OPTIMAL ELECTRONIC DEVICE DESIGN

by

Kelly C. Magruder

A Dissertation Presented to the
FACULTY OF THE GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
DOCTOR OF PHILOSOPHY
(ELECTRICAL ENGINEERING)

August 2011

Copyright 2011 Kelly C. Magruder

Table of Contents

List of Figures v

Abstract xv

Chapter 1: Introduction 1
1.1 Basic optimal design algorithm 4
1.2 Improved physical models 5
1.2.1 First-order perturbation theory 7
1.2.2 Going beyond perturbation theory 11
1.3 Organization of thesis 12

Chapter 2: Calculation of the self-consistent potential 14
2.1 Prescription for calculating the self-consistent potential 16
2.2 Solution of the Schrödinger equation 17
2.2.1 Propagation matrix method 19
2.2.2 Scattering state wave functions 21
2.2.3 Quasi-bound state wave functions 23
2.3 Calculation of current density 25
2.4 Calculation of the electron density 27
2.4.1 Charge neutrality in the contacts 27
2.4.2 Scattering state electron density 30
2.4.3 Quasi-bound state electron density 30
2.5 Solution of Poisson equation 32
2.5.1 Solution method 34
2.5.2 Predictor-corrector scheme 37
2.6 Measures taken to enhance stability and accuracy 38
2.6.1 Sufficient sampling of electron transmission resonances . . . 38
2.6.2 Energy vector for current density calculation 39
2.6.3 Accuracy of quasi-bound state energies 39
2.6.4 Spatial requirements for slowly decaying quasi-bound states 40
2.6.5 Dynamic slowing of the solver 40

2.6.6 Absolute error check . . . 42
2.7 Conclusion and future work . . . 42

Chapter 3: Optimal electronic device design 46
3.1 $Al_{\xi}Ga_{1-\xi}As$ material system . . . 46
3.2 Optimal design of semiconductor heterostructure devices . . . 48
3.2.1 The adjoint method . . . 50
3.2.2 Derivative of the cost function . . . 51
3.3 Extent and limit of control for linear objective . . . 53
3.3.1 The single-barrier heterostructure diode . . . 54
3.3.2 Chosen device structure for optimization . . . 55
3.3.3 Optimization for linear objective . . . 57
3.3.4 Results for linear objective function . . . 58
3.3.5 Feasibility of design . . . 62
3.3.6 Sensitivity analysis . . . 65
3.4 Application of customized, optimally designed diode in an RF switching mixer . . . 67
3.4.1 Ideal square-law mixer . . . 68
3.4.2 Improved performance of switching mixer with a customized diode . . . 71
3.4.3 Voltage gain vs. power gain . . . 77
3.4.4 Optimal design of parametrically defined diode . . . 78
3.4.5 Switching mixer utilizing optimized diode . . . 80
3.5 Conclusion and future work . . . 81

Chapter 4: Inelastic electron transport 83
4.1 Quantum model of electron-phonon interaction . . . 85
4.1.1 Solution to Schrödinger equation using the propagation matrix method . . . 90
4.1.2 Conservation of current in the presence of inelastic scatterers . . . 92
4.1.3 Strong coupling regime . . . 94
4.2 Introduction to unitary feedback . . . 95
4.3 Quantum and semi-classical inelastic electron transport . . . 97
4.3.1 Simple first-order perturbation theory . . . 97
4.3.2 Inelastic scattering in an otherwise constant background potential . . . 102
4.3.3 Condition when perturbation solution appears similar to exact calculation . . . 103
4.3.4 Threshold at which perturbation solution appears similar to exact calculation . . . 106
4.4 Conclusion and future work . . . 114

Chapter 5: Conclusion 116

Bibliography 118

List of Figures

1.1 (a) Non-equilibrium electron transistor (NET) featuring material choices that isolate equilibrium and non-equilibrium electrons in the base region [22]. (b) InP heterostructure bipolar transistor (HBT) featuring a chirped superlattice in the collector. The superlattice is designed to prevent electrons from scattering into the L-band minimum, and is effective only because of the wave nature of the electron in a non-equilibrium transport regime [38]. 2

1.2 (a) Optimized potential profile and (b) transmission probability of an $\mathrm{Al}_{\xi}\mathrm{Ga}_{1-\xi}\mathrm{As}$ heterostructure tunnel diode optimized for linear transmission as a function of V_{bias} for an electron of energy $E = 26$ meV and effective electron mass $m = 0.07 \times m_0$ [39]. The optimal potential was found through an exhaustive search of the design space in which physical constraints on the potential barrier energy, $V(x)$, were $0 \leq V(x) \leq 0.3$ eV discretized into 1 meV increments. The n-type doping concentration is $N_{\mathrm{D}} = 1 \times 10^{18}$ cm^{-3}. (c) Optimized potential profile and (d) current density of $\mathrm{Al}_{\xi}\mathrm{Ga}_{1-\xi}\mathrm{As}$ heterostructure tunnel diode optimized for quadratic current-voltage characteristic [40]. The design was discovered using the Newton-Raphson method to minimize cost. Simulation parameters are electron effective mass $m = 0.07 \times m_0$, n-type doping concentration $N_{\mathrm{D}} = 1 \times 10^{18}$ cm^{-3}, and temperature $T = 300$ K. 3

1.3 Flow chart showing basic optimal design algorithm. 5

1.4 Diagram showing first-order approximation for Fermi's golden rule, where the probability of an electron occupying state $|\mathbf{k}\rangle$ is $P_{\mathbf{k}}$. Fermi's golden rule assumes that the electron is weakly perturbed by the interaction, decoupling the scattering states from one another and violating unitarity. 8

1.5 Diagram depicting interaction of an electron wave packet and LO phonons having energy $\hbar\omega_{\mathrm{LO}}$. In a polar semiconductor such as GaAs LO phonons displace lattice ions from their equilibrium positions (dashed line) such that electric dipoles are created which enable the lattice and electron to exchange momentum and energy. 9

1.6 Calculated inelastic scattering rate for an electron of energy $E_{\mathbf{k}}$ interacting with LO phonons in GaAs, with scattering rates for an electron with energy $E_{\mathbf{k}} = 100$ meV indicated. At temperature $T = 0$ K, no scattering may occur until the electron has energy $E_{\mathbf{k}} > \hbar\omega_{\mathrm{LO}}$. Simulation parameters are electron effective mass $m = 0.07 \times m_0$, LO phonon energy $\hbar\omega_0 = 36$ meV, TO phonon energy $\hbar\omega_{\mathrm{TO}} = 33$ meV, high frequency permittivity $\epsilon_\infty = 11.1 \times \epsilon_0$, and low frequency permittivity $\epsilon_{\mathrm{DC}} = 13.2 \times \epsilon_0$. 11

2.1 (a) Potential profile of $Al_{\xi}Ga_{1-\xi}As$ resonant tunneling diode calculated both self-consistently and using the depletion approximation. Storage of charge within the structure results in band bending and a significant increase in potential energy. (b) Electron concentration for potentials of part (a) and background doping concentration profile, which has been smoothed using an error function profile. Simulation parameters are electron effective mass $m = 0.07 \times m_0$, temperature $T = 300$ K, n-type doping concentration $N_{\mathrm{D}} = 1 \times 10^{18}$ cm^{-3}, dopant diffusion parameter $L_{\mathrm{d}} = 1$ nm, and $V_{\mathrm{bias}} = 0.1$ V. . . 15

2.2 Schematic representation of electrons injected into device region from contacts maintaining thermal distributions of electrons in the presence of an applied voltage bias, V_{bias} . Also indicated is occupation of electrons in accumulation region quasi-bound states. The emitter and collector contacts are both n-type doped with doping concentration N_{D} and characterized by chemical potentials μ_{L} and μ_{R}, respectively at temperature T. V_{bias} causes the chemical potentials far away from the device region to differ by $\mu^{\mathrm{R}} = \mu^{\mathrm{L}} - eV_{\mathrm{bias}}$. 18

2.3 Diagram of elastic scattering from a step change in potential occurring at position $x = 0$. The propagation matrix for region j accounts for free electron propagation across length L_j and scattering at the interface between regions j and $j+1$. Quantities a_j, b_j, c_j, and d_j are wave function coefficients associated with region j as shown. 20

2.4 Diagram of boundary condition for electrons injected from each of the contacts. Wave function coefficients A and B are associated with the injecting contact, while coefficients C and D are associated with the opposite contact. Boundary conditions for two-calculation and one-calculation solution methods are shown. 22

2.5 (a) Electron probability density injected from collector contact for potential profile containing a rectangular potential barrier having a 0.25 eV conduction band offset and a thickness of 5 nm. The quasi-bound state energy is $E^{\mathrm{B}} = -33.7$ meV. (b) Diagram of electron injection into the accumulation region to find quasi-bound state resonances and calculate quasi-bound state wave functions. Boundary conditions for each calculation are indicated, and the wave function shown is that of the quasi-bound state of part (a) offset in energy by E^{B}. The applied voltage bias is $V_{\mathrm{bias}} = 0.35$ eV, and simulation parameters are electron effective mass $m = 0.07 \times m_0$, temperature $T = 300$ K, n-type doping concentration $N_{\mathrm{D}} = 1 \times 10^{18}$ cm^{-3}, and dopant diffusion parameter $L_{\mathrm{d}} = 1$ nm. 24

2.6 (a) Charge density calculated without using drifted Fermi distributions for the case of a strong ($T(E_{\mathrm{F}} = 52\ \mathrm{meV}) = 3.1 \times 10^{-9}$) potential barrier having a 0.35 eV conduction band offset and 9 nm thickness subject to an applied voltage bias $V_{\mathrm{bias}} = 0.25$ V. (b) Same as (a) but for a weaker ($T(E_{\mathrm{F}} = 52\ \mathrm{meV}) = 3.7 \times 10^{-3}$) potential barrier 3 nm in thickness. The potentials have been calculated using the depletion approximation, and simulation parameters are electron effective mass $m = 0.07 \times m_0$, temperature $T = 300$ K, n-type impurity concentration $N_{\mathrm{D}} = 1 \times 10^{18}$ cm^{-3}, and dopant diffusion parameter $L_{\mathrm{d}} = 1$ nm. 28

2.7 Calculated electron density for potential profile of Fig. 2.6(b) with drifted Fermi distributions used to neutralize the contacts. Calculated values of $k^{\alpha}_{\mathrm{D},\perp}$ are indicated in the Fig., and the potential barrier has a 0.35 eV conduction band offset and is 3 nm in thickness. Simulation parameters are electron effective mass $m = 0.07 \times m_0$, temperature $T = 300$ K, n-type doping concentration $N_{\mathrm{D}} = 10^{18}$ cm^{-3}, dopant diffusion parameter $L_{\mathrm{d}} = 1$ nm, and $V_{\mathrm{bias}} = 0.25$ V. 29

2.8 Diagram depicting scattering state electrons inelastically scattering into a quasi-bound state in the accumulation region, with this scattering characterized by a scattering rate $\frac{1}{\tau_{\text{in}}}$. Electrons within the quasi-bound state may escape the bound state by tunneling through the potential barrier at a rate characterized by $\frac{1}{\tau_{\text{tun}}}$. Assuming $\frac{1}{\tau_{\text{in}}} \gg \frac{1}{\tau_{\text{tun}}}$, the quasi-bound state becomes an electron reservoir which will be in thermal equilibrium with the left-hand contact and characterized by chemical potential μ^{L}. 32

2.9 Diagram showing discretization scheme and derivation for the discretized Poisson equation. Potential ϕ and charge density ρ are defined at points between regions. The regions are separated by a length dx and are characterized by permittivity ϵ and electric field E. 33

2.10 Difference in electron probability density injected from left-hand contact between two potentials that differ by a $\Delta\phi = +1$ meV perturbation at position $x = 50$ nm. The inset shows the effect of $\Delta\phi$ in the calculated charge densities. The primes indicate quantities calculated with the perturbation. Simulation parameters are electron effective mass $m = 0.07 \times m_0$, temperature $T = 300$ K, n-type impurity concentration $N_{\text{D}} = 1 \times 10^{18}$ cm^{-3}, and dopant diffusion parameter $L_{\text{d}} = 1$ nm. 36

2.11 Flow chart displaying algorithm for calculating the self-consistent potential. 44

2.12 Flow chart describing driver for the self-consistent solver. This algorithm provides error handling and calculates the current-voltage characteristic. 45

3.1 Band structure for (a) GaAs and (b) AlAs calculated using an sp^3s^* model [45]. Band gap energies, E_{g}, are indicated. 47

3.2 (a) Tunnel-diode consisting of an undoped AlAs layer of length L_0 between two GaAs contacts subject to an applied voltage bias $V_{\text{bias}} = 0.25$ V. Also shown are current-voltage characteristics through the device of part (a) in (b) linear and (c) log scales for several values of L_0. Current is exponential for $|V_{\text{bias}}| > \mu$ ($\mu = 38.9$ meV) and quasi-linear for $|V_{\text{bias}}| < \mu$. Simulation parameters are electron effective mass $m = 0.07 \times m_0$, temperature $T = 300$ K, and n-type doping concentration $N_{\text{D}} = 1 \times 10^{18}$ cm^{-3}. 56

3.3 Randomly configured $\text{Al}_\xi\text{Ga}_{1-\xi}\text{As}$ heterostructure diode displaying features selected for the optimized potential profiles. An AlAs layer having a thickness L_0 is the primary current limiter. The following series of $\text{Al}_\xi\text{Ga}_{1-\xi}\text{As}$ layers fine-tune the current-voltage characteristic, with the number of $\text{Al}_\xi\text{Ga}_{1-\xi}\text{As}$ layers, N, used as an auxiliary parameter to set the length of the device. The $\text{Al}_\xi\text{Ga}_{1-\xi}\text{As}$ layers have a fixed thickness $L_\delta = 4 \times \delta$ and $L_0 = n \times \delta$, where n is a positive integer and $\delta = 0.2827$ nm is one monolayer of $\text{Al}_\xi\text{Ga}_{1-\xi}\text{As}$. Total thickness of the active region is $L = L_0 + NL_\delta$. 57

3.4 Overlay of simulated current through the most optimal design for each r, with corresponding N indicated. Currents for $r > 10$ k$\Omega \times \mu\text{m}^2$ have been scaled so that all curves overlap. Simulation parameters are electron effective mass $m = 0.07 \times m_0$, temperature $T = 300$ K, doping concentration $N_\text{D} = 1 \times 10^{18}$ cm^{-3} 59

3.5 (a) Optimized potential profile for a diode characterized by $r = 1$ M$\Omega \times \mu\text{m}^2$ and $N = 12$. (b) Error in differential resistance relative to r. Simulation parameters are electron effective mass $m = 0.07 \times m_0$, temperature $T = 300$ K, doping concentration $N_\text{D} = 1 \times 10^{18}$ cm^{-3} 59

3.6 Spectral intensity and total harmonic distortion (THD) for optimized diode shown in Fig. 3.5. The input voltage is a sinusoidal wave of frequency ω_0 and amplitude V_p with constant offset voltage $V_\text{bias} = 0.25$ V. Suppression of the second harmonic for $V_p \leq 0.1$ V results from odd symmetry in the differential resistance about $V_\text{bias} = 0.25$ V. Total harmonic distortion of less than -65 dB indicates greater than 10 bit dynamic range. 60

3.7 Deviation from objective as a function of the number of $\text{Al}_\xi\text{Ga}_{1-\xi}\text{As}$ layers, N. The trend indicates that a natural value for auxiliary parameter N exists at $N = 10$. 61

3.8 Relative error in differential resistance for (a) $N = 7$, (b) $N = 10$, (c) $N = 12$, and (d) $N = 15$ $\text{Al}_\xi\text{Ga}_{1-\xi}\text{As}$ layers. As device length increases, the number and strength of the resonances increases, causing greater difficulty controlling the current-voltage characteristic. Designs featuring $N = 5$ and $N = 17$ designs are not shown as much of the error extends beyond the scale shown. 63

3.9 (a) Potential profile of device of Fig. 3.5 with exponential decaying potential manually fit to the average potential energy of the $Al_\xi Ga_{1-\xi}As$ layers. Also shown are fits used for the $N = 7$ and $N = 10$ designs of $r = 1$ M$\Omega \times \mu$m^2. (b) Current calculated using exponentially decaying potential profiles of part (a). Proximity of curves to the objective is shows that the physical mechanism responsible for setting the rough features in the current is electron tunneling. Potential well features added by the optimization algorithm fine-tune the current. Simulation parameters are electron effective mass $m = 0.07 \times m_0$, temperature $T = 300$ K, doping concentration $N_\mathrm{D} = 1 \times 10^{18}$ cm^{-3} 64

3.10 (a) Histogram showing sensitivity of the optimal design of Fig. 3.5 to errors in $Al_\xi Ga_{1-\xi}As$ layer thickness by calculating Δ with a single $Al_\xi Ga_{1-\xi}As$ layer perturbed by $\pm 1 \times \delta$ or $\pm 2 \times \delta$. (b) Data from part (a) plotted by $Al_\xi Ga_{1-\xi}As$ layer. The inset shows the potential profile for the device simulated, indicating the $Al_\xi Ga_{1-\xi}As$ layer numbers. Simulation parameters are electron effective mass $m = 0.07 \times m_0$, temperature $T = 300$ K, doping concentration $N_\mathrm{D} = 1 \times 10^{18}$ cm^{-3} 65

3.11 Histogram showing sensitivity of optimal design to alloy fluctuations and error in Al alloy concentration. In each simulation all $Al_\xi Ga_{1-\xi}As$ layers were perturbed by a random potential energy uniformly distributed between ± 3 meV. Simulation parameters are electron effective mass $m = 0.07 \times m_0$, temperature $T = 300$ K, doping concentration $N_\mathrm{D} = 1 \times 10^{18}$ cm^{-3} 66

3.12 Circuit diagram of full-wave rectifier which may be used as a switching mixer to mix RF and local oscillator (LO) signals. The output signal, V_out, is taken across the load resistance, R_L, and the resistance of the source is R_S. 68

3.13 Spectral intensity of V_out of the switching mixer shown in Fig. 3.12 simulated using LTSPICE. Although the mixing has been accomplished, many spurious harmonics are produced. The IF output intensity is -40 dBm and the intensity at frequency $f = 2f_\mathrm{IF}$ is -82 dBm. Simulation parameters are LO power $P_\mathrm{LO} = 0$ dBm (-10 dBV) and frequency $f_\mathrm{LO} = 1$ GHz, RF power $P_\mathrm{RF} = -30$ dBm (-40 dBV) and frequency $f_\mathrm{RF} = 900$ MHz, source resistance $R_\mathrm{S} = 50$ Ω, and load resistance $R_\mathrm{L} = 600$ Ω. 70

3.14 (a) Diagram showing a rectified sinusoidal input to a perfect square-law device yields a squared sine wave. (b) Frequency components of sign function and square of the sign function. The square-law transfer characteristics redistributes the power from an infinite set of harmonics into only the DC component. 72

3.15 Circuit diagram of switching mixer utilizing diode characterized by a parametrically defined current-voltage characteristic to eliminate unwanted harmonics. Inputs to the circuit are RF signal of amplitude V_{RF} and frequency ω_{RF} and LO signal of amplitude V_{LO} and frequency ω_{LO}. The source and load resistances are R_{S} and R_{L}, respectively. 73

3.16 Parametrically defined diode current-voltage characteristic given by the equations in the Fig. A feasible design region is indicated bounded by maximum source voltage, $V_{\mathrm{s,max}}$, where the slope $\partial I/\partial V_{\mathrm{d}} = \infty$. Simulation parameters are $\alpha = 6.9$ mA/V^2, source resistance $R_{\mathrm{S}} = 50\ \Omega$, and load resistance $R_{\mathrm{L}} = 50\ \Omega$. 74

3.17 Comparison of current-voltage characteristics for series circuits with and without the parametrically defined diode. Maximum source voltage, $V_{\mathrm{s,max}}$, is defined as the point at which the slope of the parabola matches that of the line. Parameters for the Fig. are source resistance $R_{\mathrm{S}} = 50\ \Omega$, load resistance $R_{\mathrm{L}} = 50\ \Omega$, and $\alpha = 6.9$ mA/V^2, resulting in $V_{\mathrm{s,max}} = 0.73$ V. 75

3.18 Spectral intensity of output of switching mixer containing parametrically defined diode. The customized current-voltage characteristic of the ideal diode eliminates all spurious harmonics above the second-order terms. The IF power is -49 dBm, for a power gain of -19 dB. Simulation parameters are LO power $P_{\mathrm{LO}} = 0$ dBm and frequency $f_{\mathrm{LO}} = 1$ GHz, RF power $P_{\mathrm{RF}} = -30$ dBm and frequency $f_{\mathrm{RF}} = 900$ MHz, source resistance $R_{\mathrm{S}} = 50\ \Omega$, and load resistance $R_{\mathrm{L}} = 50\ \Omega$. The parametrically defined diode is characterized by $\alpha = 6.9$ mA/V^2. 77

3.19 Simulated current density through a Al$_\xi$Ga$_{1-\xi}$As heterostructure tunnel barrier diode optimized for current shown in Fig. 3.16 given a cross-sectional area of $A = 1.6\ \mu$m^2. Shown in the inset is the conduction band potential profile for the device. Simulation parameters are temperature $T = 300$ K, electron effective mass $m = 0.07 \times m_0$, n-type doping concentration $n = 1 \times 10^{19}$ cm^{-3}. The objective function is characterized by $\alpha = 1.06$ mA/V^2, and $R_{\mathrm{tot}} = 100\ \Omega$. 79

3.20 (a) Spectral intensity of output of switching mixer using optimized parametrically defined diode. The IF voltage gain for this mixer is -14.7 dBV and the power in the fourth order harmonic at frequency $f = 2f_{\mathrm{IF}}$ is -114.3 dBV. (b) Spectral intensity of output of switching mixer without parametrically defined diode. The IF voltage gain for this mixer is -4.6 dBV and the power in the fourth order harmonic at frequency $f = 2f_{\mathrm{IF}}$ is -86.7 dBV. Simulation parameters are LO power $P_{\mathrm{LO}} = 0$ dBm and frequency $f_{\mathrm{LO}} = 1$ GHz, RF power $P_{\mathrm{RF}} = -30$ dBm and frequency $f_{\mathrm{RF}} = 900$ MHz, source resistance $R_{\mathrm{S}} = 50\ \Omega$, and load resistance $R_{\mathrm{L}} = 600\ \Omega$, and $\alpha = 1.06$ mA/V^2. 80

4.1 Diagram depicting coherent and incoherent regimes for interaction between an electron and optical phonon. The electron and phonon may interact in a coherent manner until the optical phonon dissipates into acoustic phonons. 84

4.2 Reflection and transmission due to an inelastic system. An electron initially in state $|k_0\rangle$ with energy E_0 inelastically scatters into final state $|k_n\rangle$ with energy $E_n = E_0 - n\hbar\omega_0$. Positive n denotes net phonon emission and the phonon energy is $\hbar\omega_0$. 85

4.3 Diagram of a potential step with Einstein phonons localized at position $x = x_0$ (dashed line). Electron waves are incident from the left and right with amplitudes a_n and d_n, respectively. Waves will be scattered from position $x = x_0$ to the left and to the right having amplitudes b_n and c_n, respectively. The electron-phonon coupling constant is g_1, the static amplitude of the delta barrier is g_0, the phonon energy is $\hbar\omega_0$, and the temperature is $T = 0$ K. 87

4.4 (a) A rectangular potential barrier containing Einstein phonons (dashed line). The phonons are centered within the barrier and the electron injected with energy E_0 has plane wave components incident from $\pm\infty$. (b) Probability of an electron of energy E_0 exiting the right hand side of (a) in channel n. The potential barrier has energy $U_0 = 0.1$ eV and length $L = 1$ nm, the phonons have energy $\hbar\omega_0 = 36$ meV, and the coupling constant is $g_1 = 0.05$ eV nm. The effective electron mass is $m = 0.07 \times m_0$, the number of inelastic channels included in the simulation is $M = 11$, the temperature is $T = 0$ K, and the simulation converged with a relative error of less than 10^{-10}. 96

4.5 (a) Calculated total inelastic electron transmission in the presence of inelastic phonon scattering (solid) and total elastic transmission in the absence of any phonon scattering (dashed). (b) Transmission due to each of the inelastic channels. The potential contains Einstein phonons in an otherwise constant background potential, $U(x) = 0$. The model parameters are $m = 0.07 \times m_0$, $\hbar\omega_0 = 36$ meV, $g_0 = 0.2$ eV nm, $g_1 = 0.12$ eV nm, and $T = 0$ K. 96

4.6 (a) Transmission coefficient calculated using quantum model and Born series to selected orders. Near phonon thresholds the perturbative solution does produce qualitatively accurate predictions. Simulation parameters are effective electron mass $m = 0.07 \times m_0$, phonon energy $\hbar\omega_0 = 36$ meV, coupling strength $g_1 = 0.05$ eV nm, and $N = 15$ inelastic channels in the quantum simulation. 101

4.7 Transmission through Einstein phonons located in a constant background potential $U(x) = 0$ eV with coupling constants of $g_1 = 0.008$ eV nm (solid) and $g_1 = 0.08$ eV nm (dashed) using the (a) exact quantum solution and (b) perturbative solution. The electron of effective mass $m = 0.07 \times m_0$ is injected from $x = -\infty$ with energy E_0, the phonon energy is $\hbar\omega_0 = 36$ meV, and $M = 11$ inelastic channels were included in the simulation. The inset shows the features of the transmission about $E_0 = \hbar\omega_0$ with a coupling constant of $g_1 = 0.008$ eV nm on a fine energy scale. 103

4.8 (a) Quantum and (b) first-order perturbative solutions for the transmission of an electron of energy E_0 injected from $x = -\infty$ through the potential shown in Fig. 4.4(a). The potential barrier has energy $U_0 = 0.25$ eV and length L, $g_1 = 0.05$ eV nm, $\hbar\omega_0 = 36$ meV, the effective electron mass is $m = 0.07 \times m_0$, and $M = 11$ inelastic channels were included in the simulation. 104

4.9 Quantum (solid) and perturbative (dashed) transmission from Fig. 4.8 shown on a fine scale for potential barrier lengths of (a) $L = 1$ nm and (b) $L = 5$ nm. The transmission curves have been offset by $T(\hbar\omega_0)$ so that the quantum and perturbative solutions may be compared. The potential barrier has energy $U_0 = 0.25$ eV and length L, and the phonon is characterized by $g_1 = 0.05$ eV nm and $\hbar\omega_0 = 36$ meV. $M = 11$ inelastic channels were included in the simulation and the effective electron mass is $m = 0.07 \times m_0$. 105

4.10 Transmission coefficient at which $\partial T/\partial E_0 = 0$ at energy $E_0 = \hbar\omega_0$ for three different phonon locations. In the simulation the potential barrier energy $U_0 > \hbar\omega_0$ is fixed and the potential barrier length L is swept until $\partial T/\partial E_0 = 0$ at energy $E_0 = \hbar\omega_0$. The simulation used a coupling constant of $g_1 = 0.008$ eV nm, phonon energy $\hbar\omega_0 = 36$ meV, electron effective mass $m = 0.07 \times m_0$, and $M = 11$. 113

Abstract

A physical model of electron transport through nanoscale semiconductor heterostructures is developed featuring self-consistent solution of the Schrödinger and Poisson equations for the potential. Using this model heterostructure tunnel diodes featuring customized current-voltage characteristics are designed using principles of optimal design. Control over electron transport in these devices, having lengths in the tens of nanometers, is possible by exploiting the wave nature of the electron.

Limits of this control are explored using a linear objective function for optimal design of $Al_{\xi}Ga_{1-\xi}As$ heterostructure tunnel diodes. Device optimizations demonstrate that greater than 65 dB dynamic range over a 0.4 V peak-to-peak voltage swing may be achieved in a device that is only 17 nm thick. It is also shown how material choice and device lengths can limit achievable linearity.

The ability to design heterostructure diodes featuring customized current-voltage characteristics enable co-design of circuits and the devices used in those circuits. As an example, an $Al_{\xi}Ga_{1-\xi}As$ heterostructure tunnel diode is customized using principles of optimal design for use in an RF mixer by parametrically defining the current-voltage characteristic. It is shown that the customized current-voltage characteristic of this device is able to reduce spurious frequency components created by a switching mixer. In particular, fourth-order harmonic at twice the intermediate frequency is reduced by 28 dB, reducing by 4 the number of filter poles necessary for a fixed dynamic range.

Development of the physical model is possible by considering coherent scattering between an electron and phonon. A simple quantum model that captures the physics of this scattering regime is used to determine when predictions of first-order perturbation theory are qualitatively accurate with respect to a fully quantum model. It is found that small matrix elements cannot be used to justify perturbative methods. Only when feedback effects occur predominantly in the reflection coefficient do perturbative predictions appear to emulate quantum solutions. Situations are demonstrated in which perturbation theory cannot emulate quantum predictions even in the presence of weak coupling and large elastic scattering strengths.

Chapter 1

Introduction

Control over the input-output characteristics of nano-scale heterostructure devices is possible by exploiting the wave nature of the electron in precisely defined potential profiles. Through high-precision crystal growth methods structures shorter than a characteristic inelastic scattering mean free path, λ_{in}, are accessible allowing electrons to propagate through the structure coherently. Coherence in a non-equilibrium electron transport regime enables the potential profile of a heterostructure device to control output behavior through manipulation of electron transmission resonances via conduction band offsets controlled through material choice.

Two examples of devices exploiting degrees of freedom available through heterostructure material choice and non-equilibrium electron transport are shown in Fig. 1.1. The non-equilibrium electron transistor (NET) shown in part (a) uses large conduction band offsets to energetically separate equilibrium and non-equilibrium electron distributions in the base [22]. This feature reduces electron-electron scattering that may otherwise occur in the base region, increasing current gain. Part (b) shows an InP heterostructure bipolar transistor (HBT) [38]. Here designers have created a chirped superlattice in the collector to reduce the probability of electrons scattering into the L-band minimum. Its effectiveness is made possible only through the wave nature of the electron.

These devices are the product of an *ad-hoc* design process. In this design process, decisions are driven by the knowledge and intuition of the designer. Because of the

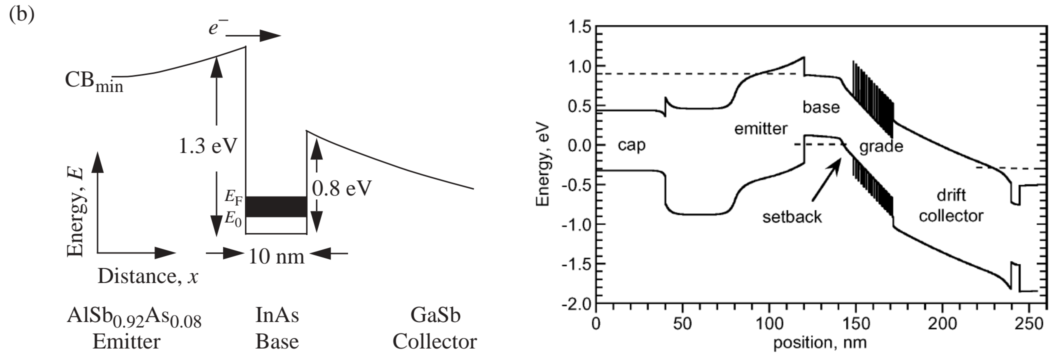

Figure 1.1: (a) Non-equilibrium electron transistor (NET) featuring material choices that isolate equilibrium and non-equilibrium electrons in the base region [22]. (b) InP heterostructure bipolar transistor (HBT) featuring a chirped superlattice in the collector. The superlattice is designed to prevent electrons from scattering into the L-band minimum, and is effective only because of the wave nature of the electron in a non-equilibrium transport regime [38].

complexity of non-equilibrium electron transport, final designs may exhibit known symmetries and operate in an intuitive manner. Of the feasible design space, it is likely that only those regions containing these features are searched. And although the final design is likely the most optimal one considered by the designer, it will be unknown whether the design is optimal either locally or globally.

An optimal design process improves upon *ad-hoc* design by executing a systematic search of the design space for those solutions that most closely match a desired objective. This search may find locally optimal solutions that exhibit broken symmetries and behave in a non-intuitive fashion. Complexity can improve results by providing multiple, locally optimum solutions in a non-convex solution space. Through analysis of these designs, additional physical insight may be possible. This insight can be passed back into the optimization through, for example, improved physical models or more natural objective functions.

Proof-of-principle demonstrations of control over electron transport in nanoscale heterostructure tunnel diodes using optimal design are shown in Fig. 1.2. The

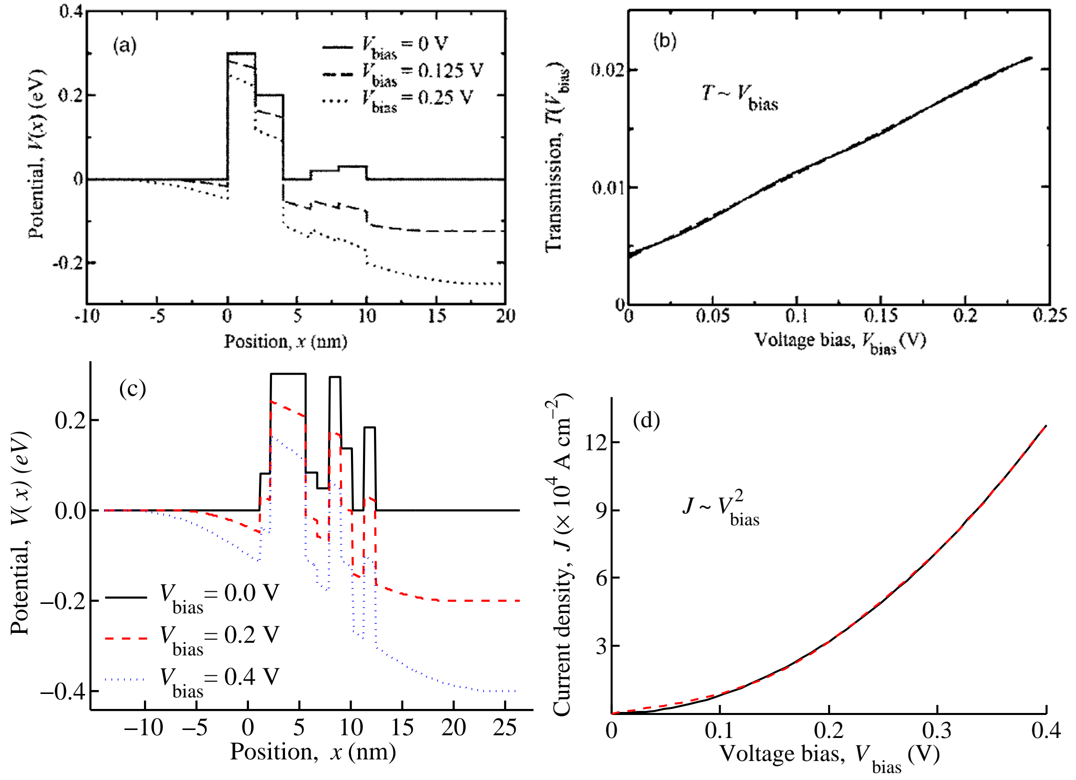

Figure 1.2: (a) Optimized potential profile and (b) transmission probability of an $Al_{\xi}Ga_{1-\xi}As$ heterostructure tunnel diode optimized for linear transmission as a function of $V_{\rm bias}$ for an electron of energy $E = 26$ meV and effective electron mass $m = 0.07 \times m_0$ [39]. The optimal potential was found through an exhaustive search of the design space in which physical constraints on the potential barrier energy, $V(x)$, were $0 \leq V(x) \leq 0.3$ eV discretized into 1 meV increments. The n-type doping concentration is $N_{\rm D} = 1 \times 10^{18}$ cm^{-3}. (c) Optimized potential profile and (d) current density of $Al_{\xi}Ga_{1-\xi}As$ heterostructure tunnel diode optimized for quadratic current-voltage characteristic [40]. The design was discovered using the Newton-Raphson method to minimize cost. Simulation parameters are electron effective mass $m = 0.07 \times m_0$, n-type doping concentration $N_{\rm D} = 1 \times 10^{18}$ cm^{-3}, and temperature $T = 300$ K.

$Al_{\xi}Ga_{1-\xi}As$ tunnel diode of parts (a) and (b) is optimized for linear transmission probability as a function of voltage bias, $V_{\rm bias}$, for an electron with energy $E = 26$ meV and effective electron mass $m = 0.07 \times m_0$ [39]. The $Al_{\xi}Ga_{1-\xi}As$ tunnel diode of parts (c) and (d) is optimized for a quadratic current-voltage characteristic as a function of $V_{\rm bias}$ at temperature $T = 300$ K [40]. Each of these designs displays broken symmetry and their

input-output characteristic is not readily apparent through intuition of wave mechanics. The potential well features of these potential profiles show that the optimization algorithm utilizes electron transmission resonances to control electron transport.

1.1 Basic optimal design algorithm

Optimal design is an automated iterative design process, the basic algorithm of which is shown in Fig. 1.3. A physical model must be chosen or developed that captures the essential physics of a device, enabling accurate prediction of a measurable quantity. Here the measurable quantity of interest is the current density as a function of V_{bias} , $J_{\text{sim}}(V_{\text{bias}})$. After executing a forward solve of the physical model for a design configuration, its performance is compared against the objective current density, $J_{\text{obj}}(V_{\text{bias}})$.

A metric for optimality of a given design is the cost functional, $C(p)$, where the design parameters are contained in the vector p. A common cost functional is the distance between J_{sim} and J_{obj} in the least squares sense,

$$C(p) = \sum_{j=1}^{\nu} w_j \left(J_{\text{obj}}(V_{\text{bias}}^j) - J_{\text{sim}}^j(V_{\text{bias}}^j, p) \right)^2 , \tag{1.1}$$

where w_j is a weighting factor. This form of the cost functional is chosen because it guarantees a continuous derivative at local minima, enabling the use of gradient-based methods of optimization. Since more optimal designs have lower cost, the optimization is cast as a minimization problem.

To begin the optimal design process, an initial device configuration is chosen either randomly or guided by intuition about an expected result. After evaluating the cost for this configuration, the optimization algorithm chooses the next design to be simulated. The cost for the new design is calculated, and the process is iterated until selected convergence criteria are satisfied.

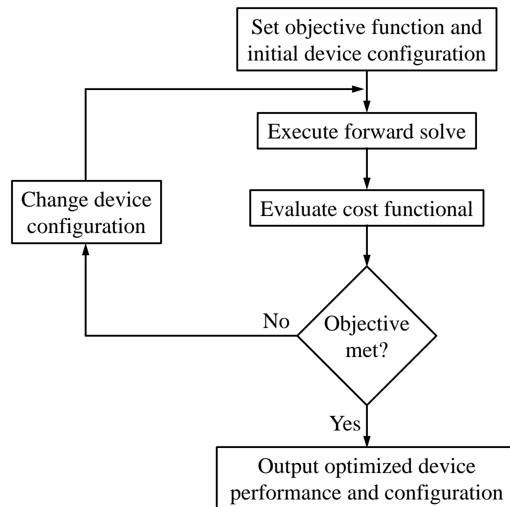

Figure 1.3: Flow chart showing basic optimal design algorithm.

This process flow provides the basic algorithm that defines the optimal design process. Further development of optimal design is possible by allowing the algorithm to adapt to the specific optimization being performed. For example, optimizations that become stuck in high cost local minima could adjust auxiliary parameters of either the objective function or the device configuration. In this way, objective functions and device configurations can be optimized together, yielding natural combinations which could provide more optimal solutions.

1.2 Improved physical models

For optimal design to be effective the physical model must capture the essential physics and accurately predict behavior of a given device configuration. Such a requirement can exclude from optimization device types for which the physical model is insufficient. As an example, consider the tunnel field effect transistor (TFET). The TFET is a promising candidate for replacement of CMOS in low-power applications as scaling limits are reached and Moore's Law comes to an end [31]. The ability for a MOSFET to drive

current at very low gate voltages is fundamentally limited by the high energy Boltzmann tail of the Fermi distribution to a subthreshold swing of $S = k_{\mathrm{B}}T\ln(10) \approx 60$ mV/dec, where T is the absolute temperature.

TFETs overcome this limit by using the band gap to truncate the high energy Boltzmann tail of the thermal distribution. This effectively lowers the operating temperature of the device but requires carriers to tunnel through the band gap. Predictions of transistors capable of subthreshold swings of $S = 15$ mV/dec with ON currents of $I_{\mathrm{ON}} = 1.3$ mA/μm [17] have yet to be achieved experimentally. Although Si TFETs have been shown that exhibit $S = 52.8$ mV/dec and $I_{\mathrm{ON}} = 12$ μA/μm [6], convincing experimental evidence has not been shown of TFETs exhibiting subthreshold swings below 50 mV/dec.. Higher I_{ON} may be achieved by using smaller band gap materials. An $\mathrm{In}_0.7\mathrm{Ga}_0.3$As TFET has been reported with $I_{\mathrm{ON}} = 50$ μA/μm but with subthreshold swing $S = 93$ mV/dec [48].

The contrast between predictions of simulation and experimental results indicate that simple effective mass models used to model TFETs are insufficient. Additional complexities such as complex band structure and reflection due to mismatch of the character of the valence and conduction band wave functions should be incorporated into the model before optimization of these devices is attempted.

Complexities such as these can yield surprising new physics that cannot be captured by simpler models. For example, it has been shown that silicon twist boundaries can yield large reflections for low energy conduction band electrons [41]. This defect in the crystal structure changes only the character of the wave function across the boundary. Absence of a velocity mismatch indicates that an effective mass model is not capable of capturing this effect.

Another example considers a many-body effect by accounting for the Coulomb interaction between a pair of electrons propagating in proximity to one another. In Ref. [34]

it was shown that in the presence of strong Coulomb interaction one electron tunneling through a potential barrier can enhance the probability of another electron tunneling through the same barrier at a later time.

To incorporate many-body effects into the Landauer model of electron transmission [19, 20] non-equilibrium Green's functions (NEGF) [27] have been developed. Although NEGFs present a formally exact solution to electron transport, the many-particle interactions make numerical solutions intractable and usually force one to resort to perturbative methods. This often reduces the solution to a random phase approximation, averaging out many quantum mechanical effects and resulting in a semi-classical approximation often interpreted in terms of tunneling and scattering rates for localized particles [8, 29].

Resorting to perturbative methods can eliminate physical effects associated with the scattering mechanisms being modeled. As an explicit example, in the next section I calculate scattering rates associated with the electron-phonon interaction using first-order perturbation theory and analyze the results in terms of lost physics.

1.2.1 First-order perturbation theory

The rate at which an electron in initial state $|\mathbf{k}\rangle$ with energy $E_{\mathbf{k}}$ scatters into final state $|\mathbf{k}'\rangle$ with energy $E_{\mathbf{k}'}$ due to a perturbing potential $\hat{V}$ is typically calculated using Fermi's golden rule [23],

$$\frac{1}{\tau_{\mathbf{k},\mathbf{k}'}} = \frac{2\pi}{\hbar}|\langle\mathbf{k}'|\hat{V}|\mathbf{k}\rangle|^2 D(E_{\mathbf{k}'})\delta(E_{\mathbf{k}'} - E_{\mathbf{k}} \pm \hbar\omega), \tag{1.2}$$

where $D(E_{\mathbf{k}'})$ is the density of final states and $\pm\hbar\omega$ represents energy loss or energy gain, respectively, during the interaction. For elastic scattering $\hbar\omega = 0$. Fermi's golden rule is a first-order approximation that assumes the initial state is weakly perturbed by

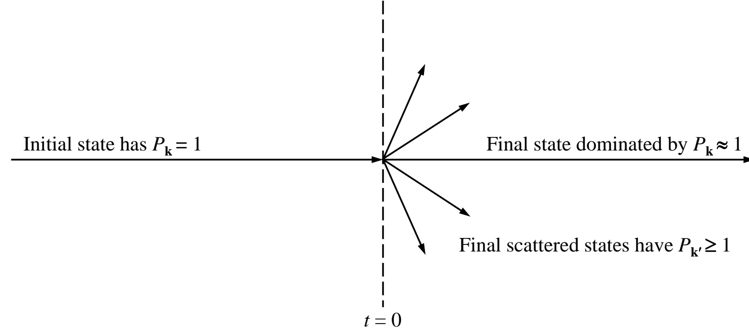

Figure 1.4: Diagram showing first-order approximation for Fermi's golden rule, where the probability of an electron occupying state $|\mathbf{k}\rangle$ is $P_{\mathbf{k}}$. Fermi's golden rule assumes that the electron is weakly perturbed by the interaction, decoupling the scattering states from one another and violating unitarity.

the interaction. This is shown pictorially in Fig. 1.4, where the probability of the electron occupying state $|\mathbf{k}\rangle$ is represented by $P_{\mathbf{k}}$.

In general, each pair of scattering states can have a non-zero matrix element coupling them. However, by assuming the incident electron is weakly perturbed, these matrix elements are dominated by interaction with the incident state, and all other matrix elements are ignored. This removes interference and feedback between scattering states and destroys unitarity.

In polar semiconductors, electron and longitudinal-optic (LO) phonons interact via the Frölich interaction. LO phonons are characterized by displacement of the anion and cation of the lattice in opposite directions from their equilibrium positions, the orientation of which relative to an interacting electron is shown in Fig. 1.5. The wave packet representation of the electron in the Fig. represents the finite coherence length of the electron due to collisions. Displacement of the ions creates an electric dipole moment which enables the electron and lattice to exchange energy and momentum via the Coulomb potential. The electron may excite and, at temperatures above $T = 0$ K, absorb real LO phonons having momentum $\hbar\mathbf{q}$ and energy $\hbar\omega_0$. These processes correspond to momentum exchange $\hbar\mathbf{k}' = \hbar\mathbf{k} \pm \hbar\mathbf{q}$ and energy exchange $E' = E \mp \hbar\omega_{\mathrm{LO}}$,

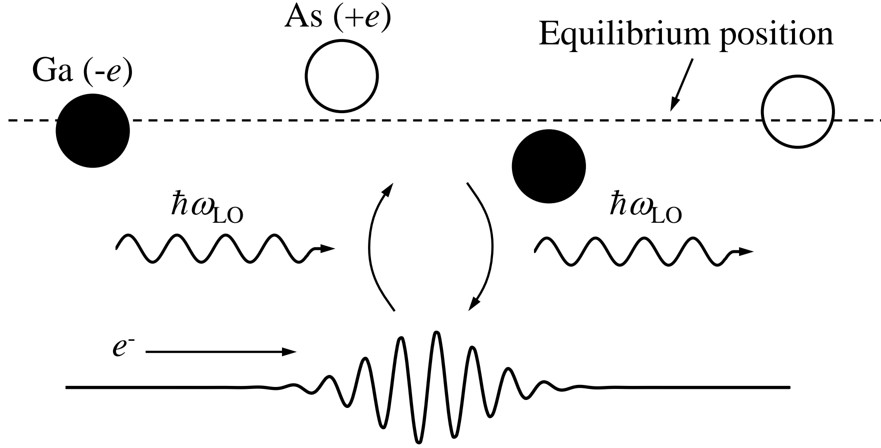

Figure 1.5: Diagram depicting interaction of an electron wave packet and LO phonons having energy $\hbar\omega_{\mathrm{LO}}$. In a polar semiconductor such as GaAs LO phonons displace lattice ions from their equilibrium positions (dashed line) such that electric dipoles are created which enable the lattice and electron to exchange momentum and energy.

where the top sign corresponds to phonon emission and the bottom sign corresponds to phonon absorption.

The perturbing potential is the Coulomb potential of the electric dipole of the ions. The matrix element coupling initial and final states is

$$\left\langle \mathbf{k} \left| \frac{-e^2}{4\pi\epsilon(\mathbf{r})} \right| \mathbf{k}' \right\rangle = \int d^3\mathbf{r} e^{-i\mathbf{k}\cdot\mathbf{r}} \frac{-e^2}{4\pi\epsilon\mathbf{r}} e^{i\mathbf{k}'\cdot\mathbf{r}} = -\frac{e^2}{\epsilon(\mathbf{q},\omega)\mathbf{q}}, \tag{1.3}$$

where the dielectric function is [42]

$$\epsilon(\mathbf{q},\omega) = \epsilon_\infty \left(1 + \frac{\omega_{\mathrm{TO}}^2 - \omega_{\mathrm{LO}}^2}{\omega^2 - \omega_{\mathrm{TO}}^2} \right). \tag{1.4}$$

Here ϵ_∞ is the high frequency permittivity, ω_{LO} is the LO phonon frequency, and ω_{TO} is the transverse optical (TO) phonon frequency. In GaAs the phonon energies are $\hbar\omega_{\mathrm{LO}} = 36$ meV and $\hbar\omega_{\mathrm{TO}} = 33$ meV and the high frequency permittivity is $\epsilon_\infty = 11.1 \times \epsilon_0$, where ϵ_0 is the permittivity of free space.

Substituting (1.3) and (1.4) into (1.2) and integrating over all final states, the scattering rate is [7]

$$\frac{1}{\tau_{\mathrm{LO}}} = \frac{e^2 m\hbar\omega_{\mathrm{LO}}}{2\pi\hbar^2\sqrt{2mE_{\mathbf{k}}}} \left(\frac{1}{\epsilon_\infty} - \frac{1}{\epsilon_{\mathrm{DC}}}\right) \left\{ g(\hbar\omega_{\mathrm{LO}}) \sinh\left(\sqrt{\frac{E_{\mathrm{k}}}{\hbar\omega_{\mathrm{LO}}}}\right) + (g(\hbar\omega_{\mathrm{LO}}) + 1) \sinh\left(\sqrt{\frac{E_{\mathrm{k}} - \hbar\omega_{\mathrm{LO}}}{\hbar\omega_{\mathrm{LO}}}}\right) \right\}, \tag{1.5}$$

where ϵ_{DC} is the low frequency permittivity and in GaAs is equal to $\epsilon_{\mathrm{DC}} = 13.2 \times \epsilon_0$. The left-hand term in the brackets corresponds physically to phonon absorption, and the right-hand term in the brackets corresponds physically to phonon emission.

Figure 1.6 shows the calculated inelastic LO phonon scattering rate at temperatures $T = 0$ K and $T = 300$ K, with the scattering rate for an electron with energy $E_{\mathbf{k}} = 100$ meV indicated. At room temperature, an electron with energy $E_{\mathbf{k}} < \hbar\omega_{\mathrm{LO}}$ may only absorb phonons. Once the electron has enough energy to excite LO phonons, the scattering rate quickly increases. For an electron with energy $E_{\mathbf{k}} = 100$ meV, the scattering rate $\frac{1}{\tau_{\mathrm{LO}}} = 9.1$ ps^{-1} indicates an inelastic mean free path $\lambda_{\mathrm{LO}} = 77.8$ nm.

At temperature $T = 0$ K, and for electron energies $E_{\mathbf{k}} \leq \hbar\omega_{\mathrm{LO}}$, (1.5) indicates that no interaction can occur between the electron and LO phonons. Physically, however, the electron can scatter off of $virtual$ phonons at these energies, and will have a finite scattering rate. Additionally, it is possible for the electron to emit a second phonon after emission of an initial phonon. However, such a process requires the one- and two-emitted phonon inelastic channels to interact. Since these matrix elements have been excluded from the calculation, Fermi's golden rule is not capable of capturing some physics of this interaction.

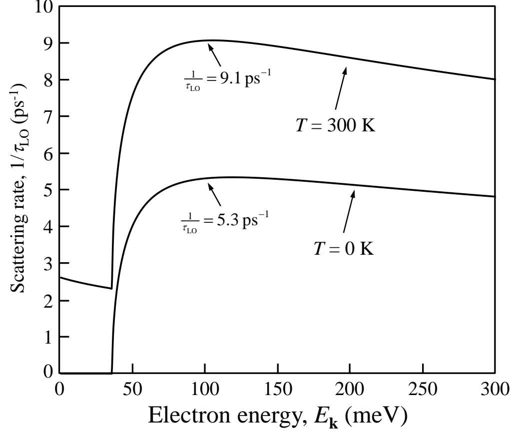

Figure 1.6: Calculated inelastic scattering rate for an electron of energy $E_{\mathbf{k}}$ interacting with LO phonons in GaAs, with scattering rates for an electron with energy $E_{\mathbf{k}} = 100$ meV indicated. At temperature $T = 0$ K, no scattering may occur until the electron has energy $E_{\mathbf{k}} > \hbar\omega_{\mathrm{LO}}$. Simulation parameters are electron effective mass $m = 0.07 \times m_0$, LO phonon energy $\hbar\omega_0 = 36$ meV, TO phonon energy $\hbar\omega_{\mathrm{TO}} = 33$ meV, high frequency permittivity $\epsilon_\infty = 11.1 \times \epsilon_0$, and low frequency permittivity $\epsilon_{\mathrm{DC}} = 13.2 \times \epsilon_0$.

1.2.2 Going beyond perturbation theory

Perturbative approaches ignore unitarity and coherent processes that could exist and be important in small nano-scale devices. Simple first-order perturbative approximations cannot access key features of non-perturbative quantum predictions. The preservation of unitarity in a self-consistent quantum model drives a feedback mechanism between all inelastic channels. Even under the condition of weak coupling, which is commonly invoked to justify perturbation theory, the incident electron must be affected by the existence of inelastic channels and there is no guarantee the predictions of first-order perturbation theory can qualitatively mimic the behavior of a non-perturbative quantum model. A theory capable of properly describing the transition from the coherent quantum transport regime to the incoherent semi-classical regime has yet to be developed. Only by starting from a quantum model can the transition from quantum to semi-classical regimes be discovered, as perturbative methods can eliminate physical effects without the possibility of recovery.

1.3 Organization of thesis

In Chapter 2 the physical model for electron transport in a non-equilibrium, non-interacting regime is presented. Within this thesis I define non-interacting to mean that no interaction occurs between each electron and any other particle in the system. Since the longest structure considered in Chapter 3 has length $L = 25$ nm, $L \ll \lambda_{\mathrm{LO}}$ and a non-equilibrium transport regime may be assumed. Using this model the self-consistent solution to the coupled Schrödinger and Poisson equations for the potential is discussed. However, scattering is required for maintaining contact neutrality as well as forming an accumulation of charge in the accumulation region [11]. Modifications to the model used in calculating the self-consistent potential necessary for creation of these features are shown.

In Chapter 3 the model developed in Chapter 2 is used to design nano-scale heterostructure diodes exhibiting custom current-voltage characteristics. A linear objective is chosen to explore the level of control possible through optimal design. The generally nonlinear current-voltage characteristic of non-equilibrium electron transport indicates that a linear objective should be difficult to achieve and so may provide insight into the physical mechanisms responsible for control. Results of optimizations for multiple device lengths and objective function slopes are compared to determine the physical mechanisms limiting control.

The ability to design devices featuring custom current-voltage characteristics enables co-design of circuits and the devices within the circuits. An example of this design approach is demonstrated in Chapter 3. Here an RF switching mixer is considered. A typical RF switching mixer produces spurious harmonics which must be filtered in the following receiver stage. The current-voltage characteristic of a two-terminal device capable of attenuating spurious harmonics is defined parametrically. A

heterostructure tunnel diode featuring this parametrically defined current-voltage characteristic is designed, and it is shown through simulation that the manufacturable device is capable of enhancing mixer functionality.

Chapter 4 studies a simple, exactly solvable quantum mechanical model of the electron-phonon interaction to discover connections with semi-classical perturbation theory. First-order perturbation theory for the quantum model is derived using the first term in the Born series. Comparison with the exact quantum model shows that in the weak coupling limit perturbation theory is not a good approximation to quantum behavior. In the presence of a strong potential barrier, predictions of first-order perturbation theory can qualitatively emulate predictions of an exactly solved quantum model. It is also shown that situations exist where, even in the presence of weak coupling and strong elastic scattering, perturbation theory may not qualitatively emulate quantum behavior.

Chapter 2

Calculation of the self-consistent potential

Interaction between electrons through the Coulomb interaction may be described by a scalar potential energy surface. The equation governing the shape of this potential, ϕ, in terms of the local electron density, n, is the Poisson equation

$$\nabla \left(\epsilon(\mathbf{r})\nabla\phi(\mathbf{r})\right) = e\left(q(\mathbf{r}) - n(\mathbf{r})\right), \tag{2.1}$$

where $q(\mathbf{r})$ represents additional charge in the system, e is the electron charge, and $\epsilon(\mathbf{r})$ is the spatially-dependent permittivity of the medium. The equation governing the steady-state spatial distribution of an electron in terms of the potential is the time-independent Schrödinger equation

$$-\frac{\hbar^2}{2m}\nabla^2\psi(\mathbf{r}) + \left(U_0(\mathbf{r}) - \phi(\mathbf{r})\right)\psi(\mathbf{r}) = E\psi(\mathbf{r}), \tag{2.2}$$

where m is the effective electron mass, $\psi(\mathbf{r})$ is the electron wave function, $U_0(\mathbf{r})$ is the conduction band edge (on-site) potential, and E is the total energy of the electron. A self-consistent potential is any potential which can simultaneously satisfy both of these equations.

As an example, consider the $Al_\xi Ga_{1-\xi}As$ resonant tunnel diode shown in Fig. 2.1(a), having two GaAs contacts with n-type doping concentrations $N_\mathrm{D} = 1\times10^{18}$ cm^{-3}. The potential barriers have 0.25 eV conduction band offsets, a 2.8 nm thickness, and are

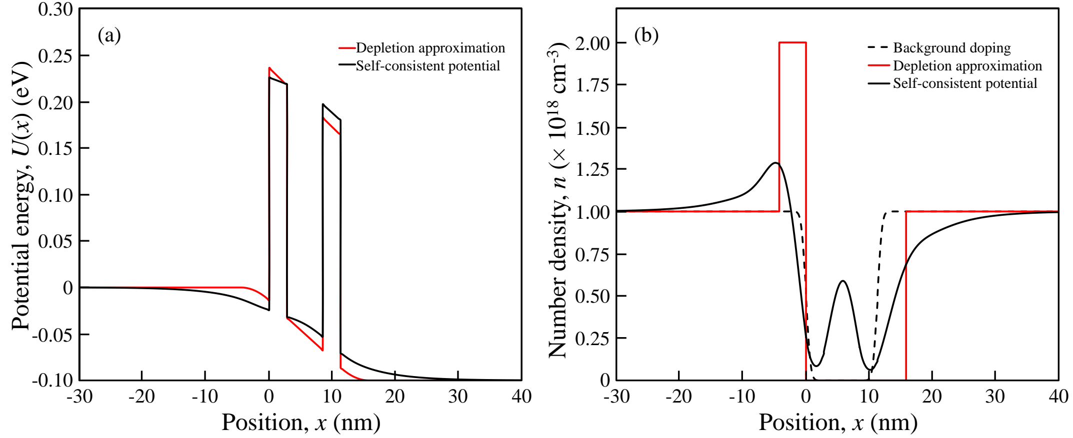

Figure 2.1: (a) Potential profile of $\text{Al}_\xi\text{Ga}_{1-\xi}\text{As}$ resonant tunneling diode calculated both self-consistently and using the depletion approximation. Storage of charge within the structure results in band bending and a significant increase in potential energy. (b) Electron concentration for potentials of part (a) and background doping concentration profile, which has been smoothed using an error function profile. Simulation parameters are electron effective mass $m = 0.07 \times m_0$, temperature $T = 300$ K, n-type doping concentration $N_\text{D} = 1 \times 10^{18}$ cm^{-3}, dopant diffusion parameter $L_\text{d} = 1$ nm, and $V_\text{bias} = 0.1$ V.

separated by a 5 nm thick layer of GaAs. The applied voltage bias $V_\text{bias} = 0.1$ eV has been chosen to align a resonant state with thermally occupied electrons in the emitter (left-hand) contact. Dopant atoms are diffused into the active region using an error function profile characterized by a diffusion parameter $L_\text{d} = 1$ nm, defined as the length on either side of the metallurgical junction over which diffusion extends.

Figure 2.1(a) shows the potential profile for this device calculated both self-consistently and using the depletion approximation. Comparison of the classical depletion approximation and quantum self-consistent potential demonstrates additional physics captured by a quantum treatment. Part (b) of the Fig. shows the charge density corresponding to these two potentials as well as the background doping profile.

In the contacts the wave nature of electrons causes the electron density to become more spatially distributed than predicted by the classical depletion approximation. As

a result, predicted accumulation and depletion region lengths are quite different and the electric field in the device region is more effectively screened.

The resonant state causes a storage of electrons between the potential barriers. These electrons repel one another, and this repulsion is captured by an increased potential energy in the vicinity of the quantum well. The changed potential will in turn modify the charge density. This interdependence of charge density and potential requires that the two quantities be solved self-consistently.

2.1 Prescription for calculating the self-consistent potential

The electron density n and potential ϕ are nonlinearly coupled, and may be calculated by iteratively solving the Schrödinger and Poisson equations. The iterative solution proceeds according to the following algorithm:

1. Calculate initial guess for potential, $\phi^{(0)}$.
2. Solve Schrödinger equation for electron wave functions, $\psi^{(m)}$: $\phi^{(m)} \rightarrow \psi^{(m)}$
3. Calculate current density, $J^{(m)}$: $\psi^{(m)} \rightarrow J^{(m)}$
4. If $\left|\frac{J^{(m)}-J^{(m-1)}}{J^{(m-1)}}\right| < \epsilon_J$, exit algorithm
5. Calculate charge density, $n^{(m)}$: $\psi^{(m)} \rightarrow n^{(m)}$
6. Solve Poisson equation for new potential, $\phi^{(m+1)}$: $n^{(m)} \rightarrow \phi^{(m+1)}$
7. Return to step 2.

In the algorithm, m is the iteration number and w is an under-relaxation parameter constrained to $0 \leq w \leq 1$. This parameter controls the speed of the solver and can

be used to avoid oscillations and instabilities, as will be discussed in section 2.6.5. As current density is the measurable quantity of interest, the convergence criterion is a maximum relative error in current density between iterations of ϵ_J, where typically $\epsilon_J = 10^{-3}$. Higher precision is also possible, but I have found this value to provide a good balance between simulation time and accuracy.

2.2 Solution of the Schrödinger equation

The physical model consists of an effective mass, non-interacting electron transport model with occupation of accumulation region quasi-bound states added phenomenologically [10]. The model is non-interacting in the sense that interactions between each individual electron and any other particle in the system have been turned off. Thus, all relaxation processes are assumed to occur outside of the simulation domain and electron transport occurs in a non-equilibrium regime. Only one-dimensional, n-type majority carrier devices are considered and the cross-sectional area is assumed to be sufficiently large that quantization effects may be ignored. For each $\mathrm{Al}_\xi\mathrm{Ga}_{1-\xi}\mathrm{As}$ layer, the relative permittivity is $\epsilon_\mathrm{r} = 13.2 - 3.1 \times \xi$ [1], the on-site potential is $U_0 = 0.8355 \times \xi$, and for simplicity an effective electron mass $m = 0.07 \times m_0$ is assumed in all layers.

Shown schematically in Fig. 2.2 is injection of charge into a heterostructure device and occupation of accumulation region quasi-bound states. It is assumed that the contacts maintain a thermal electron distribution, from which electrons in scattering states are injected into the simulation domain. Occupation of localized states in the accumulation region is necessary for an accumulation of charge in this region [11]. Since electrons in these states are not fully bound but may tunnel into a scattering state on one side, they are $quasi$-bound. A detailed discussion of the occupation of these states is deferred to section 2.4.3.

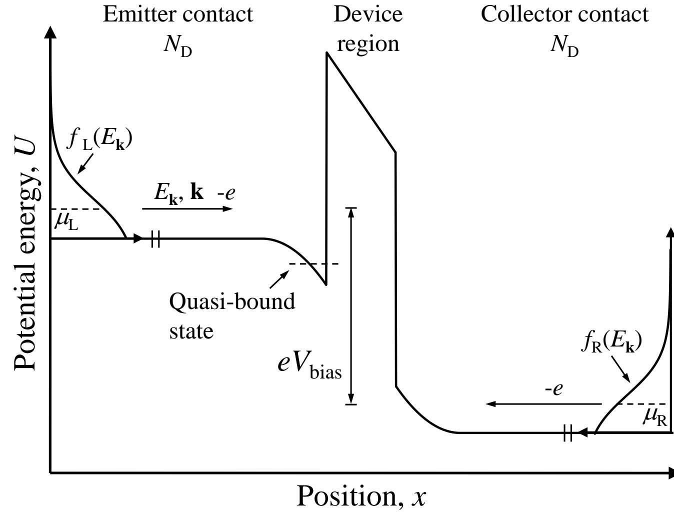

Figure 2.2: Schematic representation of electrons injected into device region from contacts maintaining thermal distributions of electrons in the presence of an applied voltage bias, V_{bias} . Also indicated is occupation of electrons in accumulation region quasi-bound states. The emitter and collector contacts are both n-type doped with doping concentration N_{D} and characterized by chemical potentials μ_{L} and μ_{R}, respectively at temperature T. V_{bias} causes the chemical potentials far away from the device region to differ by $\mu^{\text{R}} = \mu^{\text{L}} - eV_{\text{bias}}$.

Since only layered heterostructure devices are considered, the potential varies in only one dimension. The Hamiltonian is divided into components perpendicular (x-direction, indicated by $\perp$) and parallel (y- and z-directions, indicated by $\|$) to the heterostructure layer interfaces. Accordingly, in a plane wave basis the electron wave function for state $|\mathbf{k}\rangle$ with energy $E_{\mathbf{k}}$ injected from contact α is

$$\psi_{\mathbf{k}}^{\alpha}(x, y, z, t) = e^{i\mathbf{k}_{\|}\cdot(y\hat{y}+z\hat{z})}\psi_{k_{\perp}}^{\alpha}(x)e^{-i\omega t}, \tag{2.3}$$

where

$$E_{\mathbf{k}} = E_{\parallel} + E_{\perp} = \frac{\hbar^2 |\mathbf{k}_{\parallel}|^2}{2m} + \frac{\hbar^2 k_{\perp}^2(x)}{2m} + U(x) = \hbar\omega, \quad (2.4)$$

$$k_{\perp}(x) = \frac{\sqrt{2m(E_{\perp} - U(x))}}{\hbar}, \quad (2.5)$$

$$U(x) = U_0(x) - \phi(x). \quad (2.6)$$

α indicates the contact from which the carrier is injected, with $\alpha = \mathrm{E}$ for the emitter contact or $\alpha = \mathrm{C}$ for the collector contact.

The perpendicular component of the electron wave function is calculated by solving the one-dimensional time-independent Schrödinger equation

$$-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2}\psi_{k_{\perp}}^{\alpha}(x) + U(x)\psi_{k_{\perp}}^{\alpha}(x) = E_{\mathbf{k}}\psi_{k_{\perp}}^{\alpha}(x), \quad (2.7)$$

where boundary conditions appropriate for injection from contact α are used.

2.2.1 Propagation matrix method

The Schrödinger equation is solved using the propagation matrix method [23, 16]. In this method the spatial domain is discretized into $N + 1$ regions and the potential within each region of the domain is assumed constant. For each electron the convention is adopted that region $j = 0$ corresponds to the contact from which it is injected.

Figure 2.3 shows two regions of the simulation domain, with elastic scattering amplitudes a_j, b_j, c_j, and d_j at the step change in potential occurring at position $x = 0$ shown schematically. A plane wave basis is chosen for the electron wave function, yielding wave functions in regions j and $j + 1$ of

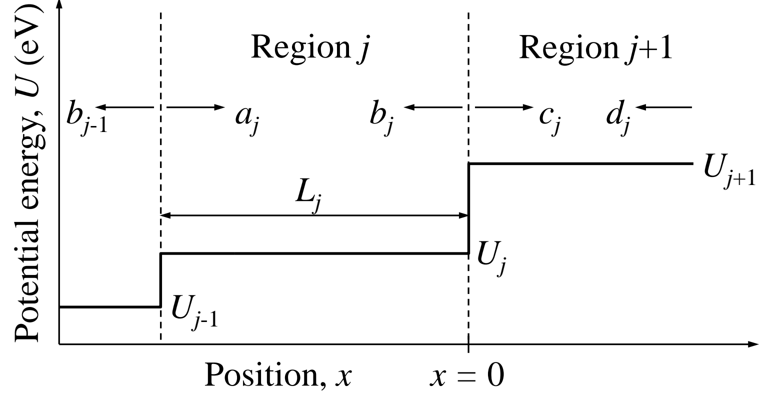

Figure 2.3: Diagram of elastic scattering from a step change in potential occurring at position $x = 0$. The propagation matrix for region j accounts for free electron propagation across length L_j and scattering at the interface between regions j and $j+1$. Quantities a_j, b_j, c_j, and d_j are wave function coefficients associated with region j as shown.

$$\begin{aligned}\psi_j(x) &= a_j e^{ik_j x} + b_j e^{-ik_j x},\\ \psi_{j+1}(x) &= c_j e^{ik_{j+1} x} + d_j e^{-ik_{j+1} x}.\end{aligned} \tag{2.8}$$

The two boundary conditions

$$\begin{aligned}\psi_j(0) &= \psi_{j+1}(0),\\ \left.\frac{\partial \psi_j(x)}{\partial x}\right|_{x=0} &= \left.\frac{\partial \psi_{j+1}(x)}{\partial x}\right|_{x=0},\end{aligned} \tag{2.9}$$

are enforced at $x = 0$.

Applying these conditions to (2.8) yields the matrix equation

$$\begin{bmatrix} a_j \\ b_j \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1+\frac{k_{j+1}}{k_j} & 1-\frac{k_{j+1}}{k_j} \\ 1-\frac{k_{j+1}}{k_j} & 1+\frac{k_{j+1}}{k_j} \end{bmatrix}\begin{bmatrix} c_j \\ d_j \end{bmatrix} = \mathbf{p}_{\mathrm{step},j}\begin{bmatrix} c_j \\ d_j \end{bmatrix}. \tag{2.10}$$

Between interfaces the electron propagates freely across a distance L_j, which is accounted for by the matrix

$$\begin{bmatrix} a_j \\ b_j \end{bmatrix} = \begin{bmatrix} e^{-ik_jL_j} & 0 \\ 0 & e^{ik_jL_j} \end{bmatrix} \begin{bmatrix} c_j \\ d_j \end{bmatrix} = \mathbf{p}_{\text{free},j} \begin{bmatrix} c_j \\ d_j \end{bmatrix}. \tag{2.11}$$

Combining (2.10) and (2.11), the propagation matrix for region j is

$$\mathbf{p}_j = \mathbf{p}_{\text{free},j}\mathbf{p}_{\text{step},j} = \frac{1}{2} \begin{bmatrix} \left(1 + \frac{k_{j+1}}{k_j}\right) e^{-ik_jL_j} & \left(1 - \frac{k_{j+1}}{k_j}\right) e^{ik_jL_j} \\ \left(1 - \frac{k_{j+1}}{k_j}\right) e^{-ik_jL_j} & \left(1 + \frac{k_{j+1}}{k_j}\right) e^{ik_jL_j} \end{bmatrix}. \tag{2.12}$$

Since $a_{j+1} = c_j$ and $b_{j+1} = d_j$, multiplying all propagation matrices together couples the electron flux into and out of the simulation domain by the equation

$$\begin{bmatrix} A \\ B \end{bmatrix} = \prod_{i=N}^{0} \mathbf{p}_i \begin{bmatrix} C \\ D \end{bmatrix} = \mathbf{P} \begin{bmatrix} C \\ D \end{bmatrix}. \tag{2.13}$$

As shown in Fig. 2.4, A represents the injected electron flux, B represents the reflected electron flux, C represents the transmitted flux, and D represents the electron flux that is back-scattered from outside of the simulation domain.

2.2.2 Scattering state wave functions

Physical boundary conditions at the edges of the simulation domain are required to solve (2.13). Physically, an electron of unit flux density is injected ($A = 1$) and it is assumed that there is no back-scattered component ($D = 0$). Using these boundary conditions the matrix $\mathbf{P}$ must be calculated twice. The first calculates C,

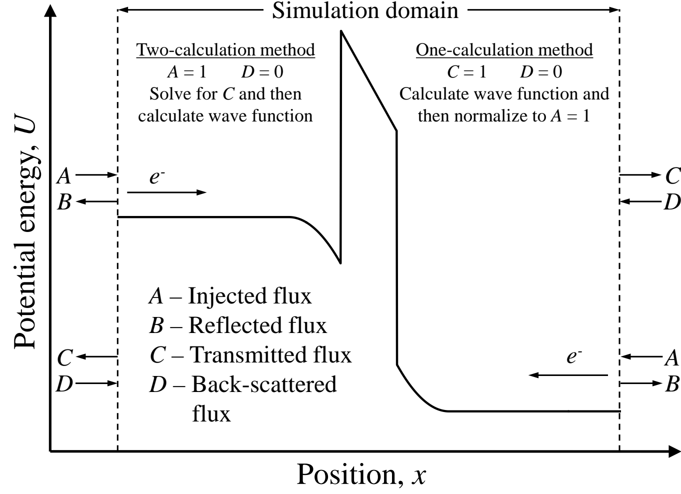

Figure 2.4: Diagram of boundary condition for electrons injected from each of the contacts. Wave function coefficients A and B are associated with the injecting contact, while coefficients C and D are associated with the opposite contact. Boundary conditions for two-calculation and one-calculation solution methods are shown.

$$\begin{bmatrix} 1 \\ B \end{bmatrix} = \begin{bmatrix} P_{1,1} & P_{1,2} \\ P_{2,1} & P_{2,2} \end{bmatrix} \begin{bmatrix} C \\ 0 \end{bmatrix} \quad \Rightarrow C = \frac{1}{P_{1,1}}, \tag{2.14}$$

and the second calculates the electron wave function

$$\begin{bmatrix} a_j \\ b_j \end{bmatrix} = \prod_{i=N}^{j} \mathbf{p}_i \begin{bmatrix} \frac{1}{P_{1,1}} \\ 0 \end{bmatrix}. \tag{2.15}$$

An alternative approach that calculates the electron wave function using only one calculation of $\mathbf{P}$ is to cast (2.13) as a terminal value problem. It is assumed that an electron of unit flux $leaves$ the simulation domain through the contact opposite the injecting contact ($C = 1$) and that no reflection occurs outside of the simulation domain ($D = 0$). Once the wave function coefficients in all regions have been calculated, the physical boundary condition $A = 1$ is enforced by normalizing the wave function to a_0. In this formalism,

$$\begin{bmatrix} a_j \\ b_j \end{bmatrix} = \frac{1}{a_0} \prod_{i=N}^{j} \mathbf{p}_i \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{2.16}$$

Boundary conditions for these two solution methods are shown in Fig. 2.4.

2.2.3 Quasi-bound state wave functions

A quasi-bound state is a resonance within the potential well of the accumulation region. By injecting electrons of unit flux through the barrier and into the well, as shown in Fig. 2.5, the density of states within the well will peak at quasi-bound state resonance energies, E_n^{B}, where n indexes the states. Figure 2.5(a) shows the electron probability density injected from the collector contact into a 5 nm thick rectangular potential barrier with conduction band offset 0.25 eV subject to $V_{\mathrm{bias}} = 0.35$ eV. The quasi-bound state has energy $E^{\mathrm{B}} = -33.7$ meV and its wave function is shown in Fig. 2.5(b).

The wave functions of Fig. 2.5(a) are not those of electrons inelastically scattered into quasi-bound states from the emitter, requiring an additional solve of the Schrödinger equation. Boundary conditions for quasi-bound state wave function are $A = D = 0$, and it is assumed that unity flux is exiting the right-hand side of the system ($B = 1$). These boundary conditions are given in Fig. 2.5(a), which shows the quasi-bound state wave function of part (a) overlayed onto the potential and offset by the quasi-bound state energy.

The quasi-bound state is composed of both scattering and bound states. While normalization for a scattering state is unity flux, normalization for a bound state is unit probability of finding the electron in space. Without a source as reference, quasi-bound state flux exiting the simulation domain cannot be assumed to equal one, and the wave function must be normalized by other means. I assume that the electron is strongly confined to the potential well, so that the probability of finding the electron within this

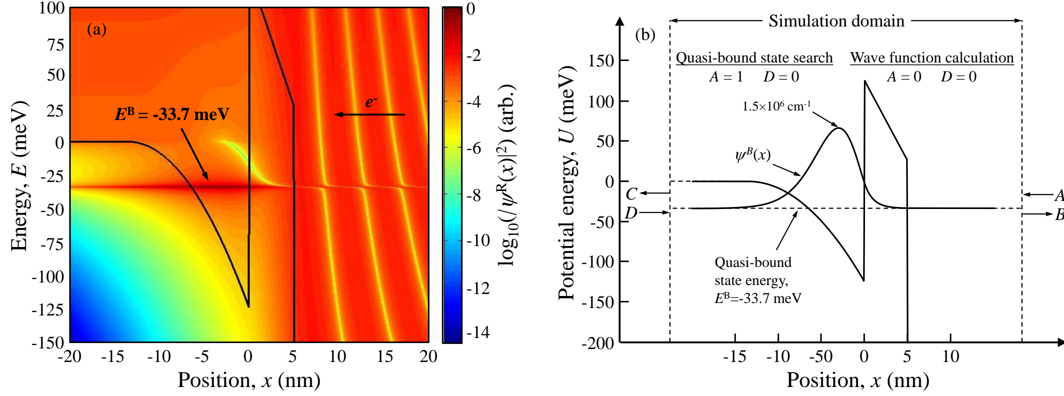

Figure 2.5: (a) Electron probability density injected from collector contact for potential profile containing a rectangular potential barrier having a 0.25 eV conduction band offset and a thickness of 5 nm. The quasi-bound state energy is $E^{\mathrm{B}} = -33.7$ meV. (b) Diagram of electron injection into the accumulation region to find quasi-bound state resonances and calculate quasi-bound state wave functions. Boundary conditions for each calculation are indicated, and the wave function shown is that of the quasi-bound state of part (a) offset in energy by E^{B}. The applied voltage bias is $V_{\mathrm{bias}} = 0.35$ eV, and simulation parameters are electron effective mass $m = 0.07 \times m_0$, temperature $T = 300$ K, n-type doping concentration $N_{\mathrm{D}} = 1 \times 10^{18}$ cm^{-3}, and dopant diffusion parameter $L_{\mathrm{d}} = 1$ nm.

region is close to unity. Letting this probability equal one, the quasi-bound states is normalized such that

$$\int_{\ell} |\psi_n^{\mathrm{B}}(x)|^2 dx = 1, \tag{2.17}$$

where ℓ is the length over which the electron is localized within the well. I have chosen ℓ to extend from the left edge of the domain to the point at which the quasi-bound state wave function exits the confining potential. In Fig. 2.5 this occurs at position $x = 5$ nm and it can be seen that this normalization yields a reasonable one-dimensional density of states in comparison to $N_{\mathrm{D}}^{1/3} = 10^6$ cm^{-1}.

2.3 Calculation of current density

Current density in the x-direction for a single electron in state $|\mathbf{k}\rangle$ is found by applying

$$\mathbf{J}_{x,\mathbf{k}} = -i\frac{e\hbar}{2m}\left(\psi_{\mathbf{k}}\frac{\partial\psi_{\mathbf{k}}^*}{\partial x} - \frac{\partial\psi_{\mathbf{k}}}{\partial x}\psi_{\mathbf{k}}^*\right), \tag{2.18}$$

to the electron wave function, $\psi_{\mathbf{k}}$. In steady state the current density must be constant throughout the device and $\mathbf{J}_{x,\mathbf{k}}$ may be evaluated at any location. However, by evaluating (2.18) in region $j = N$, additional terms in the derivative from position-dependent $\mathbf{k}$ and wave function coefficients are avoided due to the assumed zero electric field. In this region the wave function for an electron in state $|\mathbf{k}\rangle$ injected from contact α is

$$\begin{aligned}\psi_{\mathbf{k}}^{\alpha}(x, y, z, t) =& e^{i\mathbf{k}_{\parallel}\cdot(y\hat{y}+z\hat{z})}c_N^{\alpha}e^{ik_N x}e^{-i\omega t}\\ =& e^{i\mathbf{k}_{\parallel}\cdot(y\hat{y}+z\hat{z})}\frac{1}{a_0^{\alpha}}e^{ik_N x}e^{-i\omega t},\end{aligned} \tag{2.19}$$

Substituting into (2.18),

$$\mathbf{J}_{x,\mathbf{k}}^{\alpha} = e\frac{\hbar k_N}{m}\frac{1}{|a_0^{\alpha}|^2}, \tag{2.20}$$

which is the product of the electron's charge, velocity, and transmission probability. Total current density from contact α is found by summing this eigenvalue over all thermally occupied electron states. Converting this sum to an integral,

$$\mathbf{J}_x^{\alpha} = 2\sum_{\mathbf{k}}\mathbf{J}_{x,\mathbf{k}}^{\alpha}f(E_{\mathbf{k}}) = 2e\int\frac{d^2k_{\parallel}}{(2\pi)^2}\int_0^{\infty}\frac{dk_{\perp}}{2\pi}\frac{\hbar\mathbf{k}_{\perp}}{m}\frac{1}{|a_0^{\alpha}|^2}\frac{1}{1+e^{(E_{\mathbf{k}}-\mu^{\alpha})\beta}}, \tag{2.21}$$

where $\beta = \frac{1}{k_{\mathrm{B}}T}$ is the inverse thermal energy, the factor of two accounts for electron spin degeneracy, and

$$f(E_{\mathbf{k}}) = \frac{1}{1 + e^{(E_{\mathbf{k}} - \mu^{\alpha})\beta}}, \tag{2.22}$$

is the Fermi occupation factor.

Evaluation of (2.21) is simplified by converting to energy via the one- and two-dimensional density of states,

$$\mathbf{J}_x^{\alpha} = \frac{em}{2\pi^2\hbar^3} \int_0^{\infty} dE_{\parallel} \int_0^{\infty} dE_{\perp} \frac{1}{|a_0^{\alpha}|^2} \frac{1}{1 + e^{(E_{\mathbf{k}} - \mu^{\alpha})\beta}}. \tag{2.23}$$

Here the velocity term has been canceled by the one-dimensional density of states. The integral over $E_{\parallel}$ is evaluated analytically,

$$\begin{aligned} \int_0^{\infty} dE_{\parallel} \frac{1}{1 + e^{(E_{\parallel} + E_{\perp} - \mu^{\alpha})\beta}} &= \int_0^{\infty} dE_{\parallel} \frac{e^{(\mu^{\alpha} - E_{\parallel} - E_{\perp})\beta}}{1 + e^{(\mu^{\alpha} - E_{\parallel} - E_{\perp})\beta}} \\ &= \frac{1}{\beta} \ln\left(1 + e^{(\mu^{\alpha} - E_{\perp})\beta}\right), \end{aligned} \tag{2.24}$$

so that

$$\mathbf{J}_x^{\alpha} = \frac{em}{2\pi^2\hbar^3\beta} \int_0^{\infty} dE_{\perp} \frac{1}{|a_0^{\alpha}|^2} \ln\left(1 + e^{(\mu^{\alpha} - E_{\perp})\beta}\right). \tag{2.25}$$

Total current is the difference between that of each contact. From time-reversal symmetry, $\frac{1}{a_0^{\mathrm{E}}} = \frac{1}{a_0^{\mathrm{C}}} = \frac{1}{a_0}$, and the applied voltage bias will separate the chemical potentials of the two contacts by an energy eV_{bias}. Defining current flow from emitter to collector as positive current and dropping the $\perp$ notation,

$$\mathbf{J}_x = \mathbf{J}_x^{\mathrm{E}} - \mathbf{C}_x^{\mathrm{R}} = \frac{em}{2\pi^2\hbar^3\beta} \int_0^{\infty} dE \frac{1}{|a_0|^2} \ln\left(\frac{1 + e^{(\mu^{\mathrm{L}} - E)\beta}}{1 + e^{(\mu^{\mathrm{L}} - E - eV_{\mathrm{bias}})\beta}}\right). \tag{2.26}$$

2.4 Calculation of the electron density

The total electron density consists of three components,

$$n(x) = n^{\mathrm{E}}(x) + n^{\mathrm{C}}(x) + n^{\mathrm{B}}(x). \tag{2.27}$$

The first two terms represent electrons injected from the emitter and collector contacts, respectively, and the final term represents the density of electrons contained within quasi-bound states. Before discussing the calculation of these quantities, a deficiency in the non-interacting model must be corrected.

2.4.1 Charge neutrality in the contacts

Consider a region deep within the bulk of a contact, where it is assumed that the contact is neutral and the electric field is approximately zero. With this electric field, the Fermi sphere is centered on $\mathbf{k} = 0$. Half of the electrons have $k_{\perp} > 0$ and half have $k_{\perp} < 0$.

Now suppose that this region is divided in two and a potential barrier of arbitrary strength is placed between the now separated contacts. With $V_{\mathrm{bias}} = 0$ V, spatial symmetry of the transmission and reflection coefficients force contact charge neutrality. However, once symmetry is lost by finite V_{bias} , asymmetry in the transmission and reflection coefficients will eliminate charge neutrality. Of the two contacts, one will have an excess of electrons and the other will be deficient in electrons.

If the potential barrier is strong, nearly all injected electrons will be reflected by the potential barrier and forward and backward traveling waves sum to a charge density of $n \approx N_{\mathrm{D}}$. This is shown in Fig. 2.6(a), where the strong potential barrier features a transmission coefficient $T(E_{\mathrm{F}} = 52\,\mathrm{meV}) = 3.1 \times 10^{-9}$, where E_{F} is the Fermi energy. The electron deficiency in the collector contact is due to chemical potential convergence criterion $|n - N_{\mathrm{D}}| \leq 10^{-4} \times N_{\mathrm{D}}$ being greater than the electron density transmitted from

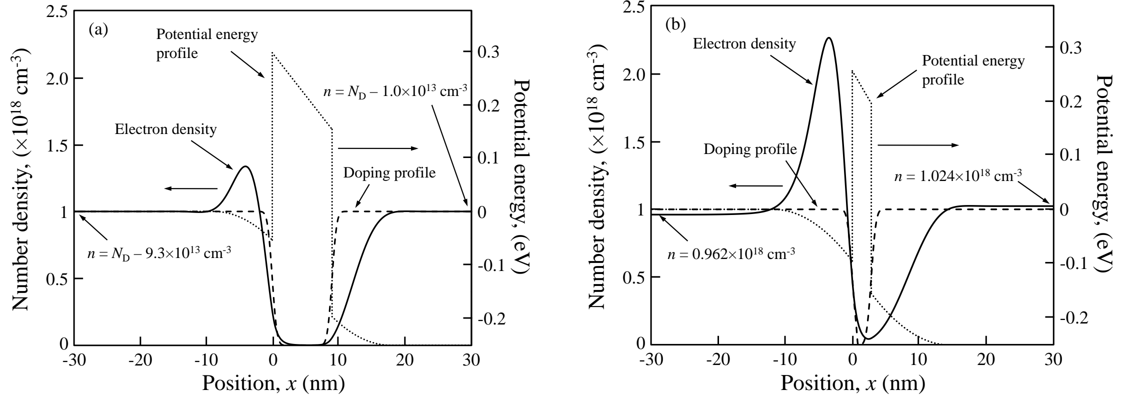

Figure 2.6: (a) Charge density calculated without using drifted Fermi distributions for the case of a strong ($T(E_{\mathrm{F}} = 52\ \mathrm{meV}) = 3.1 \times 10^{-9}$) potential barrier having a 0.35 eV conduction band offset and 9 nm thickness subject to an applied voltage bias $V_{\mathrm{bias}} = 0.25$ V. (b) Same as (a) but for a weaker ($T(E_{\mathrm{F}} = 52\,\mathrm{meV}) = 3.7\times10^{-3}$) potential barrier 3 nm in thickness. The potentials have been calculated using the depletion approximation, and simulation parameters are electron effective mass $m = 0.07 \times m_0$, temperature $T = 300$ K, n-type impurity concentration $N_{\mathrm{D}} = 1\times10^{18}$ cm^{-3}, and dopant diffusion parameter $L_{\mathrm{d}} = 1$ nm.

the emitter contact. This small error in calculated electron density will not significantly impact the self-consistent potential.

A weaker potential barrier, however, may feature a more pronounced charge imbalance. The potential barrier in Fig. 2.6(b) features transmission probability $T(E_{\mathrm{F}} = 52\ \mathrm{meV}) = 3.7 \times 10^{-3}$, causing greater charge imbalance which can lead to undesirable band bending near the contacts. The potentials in the Fig. have been calculated in the first iteration of the solver, where the depletion approximation is used and this band bending has not yet been calculated.

Figure 2.6(b) results from a transport model with no inelastic or diffusive elastic scattering, preventing the distribution of electrons injected from the contacts from responding to a charge imbalance. A model that includes these scattering effects will develop linear voltage drops within the contacts, causing injection of non-equilibrium electron

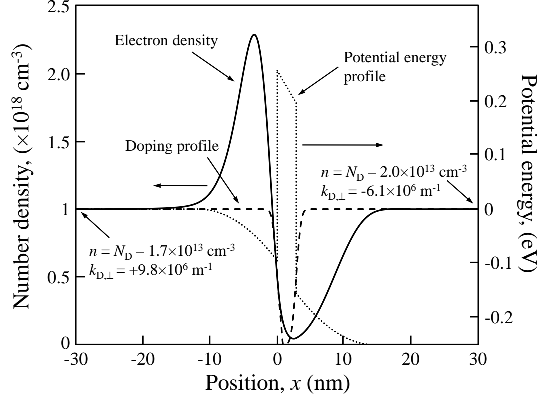

Figure 2.7: Calculated electron density for potential profile of Fig. 2.6(b) with drifted Fermi distributions used to neutralize the contacts. Calculated values of $k^{\alpha}_{\mathrm{D},\perp}$ are indicated in the Fig., and the potential barrier has a 0.35 eV conduction band offset and is 3 nm in thickness. Simulation parameters are electron effective mass $m = 0.07 \times m_0$, temperature $T = 300$ K, n-type doping concentration $N_{\mathrm{D}} = 10^{18}$ cm^{-3}, dopant diffusion parameter $L_{\mathrm{d}} = 1$ nm, and $V_{\mathrm{bias}} = 0.25$ V.

distributions. Semi-classically, this effect is captured in the Boltzmann transport equation, which predicts a shift of the Fermi sphere in the presence of scattering and an electric field. However, without a model that accounts for scattering within the contacts, an additional parameter must be introduced which can emulate the effects of scattering on injected electron distributions.

I have chosen to use a drifted Fermi distribution [36], in which the electron's kinetic energy within the Fermi-Dirac distribution is modified to

$$E^{\alpha}_{\mathrm{D},\perp} = \frac{\hbar^2 (k_\perp - k^{\alpha}_{\mathrm{D},\perp})^2}{2m} = E_\perp + \frac{\hbar^2}{2m} k^{\alpha}_{\mathrm{D},\perp} (k^{\alpha}_{\mathrm{D},\perp} - 2k_\perp), \tag{2.28}$$

where $k^{\alpha}_{\mathrm{D},\perp}$ is the drifted wave vector in the x-direction for contact α. This drifted wave vector shifts the Fermi sphere to an average electron velocity $\frac{\hbar k^{\alpha}_{\mathrm{D},\perp}}{m}$. Each contact will have separate $k^{\alpha}_{\mathrm{D},\perp}$, which will allow the contact to either inject additional electrons with $k^{\alpha}_{\mathrm{D},\perp} > 0$, or inject less electrons with $k^{\alpha}_{\mathrm{D},\perp} < 0$. These two quantities are calculated self-consistently to neutralize both contacts.

In Fig. 2.7 the simulation of Fig. 2.6(b) has been repeated with a drifted Fermi distribution. It can be seen that the contacts are now neutralized up to the $k^{\alpha}_{\mathrm{D},\perp}$ convergence criterion $|n - N_{\mathrm{D}}| \leq 10^{-4} \times N_{\mathrm{D}}$. The calculated values of $k^{\alpha}_{\mathrm{D},\perp}$ are shown in the Fig.

2.4.2 Scattering state electron density

The electron density for scattering state $|\mathbf{k}\rangle$ is given by the probability density

$$n^{\alpha}_{\mathbf{k}}(x) = |\psi^{\alpha}_{k_\perp}(x)|^2. \tag{2.29}$$

Summing over all thermally occupied states and converting to an integral,

$$n^{\alpha}(x) = 2\sum_{\mathbf{k}} n^{\alpha}_{\mathbf{k}}(x) f(E_{\mathbf{k}}) = 2\int_{-\infty}^{\infty} \frac{dk_\perp}{2\pi} \int \frac{d^2k_\parallel}{(2\pi)^2} n^{\alpha}_{\mathbf{k}}(x) f(E_{\mathbf{k}}). \tag{2.30}$$

Unlike the similar calculation for current density, there is no velocity term to cancel the $\frac{1}{\sqrt{E_\perp}}$ dependence of the one-dimensional density of states. To avoid a singularity in the integrand, the $k_\perp$ integral is not converted to energy. This results in an electron density injected from contact α of

$$n^{\alpha}(x) = \frac{m}{\pi\hbar^2\beta} \int_{-\infty}^{\infty} dk_\perp |\psi^{\alpha}_{k_\perp}(x)|^2 \ln\left(1 + e^{(\mu^{\alpha} - E^{\alpha}_{\mathrm{D},\perp})\beta}\right). \tag{2.31}$$

2.4.3 Quasi-bound state electron density

The probability of scattering state electrons occupying space in the accumulation region is reduced due to an increased kinetic energy. To develop the accumulation of charge necessary to form an electric dipole across the device region, significant charge must be stored in accumulation region quasi-bound states [11]. Accurate modeling of transients

associated with relaxation into these states requires a kinetic electron transport model. The physical model described in this chapter however simulates steady-state response of the electrons to an applied voltage bias. Occupation of accumulation region quasi-bound states in this model is intended to emulate relaxation of injected electrons into these states.

Normalization of quasi-bound state wave functions shown in section 2.2.3 yields the one-dimensional density of states in the $k_\perp$-direction. The two-dimensional density of states in the $\mathbf{k}_\parallel$-directions must be found. Figure 2.8 depicts the processes related to scattering in and tunneling out of quasi-bound states.

Electrons injected from the emitter contact may inelastically scatter into quasi-bound states in the accumulation region with scattering rate $\frac{1}{\tau_{\text{in}}}$. Electrons within quasi-bound states may escape by tunneling through the potential barrier at a rate $\frac{1}{\tau_{\text{tun}}}$. Assuming that the potential barrier sufficiently confines electrons in the potential well such that $\frac{1}{\tau_{\text{in}}} \gg \frac{1}{\tau_{\text{tun}}}$, the quasi-bound states become electron reservoirs which fill to thermal equilibrium with the emitter contact, characterized by chemical potential μ^{L}.

With this assumption, the quasi-bound state electron density is

$$n^{\text{B}}(x) = 2\sum_{\mathbf{k}} |\psi^{\text{B}}_{\mathbf{k}}(x)|^2 f(E_{\mathbf{k}}) = \frac{m}{\pi\hbar^2\beta}\sum_{n} |\psi^{\text{B}}_{n}|^2 \ln\left(1 + e^{(\mu^{\text{L}} - E^{\text{B}}_{n})\beta}\right). \tag{2.32}$$

As an example, consider the quasi-bound state shown in Fig. 2.5. This state features a maximum one-dimensional density of states of 1.5×10^6 cm^{-1} and an energy $E^{\text{B}} = -33.7$ meV. Substituting these values into (2.32), the maximum quasi-bound state electron density in the depletion approximation is $n^{\text{B}}_{\text{max}} = 3.3 \times 10^{18}$ cm^{-3}, and in the self-consistent potential $n^{\text{B}}_{\text{max}} = 2.3 \times 10^{18}$ cm^{-3}. These values are similar to the prediction of the depletion approximation, $n^{\text{B}}_{\text{max}} = 2 \times N_{\text{D}}$.

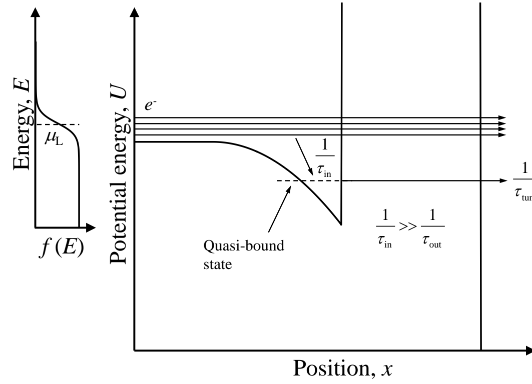

Figure 2.8: Diagram depicting scattering state electrons inelastically scattering into a quasi-bound state in the accumulation region, with this scattering characterized by a scattering rate $\frac{1}{\tau_{\text{in}}}$. Electrons within the quasi-bound state may escape the bound state by tunneling through the potential barrier at a rate characterized by $\frac{1}{\tau_{\text{tun}}}$. Assuming $\frac{1}{\tau_{\text{in}}} \gg \frac{1}{\tau_{\text{tun}}}$, the quasi-bound state becomes an electron reservoir which will be in thermal equilibrium with the left-hand contact and characterized by chemical potential μ^{L}.

2.5 Solution of Poisson equation

The Poisson equation relates the net local charge density to the local curvature of the potential. Allowing for a spatially varying permittivity, the one-dimensional Poisson equation is

$$\frac{\partial}{\partial x}\left(\epsilon(x)\frac{\partial \phi(x)}{\partial x}\right) = \rho(x) = e(N_{\text{D}}(x) - n(x)), \tag{2.33}$$

where ϵ is the permittivity and ρ is the net charge density. Care must be taken when discretizing this equation, as potential and charge density are defined at points while permittivity and electric field are defined in *regions* between these points. Fig. 2.9 shows how I have defined these quantities.

Each region is defined in one dimension as a point and the region to its right before the next point. The local potential and charge density are defined at this point, while the

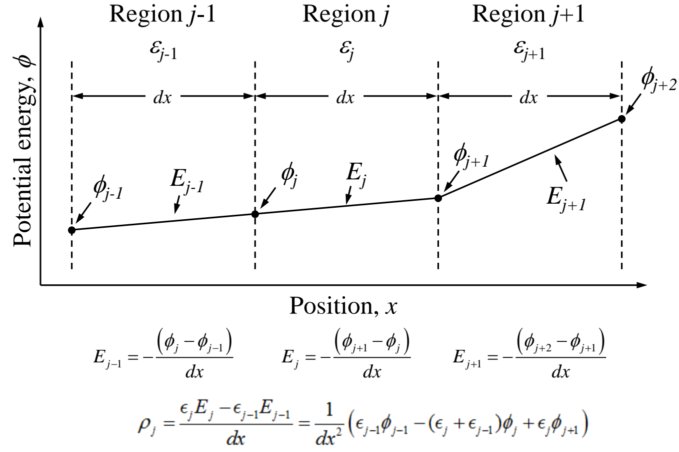

Figure 2.9: Diagram showing discretization scheme and derivation for the discretized Poisson equation. Potential ϕ and charge density ρ are defined at points between regions. The regions are separated by a length dx and are characterized by permittivity ϵ and electric field E.

permittivity and electric field E are defined between these points. Using the physical relationships $E = -\frac{\partial \phi}{\partial x}$ and $\rho = \frac{\partial}{\partial x}(\epsilon E)$, the discretized Poisson equation is

$$\frac{1}{dx^2}\left(\epsilon_{j-1}\phi_{j-1} - (\epsilon_{j-1} + \epsilon_j)\phi_j + \epsilon_j \phi_{j+1}\right) = e(N_{\mathrm{D},j} - n_j), \tag{2.34}$$

where dx is the uniform spatial discretization length. What remains is to determine the boundary conditions and the method of solution, both of which require consideration of the net charge within the simulation domain.

It is desired that the self-consistent potential both drop the entire potential energy eV_{bias} within the simulation domain and have zero electric field at the contacts. For the derivatives to match at the contacts, net charge within the simulation domain must sum to zero. However, with charge density completely specified by chemical potential μ^α, drifted wave vector $k^\alpha_{\mathrm{D},\perp}$, and injection of unit electron flux, there are no additional parameters with which charge neutrality of the device as a whole may be enforced. Thus, matrix inversion may not be used to solve the Poisson equation, as excess charge

in the simulation domain will result in a parabolic background potential over which the desired features are superimposed [12].

A gradient based method will produce better results by weighting changes in the potential to the sensitivity of the charge density to these changes. I have chosen to solve the Poisson equation using the Newton-Raphson method with Dirichlet boundary conditions, which yields self-consistent potentials exhibiting the desired characteristics and qualitatively similar to the depletion approximation.

2.5.1 Solution method

Moving all terms of (2.34) to one side,

$$\frac{1}{dx^2}\left(\epsilon_{j-1}\phi_{j-1} - (\epsilon_{j-1}+\epsilon_j)\phi_j + \epsilon_j\phi_{j+1}\right) - e(N_{\mathrm{D},j} - n_j) = 0, \qquad (2.35)$$

for which the zero may be found using the Newton-Raphson method. Letting F_j equal the left-hand side of (2.35), the equation to be solved is [18]

$$\sum_i \frac{\partial F_j^{(m)}}{\partial \phi_i}\delta\phi_i^{(m+1)} = -F_j^{(m)}, \qquad (2.36)$$

and the updated potential is $\phi^{(m+1)} = (1-w)\phi^{(m)} + w\delta\phi^{(m+1)}$, where w is an under-relaxation parameter.

The Jacobian, $\frac{\partial F_j}{\partial \phi_i}$, includes the term $\frac{n_j}{\partial \phi_i}$, corresponding to the sensitivity of the charge density to changes in potential. For the scattering state charge densities, it is

$$\frac{\partial n_j^\alpha(x)}{\partial \phi_i} = \frac{m}{\pi\hbar^2\beta}\int_{-\infty}^{\infty} dk_\perp \frac{\partial}{\partial \phi_i}\left(|\psi_{k_\perp,j}^\alpha(x)|^2 \ln\left(1 + e^{(\mu^\alpha - E_{\mathrm{D},\perp}^\alpha)\beta}\right)\right). \qquad (2.37)$$

This derivative will produce two terms, and I will first consider $\frac{\partial|\psi|^2}{\partial\phi}$. Given the non-local nature of this term and the absence of a closed form expression for the dependence of $|\psi|^2$ on ϕ, calculation of this derivative can be computationally expensive, requiring a method such as finite differences. However, from the wave nature of the electron one might expect interference effects to smooth changes to the total charge density resulting from local perturbations in potential.

To get a feel for the magnitude of this derivative, an example is provided in Fig. 2.10. Here the change in probability density due to a perturbation $\Delta\phi = 1$ meV at position $x = 50$ nm is shown on a log scale. In the Fig. a prime is used to indicate quantities calculated in the presence of the perturbation. Maximum change in $|\psi|^2$ on the order of 0.1 is a only few percent of the maximum value of $|\psi|^2 = 4$ (from constructive interference of two unity flux waves). The inset shows change in charge density due to the perturbation. When the perturbed states are summed to calculate total charge density, deconstructive interference occurs away from the perturbation. Maximum change in electron density for this perturbation is three orders of magnitude below $N_{\mathrm{D}} = 1 \times 10^{18}$ cm^{-3} and is localized around the perturbation. Because of the small magnitude of this derivative and the computational cost associated with its numerical evaluation, this derivative is ignored and set to zero.

The remaining term is computed by converting the integral of (2.37) to energy,

$$\frac{\partial n^\alpha(x)}{\partial\phi} = \frac{1}{8\pi^2\beta}\left(\frac{2m}{\hbar^3}\right)^{3/2}\int_0^\infty dE_\perp|\psi^\alpha_{k_\perp}(x)|^2\frac{\partial}{\partial\phi}\left(\frac{1}{\sqrt{E^\alpha_\perp-\phi}}\ln\left(1+e^{(\mu^\alpha-E^\alpha_{\mathrm{D},\perp})\beta}\right)\right). \tag{2.38}$$

Changing the integration variable to $E' = E_\perp - \phi$ and letting $E'_{\mathrm{D}} = \frac{\hbar^2 k_{\mathrm{D}}}{2m}\left(k^\alpha_{\mathrm{D},\perp} - 2k_\perp\right)$,

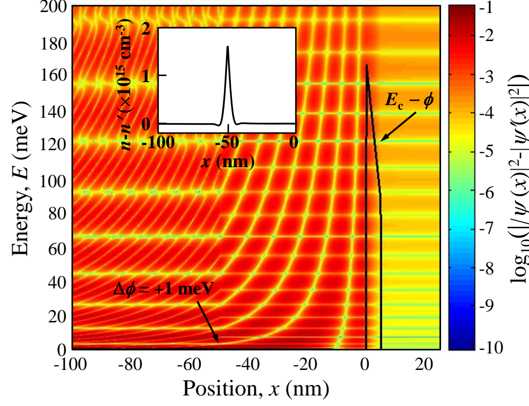

Figure 2.10: Difference in electron probability density injected from left-hand contact between two potentials that differ by a $\Delta\phi = +1$ meV perturbation at position $x = 50$ nm. The inset shows the effect of $\Delta\phi$ in the calculated charge densities. The primes indicate quantities calculated with the perturbation. Simulation parameters are electron effective mass $m = 0.07 \times m_0$, temperature $T = 300$ K, n-type impurity concentration $N_\mathrm{D} = 1 \times 10^{18}$ cm^{-3}, and dopant diffusion parameter $L_\mathrm{d} = 1$ nm.

$$
\begin{aligned}
\frac{\partial n^\alpha(x)}{\partial\phi} =& \frac{1}{8\pi^2\beta}\left(\frac{2m}{\hbar^3}\right)^{3/2}\int_0^\infty dE'|\psi^\alpha_{k_\perp}(x)|^2\frac{1}{\sqrt{E'}}\frac{\partial}{\partial\phi}\left(\ln\left(1+e^{(\mu^\alpha-E'+\phi-E'_\mathrm{D})\beta}\right)\right)\\
=& \frac{1}{8\pi^2\beta}\left(\frac{2m}{\hbar^3}\right)^{3/2}\int_0^\infty dE'|\psi^\alpha_{k_\perp}(x)|^2\frac{1}{\sqrt{E'}}\frac{\beta e^{(\mu^\alpha-E'+\phi-E'_\mathrm{D})\beta}}{1+e^{(\mu^\alpha-E'+\phi-E'_\mathrm{D})\beta}}\\
=& \frac{m}{\pi\hbar^2}\int_{-\infty}^\infty dk_\perp\frac{|\psi^\alpha_{k_\perp}(x)|^2}{1+e^{(E^\alpha_{\mathrm{D},\perp}-\mu^\alpha)\beta}},
\end{aligned}
\tag{2.39}
$$

where in the last step the integral was converted back into k-space after changing the integration variable back to $E_\perp$. Similarly for the quasi-bound states,

$$
\frac{\partial n^\mathrm{B}(x)}{\partial\phi} = \frac{m}{\pi\hbar^2}\sum_n\frac{|\psi^\mathrm{B}_n(x)|^2}{1+e^{(\mu^\mathrm{L}-E^\mathrm{B}_n)\beta}}. \tag{2.40}
$$

The wave functions in (2.39) and (2.40) are valid only for the potential $\phi^{(m)}$. Once the potential has been updated it is necessary to solve the Schrödinger equation for the new electron wave functions before again solving the Poisson equation. Additionally,

as the derivatives have been approximated strong under-relaxation $(w \leq 0.1)$ is necessary to avoid oscillations and divergent behavior. Finding the self-consistent potential with this algorithm can require on the order of one hundred iterations, which is unacceptable in an optimization procedure for which thousands of self-consistent potentials must be calculated. Fortunately, a scheme has been developed that dramatically speeds convergence to the self-consistent potential.

2.5.2 Predictor-corrector scheme

The predictor-corrector scheme was developed by Trellakis et. al. [44] to address the significant computational cost of the under-relaxation method. In this scheme, first-order perturbation theory is used to find an approximate closed form expression for the dependence of the charge density on the potential. Using this expression, the Schrödinger and Poisson equations become partially decoupled by predicting the new charge density without an additional solve of the Schrödinger equation. This allows for multiple solves of the Poisson equation within an inner loop up to a relative change in potential less than $\epsilon_{\rm P}$, and a full step to be taken at the end of each iteration of the outer loop. Once $\phi^{(m+1)}$ has been found, the Schrödinger equation is solved to correct the predicted n, allowing for the next iterative solution of the Poisson equation. Using this scheme the solver rapidly converges to the self-consistent potential, often requiring only two to five iterations, depending on desired accuracy.

For a full derivation of the following equations the reader is referred to Ref. [44], as I will only repeat their final results here. Using first-order perturbation theory the approximate expression for the charge density injected from the contacts is

$$n^{\alpha}(x,\phi) = \frac{m}{\pi\hbar^2\beta}\int_0^{\infty} dk_{\perp}|\psi^{\alpha}_{k_{\perp}}(x)|^2 \ln\left(1 + e^{(\mu^{\alpha} - E^{\alpha}_{\mathrm{D},\perp} - (\phi^{(\ell)}(x) - \phi^{(0)}(x)))\beta}\right), \quad (2.41)$$

$$n^{\mathrm{B}}(x,\phi) = \frac{m}{\pi\hbar^2\beta}\sum_n |\psi^{\mathrm{B}}_n(x)|^2 \ln\left(1 + e^{(\mu^{\mathrm{L}} - E^{\mathrm{B}}_n - (\phi^{(\ell)}(x) - \phi^{(0)}(x)))\beta}\right), \quad (2.42)$$

and for the Jacobian,

$$\frac{\partial n^{\alpha}(x,\phi)}{\partial \phi} = \frac{m}{\pi\hbar^2}\int_0^{\infty} dk_{\perp}\frac{|\psi^{\alpha}_{k_{\perp}}(x)|^2}{1 + e^{(E^{\alpha}_{\mathrm{D},\perp} - \mu^{\alpha} - (\phi^{(\ell)}(x) - \phi^{(0)}(x)))\beta}}, \quad (2.43)$$

$$\frac{\partial n^{\mathrm{B}}(x,\phi)}{\partial \phi} = \frac{m}{\pi\hbar^2}\sum_n \frac{|\psi^{\mathrm{B}}_n(x)|^2}{1 + e^{(E^{\mathrm{B}}_n - \mu^{\mathrm{L}} - (\phi^{(\ell)}(x) - \phi^{(0)}(x)))\beta}}, \quad (2.44)$$

where ℓ indexes the inner loop.

2.6 Measures taken to enhance stability and accuracy

In the following subsections I will discuss additional features intended to improve both the stability of the solver and accuracy of, ultimately, the calculated current density flowing through the self-consistent potential. At the end of this section flow charts showing algorithms for the self-consistent potential solver and its driver are shown Fig. 2.11 and Fig. 2.12, respectively.

2.6.1 Sufficient sampling of electron transmission resonances

In devices featuring resonances with small line-widths relative to the spacing of the sampled values of $k_{\perp}$, it is critical that the full width of the resonance be sampled. Otherwise, the significant charge density stored in the resonance depends on only a few sample points. As the self-consistent potential reacts to this stored charge and shifts the

energy of the resonance, large variations in stored charge are possible and oscillations likely.

To prevent this I have added the capability of dynamically altering the sampled values of $k_\perp$ to respond to the presence of strong resonances. Each iteration the transmission spectrum is analyzed for peaks. If less than ten $k_\perp$ samples fall within the full-width-at-half-maximum (Γ) of the peak, 60 additional samples, linearly spaced in k-space, are inserted into the $k_\perp$ vector between $\pm 3 \times \frac{\Gamma}{2}$.

2.6.2 Energy vector for current density calculation

If the dynamically altered $k_\perp$ vector is used for current density calculations, inconsistencies may occur. Additionally, it must be ensured that the spacing of the $E_\perp$ samples used for the current density calculation be smaller than that of V_{bias} . Otherwise, changes in V_{bias} may not open additional energy samples for current flow, resulting in a step-like current-voltage characteristic with steps occurring each time a new energy sample is opened for current flow. Therefore the current density calculation uses a separate, static energy vector from that used for the calculation of the charge density.

2.6.3 Accuracy of quasi-bound state energies

The quasi-bound states contain the majority of charge responsible for band bending, and therefore their spatial distribution is a dominant factor in the shape of the self-consistent potential. If a change in V_{bias} is on the order of the energy sample spacing, it is possible for the quasi-bound state to become stuck at a single energy sample for multiple values of V_{bias} . As a consequence, ripple-like features occur in the current-voltage characteristic. The current varies smoothly as a function of V_{bias} for several samples until the quasi-bound state shifts to a new energy sample, causing a discontinuity in the current-voltage characteristic.

The primary concern in addressing this issue is speed of the algorithm. While a fine energy spacing over the extent of the accumulation region will suffice, often only a single quasi-bound state exists within the range of V_{bias} considered in Chapter 3. Therefore I have separated the quasi-bound state search into two parts. The first is a coarse scale search over the entire energy range of the accumulation region. The second is a fine scale search about each of the bound states found in the first search.

2.6.4 Spatial requirements for slowly decaying quasi-bound states

While it is desirable to use the smallest domain size possible, quasi-bound states having energies near zero eV may feature spatial decay constants in the hundreds of nanometers. If the charge density of a quasi-bound state is too large at the edge of the simulation domain, physical requirements for zero electric field cannot be enforced. Additional space is required to allow the quasi-bound state charge density to decay to a value sufficiently small compared to N_{D}. I have chosen $10^{-4} \times N_{\mathrm{D}}$ as this threshold. When the quasi-bound state charge density at the contact exceeds this threshold, the additional length required meet this condition is estimated, and the self-consistent solver restarts with the additional space added to the left-hand side of the domain. A maximum of 400 nm of additional space is allowed. If additional space is required an error is generated and the self-consistent solver returns to the driver.

2.6.5 Dynamic slowing of the solver

Success of Newton's method depends critically upon step size, w. Choosing a value for w must weigh trade-offs between simulation time and stability. The predictor-corrector scheme is excellent in this regard, remaining stable under most conditions with $w = 1$. However, the self-consistent potential can diverge if any assumptions break down.

Most often this occurs when quasi-bound states are added or removed between iterations. Changes in potential as small as 1 meV can cause these oscillations. Supposing that the quasi-bound state charge density has a maximum of $n = 2 \times 10^{18}$ cm^{-3}, the rate of change of the charge density is $\frac{\partial n}{\partial \phi} = 2 \times 10^{21}$ cm^{-3}/eV! Under these conditions the self-consistent potential must be approached more slowly.

I have implemented algorithms that will slow the solver any time the relative error in current density, J_{rel}, increases compared to the previous iteration. The solver is always started using the predictor-corrector scheme. If available the initial guess for the potential is a previous self-consistent potential scaled to V_{bias} , otherwise the depletion approximation is used. The predictor-corrector scheme is referred to in the following and in Fig. 2.11 as mode 0. If J_{rel} increases while in mode 0, the previously calculated potential is ignored and solver switches to the single iteration solution method for the Poisson equation outlined in section 2.5, referred to in the following and in Fig. 2.11 as mode 1. When this switch occurs, the under-relaxation parameter is set to $w = 0.2$, which from experience is a slightly aggressive value. From this point on any increase in J_{rel} will cause the latest potential to be discarded and the under-relaxation parameter to be reduced by half until J_{rel} converges (whether or not the potential has converged in this case is discussed in section 2.6.6).

It is still possible for self-consistent potential to terminate without convergence. The first time this occurs the self-consistent solver is restarted in mode 1. Failure in this attempt will cause the self-consistent solver to restart in mode 1 with the depletion approximation used as the initial potential. If the solver cannot converge to the self-consistent potential in this attempt, the depletion approximation is used for calculation of the current density.

2.6.6 Absolute error check

Since the solver is able to dynamically increase the amount of under-relaxation, it is possible for J_{rel} to converge based solely on the small step size that has been taken. Thus, it is necessary to have a measure of the absolute error to ensure that the potential has also converged. The Poisson solver provides such a metric in the calculated change in potential, $\delta\phi$. The magnitude of $\delta\phi$ is related to the absolute error in the Poisson equation given $\phi^{(m)}$ and $n^{(m)}$. If the potential is in fact converging, then near convergence the Poisson equation will generate changes in the potential that are small, typically on the order of 10^{-5} eV. Therefore once J_{rel} has converged, the solver checks that the maximum value of $\delta\phi$ from the last solution of the Poisson equation is less than 10^{-2} eV. If this is not the case, an error is generated and the solver returns to the driver.

2.7 Conclusion and future work

The self-consistent potential solver described in this chapter works well with doping concentrations 10^{18} cm$^{-3} \leq n \leq 10^{20}$ cm^{-3}, temperatures from $T = 5$ K to greater than $T = 500$ K, and for $V_{\text{bias}} \leq 0.5$ V. The weakest approximation is that of an equilibrium concentration of electrons in quasi-bound states. Like the contacts, without a proper scattering model these states cannot react to injected charge density or strength of the confining potential.

Future work may include, at least phenomenologically, scattering in the contacts and into quasi-bound states. One common method is to broaden the energy of the quasi-bound states to emulate the finite lifetime of the states. This broadening may be done somewhat self-consistently by evaluating tunneling rate through the confining potential. It may be also possible to create a model that, like the model of Chapter 4, links the electronic energy levels in a unitary fashion within the propagation matrix method. Such

an approach could allow electron flux to scatter between energy levels, changing the non-equilibrium distribution. However, the model would still be one-dimensional as the propagation matrix method assumes no change in $\mathbf{k}_{\|}$ upon scattering at an interface.

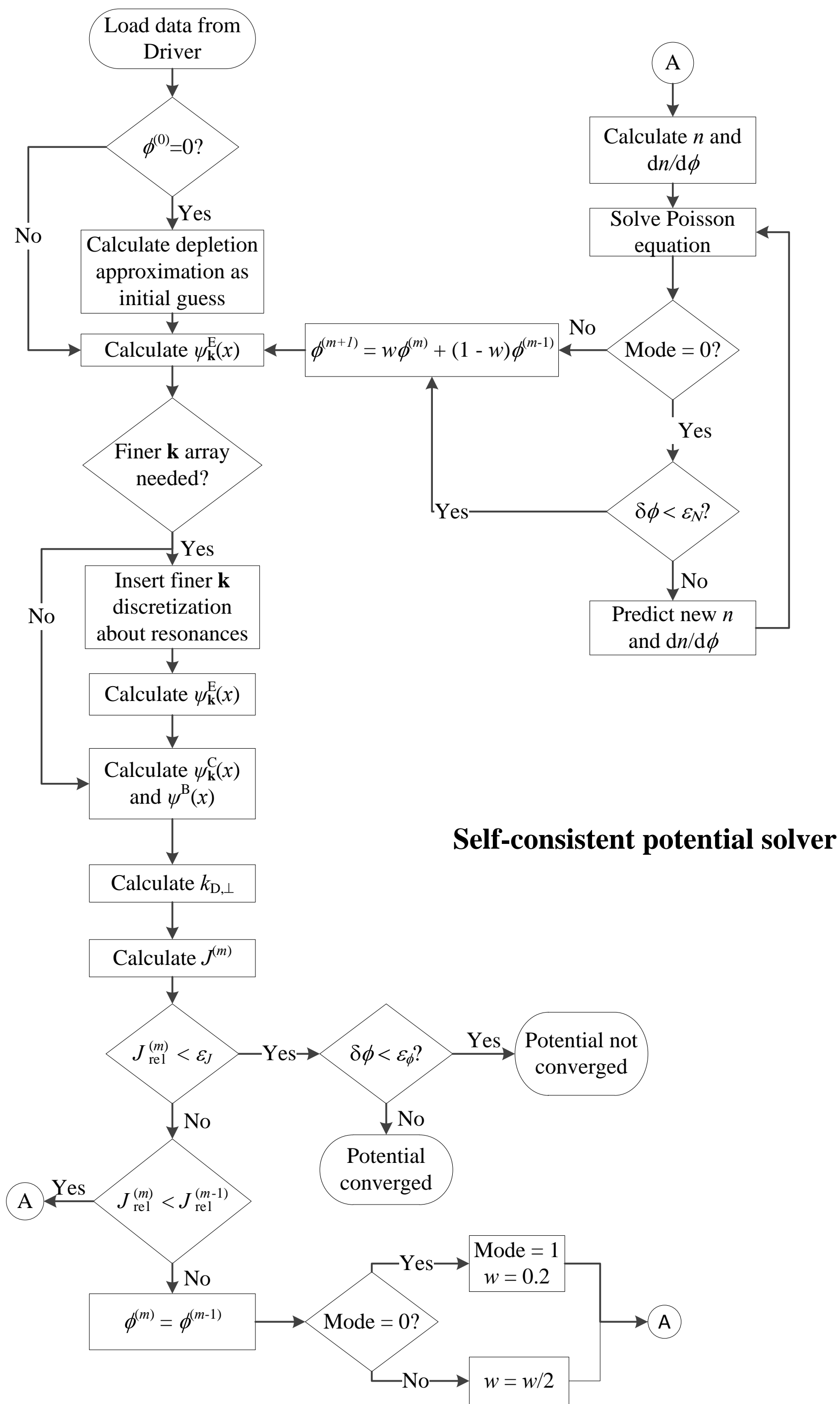

Figure 2.11: Flow chart displaying algorithm for calculating the self-consistent potential.

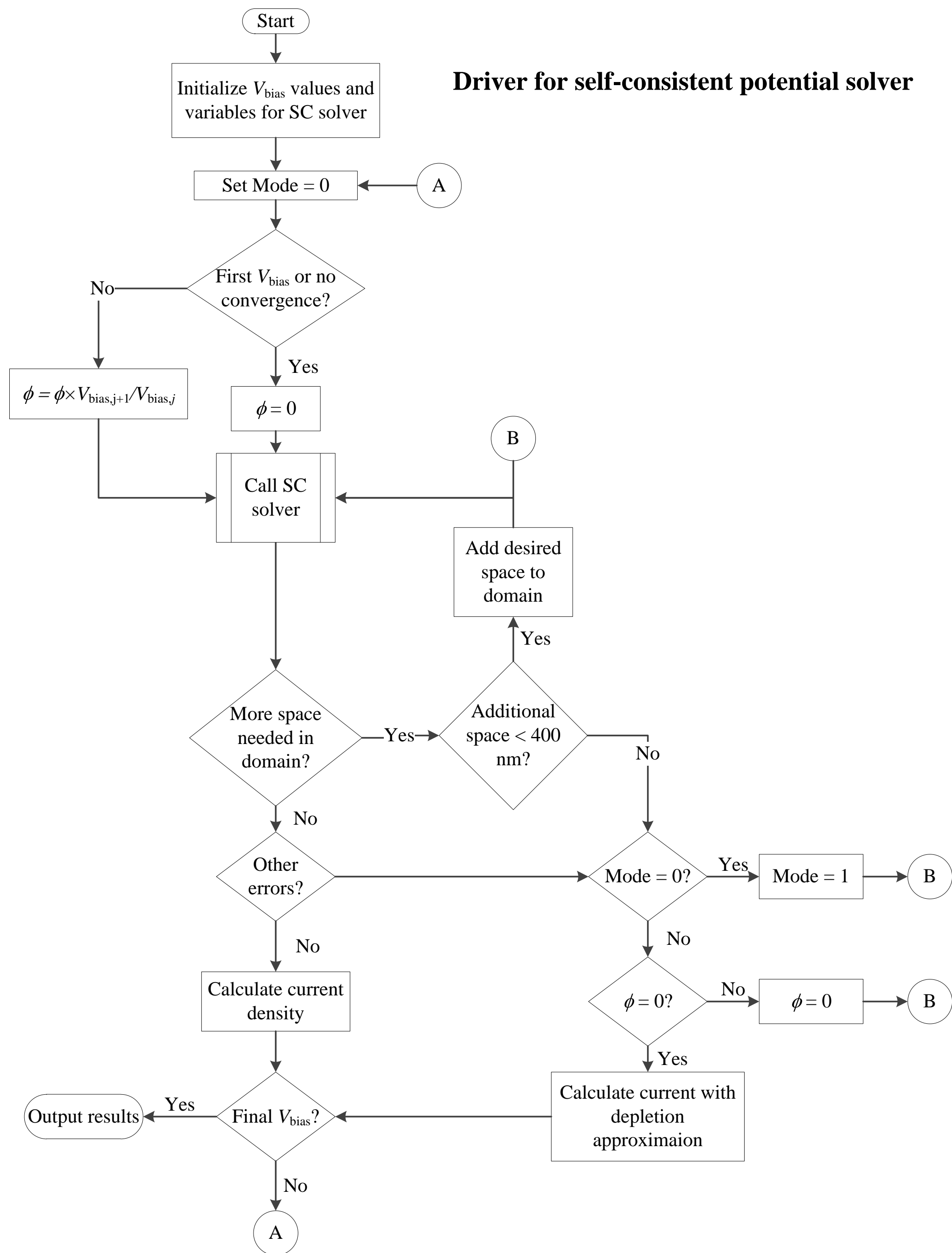

Figure 2.12: Flow chart describing driver for the self-consistent solver. This algorithm provides error handling and calculates the current-voltage characteristic.

Chapter 3

Optimal electronic device design

3.1 $\mathrm{Al}_{\xi}\mathrm{Ga}_{1-\xi}\mathrm{As}$ material system

As a prototype material system the $\mathrm{Al}_{\xi}\mathrm{Ga}_{1-\xi}\mathrm{As}$ material system is chosen for optimization of semiconductor heterostructure diodes. An advantage of this material system is the near lattice match between AlAs and GaAs, having lattice constants differing by only 0.14%. This allows one to ignore effects of strain in the physical model. Additionally, this material has been heavily studied and crystals may be grown with near atomic-layer accuracy via methods such as molecular beam epitaxy and atomic layer deposition. At temperature $T = 300$ K one monolayer in [100]-oriented GaAs is $\delta = 0.2827$ nm thick, the conduction band offset between AlAs and GaAs is 1.04 eV [46], and the relative permittivity is assumed to vary as $\epsilon_{\mathrm{r}} = 13.2 - 3.1 \times \xi$ [1].

Band structures for GaAs and AlAs calculated using an sp^3s^* model [45] are shown in Fig. 3.1. In the devices considered electrons will be injected from GaAs contacts, featuring a direct band gap. The active region of these devices will feature heterostructure layers of varying Al concentration, ξ. This parameter control the conduction band offset between each heterostructure layer and the GaAs contact as $U = 0.8355 \times \xi$ for $\xi < 0.42$ ($U < 0.35$ eV). Al alloy concentrations $\xi > 0.42$ feature the indirect band gap characteristic of AlAs. To avoid the need to account for scattering into X-band minima, only Al alloy concentrations $\xi < 0.42$ are considered.

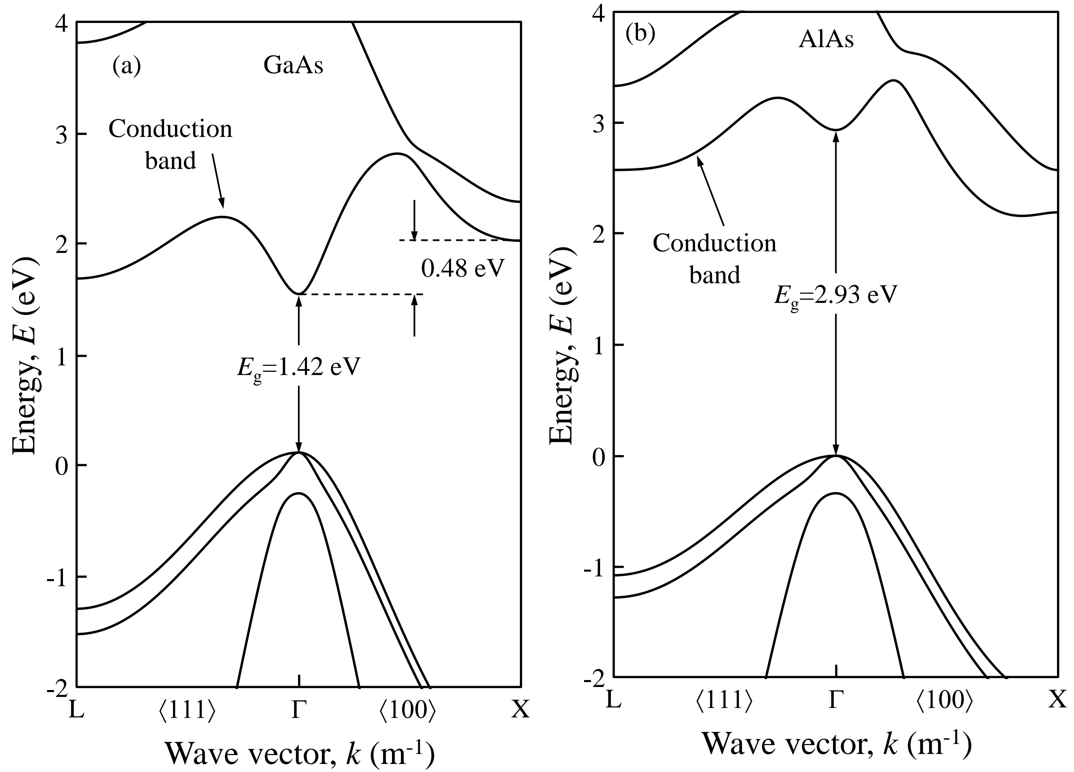

Figure 3.1: Band structure for (a) GaAs and (b) AlAs calculated using an sp^3s^* model [45]. Band gap energies, E_g, are indicated.

For simplicity a conduction band effective electron mass $m = 0.07 \times m_0$ is used in all heterostructure layers, where m_0 is the bare electron mass. The simple coherent transport model captures the essential physics of electron transport in nanoscale $Al_\xi Ga_{1-\xi}As$ heterostructures and is capable of accurately predicting qualitative features of the current-voltage characteristic. Complexities that can be added to the physical model such as material-dependent effective mass and band non-parabolicities [3] as well as scattering mechanisms [37] will produce mostly quantitative changes in calculated current. While these changes may result in optimized potential profiles that differ from those obtained, the achievable degree of optimality should not be significantly changed. Therefore, for a proof-of-principle demonstration these additional complexities provide little additional insight and are not included in the physical model.

3.2 Optimal design of semiconductor heterostructure devices

Consider the jth layer of a layered $\text{Al}_\xi\text{Ga}_{1-\xi}\text{As}$ heterostructure device. In the physical model this layer may control electron transport through both its thickness, L_j, and its conduction band offset, U_j. These design parameters for all heterostructure layers may be grouped into the vector p, containing P elements. As a measure of the optimality of a set of design parameters p, the cost function is defined as

$$C(p) = \sum_{j=1}^{\nu} w_j \left(J_{\text{obj}}(V_{\text{bias}}^j) - J_{\text{sim}}(V_{\text{bias}}^j, p) \right)^2, \tag{3.1}$$

where w_j is a weighting factor. This form of the cost function represents a least squares measure of the difference between the simulated current density, J_{sim}, and the objective function, J_{obj}. It has been chosen to guarantee a continuous derivative for local parabolic minima, necessary for gradient-based methods of optimization.

Because more optimal designs are associated with lower cost, the optimal design problem is formulated as a minimization problem,

$$\min_p C, \tag{3.2}$$

subject to the forward solve

$$J_{\text{sim}}(V_{\text{bias}}^j) = \frac{em}{2\pi^2\hbar^3\beta} \sum_{r=1}^{R} T(V_{\text{bias}}^j, E^r) \ln \left(\frac{1 + e^{(\mu - E^r)\beta}}{1 + e^{(\mu - E^r - V_{\text{bias}}^j)\beta}} \right) dE, \tag{3.3}$$

$$T(V_{\text{bias}}^j, E^r) = \frac{1}{\left| a_0(V_{\text{bias}}^j, E^r) \right|^2}, \tag{3.4}$$

where a discrete sum necessary for numerical evaluation has been shown explicitly. In this expression j indexes $V_{\rm bias}$, r indexes energy, and dE is the separation between energy samples. The design parameters are subject to the constraints

$$0 \leq U_j \leq 0.35 \text{ eV}, \tag{3.5}$$

$$L_j = n \times \delta, \;\; n \in N^+. \tag{3.6}$$

Sufficient conditions for a locally optimal device configuration, p^*, are

$$\nabla_p C|_{p^*} = 0, \tag{3.7}$$

and

$$\mathbf{H}_p(C)|_{p^*} > 0, \tag{3.8}$$

if no constraints are active, where $\mathbf{H}_p$ is the Hessian operator which is a matrix of second order partial derivatives with respect to the design parameters p. Active constraints occur when a design parameter approaches one of its constraints, so that the device configuration is at the edge of the feasible solution space. When this occurs, variations of these conditions can be satisfied.

If the gradient $\nabla_p C$ is known, the cost function may be minimized efficiently using Newton's method, for which successive approximations for (3.7) are

$$p_{n+1} = p_n + \frac{\nabla_p C}{\mathbf{H}_p C}. \tag{3.9}$$

The $fmincon$ program from the optimization toolbox in MATLAB is used to implement the Newton method. If any constraints are active the optimization algorithm solves a quadratic programming problem to determine the search direction.

The Newton method requires evaluation of the gradient $\nabla_p C$. Although $\nabla_p C$ may be evaluated using a finite difference approximation, this method is computationally expensive, requiring P additional forward solves. Since each optimization will likely require several hundred forward solves and the calculation of the self-consistent potential alone is computationally costly, efficient calculation of the gradient is necessary for the local optimum to be found in a reasonable amount of time. The adjoint method provides an efficient method for evaluating $\nabla_p C$.

3.2.1 The adjoint method

The forward solve may be cast in terms of a linear system

$$\mathbf{L}(p) \cdot x = b, \tag{3.10}$$

$$J_{\text{sim}}(V_{\text{bias}}^{j}, p) = W_j(x), \tag{3.11}$$

where matrix $\mathbf{L}(p)$ contains the device physics, vector x contains transmitted wave function amplitudes, and vector b contains boundary conditions.

The P-dimensional gradient of C may be evaluated with the chain rule. For the ℓth design parameter, the derivative of C is

$$\partial_{p_\ell} C = -\sum_{j=1}^{\nu} 2w_j \left(J_{\text{obj}}(V_{\text{bias}}^{j}) - J_{\text{sim}}(V_{\text{bias}}^{j}, p) \right) \partial_x W_j(x) \partial_{p_\ell} x. \tag{3.12}$$

While $\partial_x W_j(x)$ may be computed explicitly, $\partial_{p_\ell} x$ can be difficult to evaluate directly. The propagation matrix method of the physical model calculates x by multiplying together a series of matrices. Evaluating the derivative of this product may be computationally expensive. This problem may be circumvented by defining an adjoint equation

$$\mathbf{L}^T(p) \cdot h = \sum_{j=1}^{\nu} 2w_j \left(J_{\text{obj}}(V_{\text{bias}}^j) - J_{\text{sim}}(V_{\text{bias}}^j, p) \right) \partial_x W_j^T(x), \tag{3.13}$$

solved in the same manner as the forward solve. Using the derivative of the forward solve

$$(\partial_{p_\ell} \mathbf{L}(p)) \cdot x + \mathbf{L}(p) \cdot (\partial_{p_\ell} x) = \partial_{p_\ell} b. \tag{3.14}$$

and substituting (3.14) and (3.13) into (3.12), the derivative of C is

$$\partial_{p_\ell} C = h^T \cdot (\partial_{p_\ell} \mathbf{L}(p) \cdot x - \partial_{p_\ell} b) \,. \tag{3.15}$$

The adjoint method calculates a gradient efficiently because, rather than computing P additional forward solves for a finite differences method evaluation of the adjoint variable h requires the same computational cost as only one forward solve. x is calculated during the forward solve and the derivatives $\partial_{p_\ell} \mathbf{L}(p)$ are simple functions that are determined analytically in the next section. Thus, for the additional computational cost of a single forward solve the gradient of C with respect to all P design parameters may be computed.

3.2.2 Derivative of the cost function

For the physical model, application of the adjoint method yields a derivative [26, 24]

$$\frac{\partial C}{\partial p_\ell} = \mathrm{Re}\left[\sum_{j=1}^{\nu}\sum_{r=1}^{R}\sum_{i=0}^{N}(\beta_{i+1}^{j,r})^T \frac{\partial \mathbf{p}_i^{j,r}}{\partial p_\ell}\alpha_{i+1}^{j,r}\right], \tag{3.16}$$

where $\beta_{i+1}^{j,r}$ is the adjoint variable calculated by solving

$$\begin{aligned}\beta_{i+1}^{j,r} =& (\mathbf{p}_{i-1}^{j,r})^T\beta_i^{j,r} \\ &+ 4\delta_{i,0}w_j\left(J_{\text{obj}}^j - \sum_{r'=1}^{R} c^{j,r'}\frac{1}{|a_0^{j,r'}|^2}\right)c^{j,r}\frac{1}{|a_0^{j,r'}|^4}\begin{bmatrix}\left(a_0^{j,r'}\right)^* \\ 0\end{bmatrix}.\end{aligned}$$

The initial condition for $\beta_{i+1}^{j,r}$ is

$$\beta_0^{j,r} = \begin{bmatrix}0\\0\end{bmatrix}, \tag{3.17}$$

and in (3.16)

$$\alpha_i^{j,r} = \begin{bmatrix}a_i^{j,r}\\ b_i^{j,r}\end{bmatrix}. \tag{3.18}$$

In these equations i indexes position, r indexes energy, and j indexes V_{bias} . Evaluation of the derivatives $\frac{\partial \mathbf{p}_j}{\partial L_\ell}$ and $\frac{\partial \mathbf{p}_j}{\partial U_\ell}$ ideally include changes in the self-consistent potential with respect to changes in the design parameters. However, this derivative cannot be calculated explicitly. Since the optimized potential profiles feature only broad electron transmission resonances the charge density, and hence the self-consistent potential, should not be sensitive to small changes in design parameters. Ignoring the effect should at worst slow convergence of the optimization algorithm.

The derivative $\frac{\partial \mathbf{p}_j}{\partial L_\ell}$ is considered first. Perturbing a heterostructure layer thickness will perturb only the complex exponential terms in the propagation matrix. Therefore,

$$\frac{\partial \mathbf{p}_j}{\partial L_\ell} = i\frac{k_j}{2} \begin{bmatrix} -\left(1+\frac{k_{j+1}}{k_j}\right)e^{-ik_jL_j} & \left(1-\frac{k_{j+1}}{k_j}\right)e^{ik_jL_j} \\ -\left(1-\frac{k_{j+1}}{k_j}\right)e^{-ik_jL_j} & \left(1+\frac{k_{j+1}}{k_j}\right)e^{ik_jL_j} \end{bmatrix}. \tag{3.19}$$

The derivative $\frac{\partial \mathbf{p}_j}{\partial U_\ell}$ is made somewhat more complicated by the fact that either k_j or k_{j+1}, or both, in $\mathbf{p}_j$ may depend upon U_ℓ. First noting that

$$\frac{\partial k_j}{\partial U_\ell} = \frac{\sqrt{2m}}{\hbar}\frac{\partial}{\partial U_j}\sqrt{E-U_j} = -\frac{m}{2\hbar^2}\frac{1}{k_j}, \tag{3.20}$$

the derivatives are

$$\frac{\partial \mathbf{p}_{j-1}}{\partial U_\ell} = \frac{m}{2\hbar^2 k_{j-1}k_j} \begin{bmatrix} -e^{-ik_jL_j} & e^{ik_jL_j} \\ e^{-ik_jL_j} & -e^{ik_jL_j} \end{bmatrix}, \tag{3.21}$$

$$\frac{\partial \mathbf{p}_j}{\partial U_\ell} = \frac{m}{2\hbar^2 k_j^2} \begin{bmatrix} \left(iL_j\left(k_j+k_{j+1}\right)+\frac{k_{j+1}}{k_j}\right)e^{-ik_jL_j} & \left(-iL_j\left(k_j-k_{j+1}\right)-\frac{k_{j+1}}{k_j}\right)e^{ik_jL_j} \\ \left(iL_j\left(k_j-k_{j+1}\right)-\frac{k_{j+1}}{k_j}\right)e^{-ik_jL_j} & \left(-iL_j\left(k_j+k_{j+1}\right)+\frac{k_{j+1}}{k_j}\right)e^{ik_jL_j} \end{bmatrix}. \tag{3.22}$$

When both k_j and k_{j+1} depend on U_j, the derivative is the sum of (3.21) and (3.22).

3.3 Extent and limit of control for linear objective

Because of the importance of Ohmic (linear) current-voltage behavior in analog and mixed signal circuit design, the extent to which it is possible to achieve such behavior in a tunnel diode with non-equilibrium electron transport through an active region just a few nm thick is explored. Expected difficulty linearizing device behavior that, as shown in the next section, generally exhibits nonlinear behavior should provide insight into

physical mechanisms responsible for controlling the current-voltage characteristic and be a suitable test for optimal design.

Mechanisms responsible for linear current-voltage behavior in optimally designed tunnel diodes are different from those of conventional resistors and nano-wires. The former relies on near-equilibrium electron transport and dissipative processes involving the emission of phonons [49]. The latter relies on dimensional confinement and non-equilibrium electron transport with no scattering giving a maximum electron conductance per spin of $G_n = e^2/2\pi\hbar$. In addition, inability to create low impedance contacts to nanowires has limited their utility and application to electronic devices. Depositing metal contacts directly onto semiconductor nanowires effectively forms a Schottky-barrier [25] and failure to lower the contact resistance of single-walled carbon nanotubes below tens of kΩ has driven the research community to consider use of multiple-walled and bundled nanotubes [35, 32].

The design of optimized tunnel diodes considered here do not suffer from these difficulties. Elastic electron scattering from a precisely defined potential is used to control the current-voltage characteristic of the device. The work extends previous studies of optimized electron transmission [39, 47] and current density [26] by self-consistently solving the Schrödinger and Poisson equations to obtain the potential. In addition, physical mechanisms limiting feasibility of design are explored and it is shown how material parameter choice impacts dynamic range.

3.3.1 The single-barrier heterostructure diode

In general, non-equilibrium electron transport in nano-scale devices results in a nonlinear relation between current and voltage bias. As an example, consider the heterostructure tunnel diode shown in Fig. 3.2(a). The material consists of an undoped AlAs layer

of thickness L_0 grown between two GaAs contacts with n-type impurity concentration $N_{\mathrm{D}} = 1 \times 10^{18}$ cm^{-3}.

Results of calculating the current-voltage characteristic of this device are shown in Fig. 3.2. For small bias voltage $|V_{\mathrm{bias}}| < \mu$, increasing bias opens channels for current flow. At temperature $T = 300$ K, the chemical potential in the model of the GaAs contacts is $\mu = 38.9$ meV. In this regime, electron transmission probability through the tunnel barrier is weakly dependent on V_{bias} and the current-voltage characteristic is quasi-linear. Once $|V_{\mathrm{bias}}| > \mu$, the majority of current carrying states have been accessed. The slope $\frac{\partial J}{\partial V_{\mathrm{bias}}}$ decreases as changing current is now caused primarily by V_{bias} lowering the energy of the AlAs layer relative to the energy of conduction band electrons in the cathode. This change in physical mechanisms controlling electron transport should be difficult to linearize, providing a suitable test of achievable optimality.

The log scale plot of Fig. 3.2(c) shows that current exhibits an exponential dependence in this regime, with current scaling exponentially with L_0. Although a single potential barrier can scale current, it does not offer significant control over the current-voltage characteristic.

3.3.2 Chosen device structure for optimization

Control necessary to linearize the current-voltage characteristics shown in Fig. 3.2 may be achieved by introducing Al$_{\xi}$Ga$_{1-\xi}$As layers with varying composition. From previous optimization efforts, such as Fig. 1.2, it is observed that optimal designs favor a large potential barrier near the cathode to set the scale for current flow. Therefore a device structure is chosen consisting of one AlAs layer of thickness L_0 followed by N Al$_{\xi}$Ga$_{1-\xi}$As layers, each of thickness $L_{\delta} = 4 \times \delta$. Total thickness of the active region is $L = L_0 + NL_{\delta}$.

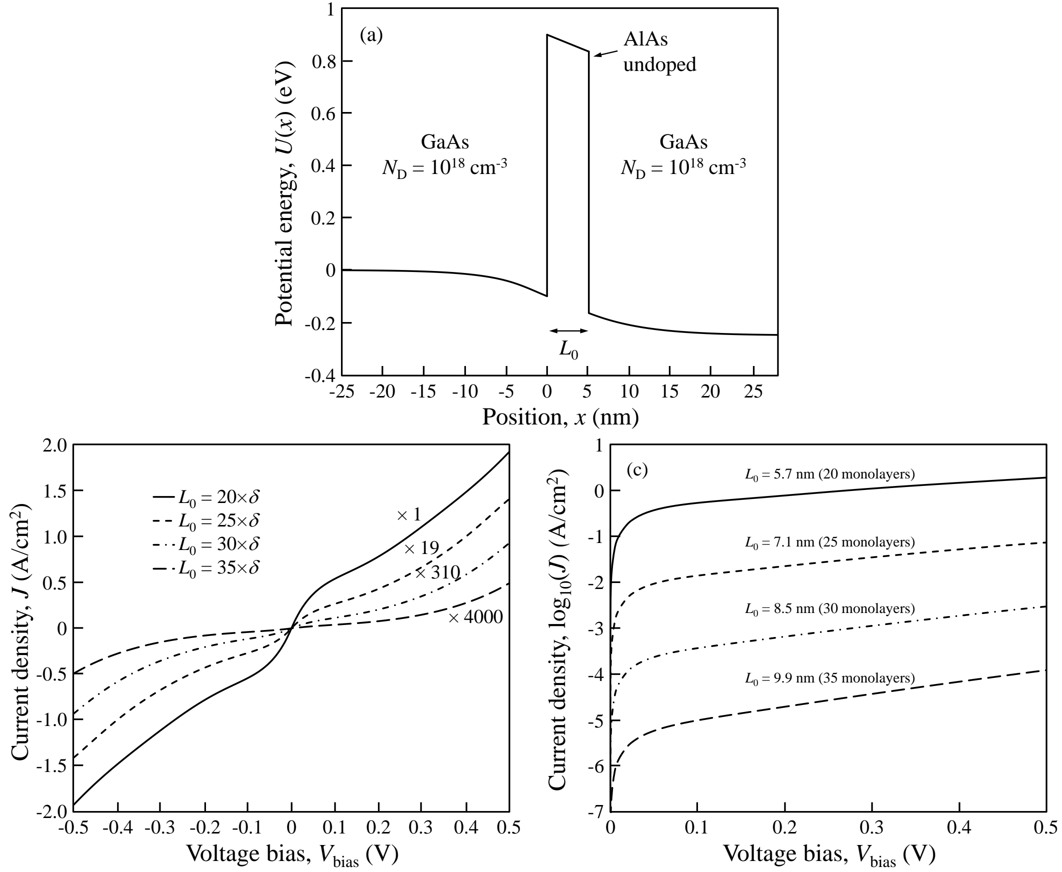

Figure 3.2: (a) Tunnel-diode consisting of an undoped AlAs layer of length L_0 between two GaAs contacts subject to an applied voltage bias $V_{\mathrm{bias}} = 0.25$ V. Also shown are current-voltage characteristics through the device of part (a) in (b) linear and (c) log scales for several values of L_0. Current is exponential for $|V_{\mathrm{bias}}| > \mu$ ($\mu = 38.9$ meV) and quasi-linear for $|V_{\mathrm{bias}}| < \mu$. Simulation parameters are electron effective mass $m = 0.07 \times m_0$, temperature $T = 300$ K, and n-type doping concentration $N_{\mathrm{D}} = 1 \times 10^{18}$ cm^{-3}.

An example is shown in Fig. 3.3, where conduction band offsets of the $\mathrm{Al}_\xi\mathrm{Ga}_{1-\xi}\mathrm{As}$ layers have been chosen randomly. Voltage bias is restricted to the polarity shown in Fig. 3.2(a). In this orientation the AlAs layer is the primary current limiter, chosen for its ability to scale the current by orders of magnitude with a change in L_0 of only a few monolayers. The following series of $\mathrm{Al}_\xi\mathrm{Ga}_{1-\xi}\mathrm{As}$ layers of fixed thickness fine-tune

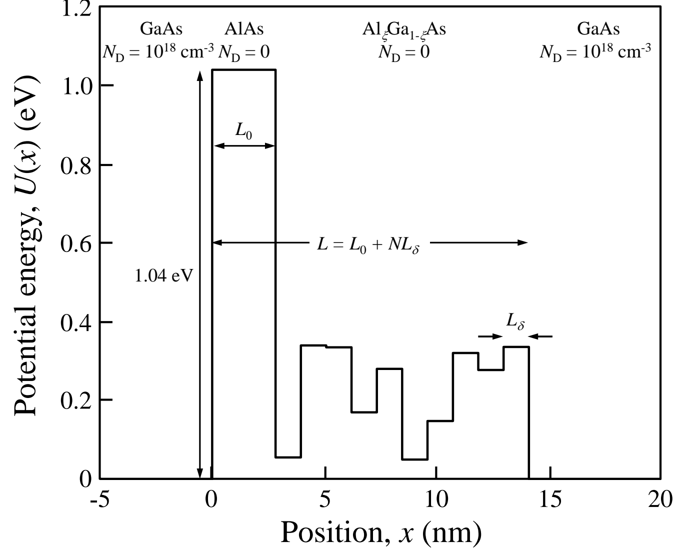

Figure 3.3: Randomly configured Al$_\xi$Ga$_{1-\xi}$As heterostructure diode displaying features selected for the optimized potential profiles. An AlAs layer having a thickness L_0 is the primary current limiter. The following series of Al$_\xi$Ga$_{1-\xi}$As layers fine-tune the current-voltage characteristic, with the number of Al$_\xi$Ga$_{1-\xi}$As layers, N, used as an auxiliary parameter to set the length of the device. The Al$_\xi$Ga$_{1-\xi}$As layers have a fixed thickness $L_\delta = 4 \times \delta$ and $L_0 = n \times \delta$, where n is a positive integer and $\delta = 0.2827$ nm is one monolayer of Al$_\xi$Ga$_{1-\xi}$As . Total thickness of the active region is $L = L_0 + NL_\delta$.

features in the current to match the objective as closely as possible. The GaAs contacts have an n-type doping concentration $N_\mathrm{D} = 1 \times 10^{18}$ cm^{-3}.

3.3.3 Optimization for linear objective

To reduce simulation time, each optimization consists of two sequential optimizations. The first optimization samples 5 linearly spaced V_bias points to establish an estimate of the optimal profile. The result of this optimization is then passed to the second optimization which samples 25 linearly spaced points V_bias points.

The optimization algorithm treats L_0 as a continuous quantity. To enforce its discrete nature, the the physical model rounds L_0 to the nearest integer multiple of δ. Once L_0 has changed sufficiently it is rounded to a new value, causing a large change in the

current density. The optimization then effectively restarts, as the set of $\mathrm{Al}_\xi\mathrm{Ga}_{1-\xi}\mathrm{As}$ conduction band offsets $\{U_j\}$ were optimized for the previous value of L_0. To prevent this from occurring in the slower of the two optimizations, L_0 is fixed during the second optimization.

The linear objective function is defined as $J_{\mathrm{obj}} = V_{\mathrm{bias}}/r$, where r defines the slope of the objective function with units of $\Omega \times \mu\mathrm{m}^2$. For a device of cross-sectional area A characterized by current density $J = V_{\mathrm{bias}}/r$, the resistance is $R = r/A$. The optimization is limited to $0 \leq V_{\mathrm{bias}} \leq 0.5$ V, and auxiliary parameters for optimization are r and the number of $\mathrm{Al}_\xi\mathrm{Ga}_{1-\xi}\mathrm{As}$ layers N. Initial device configurations consist of a random set $\{U_j\}$ and an initial guess for L_0 that ensures $J(V_{\mathrm{bias}} = 0.5$ V) is of the same order of magnitude as the objective function.

3.3.4 Results for linear objective function

Optimizations were completed using $N = 5$, 7, 10, 12, 15, and 17 layers of $\mathrm{Al}_\xi\mathrm{Ga}_{1-\xi}\mathrm{As}$ and values of r ranging from $r = 10$ k$\Omega \times \mu\mathrm{m}^2$ to $r = 100$ M$\Omega \times \mu\mathrm{m}^2$ by factors of 10. Designs for $r = 1$ k$\Omega \times \mu\mathrm{m}^2$ also yielded linear behavior, but the 2-3 monolayer thickness of the AlAs layer in the designs violated the assumption of an equilibrium distribution of electrons in accumulation region quasi-bound states [11, 10]. Resistances equal to or less than this value may be accessed by increasing N_{D}.

Fig. 3.4 shows current-voltage characteristics of the most optimal design for each r. The currents have been scaled such that the objective functions are equal and the curves overlap. It can be seen that device designs spanning 4 orders of magnitude in r display essentially linear behavior over the optimization range.

The potential profile and relative error in r for an optimally designed device featuring $N = 12$ $\mathrm{Al}_\xi\mathrm{Ga}_{1-\xi}\mathrm{As}$ layers and $r = 1$ $\Omega \times \mu\mathrm{m}^2$is shown in Fig. 3.5. The linearity of this design is quantified by calculating total harmonic distortion (THD). A sinusoidal voltage

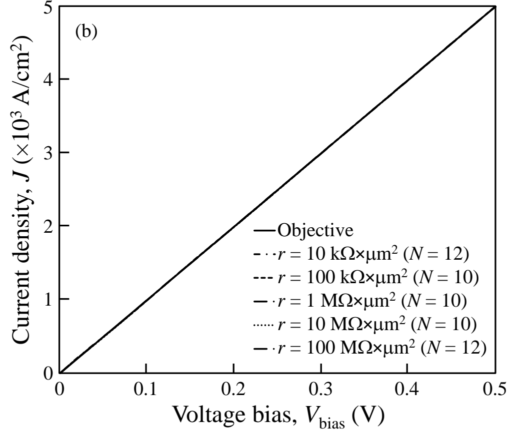

Figure 3.4: Overlay of simulated current through the most optimal design for each r, with corresponding N indicated. Currents for $r > 10\ \mathrm{k}\Omega \times \mu\mathrm{m}^2$ have been scaled so that all curves overlap. Simulation parameters are electron effective mass $m = 0.07 \times m_0$, temperature $T = 300$ K, doping concentration $N_\mathrm{D} = 1 \times 10^{18}$ cm^{-3} .

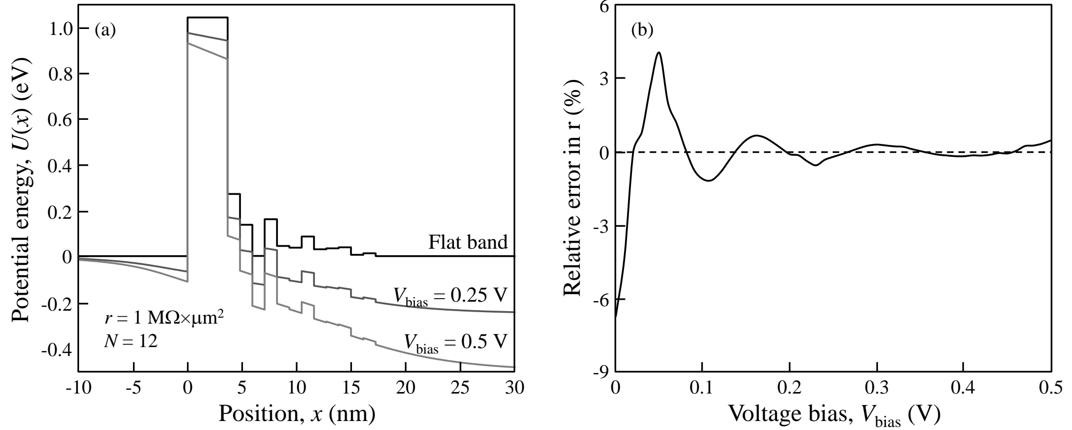

Figure 3.5: (a) Optimized potential profile for a diode characterized by $r = 1\ \mathrm{M}\Omega \times \mu\mathrm{m}^2$ and $N = 12$. (b) Error in differential resistance relative to r. Simulation parameters are electron effective mass $m = 0.07 \times m_0$, temperature $T = 300$ K, doping concentration $N_\mathrm{D} = 1 \times 10^{18}$ cm^{-3} .

of frequency ω_0 having constant offset voltage $V_\mathrm{bias} = 0.25$ V and peak amplitude V_p ($V_\mathrm{pp} = 2V_\mathrm{p}$) is applied to the diode, resulting in the distortion curves shown in Fig. 3.6. Suppression of the second harmonic for $V_p \leq 0.1$ V results from odd symmetry in the differential resistance about $V_\mathrm{bias} = 0.25$ V. Excellent linearity is achieved, with THD ≤ -65 dB for $V_\mathrm{p} \leq 0.2$ V indicating greater than 10 bit dynamic range for signals

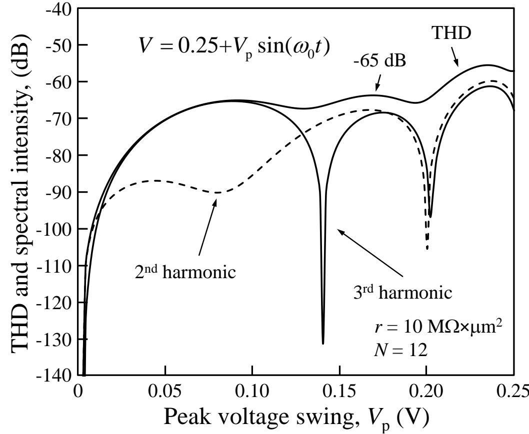

Figure 3.6: Spectral intensity and total harmonic distortion (THD) for optimized diode shown in Fig. 3.5. The input voltage is a sinusoidal wave of frequency ω_0 and amplitude V_p with constant offset voltage $V_\mathrm{bias} = 0.25$ V. Suppression of the second harmonic for $V_p \leq 0.1$ V results from odd symmetry in the differential resistance about $V_\mathrm{bias} = 0.25$ V. Total harmonic distortion of less than -65 dB indicates greater than 10 bit dynamic range.

Table 3.1: Deviation from objective, Δ, of the most optimal design for each N.

N	5	7	10	12	15	17
Δ	-4.23	-5.09	-5.62	-5.58	-4.97	-4.72

up to $V_\mathrm{pp} = 0.4$ V. Decreasing the relative error for convergence of the self-consistent potential to $\epsilon_\mathrm{J} = 10^{-5}$ produced similar results, indicating the existence of limits to achievable optimality for the system.

The device of Fig. 3.5 was achieved by weighting the cost function toward more nonlinear regions of J. An initial result of optimization for auxiliary parameter combination $N = 12$ and $r = 1$ M$\Omega \times \mu$m^2 featured a THD of -54 dB at $V_\mathrm{p} = 75$ mV, resulting from greater error in r for $V_\mathrm{bias} < 0.2$ V than that shown in Fig. 3.5. The design was further optimized using a weighting factor $w_j = 5$ for $0.1 \leq V_\mathrm{bias} \leq 0.25$ V, $w_j = 2$ for $0.25 < V_\mathrm{bias} \leq 0.4$ V, and unity weighting elsewhere. This demonstrates that control may be improved by applying greater weight to less controlled regions of J.

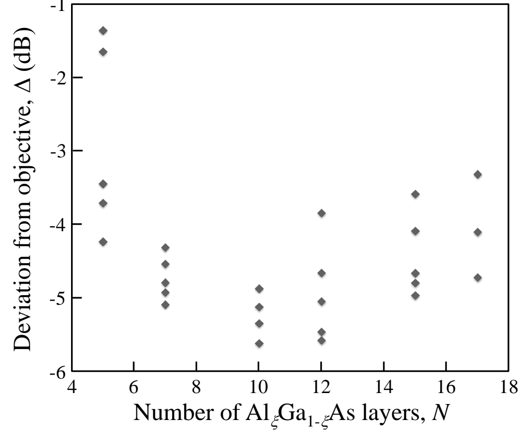

Figure 3.7: Deviation from objective as a function of the number of $\mathrm{Al}_\xi\mathrm{Ga}_{1-\xi}\mathrm{As}$ layers, N. The trend indicates that a natural value for auxiliary parameter N exists at $N = 10$.

To compare optimality independent of the magnitude of current, it is convenient to define the deviation from the objective as

$$\Delta = \log_{10}(1 - R_{\mathrm{d}}^2), \tag{3.23}$$

where R_{d}^2 is the coefficient of determination between the objective and simulated current density, constrained to $0 \leq R_{\mathrm{d}}^2 \leq 1$ with $R_{\mathrm{d}}^2 = 1$ corresponding to $J_{\mathrm{sim}} = J_{\mathrm{obj}}$. The logarithmic nature of this metric coupled with $R_{\mathrm{d}}^2 \approx 1$ for the designs make this a sensitive measure of optimality. Presented in Table 3.1 is Δ for the most optimal design of each N. As indicated in Fig. 3.4, $N = 10$ and $N = 12$ were the most natural values of auxiliary parameter N. The deviation from objective of the design of Fig. 3.5 is $\Delta = -5.58$.

Deviation from objective for all auxiliary parameter combinations is shown in Fig. 3.7 with respect to N. A clear trend in Δ as a function of N may be seen, with $N = 10$ as the most natural auxiliary parameter. This trend indicates that physical mechanisms related to device length limit the degree of achievable linearity.

While the most optimal value for the auxiliary parameter is $N = 10$, the most linear device has been shown to have $N = 12$. This results from optimization of the current density effectively optimizing the average slope of the current-voltage characteristic. To optimize for linearity the cost function would need to contain a $\frac{\partial J}{\partial V_{\text{bias}}}$ term. This form for the cost function was not chosen because it could result in linear regions in the current-voltage characteristic that do not extrapolate to the origin (non-Ohmic behavior). It is possible however to optimize the value of the current density and the slope simultaneously. Weighting the optimization towards one of these quantities could allow for additional control over device behavior.

3.3.5 Feasibility of design

Control over device behavior is accessible through manipulation of quantum mechanical effects such as tunneling and resonant electron transmission. The extent to which these features can be accessed by thermally distributed electrons in the electrodes and controlled by the design parameters for the given range of V_{bias} and N determine feasibility of a given design.

For example, optimal $\text{Al}_\xi\text{Ga}_{1-\xi}\text{As}$ region lengths of $NL_0 \approx 7-12$ nm restrict access to low-lying resonances ($< \mu$) at small V_{bias} . Low-energy resonances are necessary for control in the low bias regime, and access to these resonant states could be made available by including $\text{In}_\xi\text{Ga}_{1-\xi}\text{As}$ alloys to allow $U_j \leq 0$ eV. Absence of these resonances is apparent in Fig. 3.8, which shows the relative error in r for the optimal designs sorted by N. It can be seen that in all designs negative error in r occurs as $V_{\text{bias}} \to 0$ V and error in r increases as V_{bias} decreases.

Naively, one might expect that longer devices, with more degrees of freedom and access to additional resonances in the electron transmission spectrum, would have more control and hence higher optimality. However, for large values of N this is not the case

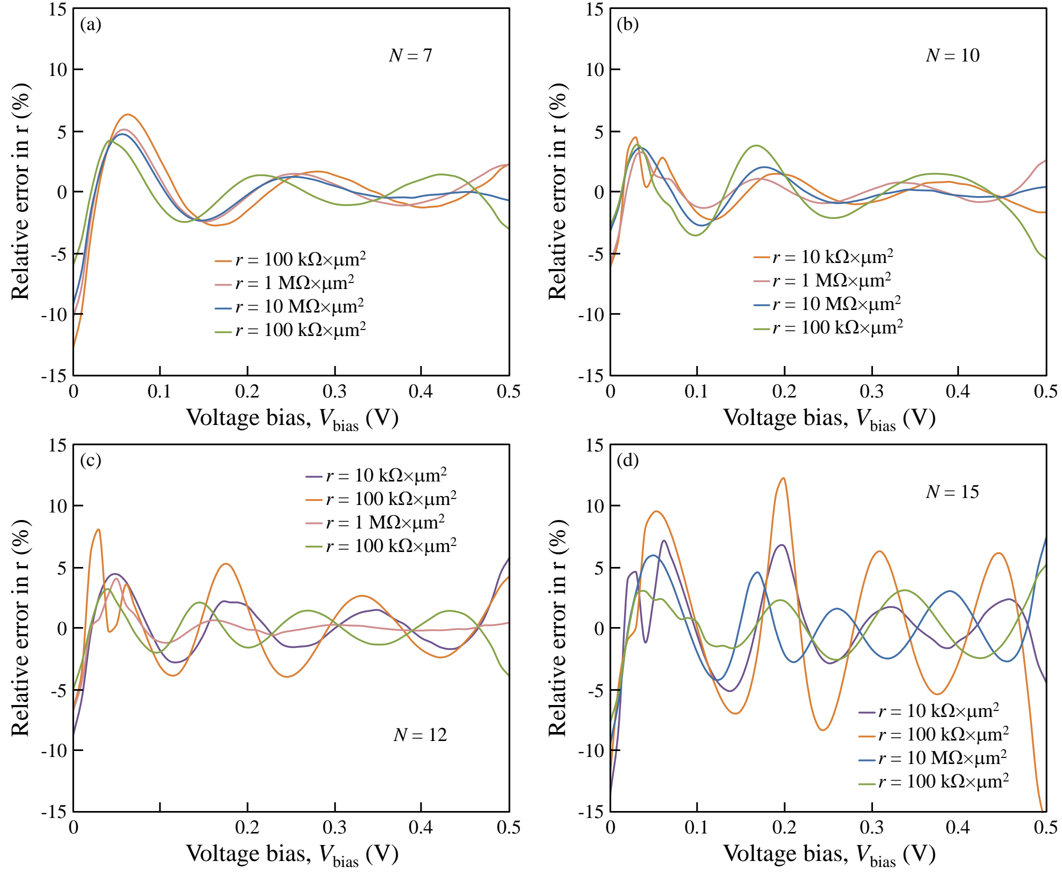

Figure 3.8: Relative error in differential resistance for (a) $N = 7$, (b) $N = 10$, (c) $N = 12$, and (d) $N = 15$ $Al_{\xi}Ga_{1-\xi}As$ layers. As device length increases, the number and strength of the resonances increases, causing greater difficulty controlling the current-voltage characteristic. Designs featuring $N = 5$ and $N = 17$ designs are not shown as much of the error extends beyond the scale shown.

in part due to a reduced ability for V_{bias} to control resonances with a necessarily smaller electric field.

As an analogy, consider reflections from the edges of the active region as a Fabry-Perot resonator. Increasing the length of the device will change electron transmission resonances by narrowing the linewidth and increasing peak-to-valley ratio. This effect is seen in Fig. 3.8, which shows both increased number of oscillations and increased strength of oscillations about the objective r as N increases. Scattering from the $Al_{\xi}Ga_{1-\xi}As$ layers is not sufficient to smooth out these stronger resonances.

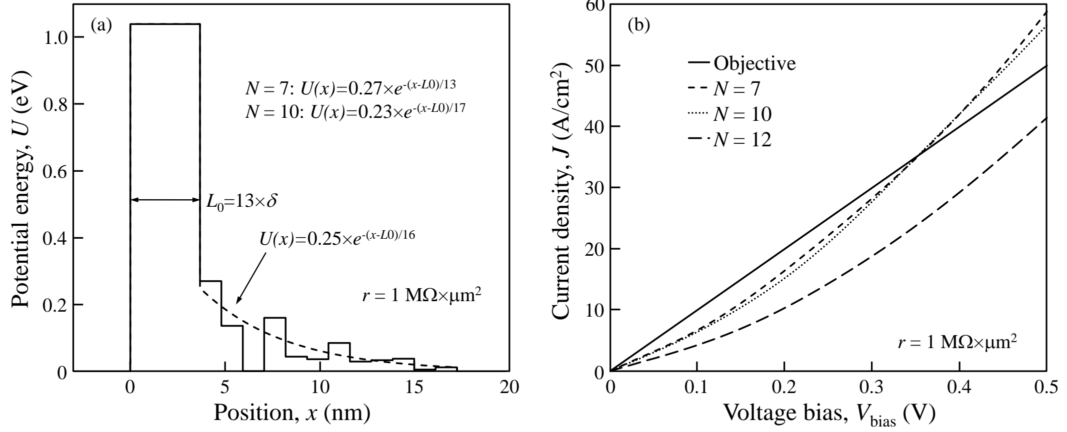

Figure 3.9: (a) Potential profile of device of Fig. 3.5 with exponential decaying potential manually fit to the average potential energy of the $Al_\xi Ga_{1-\xi}As$ layers. Also shown are fits used for the $N = 7$ and $N = 10$ designs of $r = 1$ M$\Omega \times \mu\text{m}^2$. (b) Current calculated using exponentially decaying potential profiles of part (a). Proximity of curves to the objective is shows that the physical mechanism responsible for setting the rough features in the current is electron tunneling. Potential well features added by the optimization algorithm fine-tune the current. Simulation parameters are electron effective mass $m = 0.07 \times m_0$, temperature $T = 300$ K, doping concentration $N_\text{D} = 1 \times 10^{18}$ cm^{-3} .

The potential profile of Fig. 3.5 shows decreasing average U_j along the device length. This feature is common to many $N = 7$-12 optimized designs, and is chosen to set the rough features of the current via electron tunneling. To show this, Fig. 3.9(a) plots the device of Fig. 3.5 with the average conduction band offset manually fit to an exponential decay. Equations for the exponentially decaying potentials of the $N = 7$ and $N = 10$ designs for $r = 1$ M$\Omega \times \mu\text{m}^2$ are given, and current through these potentials is shown in Fig. 3.9(b). Proximity of the curves to the objective shows that the physical mechanism responsible for setting gross features of the current-voltage characteristic is electron tunneling. Potential well features are added by the optimization algorithm to form electron transmission resonances that further linearize and fine-tune this current.

3.3.6 Sensitivity analysis

Sensitivity to errors in $\mathrm{Al}_\xi\mathrm{Ga}_{1-\xi}\mathrm{As}$ layer thickness is shown in Fig. 3.10(a). The Fig. shows Δ calculated in the presence of $\pm 1 \times \delta$ or $\pm 2 \times \delta$ perturbations in the thickness of individual $\mathrm{Al}_\xi\mathrm{Ga}_{1-\xi}\mathrm{As}$ layers for the design considered in Fig. 3.5. As shown in Fig. 3.10(b), skew in the distribution is because electron scattering strength of the first four layers from the AlAs layer are significant. These layers are responsible for all bin counts in which $\Delta > -2$. The third and fourth layers are most responsible for formation of transmission resonances because they are located near a quarter of the characteristic electron (Fermi) wavelength in the system. The two leftmost layers help the AlAs potential barrier set the scale of the current.

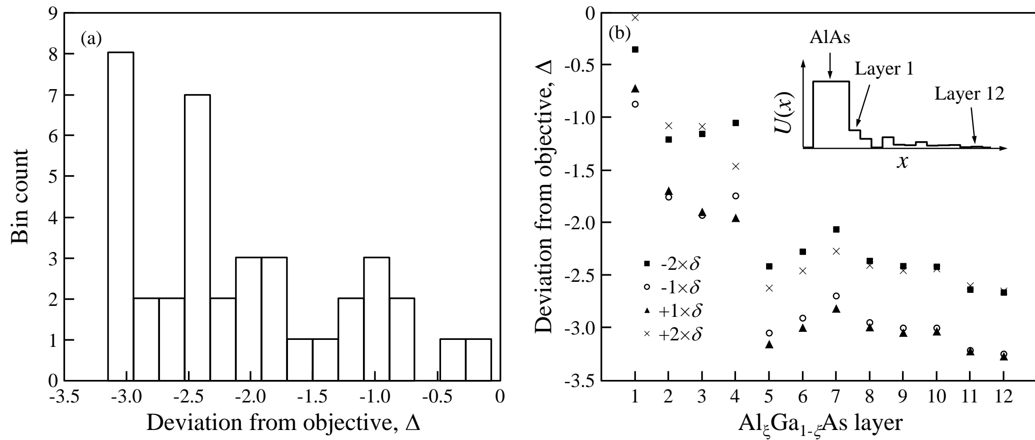

Figure 3.10: (a) Histogram showing sensitivity of the optimal design of Fig. 3.5 to errors in $\mathrm{Al}_\xi\mathrm{Ga}_{1-\xi}\mathrm{As}$ layer thickness by calculating Δ with a single $\mathrm{Al}_\xi\mathrm{Ga}_{1-\xi}\mathrm{As}$ layer perturbed by $\pm 1 \times \delta$ or $\pm 2 \times \delta$. (b) Data from part (a) plotted by $\mathrm{Al}_\xi\mathrm{Ga}_{1-\xi}\mathrm{As}$ layer. The inset shows the potential profile for the device simulated, indicating the $\mathrm{Al}_\xi\mathrm{Ga}_{1-\xi}\mathrm{As}$ layer numbers. Simulation parameters are electron effective mass $m = 0.07 \times m_0$, temperature $T = 300$ K, doping concentration $N_\mathrm{D} = 1 \times 10^{18}$ cm^{-3} .

Sensitivity to alloy fluctuations and errors in Al concentration are simulated by perturbing U_j of all $\mathrm{Al}_\xi\mathrm{Ga}_{1-\xi}\mathrm{As}$ layers by a random potential energy uniformly distributed within ± 3 meV. A histogram of 200 of these simulations for the design considered in

Fig. 3.5 is shown in Fig. 3.11. Skew of this distribution towards smaller Δ may be a result of the strongest scattering occurring in only four of the twelve $Al_{\xi}Ga_{1-\xi}As$ layers.

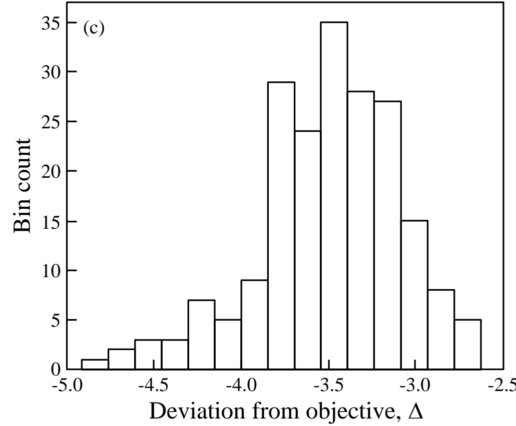

Figure 3.11: Histogram showing sensitivity of optimal design to alloy fluctuations and error in Al alloy concentration. In each simulation all $Al_{\xi}Ga_{1-\xi}As$ layers were perturbed by a random potential energy uniformly distributed between ± 3 meV. Simulation parameters are electron effective mass $m = 0.07 \times m_0$, temperature $T = 300$ K, doping concentration $N_{\text{D}} = 1 \times 10^{18}$ cm^{-3} .

Since the AlAs layer is the primary current limiter, the optimized design displays enhanced sensitivity to errors in L_0 thickness. Current through the design of Fig. 3.5 is scaled roughly by a factor of two for each single monolayer change in L_0 between $\pm 2 \times \delta$. Approximately linear behavior is preserved however, with a maximum deviation from objective of $\Delta = -2.65$ relative to a linear fit.

Changes in the Fermi-Dirac distribution with temperature will alter the current-voltage characteristic of the tunnel diode. Linearity of the current-voltage characteristic of the design in Fig. 3.5 as a function of temperature is evaluated by computing the deviation from objective with respect to a linear fit forced through the origin. Approximately linear current-voltage behavior over the temperature range 250 K$\leq T \leq$350 K is obtained with a maximum deviation from objective $\Delta = -2.95$ at temperature $T = 350$ K. Variation of r over this temperature range is exponential with approximately $r(T) = (3.21 \text{ M}\Omega \times \mu\text{m}^2) \times \exp(-T/257.3)$.

3.4 Application of customized, optimally designed diode in an RF switching mixer

The capabilities of electronic circuits are constrained by the limited set of unique electronic components available. A circuit designer attempts to find an optimal combination of available components that yields the point in the feasible solution space closest to the desired behavior. When some components are not ideally suited to an application, additional circuitry is necessary to correct or mitigate any undesirable behaviors.

To enhance the design of electronic circuits I propose a design methodology where, in addition to the circuit topology, circuit designers have access to fabrication processes and are able to create semiconductor devices that are customized for specific applications. Nano-scale electronic devices operating in a non-equilibrium electron transport regime can access a wide range of current-voltage characteristics by utilizing quantum phenomena such as tunneling and resonant electron transmission. The feasibility of this approach is made possible by availability of atomically precise crystal growth methods such as molecular beam epitaxy.

The large number of degrees of freedom available in material choice coupled with the complexity of non-equilibrium electron transport makes an $ad-hoc$ design approach difficult. To overcome these difficulties optimal design methods provide an effective and efficient approach to device design. In the following, a semiconductor heterostructure tunnel diode will be presented that has been customized for use in an RF mixer to reduce spurious mixing harmonics.

Inexpensive compute resources and direct access to semiconductor fabrication processes presents opportunities for device and circuit co-design. This approach is illustrated, in which device physics and circuit function are treated on an equal footing, by considering design of a RF mixer.

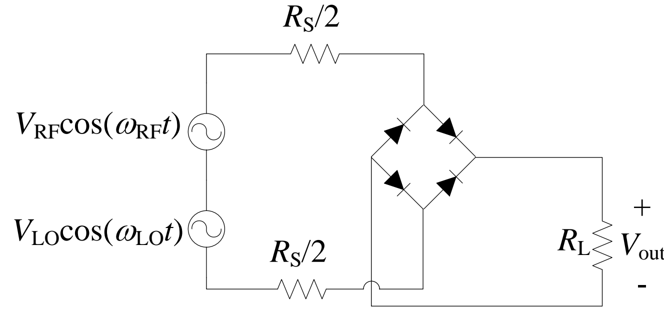

Figure 3.12: Circuit diagram of full-wave rectifier which may be used as a switching mixer to mix RF and local oscillator (LO) signals. The output signal, V_{out}, is taken across the load resistance, R_{L}, and the resistance of the source is R_{S}.

3.4.1 Ideal square-law mixer

In an RF receiver, an RF input signal of amplitude V_{RF} and frequency $\omega_{\mathrm{RF}} = 2\pi f_{\mathrm{RF}}$ is mixed with a local oscillator (LO) signal of amplitude V_{LO} and frequency $\omega_{\mathrm{LO}} = 2\pi f_{\mathrm{LO}}$ to extract a desired signal from a carrier. The LO is tuned to select a desired intermediate frequency (IF) of amplitude V_{IF} and frequency $\omega_{\mathrm{IF}} = |\omega_{\mathrm{LO}} - \omega_{\mathrm{RF}}|$. A common method of accomplishing this mixing is by using the LO signal to switch the phase of the RF signal [21]. A full-wave rectifier similar to that shown in Fig. 3.12 may be used, using either diodes as shown or transistors as switching devices. To illustrate the design concept the architecture of Fig. 3.12 is adopted. In the Fig. the series resistance of the voltage source is R_{S} and the output signal, V_{out}, is taken across the load resistance, R_{L}.

The ideal output of the switching mixer is $V_{\mathrm{out}}(t) = \frac{R_{\mathrm{L}}}{R_{\mathrm{S}}+R_{\mathrm{L}}}|V_{\mathrm{s}}(t)|$, where $V_{\mathrm{s}}(t) = V_{\mathrm{LO}}\sin(\omega_{\mathrm{LO}}t) + V_{\mathrm{RF}}\sin(\omega_{\mathrm{RF}}t)$ is the total source voltage. Typically the received RF signal is weak, requiring strong LO to drive the diodes. If $V_{\mathrm{LO}} \gg V_{\mathrm{RF}}$,

$$\begin{aligned} V_{\mathrm{out}}(t) \approx & \frac{R_{\mathrm{L}}}{R_{\mathrm{S}} + R_{\mathrm{L}}}(V_{\mathrm{LO}}\sin(\omega_{\mathrm{LO}}t)\mathrm{sgn}(\sin(\omega_{\mathrm{LO}}t)) \\ & + V_{\mathrm{RF}}\sin(\omega_{\mathrm{RF}}t)\mathrm{sgn}(\sin(\omega_{\mathrm{LO}}t))). \end{aligned} \tag{3.24}$$

Mixing of RF and LO signals occurs through the second term in this expression. Expanding the sign function in terms of its Fourier series,

$$\mathrm{sgn}(\sin(\omega_{\mathrm{LO}}t)) = \frac{4}{\pi}\sum_{n=0}^{\infty}\frac{1}{2n+1}\sin\left((2n+1)\omega_{\mathrm{LO}}t\right), \tag{3.25}$$

shows that an infinite set of harmonics are introduced into the output of the mixer. Presence of high-order harmonics will require filtering of the output signal.

The $n = 0$ term in (3.25) mixes the fundamental LO frequency with f_{RF}. Setting $R_{\mathrm{S}} = R_{\mathrm{L}}$ for maximum power transfer, the mixing term producing the IF is

$$\frac{1}{2}\frac{4}{\pi}V_{\mathrm{RF}}\sin(\omega_{\mathrm{LO}}t)\sin(\omega_{\mathrm{RF}}t), \tag{3.26}$$

In this expression the $\frac{1}{2}$ is due to voltage division, the IF voltage is

$$V_{\mathrm{IF}}(t) = \frac{V_{\mathrm{RF}}}{\pi}\cos\left((\omega_{\mathrm{LO}} - \omega_{\mathrm{RF}})t\right), \tag{3.27}$$

and power gain of approximately -10 dB is independent of V_{LO}.

The spectral intensity of the switching mixer output voltage, assuming ideal switching, is shown in Fig. 3.13. Parasitics are ignored in this proof-of-principle calculation, so that although frequencies are assigned the response is independent of frequency. Assuming ideal switching ignores finite rise-times which will be present in any physical system. Finite rise-times may be modeled as attenuation of high-frequency components of an ideal square-wave produced by ideal switching. Since the low frequency components of a square-wave produce the majority of power in the IF, finite rise-time effects should not significantly impact the mixer voltage gain provided the 10% to 90% rise-time is less than $350\ \mathrm{ps}/f_{\mathrm{LO}}(\mathrm{GHz})$. Simulations of the switching mixer with a 350 ps rise-time show a decrease in the IF power of less than 0.5 dB.

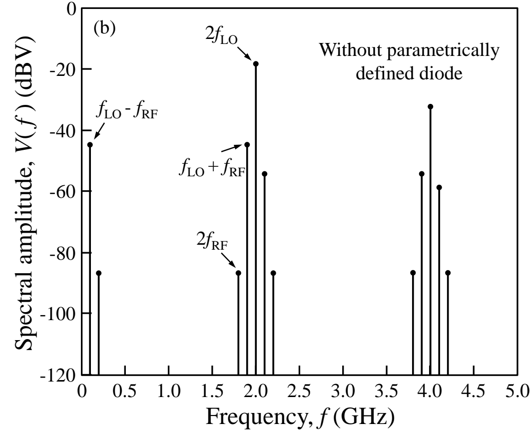

Figure 3.13: Spectral intensity of V_{out} of the switching mixer shown in Fig. 3.12 simulated using LTSPICE. Although the mixing has been accomplished, many spurious harmonics are produced. The IF output intensity is -40 dBm and the intensity at frequency $f = 2f_{\mathrm{IF}}$ is -82 dBm. Simulation parameters are LO power $P_{\mathrm{LO}} = 0$ dBm (-10 dBV) and frequency $f_{\mathrm{LO}} = 1$ GHz, RF power $P_{\mathrm{RF}} = -30$ dBm (-40 dBV) and frequency $f_{\mathrm{RF}} = 900$ MHz, source resistance $R_{\mathrm{S}} = 50\ \Omega$, and load resistance $R_{\mathrm{L}} = 600$ Ω.

The simulation of Fig. 3.13 uses LO power $P_{\mathrm{LO}} = 0$ dBm and frequency $f_{\mathrm{LO}} = 1$ GHz, RF power $P_{\mathrm{RF}} = -30$ dBm and frequency $f_{\mathrm{RF}} = 900$ MHz, source resistance $R_{\mathrm{S}} = 50\ \Omega$, and load resistance $R_{\mathrm{L}} = 50\ \Omega$. In this configuration the power contained in the IF is -40 dBm, for a simulated -10 dBm power gain. Harmonics produced include even-order LO terms and harmonics at frequencies $(2n+1)f_{\mathrm{LO}} \pm f_{\mathrm{RF}}$, from (3.24), as well as additional terms due to ideal mixing. The fourth order mixing term at frequency $f = 2f_{\mathrm{IF}} = 200$ MHz, containing -82 dBm of power, can be difficult to filter using conventional techniques, reducing the dynamic range of the switching mixer. Reduction of this harmonic by an additional 48 dB without reducing the power gain of the mixing term would require at least an 8 pole filter, whereas a receiver may contain fourth- or fifth-order filters [15, 43].

3.4.2 Improved performance of switching mixer with a customized diode

It is desirable to reduce the amplitudes of the harmonics in order to reduce filtering requirements necessary to achieve a fixed dynamic range. One way these harmonics could be reduced is by squaring the output of the switching mixer. Consider the LO output signal $V_{\mathrm{LO}}\sin(\omega_{\mathrm{LO}}t)\mathrm{sgn}\left(\sin(\omega_{\mathrm{LO}}t)\right)$. Fig. 3.14(a) shows the effects of squaring this signal in the time-domain. The square-law transfer characteristic converts a rectified sine wave into a pure sine wave by reshaping features in the signal, including smoothing discontinuities in the derivative at $V_{\mathrm{bias}} = 0$.

Fig. 3.14(b) shows the effects of squaring a sign function in the frequency domain. The square wave signal of a sign function contains the frequency components given in (3.25), and the power contained in these frequency components is shown as dashed lines in the Fig. A square-law transfer characteristic mixes these frequencies such that all harmonics are canceled and power is contained in only the zero frequency term. Thus, a component designed to square the output of the mixer may be able to reduce spurious harmonics in the output signal.

Squaring (3.24) and setting $R_{\mathrm{S}} = R_{\mathrm{L}}$,

$$V_{\mathrm{out}}(t) = \frac{1}{4}V_{\mathrm{LO}}^2\sin^2(\omega_{\mathrm{LO}}t) + \frac{1}{2}V_{\mathrm{LO}}V_{\mathrm{RF}}\cos(\omega_{\mathrm{LO}}t)\cos(\omega_{\mathrm{RF}}t) + \frac{1}{4}V_{\mathrm{RF}}^2\sin^2(\omega_{\mathrm{RF}}t), \tag{3.28}$$

having the desired mixing term $V_{\mathrm{IF}}(t) = \frac{1}{4}V_{\mathrm{LO}}V_{\mathrm{RF}}\cos\left((\omega_{\mathrm{LO}} - \omega_{\mathrm{RF}})t\right)$, and only second-order spurious harmonics. Additionally the IF amplitude now depends on V_{LO}, allowing for variable mixer gain.

To accomplish the squaring action a customized semiconductor heterostructure tunnel diode is designed and the diode placed in series with the load. For the voltage across

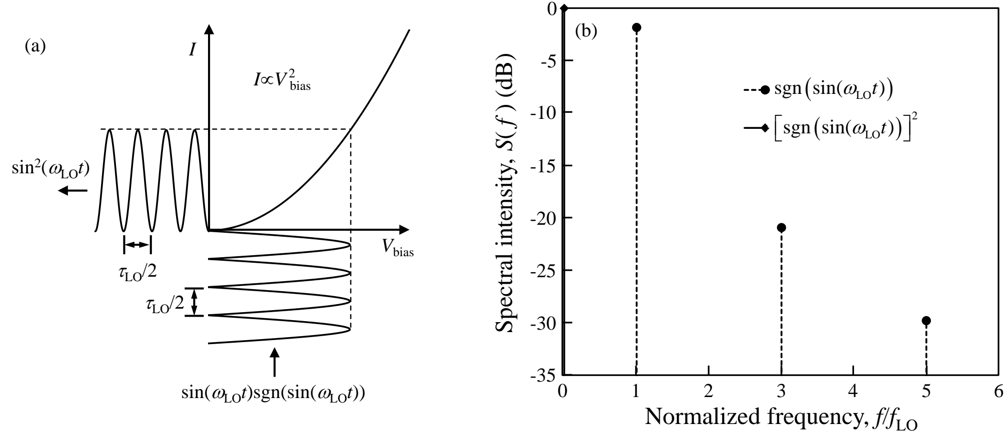

Figure 3.14: (a) Diagram showing a rectified sinusoidal input to a perfect square-law device yields a squared sine wave. (b) Frequency components of sign function and square of the sign function. The square-law transfer characteristics redistributes the power from an infinite set of harmonics into only the DC component.

R_L to be proportional to the square of the rectified signal, it is desired that the current through the load equal

$$I(t) = \alpha |V_\mathrm{s}(t)|^2, \tag{3.29}$$

where α is a constant coefficient. Naively, one may expect a diode featuring a quadratic current-voltage characteristic to satisfy (3.29). However, summing voltages around a circuit with such a device yields

$$|V_\mathrm{s}(t)| - \sqrt{\frac{I(t)}{\alpha}} - I(t)\left(R_\mathrm{S} + R_\mathrm{L}\right) = 0, \tag{3.30}$$

resulting in current that is not proportional to $|V_\mathrm{s}(t)|^2$.

To satisfy (3.29) I will not use a conventional approach of designing a circuit in which currently existing (and often active) circuit elements are configured to most closely approximate the desired behavior. Rather, a single diode featuring a customized current-voltage characteristic is designed to enable the switching mixer to satisfy (3.29).

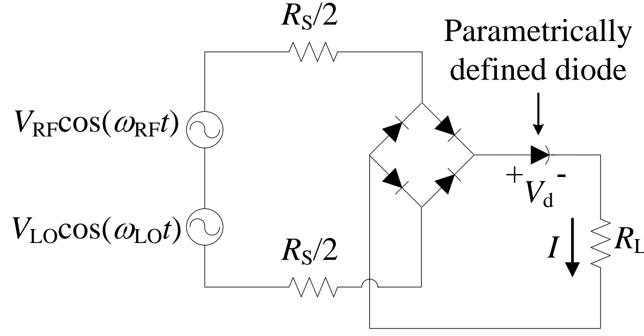

Figure 3.15: Circuit diagram of switching mixer utilizing diode characterized by a parametrically defined current-voltage characteristic to eliminate unwanted harmonics. Inputs to the circuit are RF signal of amplitude V_{RF} and frequency ω_{RF} and LO signal of amplitude V_{LO} and frequency ω_{LO}. The source and load resistances are R_{S} and R_{L}, respectively.

Suppose a diode can be designed that is capable of satisfying (3.29). Given this assumption the voltage across the diode must be

$$V_{\mathrm{d}}(t) = |V_{\mathrm{s}}(t)| - \alpha |V_{\mathrm{s}}(t)|^2 (R_{\mathrm{S}} + R_{\mathrm{L}}). \tag{3.31}$$

The required diode current-voltage characteristic is defined parametrically in terms of (3.29) and (3.31).

Defining the current-voltage characteristic of a device parametrically to satisfy the requirements of an application presents a new method of circuit design. In this method, both circuit topology and device current-voltage characteristics are co-designed for a specific application. This approach is made feasible by efficient design of heterostructure tunnel diodes using principles of optimal design.

Because the current-voltage characteristic is defined parametrically, the customized diode is not guaranteed to be physically feasible. An example current-voltage characteristic is shown in Fig. 3.16 for $\alpha = 6.9$ mA/V^2, source resistance $R_{\mathrm{S}} = 50\ \Omega$, and load resistance $R_{\mathrm{L}} = 50\ \Omega$. It can be seen that the parametrically defined current-voltage

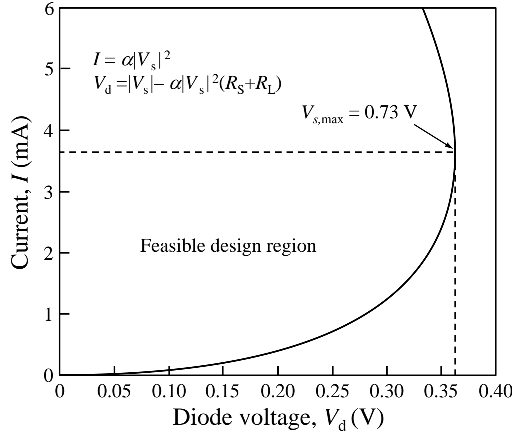

Figure 3.16: Parametrically defined diode current-voltage characteristic given by the equations in the Fig. A feasible design region is indicated bounded by maximum source voltage, $V_{\mathrm{s,max}}$, where the slope $\partial I/\partial V_{\mathrm{d}} = \infty$. Simulation parameters are $\alpha = 6.9$ mA/V^2, source resistance $R_{\mathrm{S}} = 50\ \Omega$, and load resistance $R_{\mathrm{L}} = 50\ \Omega$.

characteristic exhibits multi-valued current, which is potentially unstable, as well as an infinite slope. However, over the range $V_{\mathrm{s}} < V_{\mathrm{s,max}}$ a design is feasible, as indicated in the Fig.

Designing a tunnel diode capable of achieving the large slope in the current-voltage characteristic near $V_{\mathrm{s}} = V_{\mathrm{s,max}}$ may be difficult. Designs that can capture this feature may include physical effects such as sharp electron transmission resonances or may operate at low temperatures to take advantage of a step-like Fermi-Dirac distribution. For a proof-of-principle demonstration I have chosen to ease the design by optimizing over a smaller range of source voltage at the expense of reducing the IF voltage gain from the theoretical maximum.

Physical understanding of features in the parametrically defined current-voltage characteristic can be gained by comparing operation of a switching mixer with and without the parametrically defined diode. Figure 3.17 compares the current-voltage characteristics of the two mixers with $R_{\mathrm{S}} = 50\ \Omega$, $R_{\mathrm{L}} = 50\ \Omega$, and $\alpha = 6.9$ mA/V^2. Because a passive device is considered, the voltage across the diode and the current through the

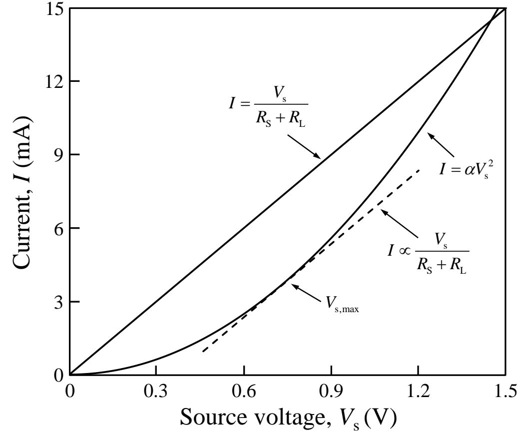

Figure 3.17: Comparison of current-voltage characteristics for series circuits with and without the parametrically defined diode. Maximum source voltage, $V_{\mathrm{s,max}}$, is defined as the point at which the slope of the parabola matches that of the line. Parameters for the Fig. are source resistance $R_{\mathrm{S}} = 50\ \Omega$, load resistance $R_{\mathrm{L}} = 50\ \Omega$, and $\alpha = 6.9$ mA/V^2, resulting in $V_{\mathrm{s,max}} = 0.73$ V.

diode must have the same sign. Therefore the current through the mixer with the diode must be less than that of the mixer without the diode, due to a finite, positive voltage drop across the diode reducing the voltage across the resistors.

The parametrically defined diode can be modeled as a voltage dependent resistor, labeled R_{d}. To satisfy (3.29), the net series resistance of mixer with the diode must be

$$R(V_{\mathrm{s}}) = \frac{1}{\alpha V_{\mathrm{s}}} = R_{\mathrm{D}}(V_{\mathrm{s}}) + R_{\mathrm{S}} + R_{\mathrm{L}}. \tag{3.32}$$

Comparing the slopes of the curves in Fig. 3.17, the diode increases the series resistance for $V_{\mathrm{s}} \leq V_{\mathrm{s,max}}$ such that $R(V_{\mathrm{s}}) > R_{\mathrm{S}}+R_{\mathrm{L}}$. At voltage $V_{\mathrm{s,max}}$, the two mixers have equal resistances, requiring $R_{\mathrm{d}} = 0\ \Omega$ and necessitating the infinite slope in Fig. 3.16. As V_{s} increases further, the resistance of the diode decreases to facilitate the quadratically increasing current.

$V_{\mathrm{s,max}}$ is defined as the source voltage at which $\frac{\partial I}{\partial V_{\mathrm{d}}} = \infty$, or $\frac{\partial V_{\mathrm{d}}}{\partial V_{\mathrm{s}}} = 0$. Thus,

$$V_{\mathrm{s,max}} = \frac{1}{2\alpha(R_{\mathrm{S}} + R_{\mathrm{L}})}, \tag{3.33}$$

resulting in a maximum voltage across the diode of

$$V_{\mathrm{d,max}} = \frac{1}{4\alpha(R_{\mathrm{S}} + R_{\mathrm{L}})}. \tag{3.34}$$

To determine the performance limits of a switching mixer utilizing a parametrically defined diode, it is assumed that $V_{\mathrm{RF}} \ll V_{\mathrm{LO}}$ such that $V_{\mathrm{LO}} \approx V_{\mathrm{max}}$. Then substituting the rectified source voltage

$$\begin{aligned} |V_{\mathrm{s}}(t)| =& V_{\mathrm{LO}} \sin(\omega_{\mathrm{LO}} t) \mathrm{sgn}(\sin(\omega_{\mathrm{LO}} t)) \\ &+ V_{\mathrm{RF}} \sin(\omega_{\mathrm{RF}} t) \mathrm{sgn}(\sin(\omega_{\mathrm{LO}} t)), \end{aligned} \tag{3.35}$$

into (3.29), the IF current is

$$\begin{aligned} I_{\mathrm{RF}} =& \alpha V_{\mathrm{LO}} V_{\mathrm{RF}} \cos\left((\omega_{\mathrm{LO}} - \omega_{\mathrm{RF}})t\right) \\ =& \frac{V_{\mathrm{RF}}}{2(R_{\mathrm{S}} + R_{\mathrm{L}})} \cos\left((\omega_{\mathrm{LO}} - \omega_{\mathrm{RF}})t\right), \end{aligned} \tag{3.36}$$

where the expression for (3.33) has been substituted for V_{LO}. The average IF power delivered to the load is

$$P_{\mathrm{IF}} = \frac{1}{2} I_{\mathrm{IF}}^2 R_{\mathrm{L}} = \frac{V_{\mathrm{RF}}^2}{8R_{\mathrm{t}}^2} R_{\mathrm{L}} = \frac{1}{16} \frac{V_{\mathrm{RF}}^2}{2R_{\mathrm{S}}}. \tag{3.37}$$

Since $P_{\mathrm{RF}} = \frac{V_{\mathrm{RF}}^2}{2R_{\mathrm{S}}}$, maximum power gain is -12 dB. Restricting the value of α, so that the parametric diode operates with $V_{\mathrm{s}} < V_{\mathrm{s,max}}$, reduces the power gain.

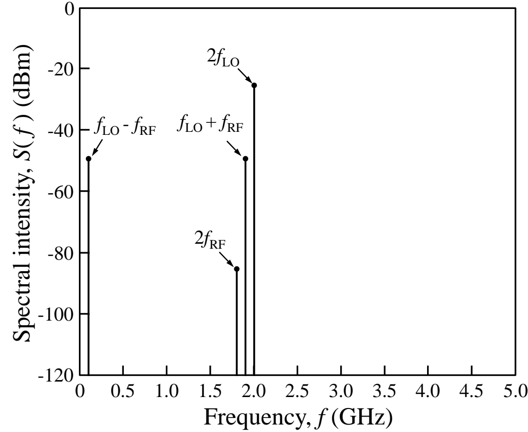

Figure 3.18: Spectral intensity of output of switching mixer containing parametrically defined diode. The customized current-voltage characteristic of the ideal diode eliminates all spurious harmonics above the second-order terms. The IF power is -49 dBm, for a power gain of -19 dB. Simulation parameters are LO power $P_{\mathrm{LO}} = 0$ dBm and frequency $f_{\mathrm{LO}} = 1$ GHz, RF power $P_{\mathrm{RF}} = -30$ dBm and frequency $f_{\mathrm{RF}} = 900$ MHz, source resistance $R_{\mathrm{S}} = 50\ \Omega$, and load resistance $R_{\mathrm{L}} = 50\ \Omega$. The parametrically defined diode is characterized by $\alpha = 6.9$ mA/V^2.

The output power of a switching mixer utilizing a parametrically defined diode simulated using LTSPICE is shown in Fig. 3.18. For comparison with Fig. 3.13, the mixers in both Figs. use identical simulation parameters. It can be seen that the parametric diode has reduced spurious harmonics to only second-order harmonics and IF power gain is -19 dB. With the ideal parametrically defined diode the lowest frequency spurious harmonic occurs at frequency $2f_{\mathrm{RF}}$, easing the requirements of the filter and increasing dynamic range.

3.4.3 Voltage gain vs. power gain

With power gain of -19 dB, reduction of the harmonics has come at the expense of a 9 dB decrease in power gain. However, it is often the case that a mixer is followed by an amplifier featuring a high input impedance. In this case the *voltage gain* of a mixer is a more suitable figure of merit.

For the two mixers considered, the voltage gain is

$$A_{\rm v} = \frac{2}{\pi}\frac{R_{\rm L}}{R_{\rm L}+R_{\rm S}}, \tag{3.38}$$

without the parametrically defined diode and

$$A_{\rm v} = \frac{1}{2}\alpha V_{\rm LO} R_{\rm L}, \tag{3.39}$$

with the parametrically defined diode. Substituting $V_{\rm s,max}$ for $V_{\rm LO}$, the maximum voltage gain with the parametrically defined diode is

$$A_{\rm v,max} = \frac{1}{2}\frac{R_{\rm L}}{R_{\rm L}+R_{\rm S}}. \tag{3.40}$$

In the following, it will be assumed that the load resistance is $R_{\rm L} = 600\ \Omega$, and the input power is referred to source resistance of $R_{\rm S} = 50\ \Omega$. With these values the switching mixer without a parametrically defined diode features a voltage gain of -4.6 dBV, the maximum voltage gain with the parametrically defined diode is -6.7 dBV. The voltage gain of the mixer in Fig. 3.18 is -14.5 dBV.

3.4.4 Optimal design of parametrically defined diode

A heterostructure tunnel diode optimized for the current of the operating region of Fig. 3.16 is shown in Fig. 3.19. The diode features $N = 10$ $\mathrm{Al}_{\xi}\mathrm{Ga}_{1-\xi}\mathrm{As}$ layers, AlAs layer thickness $L_0 = 5 \times \delta$, and n-type doping concentration $n = 1 \times 10^{19}$ cm^{-3}. The objective function is characterized by $\alpha = 1.06$ mA/V^2, and the cross-sectional area $A = 1.6\ \mu$m^2 scales the current to that of Fig. 3.16. This value of $\alpha = 1.06$ mA/V^2 with $R_{\rm L} + R_{\rm S} = 650\ \Omega$ produces an equivalent current-voltage to that of $\alpha = 6.9$ mA/V^2 with $R_{\rm L} + R_{\rm S} = 100$ used previously.

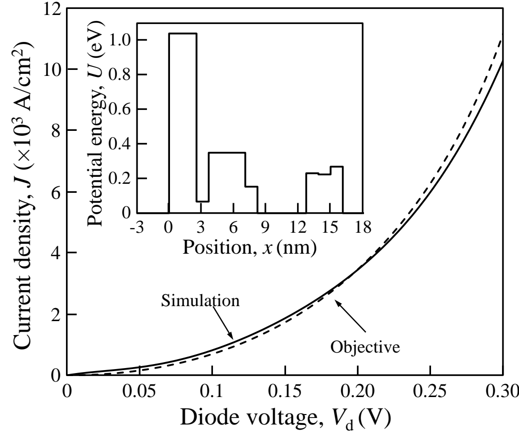

Figure 3.19: Simulated current density through a $\mathrm{Al}_{\xi}\mathrm{Ga}_{1-\xi}\mathrm{As}$ heterostructure tunnel barrier diode optimized for current shown in Fig. 3.16 given a cross-sectional area of $A = 1.6\ \mu\mathrm{m}^2$. Shown in the inset is the conduction band potential profile for the device. Simulation parameters are temperature $T = 300$ K, electron effective mass $m = 0.07 \times m_0$, n-type doping concentration $n = 1 \times 10^{19}$ cm^{-3}. The objective function is characterized by $\alpha = 1.06$ mA/V^2, and $R_{\mathrm{tot}} = 100\ \Omega$.

Shown in the inset is the conduction band potential profile for the device. Physical features in this potential profile, such as the two potential wells, are selected by the optimization algorithm to control propagating electrons. Electron transmission resonances most significantly enhance electron transmission probability when their energies are near those of occupied electron states. The majority of these occupied electron states reside below the Fermi energy $E_{\mathrm{F}} = 52$ meV. The wider potential well at position $x = 10$ nm will feature lower energy resonances than those of the narrower potential well at position $x = 3$ nm. Thus, as V_{d} lowers the potential of the right-hand contact, the lower energy resonances of the wider potential well will be pulled through the occupied electron states first. The resonances control the increase of current at low bias. The resonances of the narrower potential are pulled through the occupied electron states at higher V_{d}, helping increase the current at a faster rate.

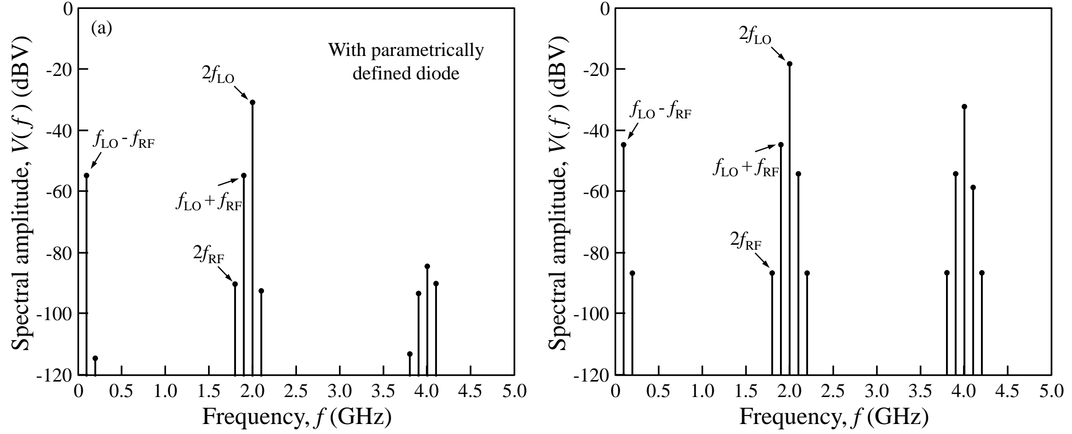

Figure 3.20: (a) Spectral intensity of output of switching mixer using optimized parametrically defined diode. The IF voltage gain for this mixer is -14.7 dBV and the power in the fourth order harmonic at frequency $f = 2f_{\mathrm{IF}}$ is -114.3 dBV. (b) Spectral intensity of output of switching mixer without parametrically defined diode. The IF voltage gain for this mixer is -4.6 dBV and the power in the fourth order harmonic at frequency $f = 2f_{\mathrm{IF}}$ is -86.7 dBV. Simulation parameters are LO power $P_{\mathrm{LO}} = 0$ dBm and frequency $f_{\mathrm{LO}} = 1$ GHz, RF power $P_{\mathrm{RF}} = -30$ dBm and frequency $f_{\mathrm{RF}} = 900$ MHz, source resistance $R_{\mathrm{S}} = 50\ \Omega$, and load resistance $R_{\mathrm{L}} = 600\ \Omega$, and $\alpha = 1.06$ mA/V^2.

3.4.5 Switching mixer utilizing optimized diode

LTSPICE simulations of the output spectral amplitude for a switching mixer utilizing the optimized diode is shown in Fig. 3.20(a). Because the load resistance has been increased and the voltage gain is of interest, Fig. 3.20(b) shows the output spectral amplitude of the switching mixer without the parametrically defined diode. Although non-idealities have introduced additional harmonics compared with Fig. 3.18, many harmonics produced by the switching action remain suppressed. In particular, the -114 dBV amplitude of the $2f_{\mathrm{IF}}$ harmonic is 28 dBV less than that produced by the switching mixer, reducing the number of poles necessary for 98 dB dynamic range by 4. The IF voltage gain with the parametrically defined diode of -14.7 dBV is only 0.2 dBV less than the ideal parametrically defined diode.

3.5 Conclusion and future work

In conclusion, the design of electronic heterostructure devices can utilize atomic layer precision in semiconductor fabrication to mimic an Ohmic (linear) current-voltage characteristic over a specific range of V_{bias} . Even though electron transport is dominated by tunneling, it is possible to manipulate low energy electron scattering and resonances to approach a desired linear current-voltage behavior. However, practical and materials constraints limit the degree of linearity that may be achieved. Here, it was demonstrated how greater than 62 dB (10 bit) dynamic range may be obtained in design of an optimized semiconductor heterostructure tunnel device that is only 17 nm thick. This type of control, in which there is no inelastic electron scattering in the active region, is only possible because of the wave nature of electrons.

It has also been shown that, using principles of optimal design, it is possible to design a semiconductor heterostructure tunnel diode customized for a specific application. Here a switching mixer is considered and it was shown that a diode featuring a parametrically defined current-voltage characteristic can reduce the number of spurious harmonics produced by this mixer. The optimized diode was shown to reduce the amplitude of the $2f_{\text{IF}}$ harmonic by 28 dB, increasing the dynamic range of the mixer. The ability to customize device behavior for specific applications is a new approach to circuit design. Using optimal design to discover devices that behave in new ways, such as featuring a parametrically defined current-voltage characteristic, has the potential to enhance the functionality of many currently existing circuits, and may enable access to previously unavailable circuit behaviors.

Optimization with respect to temperature variation provides a challenging direction for future work. In the physical model used only the Fermi-Dirac distribution varies as a function of temperature. This variation affects both current and the self-consistent potential. However, changes in potential are not strong enough to be used to control the

temperature dependence of current. A more advanced model may incorporate additional scattering mechanisms such as electron-electron scattering, in which scattering rates increase with temperature. This would tend to increase the resistance with temperature, and it may be possible to access optimal designs in which no change in resistance occurs over a desired temperature range.

Finally, the trends in Fig. 3.7 could be the result of a fundamental physical mechanism. Future work should include further data collection to develop statistically significant trends. It is expected that this data will confirm the existence of an optimal value for N. An investigation of the physical mechanism responsible for this trend could show mathematically the limits to optimal design in these heterostructure diodes. Similar work has been conducted in which the limit to the performance of linear optical components was found in terms of only their size, shape, and dielectric constants. [28]

Chapter 4

Inelastic electron transport

In Chapter 1, loss of physics from the use of perturbative methods on the Frölich interaction was discussed. In using Fermi's golden rule to calculate the inelastic electron scattering rate due to electron-phonon interactions, feedback effects were ignored between the individual scattering states. Necessary for this feedback is retention of the emitted phonons within the model, allowing processes involving multiple phonon emissions and absorptions to contribute to the solution. Since first-order perturbation theory does not keep track of the phonons, the electron and phonons interact in an incoherent manner.

To obtain additional insight into the physics of the electron-phonon interaction it is necessary to consider a quantum model in which the electron and phonons may interact coherently. The relationship between the incoherent and coherent scattering regimes may be visualized as in the diagram in Fig. 4.1. Before an emitted optical phonon dissipates into acoustic phonons the electron and phonon may interact in a coherent manner. This regime is represented by the shaded circle, with an artificial boundary drawn to guide the eye.

In a perturbative formalism, it is assumed that each scattering event occurs only after the phonon has dissipated into a bath. A quantum model should be able to transition between the coherent and incoherent regimes, providing additional physical insight into perturbative methods. In this chapter I present one such connection. The electron-phonon interaction is solved with a simple, exactly solvable quantum model, and it is shown that under certain circumstances predictions of semi-classical perturbation theory can emulate the predictions of an exact quantum solution.

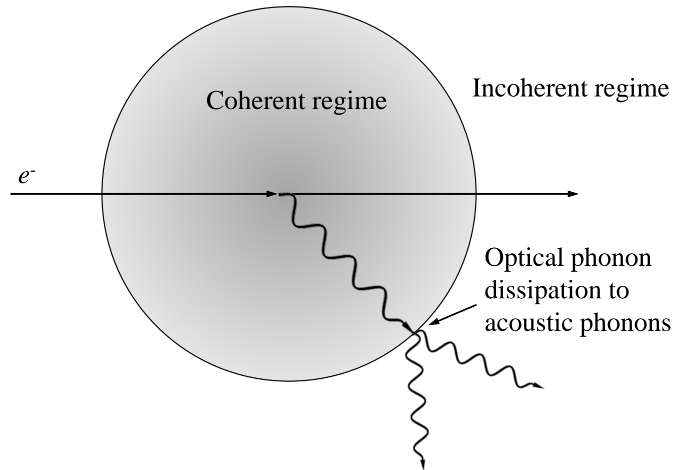

Figure 4.1: Diagram depicting coherent and incoherent regimes for interaction between an electron and optical phonon. The electron and phonon may interact in a coherent manner until the optical phonon dissipates into acoustic phonons.

To solve the inelastic scattering problem non-perturbatively an established prototype model [13, 2, 14, 9] is used that considers the interaction between an electron and a localized vibrational mode (an Einstein phonon). This model has been used to investigate feedback-related features such as fano-like resonances in the transmission [4], impurity band formation in quantum wells [33], and inelastic effects on wave packet propagation [30]. The model has a very rich solution space which may be attributed to treatment of the electron-phonon interaction as a coherent and unitary process even though it does not include many-body and some self-consistent effects such as energy level shifts, collisional broadening, and electron-electron interactions. In the following, it is assumed that it is more important that an exact solution be obtained to a simpler model than a perturbative solution be obtained to a more complex model, such as Ref. 11.

4.1 Quantum model of electron-phonon interaction

Consider the inelastic system shown in Fig. 4.2, where the potential barrier is an inelastic "black box". Only one-dimensional inelastic systems are considered, so that an electron in initial state $|k_0\rangle$ with energy E_0 enters from the left and inelastically scatters into a final state $|k_n\rangle$ with energy E_n. When the inelastic collision occurs, the electron may either emit a phonon or absorb a phonon of energy $\hbar\omega_0$ and momentum q, provided the system is not at absolute zero. If the electron has energy $E_0 < \hbar\omega_0$, then only virtual phonons will be emitted.

To conserve energy and momentum, the electron's final energy is $E_n = E_0 - n\hbar\omega_0$, where n is the number of phonons that have been excited, and it's final momentum is $k_n = k_0 - \sum_j q_j$, where the sum is over all phonons the electron has interacted with. The conventions for n and q are such that positive n denotes a net phonon emission and $-q_j$ is the momentum that the jth phonon carries away. At temperature $T = 0$ K, the existence of a ground state restricts $n \geq 0$.

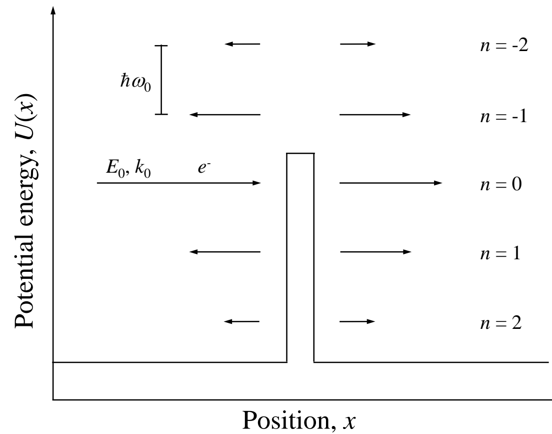

Figure 4.2: Reflection and transmission due to an inelastic system. An electron initially in state $|k_0\rangle$ with energy E_0 inelastically scatters into final state $|k_n\rangle$ with energy $E_n = E_0 - n\hbar\omega_0$. Positive n denotes net phonon emission and the phonon energy is $\hbar\omega_0$.

Inelastic scattering between an electron and dispersionless phonons is modeled by Einstein phonons linearly coupled to the electron wave function. One-dimensional electron transport in the x-direction in a semiconductor is considered. It is assumed that the electron density is sufficiently small that electron-electron interactions may be ignored. For an electron of effective mass m interacting with Einstein phonons, the Hamiltonian is [13]

$$\hat{H} = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + U(x) + \hbar\omega_0 \hat{b}^\dagger \hat{b} + \left(g_0 + g_1(\hat{b}^\dagger + \hat{b})\right)\delta(x - x_0), \tag{4.1}$$

where the first two terms are kinetic energy and potential energy of the electron respectively, the third term is the total energy stored in the phonons, and the final term couples the electron to the Einstein phonons located at position $x = x_0$. The creation and annihilation operators for a phonon of energy $\hbar\omega_0$ are $\hat{b}^\dagger$ and $\hat{b}$ respectively. The static amplitude of the delta is g_0, and the electron-phonon coupling constant is g_1, both having units of energy times distance. It is assumed that a conduction band electron interacts with longitudinal-optic (LO) phonons via the Frölich interaction in the semiconductor GaAs. The effective electron mass is $m = 0.07 \times m_0$, where m_0 is the bare electron mass, and the phonon energy is $\hbar\omega_0 = 36$ meV. Because the phonons are not lost to an external bath, there is no mechanism for dissipation and hence no concept of temperature within the model. However, the initial condition of the system will have zero excited phonons and it is in this respect that the temperature will be considered $T = 0$ K.

To solve the Schrödinger equation the wave function is expanded in terms of the oscillator basis

$$\langle x|\Psi\rangle = \sum_{n=0}^{n=\infty} \psi_n(x)|n\rangle, \tag{4.2}$$

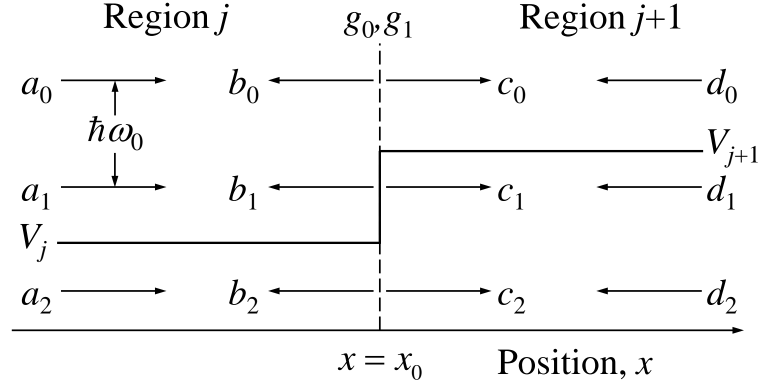

Figure 4.3: Diagram of a potential step with Einstein phonons localized at position $x = x_0$ (dashed line). Electron waves are incident from the left and right with amplitudes a_n and d_n, respectively. Waves will be scattered from position $x = x_0$ to the left and to the right having amplitudes b_n and c_n, respectively. The electron-phonon coupling constant is g_1, the static amplitude of the delta barrier is g_0, the phonon energy is $\hbar\omega_0$, and the temperature is $T = 0$ K.

with the number of excited phonons given by n. Here it is assumed that the electron, ψ, and phonon, $|n\rangle$, wave functions are separable. The domain of interest is discretized in space into $N + 1$ regions and the static potential $U(x)$ is approximated as a series of potential steps. The diagram in Fig. 4.3 shows two regions of the domain separated by a potential step and Einstein phonons localized at position $x = x_0$. In a plane wave basis, the electron wave functions of channel n in regions j and $j + 1$ are

$$\psi_n^j(x) = a_n e^{ik_n^j x} + b_n e^{-ik_n^j x}, \tag{4.3}$$

$$\psi_n^{j+1}(x) = c_n e^{ik_n^{j+1} x} + d_n e^{-ik_n^{j+1} x}, \tag{4.4}$$

where

$$k_n^j = \frac{\sqrt{2m\left(E_n - U_j\right)}}{\hbar}. \tag{4.5}$$

For the quantum harmonic oscillator,

$$\hat{b}^\dagger \hat{b} \psi_n |n\rangle = n\psi_n |n\rangle, \tag{4.6}$$

$$\hat{b}^\dagger \psi_n |n\rangle = \sqrt{n+1}\psi_{n+1} |n+1\rangle, \tag{4.7}$$

$$\hat{b} \psi_n |n\rangle = \sqrt{n}\psi_{n-1} |n-1\rangle. \tag{4.8}$$

Applying these operators to the wave function for channel n in the Schrödinger equation and rearranging,

$$\begin{aligned} \frac{\hbar^2}{2m} \frac{\partial^2 \psi_n(x)}{\partial x^2} = & \left(n\hbar\omega_0 + U(x) - E_0\right) \psi_n(x) \\ & + \delta(x) \left(\sqrt{n} g_1 \psi_{n-1}(x) + g_0 \psi_n(x) + \sqrt{n+1} g_1 \psi_{n+1}(x)\right). \end{aligned} \tag{4.9}$$

As only the electron wave function is of interest here, notation for the phonon waves function are omitted. Integrating over the position range $x = x_0^-$ to $x = x_0^+$,

$$\begin{aligned} \left.\frac{d\psi_n^{j+1}(x)}{dx}\right|_{x=x_0^+} - \left.\frac{d\psi_n^{j}(x)}{dx}\right|_{x=x_0^-} = & \frac{2m}{\hbar^2} \left(\sqrt{n} g_1 (c_{n-1} + d_{n-1}) + g_0 (c_n + d_n)\right. \\ & \left. + \sqrt{n+1} g_1 (c_{n+1} + d_{n+1})\right). \end{aligned} \tag{4.10}$$

Here continuity of the wave function across the interface, $\psi_n^j(x_0) = \psi_n^{j+1}(x_0)$, causes the integration of the first term on the right hand side of (4.9) to equal zero. Substituting the wave function into (4.10) and enforcing continuity of the wave function at position $x = x_0$ yields

$$
\begin{aligned}
a_n =& i\frac{mg_1}{\hbar^2 k_n^j}\sqrt{n}(c_{n-1}+d_{n-1}) + \left[\frac{1}{2}\left(1+\frac{k_n^{j+1}}{k_n^j}\right)+i\frac{mg_0}{\hbar^2 k_n^j}\right]c_n \\
&+ \left[\frac{1}{2}\left(1-\frac{k_n^{j+1}}{k_n^j}\right)+i\frac{mg_0}{\hbar^2 k_n^j}\right]d_n + i\frac{mg_1}{\hbar^2 k_n^j}\sqrt{n+1}(c_{n+1}+d_{n+1}),
\end{aligned}
\tag{4.11}
$$

and

$$
\begin{aligned}
b_n =& -i\frac{mg_1}{\hbar^2 k_n^j}\sqrt{n}(c_{n-1}+d_{n-1}) + \left[\frac{1}{2}\left(1-\frac{k_n^{j+1}}{k_n^j}\right)-i\frac{mg_0}{\hbar^2 k_n^j}\right]c_n \\
&+ \left[\frac{1}{2}\left(1+\frac{k_n^{j+1}}{k_n^j}\right)-i\frac{mg_0}{\hbar^2 k_n^j}\right]d_n - i\frac{mg_1}{\hbar^2 k_n^j}\sqrt{n+1}(c_{n+1}+d_{n+1}).
\end{aligned}
\tag{4.12}
$$

Letting

$$
\begin{aligned}
\alpha_n^j &= i\frac{mg_1}{\hbar^2 k_n^j}\sqrt{n}, & \beta_n^j &= i\frac{mg_1}{\hbar^2 k_n^j}\sqrt{n+1}, \\
\gamma_{n,ac}^j &= \frac{1}{2}\left(1+\frac{k_n^{j+1}}{k_n^j}\right)+i\frac{mg_0}{\hbar^2 k_n^j}, & \gamma_{n,ad}^j &= \frac{1}{2}\left(1-\frac{k_n^{j+1}}{k_n^j}\right)+i\frac{mg_0}{\hbar^2 k_n^I}, \\
\gamma_{n,bc}^j &= \frac{1}{2}\left(1-\frac{k_n^{j+1}}{k_n^j}\right)-i\frac{mg_0}{\hbar^2 k_n^j}, & \gamma_{n,bd}^j &= \frac{1}{2}\left(1+\frac{k_n^{j+1}}{k_n^j}\right)-i\frac{mg_0}{\hbar^2 k_n^I},
\end{aligned}
$$

(4.11) and (4.12) become

$$
a_n = \alpha_n^j\left(c_{n-1}+d_{n-1}\right) + \gamma_{n,ac}^j c_n + \gamma_{n,ad}^j d_n + \beta_n^j\left(c_{n+1}+d_{n+1}\right), \tag{4.13}
$$

$$
b_n = -\alpha_n^j\left(c_{n-1}+d_{n-1}\right) + \gamma_{n,bc}^j c_n + \gamma_{n,bd}^j d_n - \beta_n^j\left(c_{n+1}+d_{n+1}\right). \tag{4.14}
$$

This recursion relation can be solved in matrix form. At temperature $T = 0$ K, the matrix equation is

$$\begin{bmatrix} a_0 \\ b_0 \\ a_1 \\ b_1 \\ \vdots \end{bmatrix} = \begin{bmatrix} \gamma^j_{0,ac} & \gamma^j_{0,ad} & \beta^j_0 & \beta^j_0 & 0 & 0 & \dots \\ \gamma^j_{0,bc} & \gamma^j_{0,bd} & -\beta^j_0 & -\beta^j_0 & 0 & 0 & \\ \alpha^j_1 & \alpha^j_1 & \gamma^j_{1,ac} & \gamma^j_{1,ad} & \beta^j_1 & \beta^j_1 & \\ -\alpha^j_1 & -\alpha^j_1 & \gamma^j_{1,bc} & \gamma^j_{1,bd} & -\beta^j_1 & -\beta^j_1 & \\ \vdots & & & & & & \ddots \end{bmatrix} \begin{bmatrix} c_0 \\ d_0 \\ c_1 \\ d_1 \\ \vdots \end{bmatrix}, \tag{4.15}$$

$$\mathbf{a} = \mathbf{p}^{\text{step}}_j \mathbf{c}. \tag{4.16}$$

4.1.1 Solution to Schrödinger equation using the propagation matrix method

The above matrix equation is the step matrix for a system containing Einstein phonons. Multiplying by the free propagation matrix,

$$\mathbf{p}_j = \begin{bmatrix} \gamma^j_{0,ac} e^{-ik^j_0 L_j} & \gamma^j_{0,ad} e^{ik^j_0 L_j} & \beta^j_0 & \beta^j_0 & 0 & 0 & \dots \\ \gamma^j_{0,bc} e^{-ik^j_0 L_j} & \gamma^j_{0,bd} e^{ik^j_0 L_j} & -\beta^j_0 & -\beta^j_0 & 0 & 0 & \\ \alpha^j_1 & \alpha^j_1 & \gamma^j_{1,ac} e^{-ik^j_1 L_j} & \gamma^j_{1,ad} e^{ik^j_1 L_j} & \beta^j_1 & \beta^j_1 & \\ -\alpha^j_1 & -\alpha^j_1 & \gamma^j_{1,bc} e^{-ik^j_1 L_j} & \gamma^j_{1,bd} e^{ik^j_1 L_j} & -\beta^j_1 & -\beta^j_1 & \\ \vdots & & & & & & \ddots \end{bmatrix}. \tag{4.17}$$

If a step change in potential occurs at position $x \neq x_0$, $g_0 = g_1 = 0$, and the elastic propagation matrix of section 2.2.1 is left with each of the channels uncoupled. Multiplying the propagation matrices for all $N + 1$ regions of the spatial domain yields the matrix equation

$$\mathbf{a} = \prod_{j=N}^{0} \mathbf{p}_j \mathbf{c} = \mathbf{P}\mathbf{c}. \tag{4.18}$$

The Schrödinger equation for a system containing an arbitrary potential and Einstein phonons may now be solved. At zero temperature, there are no phonons present for the incident phonon to absorb, so $a_n = \delta_{0,n}$. Additionally, it is assumed that there are no electrons incident from the right, making $d_n = 0$ for all n. Then

$$\begin{bmatrix} 1 \\ b_0 \\ 0 \\ b_1 \\ \vdots \end{bmatrix} = \mathbf{P} \begin{bmatrix} c_0 \\ 0 \\ c_1 \\ 0 \\ \vdots \end{bmatrix}. \tag{4.19}$$

The boundary condition $d_n = 0$ decouples the equations containing the known a_n and the unknown b_n. Thus, (4.19) can be decomposed into the two matrix equations

$$\begin{bmatrix} 1 \\ 0 \\ \vdots \end{bmatrix} = \begin{bmatrix} P_{1,1} & P_{1,3} & \dots \\ P_{3,1} & P_{3,3} & \\ \vdots & & \ddots \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \end{bmatrix}, \tag{4.20}$$

and

$$\begin{bmatrix} b_0 \\ b_1 \\ \vdots \end{bmatrix} = \begin{bmatrix} P_{1,2} & P_{1,4} & \dots \\ P_{3,2} & P_{3,4} & \\ \vdots & & \ddots \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \end{bmatrix}. \tag{4.21}$$

Using matrix inversion, c_n can be calculated from (4.20), and b_n is calculated by substituting c_n into (4.21).

4.1.2 Conservation of current in the presence of inelastic scatterers

Since each inelastic channel will have separate reflection and transmission coefficients, proper normalization is required to ensure that unitarity is preserved. That is, the electron flux density, and hence current density, entering and leaving the system must be equal, or

$$T + R = \sum_{n=0}^{\infty} (R_n + T_n) = 1, \tag{4.22}$$

where R, T, R_n and T_n are the total reflection, total transmission, channel n reflection, and channel n transmission coefficients, respectively.

The current density is given by

$$\mathbf{J} = -i\frac{e\hbar}{2m}\left(\Psi^*(x,t)\frac{\partial\Psi(x,t)}{\partial x} - \frac{\partial\Psi^*(x,t)}{\partial x}\Psi(x,t)\right), \tag{4.23}$$

where the time-dependent state function is

$$\Psi(x,t) = \sum_{n=0}^{\infty} \psi_n(x)|n\rangle e^{-iE_n t/\hbar}. \tag{4.24}$$

Substituting into (4.23) and evaluating in a region having zero electric field,

$$\mathbf{J} = -i\frac{e\hbar}{2m}\sum_{n=0}^{\infty}\sum_{n'=0}^{\infty}\langle n'|n\rangle\left(\psi_{n'}^*\frac{\partial\psi_n(x)}{\partial x}e^{-i(E_n-E_{n'})t/\hbar} - \frac{\partial\psi_{n'}^*(x)}{\partial x}\psi_n e^{-i(E_{n'}-E_n)t/\hbar}\right). \tag{4.25}$$

Since the phonon states are orthonormal, the off-diagonal matrix elements are zero. Furthermore, decaying states do not carry any net current. Thus, (4.25) reduces to

$$\mathbf{J} = \frac{e\hbar}{m}\sum_{n=0}^{n_{\max}} k_n\left(|a_n|^2 - |b_n|^2\right), \tag{4.26}$$

where

$$n_{\max} = \left\lfloor \frac{E_0 - U}{\hbar\omega_0} \right\rfloor, \tag{4.27}$$

is the maximum number of real phonons that can be excited. Thus, the orthogonality of the phonon states causes the current due to each of the inelastic channels to be independent.

Again assuming temperature $T = 0$ K and no electrons incident from the right, the current densities at the left and right hand sides of the system are

$$\mathbf{J}^{\mathrm{L}} = e\frac{\hbar k_0^{\mathrm{L}}}{m} - \frac{e\hbar}{m}\sum_{n=0}^{n_{\max}^{\mathrm{L}}} k_n^{\mathrm{L}} |b_n|^2, \tag{4.28}$$

and

$$\mathbf{J}^{\mathrm{R}} = \frac{e\hbar}{m}\sum_{n=0}^{n_{\max}^{\mathrm{R}}} k_n^{\mathrm{R}} |c_n|^2, \tag{4.29}$$

where superscripts L and R indicate that the variable is to be evaluated using the potential at the left or right hand side of the system, respectively. Setting these two equations equal to each other gives the unitarity condition

$$1 = \sum_{n=0}^{n_{\max}^{\mathrm{L}}} \frac{k_n^{\mathrm{L}}}{k_0^{\mathrm{L}}} |b_n|^2 + \sum_{m=0}^{m_{\max}^{\mathrm{L}}} \frac{k_m^{\mathrm{R}}}{k_0^{\mathrm{L}}} |c_m|^2. \tag{4.30}$$

Comparing with (4.22), the transmission and reflection coefficients are

$$T(E_0) = \sum_{n=0}^{n_{\max}^{\mathrm{R}}} T_n(E_0) = \sum_{n=0}^{n_{\max}^{\mathrm{R}}} \frac{k_n^{\mathrm{R}}}{k_0} |c_n(E_0)|^2, \tag{4.31}$$

$$R(E_0) = \sum_{n=0}^{n_{\max}^{\mathrm{L}}} R_n(E_0) = \sum_{n=0}^{n_{\max}^{\mathrm{L}}} \frac{k_n^{\mathrm{L}}}{k_0} |b_n(E_0)|^2. \tag{4.32}$$

4.1.3 Strong coupling regime

Numerical solutions require that only a finite number of phonons, M, be included in the simulation. However, the solution to the truncated system will not converge with increasing M if g_1 exceeds a critical value, $g_{1,\mathrm{c}}$. For $g_1 > g_{1,\mathrm{c}}$, the inelastic channels are so strongly coupled that all inelastic channels contribute significantly to the total response of the system, and hence the matrix may not be truncated for any value of M.

To determine $g_{1,\mathrm{c}}$ analytically consider the continued fraction solution

$$c_0 = \frac{1}{1 + \frac{\frac{m^2 g_1^2}{\hbar^4 k_0 k_1}}{1 + \frac{\frac{2m^2 g_1^2}{\hbar^4 k_1 k_2}}{\ddots}}}. \tag{4.33}$$

For large M the coupling element approaches

$$\lim_{N\to\infty} i\frac{\sqrt{M} m g_1}{\hbar^2 k_M} = \frac{\sqrt{m} g_1}{\hbar\sqrt{2\hbar\omega_0}}, \tag{4.34}$$

and the continued fraction expansion becomes

$$\begin{array}{l}\ddots \\ 1- \dfrac{p^2}{1 - \frac{p^2}{1 - \frac{p^2}{\ddots}}},\end{array} \tag{4.35}$$

where p is equal to (4.34). This is the continued fraction solution of the quadratic equation

$$x^2 - x + p^2 = 0 \quad \Rightarrow \quad x = 1 - \frac{p^2}{x}, \tag{4.36}$$

which diverges by oscillation if $1 - 4p^2 < 0$. Hence from (4.34) one obtains $g_{1,\mathrm{c}} = 0.14$ eV nm for GaAs.

4.2 Introduction to unitary feedback

To understand the behavior of the model the system is first considered under symmetric conditions. For illustrative purposes, a unit amplitude sinusoidal wave function is injected from both the left and right hand sides of the spatially symmetric potential illustrated in Fig. 4.4(a). The probability of an electron of energy E_0 exiting the right hand side of Fig. 4.4(a) in channel n is

$$|c'_n(E_0)|^2 = \frac{k_n}{k_0}|c_n(E_0)|^2. \tag{4.37}$$

For energies $E_0 < \hbar\omega_0$, no real phonons may be emitted and only $|c'_0|^2$ is nonzero. Once the $n = 1$ channels open, this probability must decrease to preserve unitarity when $|c'_1|^2$ becomes finite. To a lesser extent $|c'_0|^2$ is also increased by feedback from the $n = 1$ channel as the electron may emit a phonon, reflect off the step change in the potential barrier, and reabsorb the phonon at position $x = x_0$. As a result of these coherent interactions the $n = 0$ channel may not be considered an elastic channel. One may see in Fig. 4.4(b) that the feedback between the channels driven by unitarity gives rise to features in the spectrum.

Now consider a system in which the spatial symmetry is broken by injecting the electron from position $x = -\infty$. The system is composed of a constant background potential, $U(x) = 0$ with Einstein phonons characterized by $g_0 = 0.2$ eV nm, and $g_1 = 0.12$ eV nm. Figure 4.5 shows the transmission of this system compared to the elastic transmission ($g_1 = 0$ eV nm) along with the transmission due to each inelastic channel.

The oscillating phonon creates a variation in the amplitude of the delta barrier, which in turn perturbs the electron transmission. Since tunneling exponentially increases with

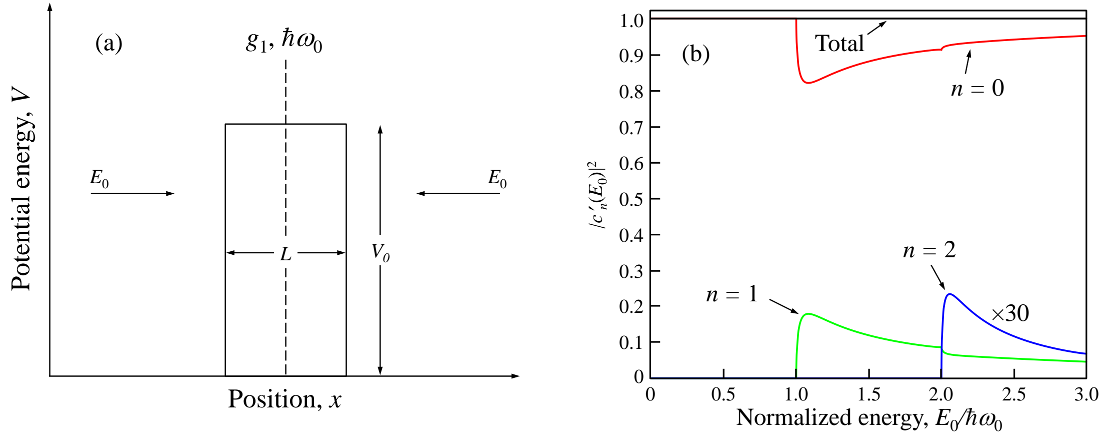

Figure 4.4: (a) A rectangular potential barrier containing Einstein phonons (dashed line). The phonons are centered within the barrier and the electron injected with energy E_0 has plane wave components incident from $\pm\infty$. (b) Probability of an electron of energy E_0 exiting the right hand side of (a) in channel n. The potential barrier has energy $U_0 = 0.1$ eV and length $L = 1$ nm, the phonons have energy $\hbar\omega_0 = 36$ meV, and the coupling constant is $g_1 = 0.05$ eV nm. The effective electron mass is $m = 0.07 \times m_0$, the number of inelastic channels included in the simulation is $M = 11$, the temperature is $T = 0$ K, and the simulation converged with a relative error of less than 10^{-10}.

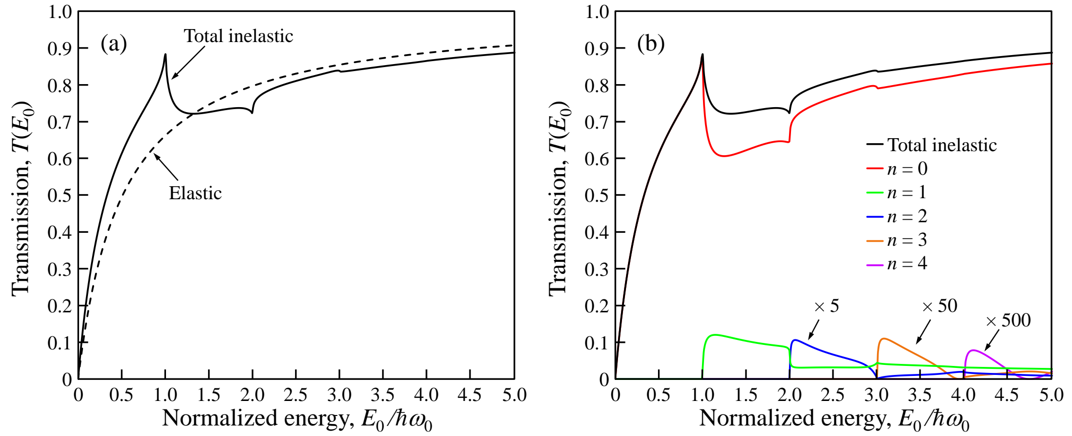

Figure 4.5: (a) Calculated total inelastic electron transmission in the presence of inelastic phonon scattering (solid) and total elastic transmission in the absence of any phonon scattering (dashed). (b) Transmission due to each of the inelastic channels. The potential contains Einstein phonons in an otherwise constant background potential, $U(x) = 0$. The model parameters are $m = 0.07 \times m_0$, $\hbar\omega_0 = 36$ meV, $g_0 = 0.2$ eV nm, $g_1 = 0.12$ eV nm, and $T = 0$ K.

decreasing potential barrier energy, the electron feels a delta barrier that is effectively lower than g_0. The result is virtual-phonon enhanced transmission, evident in Fig. 4.5 .

The cusps are a classic sign of a unitary system [13, 5], and Fig. 4.5(b) clearly shows that the cusps are due to the opening of a new inelastic channel. As a new channel opens, unitarity requires that the amplitudes of the higher energy states be reduced in order to feed into the new state. At energy $E_0/\hbar\omega_0 = 1$ in the Fig., the $n = 1$ channel is beginning to propagate. The amplitude of the $n = 0$ channel decreases as the amplitude of the $n = 1$ channel increases, creating the cusp-like appearance.

4.3 Quantum and semi-classical inelastic electron transport

The examples of section 4.2 have shown that the electron transmission coefficient calculated using a fully quantum mechanical model $decreases$ after the excitation threshold for the first real phonon. However, it will be shown that simple first-order perturbation theory predicts that the total electron transmission $increase$ at this threshold. In this section the predictions of quantum and semi-classical models of inelastic scattering are compared and the conditions under which the predictions of simple first-order perturbation theory emulate those of a fully quantum model are found. It is shown that the conditions under which this occurs is typical of inelastic scattering experiments.

4.3.1 Simple first-order perturbation theory

Simple first-order perturbation theory predictions are obtained by truncation at the first term in the Born series. In terms of the Green's function, the exact solution to the Schrödinger equation is given by

$$\Psi(x) = e^{ik_0x}|0\rangle + \int_{-\infty}^{\infty} dx' G^{(0)}(x-x')\hat{U}(x')\Psi(x'), \tag{4.38}$$

where $G^{(0)}(x-x')$ is the free particle Green's function and $\hat{U}(x)$ is the potential. Since only inelastic scattering is of interest, let $g_0 = 0$ eV nm and $U(x) = 0$ eV. Then the potential is

$$\hat{V}(x) = g_1(\hat{b}^\dagger + \hat{b})\delta(x), \tag{4.39}$$

where for simplicity $x_0 = 0$, and the Schrödinger equation for channel n is

$$\left(\frac{\partial^2}{\partial x^2} + k_n^2\right)\psi_n(x)|n\rangle = \frac{2m}{\hbar^2}\hat{U}(x)\psi_n(x)|n\rangle. \tag{4.40}$$

The Green's function is defined as the solution to

$$\left(\frac{\partial^2}{\partial x^2} + k_n^2\right)G_n^{(0)}(x-x') = \delta(x-x'), \tag{4.41}$$

where $G_n^{(0)}(x-x')$ is the propagator for an electron in channel n. The total Green's function is the sum of all channel Green's functions

$$G^{(0)}(x-x') = \sum_{n=0}^{\infty} G_n^{(0)}(x-x')|n\rangle\langle n|, \tag{4.42}$$

where the identity operator ensures that $G_n^{(0)}$ operates only on the proper channel. Substituting the Fourier transform of the channel Green's function, $\tilde{G}_n^{(0)}(q)$, into (4.41),

$$\tilde{G}_n^{(0)}(q) = \frac{1}{(q+k_n)(q-k_n)}. \tag{4.43}$$

Transforming back into real space,

$$G_n^{(0)}(x - x') = -i\frac{e^{ik_n(x-x')}}{2k_n}, \tag{4.44}$$

where the pole at $q = k_n$ has been chosen because the transmitted waves are of interest.

Substituting (4.42) and (4.44) into (4.38),

$$\Psi(x) = e^{ik_0x}|0\rangle - i\frac{2mg_1}{\hbar^2}\sum_{n=0}^{\infty}\int_{-\infty}^{\infty}dx'\frac{e^{ik_n(x-x')}}{2k_n}|n\rangle\langle n|(\hat{b}^\dagger + \hat{b})\delta(x')\Psi(x'). \tag{4.45}$$

The delta function eliminates the integral, leaving

$$\Psi = |0\rangle - i\frac{mg_1}{\hbar^2}\sum_{n=0}^{\infty}\frac{1}{k_n}|n\rangle\langle n|(\hat{b}^\dagger + \hat{b})\Psi, \tag{4.46}$$

where the complex exponentials have been omitted. These phase factors can be removed because orthogonality of the phonon states prevents interference effects between channels at positions $x \neq x_0$ and the transmission coefficient for each channel is independent of the absolute phase of the electron in that channel.

The iterative solution of (4.46) is the Born series, of which the first-order approximation is made by letting $\Psi = |0\rangle$ on the right hand side. Then

$$\Psi^{(1)} = |0\rangle - i\frac{mg_1}{\hbar^2 k_1}|1\rangle, \tag{4.47}$$

and simple first-order perturbation theory yields the transmitted wave function amplitudes

$$c_0 = 1, \qquad c_1 = -i\frac{mg_1}{\hbar^2 k_1}, \tag{4.48}$$

and transmission coefficient

$$T(E_0) = 1 + \frac{m^2 g_1^2}{\hbar^4 k_0 k_1}\theta(E_0 - \hbar\omega_0). \tag{4.49}$$

This is clearly not self-consistent, as the transmission violates unitarity and c_1 is infinite at $E_0 = \hbar\omega_0$.

The second-order Born approximation is made by substituting $\Psi^{(1)}$ into (4.46). To obtain the exact solution, this procedure is repeated $ad\ infinitum$. Considering only the $n = 0$ and $n = 1$ channels for simplicity, the exact solution is

$$\Psi = \left(|0\rangle - i\frac{mg_1}{\hbar^2 k_1}|1\rangle\right)\sum_{m=0}^{\infty}\left(-\frac{m^2 g_1^2}{\hbar^2 k_0 k_1}\right)^m \quad = \frac{1}{1 + \frac{m^2 g_1^2}{\hbar^4 k_0 k_1}}|0\rangle - i\frac{\frac{mg_1}{\hbar^2 k_1}}{1 + \frac{m^2 g_1^2}{\hbar^4 k_0 k_1}}|1\rangle. \tag{4.50}$$

Whereas the first-order solution allowed only a single transition from the $n = 0$ channel to the $n = 1$ channel, the exact solution accounts for all possible transitions between these two channels through the term in the denominator. This term normalizes the wave function amplitudes, ensuring that unitarity is preserved.

As a check (4.50) is compared to the exact solution found using the matrix equation of (4.20). With only the $n = 0$ and $n = 1$ channels considered, this matrix equation is

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & \beta_1 \\ \alpha_1 & 1 \end{bmatrix}\begin{bmatrix} c_0 \\ c_1 \end{bmatrix}, \tag{4.51}$$

which can be solved to yield

$$c_0 = \frac{1}{1 + \frac{m^2 g_1^2}{\hbar^4 k_0 k_1}}, \qquad c_1 = -i\frac{\frac{mg_1}{\hbar^2 k_1}}{1 + \frac{m^2 g_1^2}{\hbar^4 k_0 k_1}}. \tag{4.52}$$

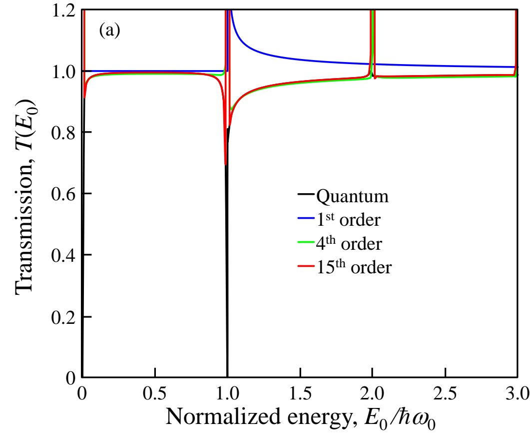

Figure 4.6: (a) Transmission coefficient calculated using quantum model and Born series to selected orders. Near phonon thresholds the perturbative solution does produce qualitatively accurate predictions. Simulation parameters are effective electron mass $m = 0.07 \times m_0$, phonon energy $\hbar\omega_0 = 36$ meV, coupling strength $g_1 = 0.05$ eV nm, and $N = 15$ inelastic channels in the quantum simulation.

As expected this is identical to (4.50). However, it is clear that for energies near $E_0 = n\hbar\omega_0$, the Born series will not converge as the wave vectors will approach zero. Thus, perturbation theory cannot produce accurate predictions for any order near phonon thresholds, where many of the unitarity-driven features occur.

This is evident in Fig. 4.6, which compares the transmission coefficients calculated with the quantum solution of (4.52) and the Born series to selected orders. It can be seen that away from the phonon thresholds the Born series approximates well the exact solution when calculated to high orders. However, near the phonon thresholds singularities occur, and perturbation theory is not a good approximation.

Simple first-order perturbation theory is usually justified on the basis of weak coupling, suggesting the g_1^2 term in the denominator can be ignored. However, it is this term that provides feedback between the channels and preserves unitarity, giving rise to important features in the exact solution.

In the following, the quantum and perturbative transmission spectra will be compared when the value of the coupling constant yields a matrix element equal to that of the Frölich interaction. The matrix element coupling initial state $|k_0\rangle$ to final state $|k_1\rangle$ is

$$|\langle k_1|\hat{H}|k_0\rangle|^2 = g_1^2|\psi_1(0)|^4. \tag{4.53}$$

For a conduction band electron of energy $E_0 = 52$ meV interacting with an LO phonon in GaAs, (4.53) is equal to the matrix element for the Frölich interaction per unit volume when $g_1 = 0.008$ eV nm.

4.3.2 Inelastic scattering in an otherwise constant background potential

When calculating inelastic scattering rates, it is commonly assumed that weak coupling leaves the incident wave unchanged by the inelastic scattering event and that simple first-order perturbation theory is an accurate approximation. Figure 4.7 shows that this is not the case. Shown in Fig. 4.7(a) is the quantum and in Fig. 4.7(b) the perturbative predictions for the transmission of an electron of energy E_0 incident from $x = -\infty$. The Einstein phonons are in a constant background potential $U(x) = 0$ eV with coupling constants of $g_1 = 0.008$ eV nm (solid) and $g_1 = 0.08$ eV nm (dashed). The inset of Fig. 4.7(a) shows the features of the quantum transmission given a coupling constant of $g_1 = 0.008$ eV nm on a fine energy scale. Although these features are resolved due to the assumed long lifetime of the electron states, it is anticipated that energy broadening due to finite lifetime effects will not qualitatively alter the results.

Rather than a decrease in transmission near $E_0 = \hbar\omega_0$ predicted in the quantum case, the transmission coefficient predicted by perturbation theory increases, violating unitarity and losing features found in the exact quantum solution. Even under conditions of weak coupling with $g_1 = 0.008$ eV nm, the quantum and perturbative solutions are dramatically different. Clearly weak coupling alone does not guarantee that the predictions of first-order perturbation theory will appear similar to the exact quantum solution.

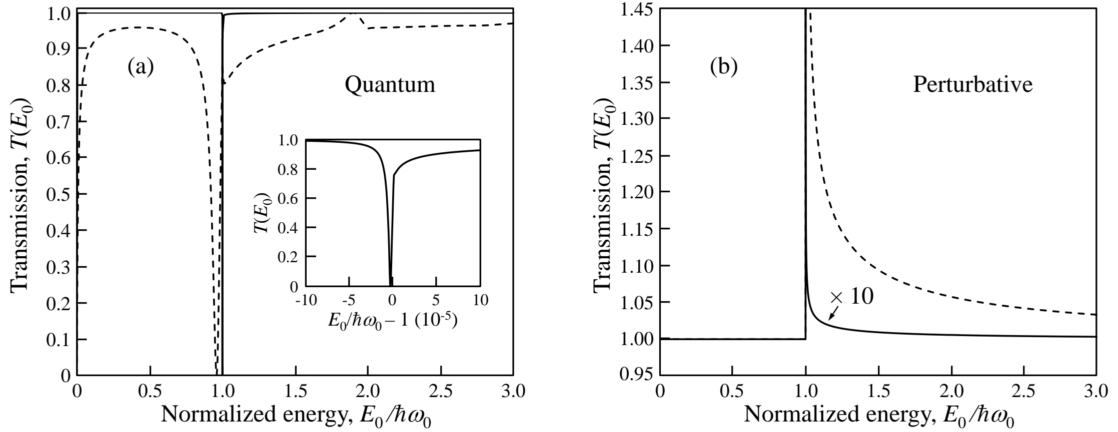

Figure 4.7: Transmission through Einstein phonons located in a constant background potential $U(x) = 0$ eV with coupling constants of $g_1 = 0.008$ eV nm (solid) and $g_1 = 0.08$ eV nm (dashed) using the (a) exact quantum solution and (b) perturbative solution. The electron of effective mass $m = 0.07 \times m_0$ is injected from $x = -\infty$ with energy E_0, the phonon energy is $\hbar\omega_0 = 36$ meV, and $M = 11$ inelastic channels were included in the simulation. The inset shows the features of the transmission about $E_0 = \hbar\omega_0$ with a coupling constant of $g_1 = 0.008$ eV nm on a fine energy scale.

4.3.3 Condition when perturbation solution appears similar to exact calculation

Consider the potential shown in Fig. 4.4(a) with an electron of energy E_0 injected from $x = -\infty$. The transmission through this system is shown for the quantum case in Fig. 4.8(a) and for the first-order perturbation theory approximation in Fig. 4.8(b). A

rectangular potential barrier of energy $U_0 = 0.25$ eV and length L, and an electron-phonon coupling constant of $g_1 = 0.05$ eV nm are used. A relatively strong coupling constant is used to enhance the features.

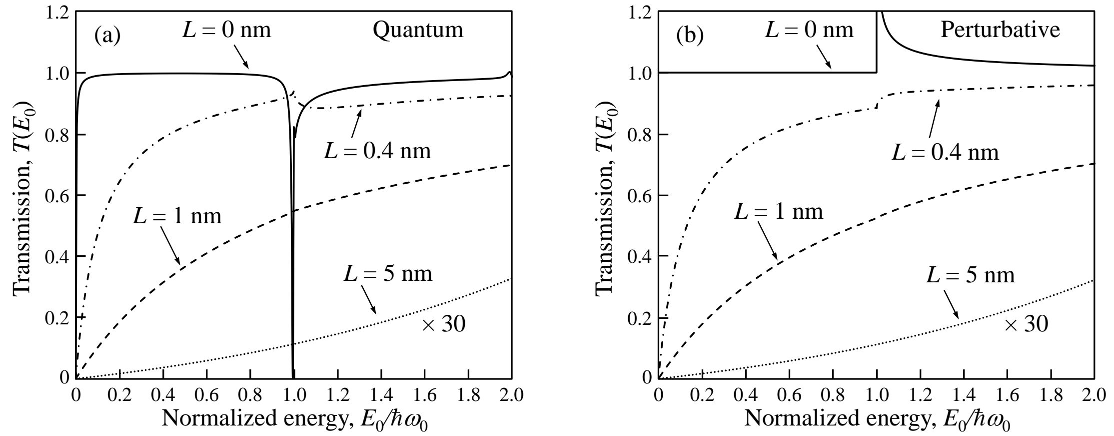

Figure 4.8: (a) Quantum and (b) first-order perturbative solutions for the transmission of an electron of energy E_0 injected from $x = -\infty$ through the potential shown in Fig. 4.4(a). The potential barrier has energy $U_0 = 0.25$ eV and length L, $g_1 = 0.05$ eV nm, $\hbar\omega_0 = 36$ meV, the effective electron mass is $m = 0.07 \times m_0$, and $M = 11$ inelastic channels were included in the simulation.

For small L features exhibited by the quantum and perturbative transmission spectra are dramatically different for energies both higher and lower than $E_0 = \hbar\omega_0$. For lower energies the quantum transmission is enhanced by the excitation of virtual phonons. For energies greater than $E_0 = \hbar\omega_0$, the $L = 0.4$ nm perturbative transmission shows the semi-classical "opening of a new channel," whereas the quantum transmission does not. However, as L increases the transmission reduces and the quantum and perturbative solutions appear qualitatively similar on the scale shown.

Near $E_0 = \hbar\omega_0$ the behavior of the two solutions become qualitatively similar for large L, as shown in Fig. 4.9. The Fig. shows the quantum (solid) and perturbative (dashed) solutions for the transmission shown in Fig. 4.8 for (a) $L = 1$ nm and (b) $L = 5$ nm on a fine energy and transmission scale. The transmission curves have been offset by $T(\hbar\omega_0)$ so that the two solutions may be compared.

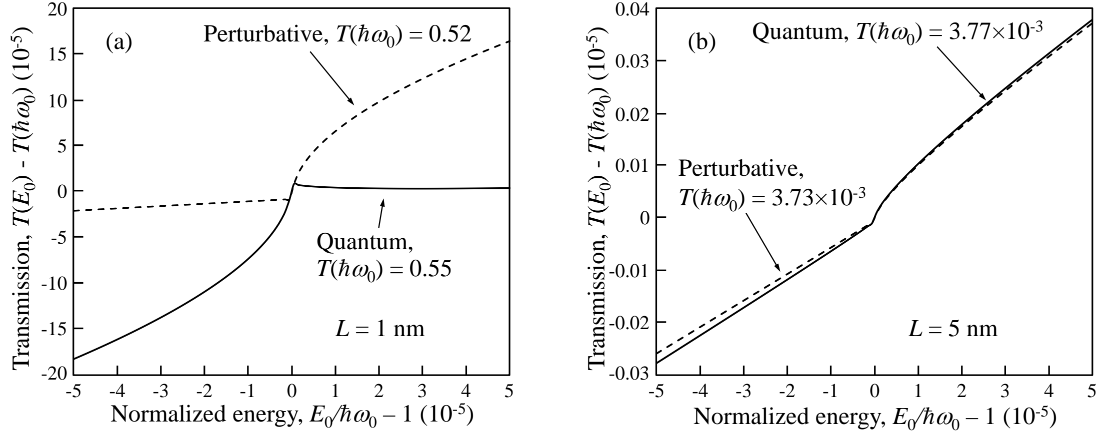

Figure 4.9: Quantum (solid) and perturbative (dashed) transmission from Fig. 4.8 shown on a fine scale for potential barrier lengths of (a) $L = 1$ nm and (b) $L = 5$ nm. The transmission curves have been offset by $T(\hbar\omega_0)$ so that the quantum and perturbative solutions may be compared. The potential barrier has energy $U_0 = 0.25$ eV and length L, and the phonon is characterized by $g_1 = 0.05$ eV nm and $\hbar\omega_0 = 36$ meV. $M = 11$ inelastic channels were included in the simulation and the effective electron mass is $m = 0.07 \times m_0$.

For $L = 5$ nm, the perturbative solution exhibits features qualitatively similar to those found in the quantum model. In both cases the transmission is well approximated by exponential tunneling through the barrier and a rapid increase occurs once the electron has enough energy to emit a real phonon. However, the reflection spectra will not be qualitatively similar since the reflection coefficient in the quantum model will reduce to preserve unitarity while the reflection coefficient in the semi-classical model will increase due to an additional channel for reflection. For the smaller barrier length, perturbation theory is not a good approximation to the quantum solution. Thus, as the strength of the potential barrier is increased the semi-classical transmission behavior predicted by first-order perturbation theory appears qualitatively similar to the behavior of the quantum model.

4.3.4 Threshold at which perturbation solution appears similar to exact calculation

As the transmission coefficient decreases due to increasing potential barrier strength, the exact quantum transmission as a function of increasing electron energy changes from a decrease at the first phonon threshold to an increase. The condition

$$\left.\frac{\partial T}{\partial E_0}\right|_{E_0=\hbar\omega_0^+} = \left.\frac{\partial T_0}{\partial E_0}\right|_{E_0=\hbar\omega_0^+} + \left.\frac{\partial T_1}{\partial E_0}\right|_{E_0=\hbar\omega_0^+} \geq 0, \tag{4.54}$$

for the exact quantum case is chosen as the condition under which the features in the first-order perturbation theory solution appear qualitatively similar to those of the exact quantum solution. The potential used in Fig. 4.8 is considered, and to simplify the expressions only the $n = 0$ and $n = 1$ channels are considered.

As a notational simplification, the energy variable is changed to $\Delta = E_0 - \hbar\omega_0$, so that the channel energies are $E_0 = \hbar\omega_0 + \Delta$ and $E_1 = \Delta$. The conditions under which (4.54) is satisfied are found by evaluating the derivatives in the limit that $\Delta \to 0$. The wave function amplitudes for the $n = 0$ and $n = 1$ channels are given by the solution to

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \mathbf{P} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix}, \tag{4.55}$$

where

$$
\begin{aligned}
P_{11} &= \cosh(\kappa_0 L) + i\frac{1}{2}\left(\frac{\kappa_0}{k_0} - \frac{k_0}{\kappa_0}\right)\sinh(\kappa_0 L),\\
P_{12} &= i\frac{mg_1}{\hbar^2 k_0}\left(\cosh(\tfrac{1}{2}\kappa_0 L) - i\frac{k_0}{\kappa_0}\sinh(\tfrac{1}{2}\kappa_0 L)\right)\left(\cosh(\tfrac{1}{2}\kappa_1 L) - i\frac{k_1}{\kappa_1}\sinh(\tfrac{1}{2}\kappa_1 L)\right),\\
P_{21} &= i\frac{mg_1}{\hbar^2 k_1}\left(\cosh(\tfrac{1}{2}\kappa_0 L) - i\frac{k_0}{\kappa_0}\sinh(\tfrac{1}{2}\kappa_0 L)\right)\left(\cosh(\tfrac{1}{2}\kappa_1 L) - i\frac{k_1}{\kappa_1}\sinh(\tfrac{1}{2}\kappa_1 L)\right),\\
P_{22} &= \cosh(\kappa_1 L) + i\frac{1}{2}\left(\frac{\kappa_1}{k_1} - \frac{k_1}{\kappa_1}\right)\sinh(\kappa_1 L),
\end{aligned} \tag{4.56}
$$

and

$$
\begin{aligned}
k_0 &= \frac{\sqrt{2m(\hbar\omega_0 + \Delta)}}{\hbar},\\
\kappa_0 &= \frac{\sqrt{2m(U_0 - \hbar\omega_0 - \Delta)}}{\hbar},\\
k_1 &= \frac{\sqrt{2m\Delta}}{\hbar},\\
\kappa_1 &= \frac{\sqrt{2m(U_0 - \Delta)}}{\hbar}.
\end{aligned} \tag{4.57}
$$

The derivative $\frac{\partial T_1}{\partial E_0}$ is considered first. The transmission coefficient for the $n = 1$ channel is

$$
T_1 = \frac{k_1}{k_0}\frac{|P_{21}|^2}{|\det(\mathbf{P})|^2}. \tag{4.58}
$$

To leading orders in Δ, the determinant has six components,

$$
\det(P) = A + iB = \left(A^-\Delta^{-1/2} + A^0 + A^+\Delta^{1/2}\right) + i\left(B^-\Delta^{-1/2} + B^0 + B^+\Delta^{1/2}\right), \tag{4.59}
$$

where

$$
\begin{aligned}
A^{-} =& \frac{g_1'}{\sqrt{\hbar\omega_0 + \Delta}} \cosh^2(\tfrac{1}{2}\kappa_0 L) \cosh^2(\tfrac{1}{2}\kappa_1 L) \\
&- \frac{g_1'\sqrt{\hbar\omega_0 + \Delta}}{U_0 - \hbar\omega_0 - \Delta} \sinh^2(\tfrac{1}{2}\kappa_0 L) \cosh^2(\tfrac{1}{2}\kappa_1 L) \\
&- \frac{\sqrt{U_0 - \Delta}}{4} \left(\frac{\kappa_0}{k_0} - \frac{k_0}{\kappa_0} \right) \sinh(\kappa_0 L) \sinh(\kappa_1 L), \\
A^{0} =& \cosh(\kappa_0 L) \cosh(\kappa_1 L) \\
&- \frac{4g_1'}{\sqrt{U_0 - \hbar\omega_0 - \Delta}\sqrt{U_0 - \Delta}} \cosh(\tfrac{1}{2}\kappa_0 L) \sinh(\tfrac{1}{2}\kappa_0 L) \cosh(\tfrac{1}{2}\kappa_1 L) \sinh(\tfrac{1}{2}\kappa_1 L), \\
A^{+} =& \frac{1}{4\sqrt{U_0 - \Delta}} \left(\frac{\kappa_0}{k_0} - \frac{k_0}{\kappa_0} \right) \sinh(\kappa_0 L) \sinh(\kappa_1 L) \\
&- \frac{g_1'}{(U_0 - \Delta)\sqrt{\hbar\omega_0 + \Delta}} \cosh^2(\tfrac{1}{2}\kappa_0 L) \sinh^2(\tfrac{1}{2}\kappa_1 L) \\
&+ \frac{g_1'\sqrt{\hbar\omega_0 + \Delta}}{(U_0 - \hbar\omega_0 - \Delta)(U_0 - \Delta)} \sinh^2(\tfrac{1}{2}\kappa_0 L) \sinh^2(\tfrac{1}{2}\kappa_1 L), \\
B^{-} =& \frac{\sqrt{U_0 - \Delta}}{2} \cosh(\kappa_0 L) \sinh(\kappa_1 L) \\
&- \frac{2g_1'}{\sqrt{U_0 - \hbar\omega_0 - \Delta}} \cosh(\tfrac{1}{2}\kappa_0 L) \sinh(\tfrac{1}{2}\kappa_0 L) \cosh^2(\tfrac{1}{2}\kappa_1 L), \\
B^{0} =& \frac{1}{2} \left(\frac{\kappa_0}{k_0} - \frac{k_0}{\kappa_0} \right) \sinh(\kappa_0 L) \cosh(\kappa_1 L) \\
&- \frac{2g_1'}{\sqrt{\hbar\omega_0 + \Delta}\sqrt{U_0 - \Delta}} \cosh^2(\tfrac{1}{2}\kappa_0 L) \cosh(\tfrac{1}{2}\kappa_1 L) \sinh(\tfrac{1}{2}\kappa_1 L) \\
&+ \frac{2g_1'\sqrt{\hbar\omega_0 + \Delta}}{(U_0 - \hbar\omega_0 - \Delta)\sqrt{U_0 - \Delta}} \sinh^2(\tfrac{1}{2}\kappa_0 L) \cosh(\tfrac{1}{2}\kappa_1 L) \sinh(\tfrac{1}{2}\kappa_1 L), \\
B^{-} =& \frac{2g_1'}{(U_0 - \Delta)\sqrt{U_0 - \hbar\omega_0 - \Delta}} \cosh(\tfrac{1}{2}\kappa_0 L) \sinh(\tfrac{1}{2}\kappa_0 L) \sinh^2(\tfrac{1}{2}\kappa_1 L) \\
&- \frac{1}{2\sqrt{U_0 - \Delta}} \cosh(\kappa_0 L) \sinh(\kappa_1 L), \\
g_1' =& \frac{m g_1^2}{2\hbar^2}.
\end{aligned}
\tag{4.60}
$$

Likewise, the numerator has two components,

$$\frac{k_1}{k_0}|P_{21}|^2 = q = q^- \Delta^{-1/2} + q^+ \Delta^{1/2}, \tag{4.61}$$

where

$$\begin{aligned}
q^- =& \frac{g_1'}{\sqrt{\hbar\omega_0 + \Delta}} \cosh^2(\tfrac{1}{2}\kappa_0 L) \cosh^2(\tfrac{1}{2}\kappa_1 L) \\
&+ \frac{g_1' \sqrt{\hbar\omega_0 + \Delta}}{U_0 - \hbar\omega_0 - \Delta} \sinh^2(\tfrac{1}{2}\kappa_0 L) \cosh^2(\tfrac{1}{2}\kappa_1 L), \\
q^+ =& \frac{g_1'}{(U_0 - \Delta)\sqrt{\hbar\omega_0 + \Delta}} \cosh^2(\tfrac{1}{2}\kappa_0 L) \sinh^2(\tfrac{1}{2}\kappa_1 L) \\
&+ \frac{g_1' \sqrt{\hbar\omega_0 + \Delta}}{(U_0 - \hbar\omega_0 - \Delta)(U_0 - \Delta)} \sinh^2(\tfrac{1}{2}\kappa_0 L) \sinh^2(\tfrac{1}{2}\kappa_1 L).
\end{aligned}$$

The derivative is then

$$\begin{aligned}
\frac{\partial T_1}{\partial E_0} &= \frac{\partial}{\partial \Delta} \frac{q}{A^2 + B^2} = \frac{(A^2 + B^2)\frac{\partial q}{\partial \Delta} - 2q\left(A\frac{\partial A}{\partial \Delta} + B\frac{\partial B}{\partial \Delta}\right)}{\left(A^2 + B^2\right)^2} \\
&= \frac{A\left(A\frac{\partial q}{\partial \Delta} - 2q\frac{\partial A}{\partial \Delta}\right) + B\left(B\frac{\partial q}{\partial \Delta} - 2q\frac{\partial B}{\partial \Delta}\right)}{\left(A^2 + B^2\right)^2}.
\end{aligned} \tag{4.62}$$

Evaluating the first term in the numerator,

$$\lim_{\Delta \to 0} A\left(A\frac{\partial q}{\partial \Delta} - 2q\frac{\partial A}{\partial \Delta}\right) = \frac{1}{2} q^- (A^-)^2 \Delta^{-5/2}. \tag{4.63}$$

From symmetry of the expressions the second term in the numerator is found by letting $A \to B$. Using these expressions, the derivative is

$$\lim_{\Delta\to 0}\frac{\partial T_1}{\partial E_0} = \frac{1}{2\Delta}\frac{q^-\Delta^{-1/2}}{(A^-\Delta^{-1/2})^2+(B^-\Delta^{-1/2})^2}\frac{(A^-\Delta^{-1/2})^2+(B^-\Delta^{-1/2})^2}{(A^-\Delta^{-1/2})^2+(B^-\Delta^{-1/2})^2}$$
$$= \frac{1}{2\Delta}T_1(\hbar\omega_0). \tag{4.64}$$

Turning now to $\frac{\partial T_0}{\partial E_0}$, the transmission coefficient for the $n=0$ channel is

$$T_0 = \frac{|P_{22}|^2}{|\det(\mathbf{P})|^2}, \tag{4.65}$$

and to leading order in Δ the numerator is

$$|P_{22}|^2 = p = p^-\Delta^{-1} + p^0 + p^+\Delta, \tag{4.66}$$

where

$$p^- = \frac{U_0-\Delta}{4}\sinh^2(\kappa_1 L),$$
$$p^0 = \cosh^2(\kappa_1 L) - \frac{1}{2}\sinh^2(\kappa_1 L),$$
$$p^+ = \frac{1}{4(U_0-\Delta)}\sinh^2(\kappa_1 L).$$

The derivative $\frac{\partial T_0}{\partial E_0}$ will have the same form as (4.62), but in this case the first term in the numerator is

$$\lim_{\Delta\to 0} A\left(A\frac{\partial p}{\partial \Delta} - 2p\frac{\partial A}{\partial \Delta}\right) = -p^-A^0A^-\Delta^{-5/2}, \tag{4.67}$$

and the derivative is

$$\lim_{\Delta\to 0}\frac{\partial T_0}{\partial E_0} = -\frac{1}{\Delta}\frac{p^-\Delta^{-1}}{(A^-\Delta^{-1/2})^2+(B^-\Delta^{-1/2})^2}\frac{\left(A^0A^-\Delta^{-1/2}+B^0B^-\Delta^{-1/2}\right)}{(A^-\Delta^{-1/2})^2+(B^-\Delta^{-1/2})^2}$$
$$= -\frac{1}{\Delta}T_0(\hbar\omega_0)\frac{\left(A^0A^-\Delta^{-1/2}+A^0B^-\Delta^{-1/2}\right)}{(A^-\Delta^{-1/2})^2+(B^-\Delta^{-1/2})^2}. \quad (4.68)$$

It can be shown that the numerator simplifies to

$$\lim_{\Delta\to 0} A^0A^- + B^0B^- = -q^-, \quad (4.69)$$

so that

$$\lim_{\Delta\to 0}\frac{\partial T_0}{\partial E_0} = -\frac{1}{\Delta}T_0(\hbar\omega_0)\frac{q^-\Delta^{-1/2}}{(A^-\Delta^{-1/2})^2+(B^-\Delta^{-1/2})^2} = -\frac{1}{\Delta}T_0(\hbar\omega_0)T_1(\hbar\omega_0). \quad (4.70)$$

Substituting (4.64),

$$\lim_{\Delta\to 0}\frac{\partial T_0}{\partial E_0} = -2T_0(\hbar\omega_0^+)\left.\frac{\partial T_1}{\partial E_0}\right|_{E_0=\hbar\omega_0^+}$$
$$= -T_0(\hbar\omega_0)\left(\left.\frac{\partial T_1}{\partial E_0}\right|_{E_0=\hbar\omega_0^+} + \left.\frac{\partial R_1}{\partial E_0}\right|_{E_0=\hbar\omega_0^+}\right), \quad (4.71)$$

where the last step comes from the spatial symmetry forcing T_1 and R_1 to be equal.

The sum of the derivatives in (4.71) gives the total rate of change of the $n = 1$ channel, and to preserve unitarity the total rate of change of the $n = 0$ channel must be equal and opposite. According to (4.71), the transmission coefficient gives the fraction of this rate of change that is due to the transmission coefficient. The remaining fraction must come from the reflection coefficient. Substituting (4.71) into (4.54), the features

in the first-order perturbation theory solution appear similar to those of the quantum solution when

$$T(\hbar\omega_0) \leq \frac{1}{2}. \tag{4.72}$$

Similarity between the quantum and perturbative transmission spectra occurs when the transmission coefficient has reduced enough for a majority of the quantum feedback effects to occur in the reflection. If $T(\hbar\omega_0)$ is very small due to a large potential barrier, the transmission contributes little to the opening of the new inelastic channel and the simple first-order perturbation approximation emulates quantum behavior. The coupling constant contributes to this condition through virtual phonon assisted tunneling. Larger coupling results in larger $T(\hbar\omega_0)$, which will require a stronger potential barrier to reach $T(\hbar\omega_0) \leq \frac{1}{2}$.

If the phonon is moved to one side of the potential barrier, the existence of a propagating state next to the phonon will enhance inelastic scattering while the reflection of only one inelastically scattered wave inside of the barrier decreases feedback from the $n = 1$ channel into the $n = 0$ channel. Both of these effects make T_0 decrease faster at the phonon threshold. In order to reduce $\frac{\partial T_0}{\partial E_0}$ sufficiently that $\frac{\partial T}{\partial E_0} > 0$, the transmission coefficient at the threshold is less than $\frac{1}{2}$. The condition becomes

$$T(\hbar\omega_0) \leq \frac{\cosh^2(\kappa_U L)}{2\left(\cosh^2(\kappa_U L) - \frac{1}{2}\sinh^2(\kappa_U L)\right)\left(\cosh^2(\kappa_\omega L) - \frac{1}{2}\left(1 - \frac{\hbar\omega_0}{U_0 - \hbar\omega_0}\right)\sinh^2(\kappa_\omega L)\right)}, \tag{4.73}$$

when the phonon is located on the right side of the barrier, and

$$T(\hbar\omega_0) \leq \frac{\cosh^2(\kappa_\omega L) + \frac{\hbar\omega_0}{U_0-\hbar\omega_0}\sinh^2(\kappa_\omega L)}{2\left(\cosh^2(\kappa_U L) - \frac{1}{2}\sinh^2(\kappa_U L)\right)\left(\cosh^2(\kappa_\omega L) - \frac{1}{2}\left(1 - \frac{\hbar\omega_0}{U_0-\hbar\omega_0}\right)\sinh^2(\kappa_\omega L)\right)}, \tag{4.74}$$

when the phonon is located on the left side of the barrier, where $\kappa_U = \sqrt{2mU_0}/\hbar$ and $\kappa_\omega = \sqrt{2m(U_0 - \hbar\omega_0)}/\hbar$.

Shown in Fig. 4.10 is the simulated transmission coefficient at which $\frac{\partial T}{\partial E_0} = 0$ at energy $E_0 = \hbar\omega_0$ when the phonon is centered (solid), placed on the right side (dashed), and placed on the left side (dotted) of the potential barrier. In the simulations, U_0 is fixed while L is swept until the condition is satisfied, with $g_1 = 0.008$ eV nm and $M = 11$. Potential barrier lengths of up to $L = 55$ nm were considered. As predicted the threshold occurs at $T(\hbar\omega_0) = \frac{1}{2}$ when the phonon is centered independent of the strength of the potential barrier and smaller transmission coefficients are required for the asymmetric cases. However, as U_0 is increased L is decreased to achieve the threshold transmission, lessening the asymmetry and causing the three cases to converge as $L \to 0$.

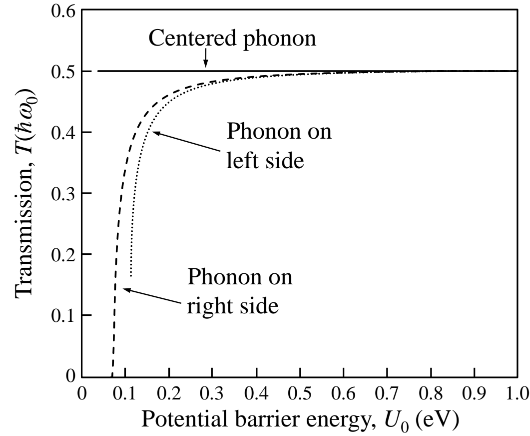

Figure 4.10: Transmission coefficient at which $\partial T/\partial E_0 = 0$ at energy $E_0 = \hbar\omega_0$ for three different phonon locations. In the simulation the potential barrier energy $U_0 > \hbar\omega_0$ is fixed and the potential barrier length L is swept until $\partial T/\partial E_0 = 0$ at energy $E_0 = \hbar\omega_0$. The simulation used a coupling constant of $g_1 = 0.008$ eV nm, phonon energy $\hbar\omega_0 = 36$ meV, electron effective mass $m = 0.07 \times m_0$, and $M = 11$.

When the phonon is placed on the left hand side of the potential barrier the transmitted waves for both the $n = 0$ and $n = 1$ channels must tunnel through the potential barrier before exiting the system. The $n = 1$ channel has a larger decay constant than the $n = 0$ channel, decreasing T_1 more than T_0 and requiring a smaller $\partial T_0/\partial E_0$ to satisfy (4.54). For potential barrier energy on the order of the phonon energy this effect becomes so great that first-order perturbation theory does not approximate well the quantum solution for any barrier length, L. In the Fig. the cutoff occurs at $V_0 = 111.4$ meV and $L = 4$ nm. One may make the argument that requiring the transmission coefficient to be small in order for perturbation theory to be a good approximation is equivalent to requiring the inelastic coupling strength be weak relative to the elastic scattering strength. However, the existence of the cutoff shows that this is not the case, as the combination of weak coupling and large elastic scattering strength are not always sufficient. Thus, even in the case of weak coupling and large potential barrier strength, conditions exist under which perturbation theory cannot appear qualitatively similar to the exact quantum solution.

4.4 Conclusion and future work

Perturbation theory, particularly Fermi's golden rule, is widely used in electron transport without any physical justification for its accuracy with respect to a quantum model. Here I have started from a simple, exactly solvable quantum model and shown a connection to semi-classical predictions of first-order perturbation theory, deriving the conditions under which its predictions may be qualitatively accurate to those of a fully quantum model. It is shown that a small coupling coefficient alone is insufficient, and that only when a majority of self-consistent feedback effects occur in the electron reflection coefficient may the predictions of first-order perturbation theory be considered qualitatively

accurate. Further progress in electron transport may be possible by repeating this process with other scattering mechanisms.

The physical model presented here models the steady-state response of an electron in an infinite lifetime plane wave state to localized Einstein phonons. Since scattering causes the lifetime of an electronic state to be finite, the finite lifetime must be accounted for in the physical model to be self-consistent. Future work on this model may include accounting for finite life-time effects in the scattering process.

Physically, a finite lifetime may be modeled by adding an imaginary component to the energy. In the time-domain this results in exponential-decay of the state, while in k-space the state is broadened into a Lorentzian distribution. Because of this broadening, it is expected that the simulated transmission spectrum will be somewhat smoothed relative to those shown here, although the qualitative features should remain unchanged.

When adding the imaginary component to the electron's energy, care must be taken to maintain self-consistency. Unitarity can be lost as the imaginary component causes the electron's wave function to decay in time, and finite lifetime of the phonon states should be accounted for as well. The added time-dependence of the interaction will require solving the time-dependent Schrödinger's equation and add significant complexity. Additional physical insight may be possible from analysis of how these effects change the threshold condition for qualitatively accurate first-order perturbation theory predictions, and whether perturbation theory is qualitatively accurate over a wider range of conditions than those found here.

Chapter 5

Conclusion

In this thesis the physical model used in previous optimal design efforts was enhanced through development of a solver for the self-consistent potential. This physical model was then applied to the optimal design of nanoscale heterostructure devices. Due to the general nonlinear behavior of non-equilibrium transport, a linear current-voltage characteristic was chosen to demonstrate the extent of control over device behavior and explore the limits of this control. Excellent linear behavior was achieved with a demonstrated dynamic range of greater than 65 dB. Limitations of control were explored by examining the results of optimal design over a range of linear objective function slopes and device lengths. It was shown that material choices limit access to low energy resonant states required to control device behavior at low voltage bias. Also shown were limits to the maximum controllable device length due to elastic scattering limits and the ability for the voltage bias to control the potential.

The ability to customize the current-voltage characteristic of semiconductor heterostructure diodes through optimal design enables co-design of electronic circuits and the devices within these circuits. This new design methodology was applied to the design of an RF switching mixer. Additional of a heterostructure diode featuring a parametrically defined current-voltage characteristic was shown to reduce spurious harmonics, easing filtering requirements and increasing dynamic range. In particular, a fourth-order harmonic at twice the intermediate frequency is reduced by 28 dB, reducing by 4 the number of filter poles necessary for a fixed dynamic range.

To progress further in the development of physical models, simple exactly solvable quantum models should be favored over advanced models that require the use of perturbation theory. It was demonstrated that simple first-order perturbation theory for the electron-phonon interaction can qualitatively predict features of an exactly solved quantum model only when transmission coefficients are sufficiently reduced. Situations in which perturbation theory cannot provide accurate predictions were also demonstrated. With these findings and new physical insight, it is hoped that the same treatment can be applied to other interactions to determine when models using perturbation theory are qualitatively approximating the exact quantum model.

This work has been completed with the intent of laying groundwork for further development and application of optimal design. Optimal design is a powerful new design technology that will enable engineers to customize device behaviors for specific applications. Further development of this design process will provide access to fundamentally new device and circuit functionalities.

Bibliography

[1] S. Adachi, J. Appl. Phys. **58**, R1 (1985).

[2] J. Bonča and S. A. Trugman, Phys. Rev. Lett. **75**, 2566 (1995).

[3] R. C. Bowen, G. Klimeck, R. K. Lake, W. R. Frensley, and T. Moise, J. Appl. Phys. **81**, 3207 (1997).

[4] T. Brandes and J. Robinson, Phys. Stat. Sol. (b) **234**, 378 (2002).

[5] G. Breit, Phys. Rev. **107**, 1612 (1957).

[6] W. Y. Choi, B. Park, J. D. Lee, and T. K. Liu, IEEE Elec. Dev. Lett. **28**, 743 (2007).

[7] E. Conwell, *High field transport in semiconductors* (Academic Press Inc., New York, 1967).

[8] J. H. Davies, S. Hershfield, P. Hyldgaard, and J. W. Wilkins, Phys. Rev. B **47**, 4603 (1993).

[9] E. G. Emberly and G. Kirczenow, Phys. Rev. B **61**, 5740 (2000).

[10] T. Fiig and A. P. Jauho, Appl. Phys. Lett. **59**, 2245 (1991).

[11] W. R. Frensley, Sol. St. Elec. **32**, 1235 (1989).

[12] W. R. Frensley, Rev. Mod. Phys. **62**, 745 (1990).

[13] B. Y. Gelfand, S. Schmitt-Rink, and A. F. J. Levi, Phys. Rev. Lett. **62**, 1683 (1989).

[14] K. Haule and J. Bonča, Phys. Rev. B **59**, 13087 (1999).

[15] D. Huang, S. Kao, and Y. Pang, in *IEEE Radio Frequency Integrated Circuits Symp.* (2007) p. 369.

[16] E. O. Kane, in *Tunneling phenomena in solids*, edited by E. Burstein and S. Lundqvist (Plenum Press, New York, 1969).

[17] Y. Khatami and K. Banerjee, IEEE Trans. Elec. Dev. **56**, 2752 (2009).

[18] R. Lake, G. Klimeck, R. C. Bowen, and D. Jovanovic, J. Appl. Phys. **81**, 7845 (1997).

[19] R. Landauer, IBM J. Res. Dev. **1**, 223 (1957).

[20] R. Landauer, Philos. Mag. **21**, 863 (1970)

[21] T. H. Lee, *The design of CMOS radio-frequency integrated circuits* (Cambridge University Press, New York, 1998).

[22] A. F. J. Levi and T. H. Chiu, Appl. Phys. Lett. **51**, 984 (1987).

[23] A. F. J. Levi, *Applied Quantum Mechanics*, 2nd ed. (Cambridge University Press, New York, 2006).

[24] A. F. J. Levi and I. G. Rosen, SIAM J. Cont. Opt. **48**, 3191 (2010).

[25] W. Lu, P. Xie, and C. M. Lieber, IEEE Trans. Elec. Dev. **85**, 2859 (2008).

[26] K. C. Magruder, P. Seliger, and A. F. J. Levi, in *Optimal Device Design*, edited by A. F. J. Levi and S. Haas (Cambridge University Press, New York, 2010).

[27] Y. Meir and N. S. Wingreen, Phys. Rev. Lett. **68**, 2512 (1992).

[28] D. A. B. Miller, J. Opt. Soc. Am. B **24**, A1 (2007).

[29] A. Mitra, I. Aleiner, and A. J. Millis, Phys. Rev. B **69**, 245302 (2004).

[30] S. Monturet and N. Lorente, Phys. Rev. B **78**, 035445 (2008).

[31] G. E. Moore, Electronics **38**, 114 (1965).

[32] A. Nieuwoudt and Y. Massoud, IEEE Trans. Elec. Dev. **55**, 2097 (2008).

[33] B. P. W. de Oliveira and S. Haas, Phys. Rev. B **79**, 155102 (2009).

[34] X. Oriols, Phys. Rev. Lett. **98**, 066803 (2007).

[35] J. J. Plombon, K. P. OBrien, F. Gstrein, V. M. Dubin, and Y. Jiao, Appl. Phys. Lett. **90**, 063106 (2007).

[36] W. Pötz, J. Appl. Phys. **66**, 2458 (1989).

[37] P. Roblin and W. Liou, Phys. Rev. B **47**, 2146 (1993).

[38] M. J. W. Rodwell, M. Le, and B. Brar, Proc. IEEE **96**, 271 (2008).

[39] P. Schmidt, S. Haas, and A. F. J. Levi, Appl. Phys. Lett. **88**, 013502 (2006).

[40] P. Seliger, *Engineering design optimization*, Ph.D. thesis, University of Southern California (2008).

[41] M. D. Stiles and D. R. Hamann, Phys. Rev. B **38**, 2021 (1988).

[42] M. A. Struscio and M. Dutta, *Phonons in nanostructures* (Cambridge University Press, New York, 2001).

[43] T. Tikka, J. Ryynnen, and K. Halonen, in *IEEE Radio and Wireless Symp.* (2008) p. 871.

[44] A. Trellakis, A. T. Galick, A. Pacelli, and U. Ravaioli, J. Appl. Phys. **81**, 7880 (1997).

[45] P. Vogl, H. P. Hjalmarson, and J. D. Dow, J. Phys. Chem. Solids **44**, 365 (1983).

[46] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, J. Appl. Phys. **89**, 5815 (2001).

[47] J. Zhang and R. L. Kosut, in *European Control Conference* (2007).

[48] H. Zhao, Y. Chen, Y. Wang, F. Zhou, and J. Lee, IEEE Elec. Dev. Lett. **31**, 1392 (2010).

[49] J. M. Ziman, *Principles of the theory of solids*, 2nd ed. (Cambridge University Press, New York, 1972).