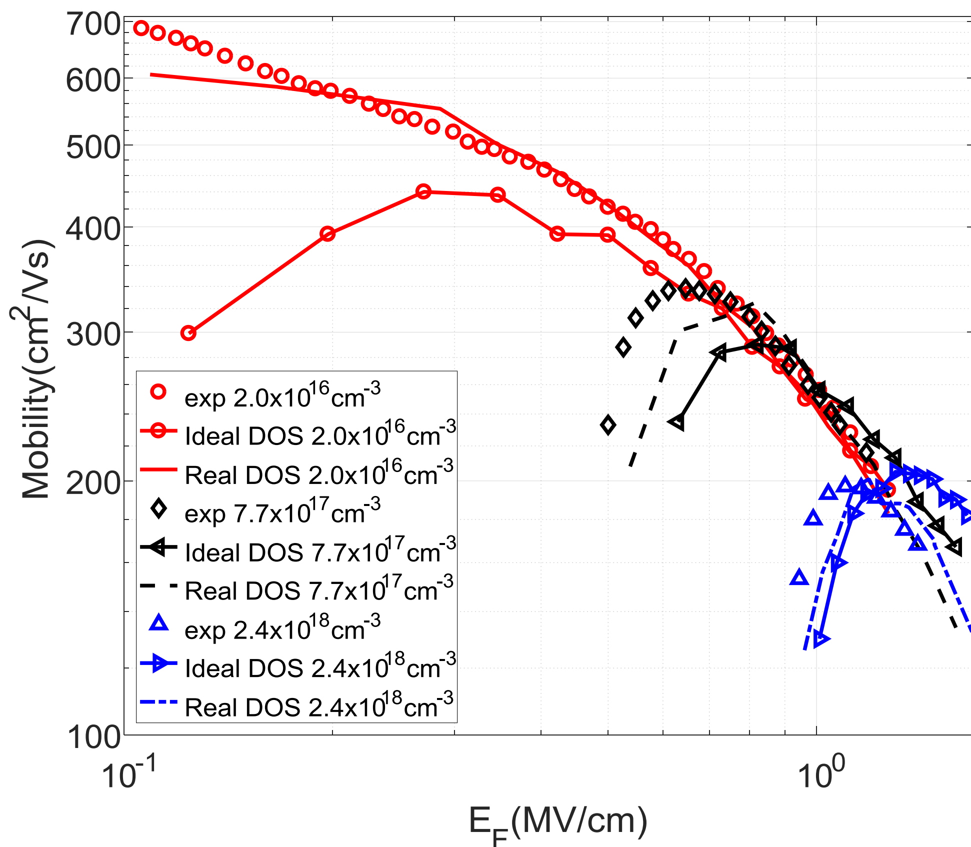

Figure 6.6 : Mobility Vs electric field comparison between the real DOS with collisional broadening of states (CBS) and the ideal DOS without CBS

The ideal DOS based mobility on the other hand, is computed based only on the ideal DOS, within the first-Born approximation (without self-consistently solving for CBS). As one would expect, the inclusion of scattering decreases the mobility of the carriers. This can be observed in the SR dominant high-field regime for all the doping concentrations.

CHAPTER 7. CONCLUSION AND FUTURE WORK

This thesis has successfully created a fully quantum mechanical nano-device simulator that can model the low-field inversion layer mobility in silicon MOS capacitors. It solves for the scattering induced collisional broadening of the states by accounting for the various scattering mechanisms present in silicon through the nEGF (near-equilibrium Green's function) approach. The novel feature in this work is the inclusion of the inelastic intervalley phonon scattering self-consistently while solving for the nEGF. In addition, this work also includes the acoustic phonon scattering, interface trap, Coulomb and surface roughness scattering to solve for the DOS and the mobility. Specifically, the Coulomb scattering (screened) and surface roughness model implemented in this work result in a very close match to the experimental mobility values for all the doping generations.

It also adopts a two-loop approach, where the outer loop solves for the self-consistency between the potential and the subband sheet charge density by solving the Poisson and the Schrödinger equation, respectively. This constitutes what we could refer to as the "macro ecosystem" of the device, where some semi-classical approximations can be made to reduce the complexity of the computation, as long as it is found reasonable to do so under the specific device conditions. For instance, as our device is considered to be in near-equilibrium, the effective mass approximation is considered to be valid.

The inner loop now solves for the nEGF (renormalization of the spectrum and the broadening of the states) self-consistently under the self-consistent Born approximation.

This results in new DOS that determines the new sheet charge density. This part involves a more precise and complete quantum mechanical treatment of the inversion layer physics on a more fundamental level. This two loop approach results in the economical way for the convergence to be achieved, instead of running two parallel loops.

The self-consistent inclusion of the inelastic intervalley phonon scattering is achieved in this inner loop by implementing a uniform energy mesh for the computation of the nEGF. This resulted in a time disadvantage, as the nEGF now had to be integrated throughout the entire energy mesh. Thus, the work employed a MPI parallelization technique that enables the user to save precious time, if enough resources are available to run the code in parallel. The unique nature of the nEGF integral equations make them extremely parallelizable. This results in a speedup of almost up to 10 times as noted by the author.

The inclusion of SCHREDV2.0 further extends the scope of the performance capabilities of the simulator. For example, it is now possible to adapt the code to different device structures like strain in silicon, different orientations in silicon and high-K devices. In addition one can also model Poly gate depletion, uniform/non-uniform doping, user defined number of valleys, partial/complete ionization of carriers and several other features. Future work could be extended to new materials like Germanium (Ge) and/or model silicon nanowires

Reference

[1] Moore, Gordon E. "Cramming more components onto integrated circuits" (1965).

[2] IntelPR., Darling Patrick "Intel 22nm 3-D Tri-Gate Transistor Technology", (2011)

[3] Roosbroeck, W. van. "Theory of the Flow of Electrons and Holes in Germanium and Other Semiconductors." Bell System Technical Journal 29.4 (1950): 560-607.

[4] Selberherr, Siegfried. "Analysis and Simulation of Semiconductor Devices." Springer Wien; New York, 1984.

[5] Jüngel, Ansgar. "Energy Transport in Semiconductor Devices." Mathematical and Computer Modelling of Dynamical Systems 16.1 (2010): 1-22.

[6] Miller, JJH, and Song Wang. "An Analysis of the Scharfetter-Gummel Box Method for the Stationary Semiconductor Device Equations." ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique 28.2 (1994): 123-40.

[7] Hess, Karl. Monte Carlo Device Simulation: Full Band and Beyond. Kluwer Academic Publishers, 1991.

[8] Jacoboni, Carlo, and Paolo Lugli. The Monte Carlo Method for Semiconductor Device Simulation. Vol. 3. Springer, 1989.

[9] Jacoboni, Carlo, and Lino Reggiani. "The Monte Carlo Method for the Solution of Charge Transport in Semiconductors with Applications to Covalent Materials." Reviews of Modern Physics 55.3 (1983): 645.

[10] Kometer, K., G. Zandler, and P. Vogl. "Lattice-Gas Cellular-Automaton Method for Semiclassical Transport in Semiconductors." Physical Review B 46.3 (1992): 1382.

[11] Zandler, G. "A Comparison of Monte Carlo and Cellular Automata Approaches for Semiconductor Device Simulation." IEEE Electron Device Letters 14.2 (1993): 77-9.

[12] Alam, Muhammad A., Mark A. Stettler, and Mark S. Lundstrom. "Formulation of the Boltzmann Equation in Terms of Scattering Matrices." Solid-state electronics 36.2 (1993): 263-71.

[13] Das, Amitava, and Mark S. Lundstrom. "A Scattering Matrix Approach to Device Simulation." Solid-State Electronics 33.10 (1990): 1299-307.

[14] Banoo, Kausar, Mark Lundstrom and R. Kent Smith. "Direct Solution of the Boltzmann Transport Equation in Nanoscale Si Devices." Simulation of Semiconductor Processes and Devices, 2000. SISPAD 2000. 2000 International Conference on.

[15] John Von Neumann. Mathematical Foundations of Quantum Mechanics. Princeton university press, 1955.

[16] Wigner, Eugene. "On the Quantum Correction for Thermodynamic Equilibrium." Physical Review 40.5 (1932): 749.

[17] Martin, Paul C., and Julian Schwinger. "Theory of Many-Particle Systems. I." Physical Review 115.6 (1959): 1342.

[18] Schwinger, Julian. "Brownian Motion of a Quantum Oscillator." Journal of Mathematical Physics 2.3 (1961): 407-32.

[19] Kadanof, Leo P., and Gordon Baym. "*Quantum Statistical Mechanics*".W. A. Benjamin, 1962.

[20] Barker, JR, and DK Ferry. "Self-Scattering Path-Variable Formulation of High-Field, Time-Dependent, Quantum Kinetic Equations for Semiconductor Transport in the Finite-Collision-Duration Regime." Physical Review Letters 42.26 (1979): 1779.

[21] Vasileska, Dragica, et al. "Calculation of the Average Interface Field in Inversion Layers using zero-temperature Green's Function Formalism." Journal of Vacuum Science & Technology B 13.4 (1995): 1841-7.

[22] Ferry, David K., and Harold L. Grubin. "Modeling of Quantum Transport in Semiconductor Devices." Solid State Physics 49 (1996): 283-448.

[23] Vasileska, Dragica, et al. "Quantum Transport Simulation of the DOS Function, Self-Consistent Fields and Mobility in MOS Inversion Layers." VLSI Design 6.1-4 (1998): 21-5.

[24] Price, PJ. "Two-Dimensional Electron Transport in Semiconductor Layers. I. Phonon Scattering." Annals of Physics 133.2 (1981): 217-39.

[25] Ridley, BK. "Hot Electrons in Low-Dimensional Structures." Reports on Progress in Physics 54.2 (1991): 169.

[26] Ferry, DK. "Hot-Electron Effects in Silicon Quantized Inversion Layers." Physical Review B 14.12 (1976): 5364.

[27] Yamada, Toshishige, et al. "In-Plane Transport Properties of Si/Si 1-x Ge x Structure and its FET Performance by Computer Simulation." Electron Devices, IEEE Transactions on 41.9 (1994): 1513-22.

[28] Ferry, David K. Semiconductors. IOP Publishing, 2013.

[29] Fischetti, Massimo V., and Steven E. Laux. "Monte Carlo Study of Electron Transport in Silicon Inversion Layers." Physical Review B 48.4 (1993): 2244.

[30] Blotekjaer, Kjell. "Transport Equations for Electrons in Two-Valley Semiconductors." Electron Devices, IEEE Transactions on 17.1 (1970): 38-47.

[31] Ferry, David K., and Carlo Jacoboni. Quantum Transport in Semiconductors. Plenum Press New York, 1992.

[32] Grasser, Tibor, and T. Grasser. *Advanced Device Modeling and Simulation*. Vol. 31. World Scientific, 2003.

[33] Lundstrom, Mark. *Fundamentals of Carrier Transport*. Cambridge University Press, 2009.

[34] Stratton, R. "Diffusion of Hot and Cold Electrons in Semiconductor Barriers." Physical Review 126.6 (1962): 2002.

[35] Abrikosov, Alekseĭ Alekseevich , Lev Petrovich Gor′kov , and Igor′ Ekhiel′evich Dzi͡aloshinskiĭ. *Quantum field theoretical methods in statistical physics*. Vol. 4. Pergamon, 1965.

[35] GC Wick; The evaluation of the collision matrix. Phys. Rev., 80 (1950), pp. 268–272

[36] Mattuck, Richard D. *A guide to Feynman diagrams in the many-body problem*. Courier Dover Publications, 2012.

[37] Mahan, Gerald D. *Many-particle physics*. Springer, 2000.

[38] Schwinger, J. "Phys.(NY) 2, 407 (1961); LV Keldysh." *Zh. Eksp. Teor. Fiz* 47 (1964): 1515.

[39] Kannan, Gokula, and Dragica Vasileska. "Schred V2. 0: Tool to model MOS capacitors." *Computational Electronics (IWCE), 2010 14th International Workshop on*. IEEE, 2010.

[40] Tan, I. H., et al. "A self‐consistent Solution of Schrödinger–Poisson Equations using a Nonuniform Mesh." Journal of Applied Physics 68 (1990): 4071.

[41] Ferry, DK. "Transport of Hot Carriers in Semiconductor Quantized Inversion Layers." Solid-State Electronics 21.1 (1978): 115-21.

[42] Stern, Frank, and WE Howard. "Properties of Semiconductor Surface Inversion Layers in the Electric Quantum Limit." Physical Review 163.3 (1967): 816.

[43] Goodnick, SM, et al. "Surface Roughness at the Si (100)-SiO 2 Interface." Physical Review B 32.12 (1985): 8171.

[44] Matsumoto, Yukio, and Yasutada Uemura. "Scattering Mechanism and Low Temperature Mobility of MOS Inversion Layers." Japanese Journal of Applied Physics 13.S2 (1974): 367.

[45] Ando, Tsuneya, Alan B. Fowler, and Frank Stern. "Electronic Properties of Two-Dimensional Systems." Reviews of Modern Physics 54.2 (1982): 437.

[46] Saguaro High Performance Computing Cluster computer at Arizona State University (HTTPS://RESEARCHCOMPUTING.ASU.EDU/)

[47] Ocotillo High Performance Computing Cluster computer at Arizona State University (HTTPS://RESEARCHCOMPUTING.ASU.EDU/)

[48] John Towns, Timothy Cockerill, Maytal Dahan, Ian Foster, Kelly Gaither, Andrew Grimshaw, Victor Hazlewood, Scott Lathrop, Dave Lifka, Gregory D. Peterson, Ralph Roskies, J. Ray Scott, Nancy Wilkins-Diehr, "XSEDE: Accelerating Scientific Discovery", Computing in Science & Engineering, vol.16, no. 5, pp. 62-74, Sept.-Oct. 2014, doi:10.1109/MCSE.2014.80

[49] Takagi, Shin-ichi, et al. "On the universality of inversion layer mobility in Si MOSFET's: Part I-effects of substrate impurity concentration." *IEEE Transactions on Electron Devices* 41.12 (1994): 2357-2362.