Ballistic Modeling of Nanowire MOSFETs

Lasse Södergren
lasse.sodergren@outlook.com

Department of Electrical and Information Technology
Lund University

Supervisor: Erik Lind

Examiner: Mats Gustafsson

June 29, 2017

Printed in Sweden
Tryckeriet i E-huset, Lund

Abstract

III-V nanowire transistors are compelling for high performance applications. The higher carrier mobility compared to Si enables high current at low operating voltages. The nanowire structure facilitates multiple gate configurations such as tri-gate or gate-all-around geometries, which has improved electrostatics compared to planar devices.

In this thesis the performance limit of a tri-gate nanowire MOSFET has been explored. Performance metrics such as g_m, g_d and f_T has been investigated with the help of a 1D non-parabolic ballistic transistor model. The non-parabolicity factor α is found from $k{\cdot}p$ theory and helps to more accurately predict the behavior of electrons which are far from the bulk band edge in an effective mass model. The ballistic current is calculated from a top of the barrier model and the charges in the channel are evaluated by integration of the density of states (DOS) multiplied by the Fermi-Dirac distribution. An attempt to approximate this calculation by shifting the Fermi level and take advantage of the fact that that the Fermi-Dirac distribution is a step function at 0 K is presented. Trans-capacitances such as C_{gs} and C_{gd} are evaluated and discussed for different approximations of the potential along the channel. Their dependence on dimensions such as nanowire width and gate length are also discussed. The model predicts non zero intrinsic capacitances even at zero gate bias because of charges present in the channel which are not a part of the current but has an influence on the capacitances. Finally the model is compared to some experimental data of a lateral InGaAs nanowire MOSFET.

Populärvetenskaplig Sammanfattning

Prestandautvecklingen av integrerade kretsar de senaste 40 åren kan främst tillskrivas nedskalningen av transistorer. Transistorn är det grundläggande byggblocket i all dagens elektronik. Transistorn fungerar som ett elektriskt relä som används för att styra strömmar. Transistorn består främst utav tre stycken elektroder som kallas gate, source och drain. Mellan source och drain finns där en kanal vars resistans kan styras utav den spänning som läggs på gate-elektroden. På detta sätt kan strömmen genom transistorn styras till att vara av eller på.

Nedskalning till nanoskala av dessa transistorer förbättrar många av dess egenskaper, så som lägre energiförbrukning, användning vid högre frekvenser och har även gjort det möjligt att integrera miljardtals transistorer på ett och samma chip i till exempel en processor (CPU). Traditionellt sett har transistorer byggts utav kisel som har många bra egenskaper för ändamålet men idag hindras fortsatt nedskalning av fysikaliska begränsningar hos materialet. Mycket forskning inriktar sig därför på utveckling av transistor utav andra material, ett av dessa materialen är III-V halvledare, till exempel InAs eller InGaAs. Dessa material har egenskaper som kan leda till snabbare transistorer och lägre energiförbrukning. III-V halvledare kan användas för att bygga så kallade nanotrådar som bara är ett tiotal nanometer i diameter. Geometrin av dessa strukturer gör det relativt lätt att bygga komponenter där gate-kontakten påverkar kanalen från flera sidor, vilket förbättrar transistorns egenskaper. När dessa nanotrådstransistorer skalas ner till väldigt korta gate-längder förväntas transportmekanismen att bli så kallad ballistisk, vilket innebär att elektronerna inte sprids utav potentialen från atomerna i kanalen, utan passerar rakt förbi. Detta är en idealistisk bild som sätter en övre gräns för hur bra egenskaper transistorn kan ha.

I det här projektet har den maximala prestandan hos nanotrådstransistorer undersökts med hjälp av fysikaliska och ballistiska modeller och simuleringar. Prestandan har undersökts beroende på olika dimensioner av nanotråden. Kapacitanser som har stor inverkan på hur snabb transistorn är, har utvärderats med hjälp av olika modeller. Simuleringsresultaten jämförs också mot uppmätt data för att kunna avgöra modellens tillförlitlighet. Modeller som beskriver hur nanotrådstransistorer fungerar är en viktig del i den fortsatta utvecklingen av dessa komponenter.

Acknowledgements

I would like to express my gratitude to my supervisor Erik Lind for his support and unlimited guidance. Many invaluable discussions took place during this thesis work which gave me great insight into the subject. I would also like to thank my family for their support and encouragement.

Table of Contents

Chapter 1

Introduction

The device component that can be found in nearly every electrical circuit is the transistor, or more specifically the Metal Oxide Semiconductor Field Effect Transistor. The MOSFET is used in many different electronic applications, such as signal and power amplification in analog circuits or as a current switch in digital circuits performing logic operations. Traditionally the MOSFET is built as a planar structure with silicon as the semiconductor, as in figure 1.1. The device

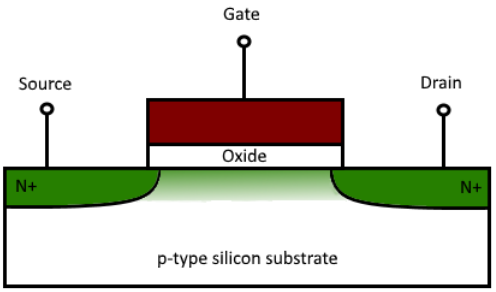

Figure 1.1: Schematic of a typical planar n-channel MOSFET, built on a p-type silicon substrate with n+ doped regions as source and drain.

consists of three contacts named gate, source and drain. The current through the channel between source and drain is modulated by the potential on the gate. The gate is electrically isolated from the channel by an insulator, often an oxide. The gate is therefore capacitively coupled to the semiconductor channel, and the amount of charge carriers in the channel can thus be controlled by the voltage applied to the gate contact. The channel can be in three different regimes depending on the gate potential, called accumulation, depletion and inversion. In accumulation the majority charge carriers are attracted to the semiconductor surface, close to the interface with the oxide. In depletion the amount of these carriers are instead vastly reduced. When the gate bias is above the threshold voltage V_T, minority charge carriers are generated and the channel is inverted, hence inversion is obtained. The amount of charge carriers available in the channel relates to

the conductivity which is modulated by changing the height of an energy barrier between the source and the drain. In a n-channel device, when the gate voltage is above the threshold voltage, the channel is inverted and electrons can pass trough from the source to the drain. When the gate voltage is below the threshold voltage, the energy barrier is very high and the current is close to zero. A well designed MOSFET has a high enough energy barrier so the device can be turned off. The off state leakage current between the source and the drain is important to minimize, especially in digital circuits where the number of transistor on the chip is very large. A well designed MOSFET should also have a good gate design which effectively can lower the energy barrier and turn the device on, with a low voltage swing. This is important to be able to operate the device at small supply voltages, which is correlated with the power dissipation of the transistor. The gate leakage current from the gate to the channel should also be minimized by using appropriately thick oxide.

The performance of a MOSFET can generally be improved by decreasing the distance between the source and the drain areas, in other word decreasing the gate length. At very small gate lengths, the close proximity between the source and the drain limits the ability of the gate to modulate the potential in the channel. These structures become susceptible to short-channel effects, such as Drain Induced Barrier Lowering (DIBL) and threshold voltage roll-off. By instead adopting a tri-gate structure where the gate contacts the channel from three sides, the short-channel effect are suppressed because of the better electrostatic control of the channel. This enables more aggressive gate length scaling which can lead to devices operating at very high frequencies.

The last 40 years the the development and performance improvement of the MOSFET has been the main focus of the semiconductor industry. This can mainly be attributed to the downscaling of the silicon MOSFETs. A shorter gate length enables the use of higher operating frequencies but also increases power density. The performance of today's digital circuits are not gated by the transistor speed but rather the power dissipation of the chip [1]. A lower supply voltage drastically decreases the power dissipation but at the same time the on-currents of the transistor degrades. By using a material in the channel with higher carrier mobility such as III-V semiconductors this issue may be solved. The high mobility of these materials enables a reduction in supply voltage without a loss of performance compared to the silicon channel device. For very scaled transistors where the transport is close to fully ballistic, the high carrier mobility is translated to high source-side carrier injection velocity [1]. For devices where the mean free path of the carriers is longer than the channel, the transport is expected to be ballistic or at least close to it [2]. In fully ballistic transport, the carriers do not scatter inside the channel, compared to diffusive transport, where scattering occurs. This means that the carriers will not change their velocity inside the channel, making the injection velocity an important parameter [3].

The inherent shape of nanowires, coupled with the possibility to mix multiple elements from semiconductors in group III and V, makes nanowires a compelling system for designing high performance MOSFETs. However, this does not come without issues, at short gate lengths thin nanowires are required to suppress the short-channel effects. Usually, shrinking the nanowires degrades the mobility and

the mean free path. This sets a high demand on the quality of the oxide and the interface between the oxide and the channel [4].

In this thesis, a ballistic model is used to investigate the performance limits of a tri-gate nanowire MOSFET structure. Metrics such as g_m, g_d and f_T will be evaluated, but the focus is on the trans-capacitances C_{gs} and C_{gd} which has a strong influence on the speed of the transistor. The amount of charges in the channel will be calculated with different approximations of the potential along the channel. The simulated capacitances are compared to measured data in order to evaluate the model. These kind simulations are an important part for further development and understanding of how the capacitances in 1D nanostructures behave. Accurate device models are critical for optimizing nanowire MOSFET fabrication and designs.

1.1 Performance Metrics

Different metrics which are used to compare the performance of MOSFETs are introduced in this section. These metrics are important to able to build accurate device models, and to understand the device performance and its use when implemented in an integrated circuit.

The threshold voltage V_T, is the minimum voltage needed on the gate contact to achieve inversion of the channel and turn the transistor on. The transistor is off when $V_{GS} < V_T$.

The transconductance g_m, is defined as the rate of change of drain current with respect to the gate voltage,

$$g_m = \frac{\delta I_D}{\delta V_{GS}}$$

This relates to how effectively the gate voltage modulates the energy barrier inside the channel. The transconductance also sets the gain of the device. The output conductance g_d, is instead the rate of change of the drain current but with respect to the drain voltage,

$$g_d = \frac{\delta I_D}{\delta V_{DS}}$$

It is generally favorable to keep the output conductance minimized. The intrinsic voltage gain is defined as $A_v = g_m/g_d$.

The transition frequency f_T, is the maximum frequency of the input signal, where the device still has current gain. The transition frequency for a well design device with low g_d and low access resistance R_s and R_d is,

$$f_T \approx \frac{g_m}{2\pi(C_{gs} + C_{gd})}.$$

Chapter 2

Model Description

2.1 Introduction

The wave function for electrons in a semiconductor with an effective mass description is obtained by solving the Schrödinger equation.

$$-\frac{\hbar^2}{2m^\star}\nabla^2\Psi(\mathbf{r}) + U(\mathbf{r})\Psi(\mathbf{r}) = E\Psi(\mathbf{r}) \tag{2.1}$$

Consider a 1D semiconductor wire where the electrons are confined in the x- and y-direction but are free to move in the z-direction. The wave function of these electrons is found by solving equation 2.1 by separation of variables [5]. If the potential energy $U(\mathbf{r})$ is constant the solutions are plane waves.

$$\Psi(\mathbf{r}) = \Phi(x,y)\Psi(z) = \Phi(x,y)\frac{1}{\sqrt{L_z}}e^{ik_z z} \tag{2.2}$$

The parameter L_z is the normalization length. With the confining potential as a square well with infinite height,

$$\Phi(x,y) = \frac{2}{\sqrt{W_x W_y}}\sin\left(\frac{nx\pi}{W_x}\right)\sin\left(\frac{my\pi}{W_y}\right). \tag{2.3}$$

The expression above leads to the kinetic energy of carriers E has a parabolic relation to the wave vector k.

$$E = \frac{\hbar^2 k^2}{2m^\star} \tag{2.4}$$

However, for electrons confined in a narrow channel by a large barrier height, it is required to include the effects of non-parabolic bands. For high applied fields or for very high carrier densities the electron energies may be far from the bulk band edge and the parabolic band description does not provide sufficient accuracy [6] [8].

The 3D non-parabolic $E-k$ dispersion is

$$E(1+\alpha E) = \frac{\hbar^2\mathbf{k}^2}{2m^\star} \tag{2.5}$$

where α is the non-parabolicity factor which can be found from $k \cdot p$ theory to be [6]

$$\alpha \approx \frac{1}{E_g}\left(1 - \frac{m^\star}{m_0}\right)^2, \tag{2.6}$$

where E_g is the direct bandgap. Following [7], for a nanowire with a hard wall quantization in the x- and y-direction, the electron wave function is required to be zero at the edges. For these boundary conditions we need $k_x = \frac{n\pi}{W_x}$ and $k_y = \frac{m\pi}{W_y}$ where W_x and W_y is the width in their respective direction and n and m is the indexes of the 1D sub-bands. Equation 2.5 can then be solved for the sub-band energies at $k_z = 0$.

$$E_{n,m} = (\gamma_{n,m} - 1)/2\alpha \tag{2.7}$$

with

$$\gamma_{n,m} = \sqrt{1 + \frac{2\alpha\hbar^2\pi^2}{m^\star}\left[\frac{n^2}{W_x^2} + \frac{m^2}{W_y^2}\right]} = \sqrt{1 + 4\alpha E^p_{n,m}} \tag{2.8}$$

where $E^p_{n,m}$ is the sub-band energies in the parabolic bands. Comparing equation 2.7 and 2.8 with equation 2.5 we find that each sub-band has their respective effective sub-band mass

$$m^\star_{n,m} = m^\star \gamma_{n,m} \tag{2.9}$$

and an effective sub-band non-parabolicity factor

$$\alpha_{n,m} = \frac{\alpha}{\gamma_{n,m}} \tag{2.10}$$

Introducing non-parabolicity in the description of the bands increases the effective sub-band mass with energy but the non-parabolicity factor decreases with higher energy sub-bands.

The 1D density of states (DOS) per unit energy and length for each non-parabolic band [7] is

$$D_{1D}(E) = \frac{\sqrt{2m^\star_{n,m}}(1 + 2\alpha_{n,m}E)}{\pi\hbar\sqrt{E(1 + \alpha_{n,m}E)}} \tag{2.11}$$

where the energy E is given with respect to the sub-band minimum $E_{n,m}$. This reduces to 1D DOS with parabolic bands if $\alpha = 0$.

The density of states multiplied by the probability that a state is occupied gives the carrier density. The probability that a state is occupied is described by the Fermi-Dirac distribution

$$f(E, E_F) = \frac{1}{e^{(E-E_F)/k_BT} + 1} \tag{2.12}$$

and the total electron line density is given by

$$n_L(E_F) = \sum_{n,m} \int_{E_0+E_{n,m}}^{\infty} D_{1D}(E - (E_0 + E_{n,m}))f(E, E_F)dE. \tag{2.13}$$

Figure 2.1: Sketch of a ballistic device with source and drain contacts which act as thermal equilibrium electron resevoirs.

2.2 Semiclassical Ballistic Current

It is assumed that strong scattering occurs in the source and drain contacts which maintains thermal equilibrium, and the contacts acts as perfect absorbers of electrons, see figure 2.1. Both the source and the drain injects a flux of electrons into the device, some electrons reflects from the potential barrier in the device and the rest enters the device and experience no scattering until they reach the opposite contact. Tunneling and quantum reflections are ignored and the electrons are treated as semiclassical particles.

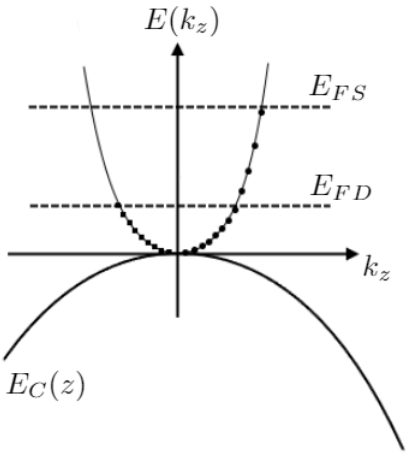

Figure 2.2: The $E-k$ dispersion at the top of the barrier shown together with the source and drain Fermi levels. All the positive k-states (circles) are occupied according to the source Fermi level and all the negative k-states (squares) are occupied according to the drain Fermi-level.

The probability that a k-state is occupied is given by an equilibrium Fermi

function [5], the system has two Fermi levels, one according to the source contact and one according to the drain contact. The energy at the top of the barrier is E_0. Because no scattering occurs, an electron can not change direction inside the device. This means that all the positive k-states with energy above E_0 must have been occupied by injection from the source. Following the same argument, all the negative k-states with energy above E_0 must have been occupied by injection from the drain side. For k-states with less energy that E_0 both positive and negative k-states are occupied, because the electrons can be reflected by the potential barrier. These electrons will not contribute to the current through the device but they will be important for the behavior of the capacitances. The position at the top of the barrier is simple because all positive k-states are filled by injection from the source and all negative k-states are filled by injection from the drain, see figure 2.2.

By using the top of the barrier model, the semiclassical current can be calculated. Starting by computing the electrostatic potential at the top of the barrier [5].

$$E_0 = -q(\alpha_G V_G + \alpha_S V_S + \alpha_D V_D) + q^2 \frac{n_L(E_{FS}) + n_L(E_{FD})}{2C_{ins}} \tag{2.14}$$

The parameters α_G, α_S and α_D are coefficients that model the short channel effects, C_{ins} is the insulator capacitance and E_{FS} and E_{FD} are the source and drain Fermi levels respectively. The density of the electrons injected from the source is half the electron line density (equation 2.13), because only the positive k-states are occupied. The same argument is used for the electrons injected from the drain, where only the negative k-states are occupied. The insulator capacitance is modeled by,

$$C_{ins} = \frac{NF_{ins}\epsilon_{ins}}{\ln\left(1 + F_{ins}\frac{T_{ins}}{H_s}\right)} - \frac{5/4\epsilon_{ins}}{\ln\left(1 + \frac{5T_{ins}}{4H_s}\right)} + \frac{5/4\epsilon_{ins}}{\ln\left(1 + \frac{5T_{ins}}{4W_s}\right)}, \tag{2.15}$$

which is an expression fitted (with fitting parameters N and F_{ins}) to a numerically calculated solution to the Laplace equation [9]. The parameter ϵ_{ins} is the dielectric constant of the insulator, T_{ins} is the thickness of the insulator, W_s and H_s are the width and height of semiconductor channel respectively. Finally the ballistic current can be calculated by

$$I_D = \frac{2qk_BT}{h} \sum_{n,m} \left[\ln\left(1 + e^{\frac{E_{FS} - E_{n,m} - E_0}{k_BT}}\right) - \ln\left(1 + e^{\frac{E_{FS} - qV_D - E_{n,m} - E_0}{k_BT}}\right) \right]. \tag{2.16}$$

2.3 Trans-Capacitances

The objective here is to calculate the number of electrons inside the channel and evaluate how this number changes with applied voltage at different contacts. The definition of trans-capacitances is

$$C_{ii} = \frac{\partial Q_i}{\partial V_i} \qquad\qquad C_{ij} = -\frac{\partial Q_i}{\partial V_j} \tag{2.17}$$

where Q is the charge, V is the voltage, i and j are indexes for the different contacts (source, drain or gate). As a first approximation a constant potential

along the channel is assumed. The total channel charge is then simply obtained by multiplying the charge at the top of the barrier with the gate length. A better approximation of the potential, $\psi(z)$ along the channel is obtained by using a simplification of the analytic model in [10]

$$\psi(z) = V_G + \frac{b_1 \sinh[k_1(L_g - z)] + c_1 \sinh[k_1 z]}{\sinh[k_1 L_g]} \tag{2.18}$$

where L_g is the length of the channel under the gate and z is the position along the channel. With V_{Bi} as the built in voltage we have

$$\begin{aligned} b_1 &= V_S + V_{Bi} - V_G \\ c_1 &= V_D + V_{Bi} - V_G \end{aligned} \tag{2.19}$$

The parameter k_1 can be found from $1/\lambda$ where λ is the natural length scale of a tri-gate MOSFET [11]

$$\lambda_{TG} = \frac{1}{\sqrt{\left(\frac{1}{\lambda_{FD}(T_s=H_s)}\right)^2 + \left(\frac{1}{\lambda_{DG}(T_s=W_s)}\right)^2}} \tag{2.20}$$

with

$$\lambda_{FD} = \sqrt{\frac{T_s}{2}\left(T_s + 2\frac{\epsilon_s T_{ins}}{\epsilon_{ins}}\right)} \tag{2.21}$$

and

$$\lambda_{DG} = \sqrt{\frac{T_s}{8}\left(T_s + 4\frac{\epsilon_s T_{ins}}{\epsilon_{ins}}\right)}. \tag{2.22}$$

Equation 2.20 is combination of the expression for the scaling length of a MOSFET with one gate and MOSFET with two gates. When the width and the height of the channel is comparable in size, both the vertical and horizontal electric fields will influence the channel simultaneously. An example of the potential shape along the channel can be seen in figure 2.3.

Using equation 2.13 the electrons injected from the source side with energies higher than E_0 as function of position z is modeled by

$$Q_s(z) = q\frac{1}{2}\sum_{n,m}\int_{\psi(z)+E_{n,m}+\Delta E_{kin}}^{\infty} D_{1D}(E - E_{n,m} - \psi(z)) f(E, E_{FS}) dE \tag{2.23}$$

with $\Delta E_{kin} = E_0 - \psi(z)$ and q is the elementary charge. The division by two is because only half of the k-states are occupied (the positive ones). Electrons with less energy than E_0 does not contribute to the current but is still important for the capacitance, these are described by

$$Q'_s(z) = q\sum_{n,m}\int_{\psi(z)+E_{n,m}}^{\psi(z)+E_{n,m}+\Delta E_{kin}} D_{1D}(E - E_{n,m} - \psi(z)) f(E, E_{FS}) dE \text{ if } z < z_0 \tag{2.24}$$

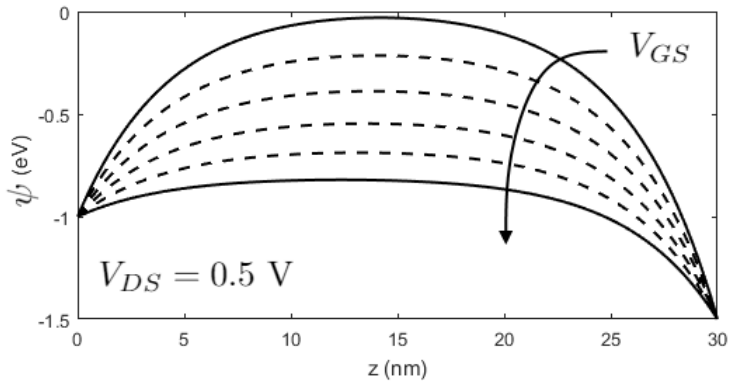

Figure 2.3: Sketch of the conduction band edge with $V_{GS} = 0...1V$ and $V_{DS} = 0.5V$ for a device with $L_g = 30nm$. Increased gate bias pulls down the energy of the top of the barrier.

and $Q'_s = 0$ for $z > z_0$ where z_0 is the position in the channel where E_0 occurs, see figure 2.4. The total charge from the source side as a function of position is the sum of the charges with energies higher and lower than E_0.

$$Q_S(z) = Q_s(z) + Q'_s(z) \tag{2.25}$$

The equations for the electrons injected from the drain side are very similar, the only difference is that the drain Fermi level $E_{FD} = E_{FS} - qV_D$ is used,

$$Q_d(z) = q\frac{1}{2}\sum_{n,m}\int_{\psi(z)+E_{n,m}+\Delta E_{kin}}^{\infty} D_{1D}(E - E_{n,m} - \psi(z))f(E, E_{FD})dE \tag{2.26}$$

$$Q'_d(z) = q\sum_{n,m}\int_{\psi(z)+E_{n,m}}^{\psi(z)+E_{n,m}+\Delta E_{kin}} D_{1D}(E - E_{n,m} - \psi(z))f(E, E_{FD})dE \text{ if } z > z_0 \tag{2.27}$$

and $Q'_d = 0$ for $z < z_0$. The total charge from the drain side as a function of position is

$$Q_D(z) = Q_d(z) + Q'_d(z) \tag{2.28}$$

To get the total number of electrons which are inside we integrate equation 2.25 and 2.28 over the total gate length as

$$Q_{S_{total}} = \int_0^{L_g} Q_S(z)dz \tag{2.29}$$

and

$$Q_{D_{total}} = \int_0^{L_g} Q_D(z)dz. \tag{2.30}$$

Finally the total charge in the channel is

$$Q_{G_{total}} = Q_{S_{total}} + Q_{D_{total}}. \tag{2.31}$$

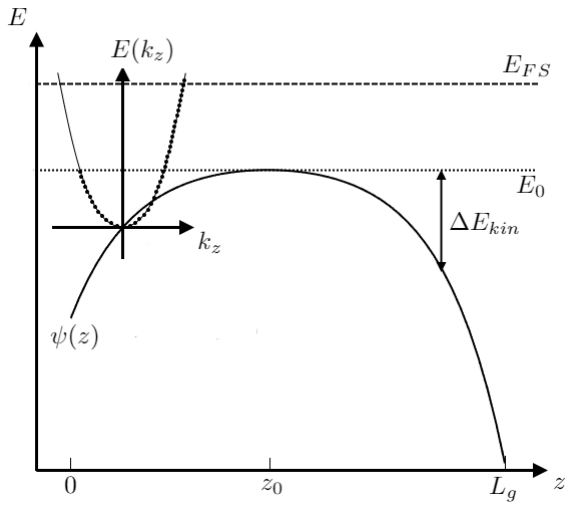

Figure 2.4: Sketch of the energy band diagram with $V_{DS} >> 0$. The source is at $z = 0$ and the drain is at $z = L_g$. The $E - k$ dispersion shows electrons injected from the source, above E_0 only the positive k-states are occupied and below E_0 both negative and positive k-states are occupied, the negative k-states are occupied by electrons which has been reflected at the potential barrier.

All the trans-capacitances can then be computed by using the definition of trans-capacitances (equation 2.17).

2.4 Fermi Level Adjustment

Computing the amount of charges with equation 2.29 and 2.30 involves solving an integral over the 1D DOS multiplied with the Fermi function several times for each position z along the channel, this is of course a heavy computational tasks which is time consuming. An attempt to reduce the amount of numerical integrals that has to be solved will now be discussed.

At $T = 0$K all states with lower energy than the Fermi level are occupied and all states with higher energy are empty. So the Fermi function is a step function at $T = 0$K, this simplifies equation 2.13 for each sub-band

$$n_L(E_F) = \int_{E_{n,m}}^{E_F} \frac{\sqrt{2m^\star_{n,m}}(1 + 2\alpha_{n,m}(E - E_{n,m}))}{\pi\hbar\sqrt{(E - E_{n,m})(1 + \alpha_{n,m}(E - E_{n,m}))}} dE. \tag{2.32}$$

and $n_L = 0$ for energies above E_F. This expression can be solved analytically to obtain

$$n_L(E_F) = \frac{2\sqrt{2m^\star_{n,m}}}{\pi\hbar}\sqrt{(E_F - E_{n,m})(1 + \alpha_{n,m}(E_F - E_{n,m}))}. \tag{2.33}$$

So in order to take advantage of this analytic expression for $T > 0$K, the Fermi level needs to be adjusted. The source side is used as an example, the adjusted source Fermi level is obtained by setting the carrier density injected from the source evaluated from equation 2.13 equal to the injected carrier density according to the analytic expression in equation 2.33 and solve for E_F

$$E_{FSa} - E_{n,m} = -\frac{1}{2\alpha_{n,m}} + \sqrt{\left(\frac{1}{2\alpha_{n,m}}\right)^2 + \frac{1}{\alpha_{n,m}}\left(\frac{\pi\hbar n_{LS}}{2\sqrt{2m^\star}}\right)^2}. \tag{2.34}$$

The carrier density originating from the source is n_{LS} which is half of the carrier density from the general expression in equation 2.13 because only the positive k-states are occupied. When the adjusted Fermi level is known the charges injected from the source can be calculated as

$$Q_s(z) = q\sum_{n,m}\frac{\sqrt{2m^\star_{n,m}}}{\pi\hbar}\Bigg[\sqrt{(E_{FSa} - E_{n,m} - \psi(z))(1 + \alpha_{n,m}(E_{FSa} - E_{n,m} - \psi(z)))} - \sqrt{(E_0 + E_{n,m})(1 + \alpha_{n,m}(E_0 + E_{n,m}))}\Bigg]. \tag{2.35}$$

Similar expressions are used for the charges injected from the drain side, but n_{LD} is used instead of n_{LS} is equation 2.34. Any approximation for charges with energy less than E_0 has not been implemented in this thesis work.

Chapter 3
Simulations & Analysis

3.1 Introduction

In practice a source Fermi level is assumed, then equation 2.13 and 2.14 are solved by multiple iterations until convergence for a given gate and drain voltage. The coefficient α_G, α_S and α_D that model the short channel effects are obtained from the rate of change of the energy at the top of the barrier from equation 2.3 at $V_{GS} = V_{DS} = 0.5$V. The fitting parameters used to calculate C_{ins} in equation 2.15 are $N = 3.05$ and $F_{ins} = 0.95$ [9]. When the energy of the top of the barrier E_0 compared to the source Fermi level is known, the drain current is then computed from equation 2.16. In order to obtain the amount of charges in the channel, the expression for the potential along the channel in equation 2.3 is fitted so the maximum potential $\psi(z_0)$ is equal to E_0. Then equation 2.29 is used to compute the source charges and for the drain charges equation 2.30 is used.

3.2 Drain Current and Transconductance

Simulations of the drain current and transconductance for a rectangular InAs nanowire MOSFET is shown in figure 3.1. The bandgap is $E_g = 0.36$ eV and the effective mass $m^\star = 0.023m_0$. Both the width and the height of the wire is 10 nm and the gate length is $L_g = 50$ nm. The insulator thickness is set to $t_{ins} = 3$ nm with a dielectric constant of $\epsilon_{ins} = 20$. The threshold voltage is adjusted so the off-current is 100 nAμm^{-1} at $V_{GS} = 0$ V. In the g_m plot, the onset where new sub-bands is starting to conduct current can be seen even for this simulation at $T = 300$ K. At lower temperatures these features would have been much sharper. Keep in mind that these are ballistic simulations where all scattering is neglected, which of course is not true for a real device. But ballistic transport sets an upper limit on what device performance can be expected from the device.

3.3 Trans-Capacitances

Simulations of C_{gs} for the same InAs nanowire MOSFET, with $L_g = 30$ nm and $V_{DS} = 0.5$ V are shown in figure 3.2. The three different data sets a) - c) corresponds to different ways to treat the charges and the potential profile inside

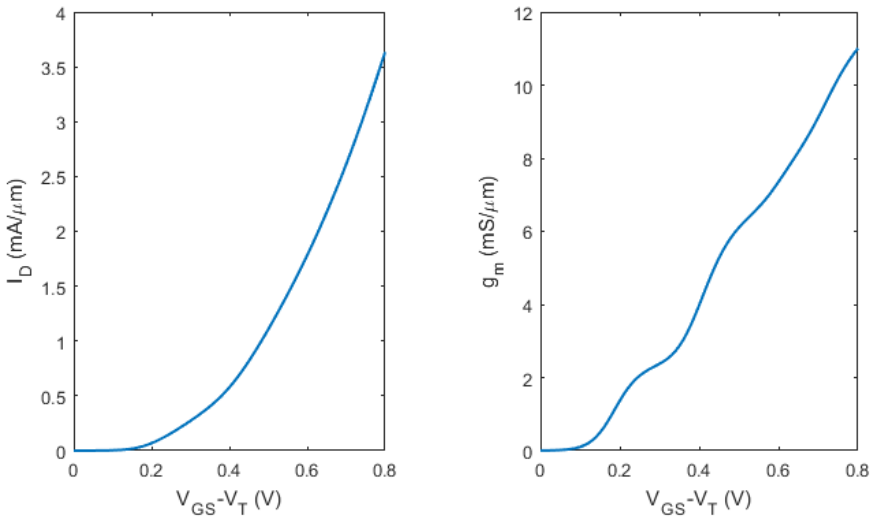

Figure 3.1: (a) Simulated ballistic drain current I_D and (b) transconductance g_m for an InAs nanowire MOSFET with $W_s = H_s = 10$ nm. Normalization is done using the total gated perimeter $(W_s + 2H_s)$.

the channel. The simplest approach is to approximate the potential along the channel to be constant, this is shown in data set a). Because of the flat potential, the onset where each sub-band starts to be occupied with electrons can clearly be seen, and the capacitance follows the 1D DOS.

In data set b) the potential is modeled as in equation 2.3, but excluding the charges with energy lower than the energy at the top of the barrier, E_0. Now the bands bend down to lower energies close to the source and drain contacts which means that the electrons will have a higher velocity, and less charges will be inside the channel. This gives a decreased capacitance compared to the first data set, and the features of the DOS are more difficult to resolve.

Finally data set c) also includes the electrons which are not a part of the current, but are still present inside the channel and will influence the capacitance. These are the electrons with energy lower than E_0. This model gives a higher capacitance at lower voltages, because charges which are positioned at the source and drain sides of the channel are still present, even below the threshold voltage. This means that the device has an intrinsic gate-source capacitance even at zero gate bias. There is two competing effects which explains the lower capacitance at higher gate voltages. With increasing gate voltage the energy at the top of the barrier is lowered and more electrons are injected into the device and filling more sub-bands. The other effect is that with higher gate voltage the potential becomes more flat and ΔE_{kin} is decreased overall which decreases the amount of charges with energy lower then E_0. These are competing effects which lower the

capacitance at high gate bias. In this model the amount of the charges with energy lower than E_0 are overestimated because depletion of the source/drain contacts are not taken into account.

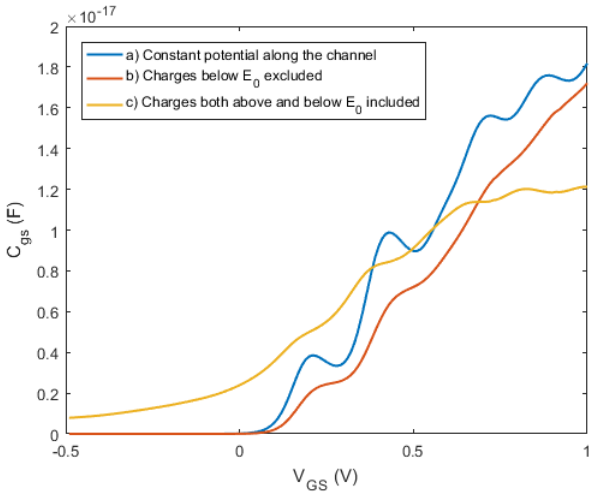

Figure 3.2: Simulated C_{gs} of an InAs nanowire MOSFET with $W_s = H_s = 10$ nm and $L_g = 30$ nm. The three data sets corresponds to different models of the potential/charges along the channel.

Simulations of C_{gd} for the same device, with the same three data sets can be seen in figure 3.3. The states with energy higher than E_0 are not filled by injection from the drain side until the top of the barrier is close in energy with drain Fermi level E_{FD}. This happens at approximately $V_{DS} + E_{11}$ for a device with zero threshold voltage and good electrostatic behavior. The small negative capacitance seen in data set a) and b) is because the top of the barrier is slightly lowered with a small increase in drain voltage. This happens because the amount of carriers injected from the drain side (negative k-states) decreases, so the amount of carriers from the source (positive k-states) has to increase to keep the device charge neutral, due to this the top of the barrier has to be lowered. For gate voltages above approximately $V_{DS}+E_{11}$ this effect is not seen because the amount of charges injected from the drain side is much larger.

The charges with energy less than E_0 are still present below $V_{GS} \approx V_{DS}+E_{11}$ as seen in data set c). This means that the intrinsic gate-drain capacitance is non-zero at $V_{GS} = 0$ V. The capacitance C_{gd} is overestimated as well because the source/drain depletion is neglected.

By using the simple expression for the transition frequency $f_T \approx g_m/2\pi(C_{gs} + C_{gd})$, the plot in figure 3.4 is computed. The oscillations seen are originating from the 1D DOS, with increasing gate voltage more sub-bands become occupied by electrons. The onset of the sub-bands, where the DOS is high, gives a large peak capacitance, hence a valley in the transition frequency. At low temperatures, with low distribution of electron energies due to thermal effects, these oscillations would have been even stronger. In data set a) and b) at voltages below threshold the f_T is not

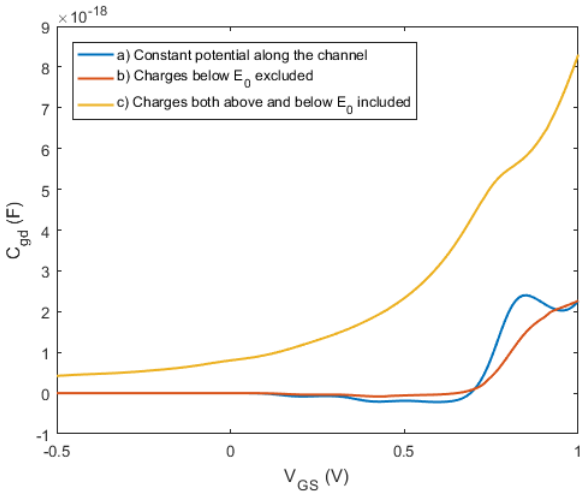

Figure 3.3: Simulated C_{gd} of an InAs nanowire MOSFET with $W_s = H_s = 10$ nm and $L_g = 30$ nm. The three data sets corresponds to different models of the potential/charges along the channel.

going towards zero because C_{gs} and C_{gd} are very small. In data set c), f_T is going towards zero at low gate voltages. Above threshold both the transconductance and $C_{gg} = C_{gs} + C_{gd}$ are increased and f_T is set by the ratio. The charges with less energy than E_0 are important to include in the estimation of the transition frequency.

3.4 Comparison with Experimental Data

Measured C_{gs} and C_{gd} data from [12] is shown in figure 3.5 together with simulated data. The device consists of 200 $In_{0.63}Ga_{0.37}As$ nanowires with dimensions, $W_s = 25$ nm, $H_s = 11$ nm and $L_g = 32$ nm. The bandgap used in the simulations is $E_g = 0.75$ eV and the effective mass used is $m^\star = 0.041 m_0$.

The simulated C_{gs} seems to agree fairly well with measured data except for the offset of approximately 5 fF. This is attributed to parasitic capacitances, $C_{gs,p}$ present in the device which are not accounted for in the simulations. An increased drain voltage means less charges injected from the drain, so the total gate charge has a higher proportion of charges which originates from the source side. The total gate charge thus has a stronger dependence on the source potential and C_{gs} is increased with V_{DS}. In the simulations, this effect is weaker at low gate voltages due to competing effect of charges below and above E_0 is more in favor of the charges below E_0 because there is very few charges above E_0 at low V_{DS}. As stated before, the amount of charges below E_0 is overestimated because the model does not take source/drain depletion areas into account.

The simulated C_{gd} does not include any parasitic capacitances which are present in the real device and estimated to be approximately $C_{gd,p} = 5$ fF. With

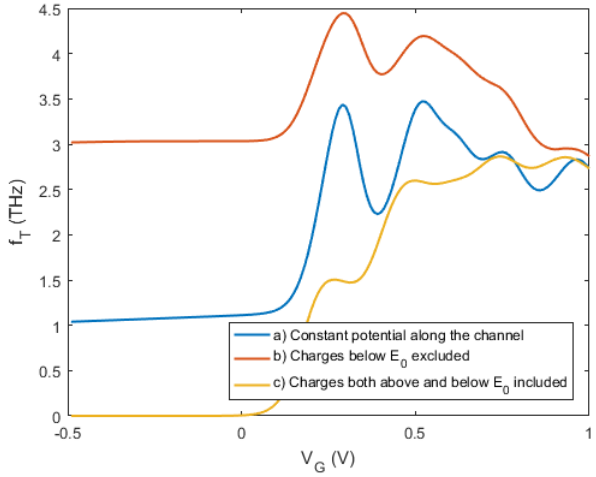

Figure 3.4: Computed f_T of an InAs nanowire MOSFET with $W_s = H_s = 10$ nm and $L_g = 30$ nm. The three data sets corresponds to different models of the potential/charges along the channel.

increased drain voltage, the gate voltage where the energy of the top of the barrier is lower than the drain Fermi level is larger. This gives a lower C_{gd} for higher drain voltages for the simulated data. The model greatly overestimated the gate voltage dependence of C_{gs}, where as in the real device C_{gd} is nearly independent of the gate voltage. This overestimation can most probably be attributed to neglecting the depletion of the drain. In a real device, not all of the charges in the channel is balanced by charges on the gate to make the device charge neutral, some of them is positioned at the source/drain. This will make the band bend at the source/drain and not all of the applied source/drain potential will shift the potential under the gate.

In figure 3.6 computed f_T is shown versus different amounts of added parasitic capacitance, $C_{gg,p}$. The carrier ballisticity is a scaling of g_m, so for 50 % carrier ballisticity, the transconductance is 50 % of the fully ballistic simulation. The parasitic capacitance has a stronger effect on the transition frequency for a high degree of ballisticity. Comparing to the measured results of $f_T = 230$ GHz [12] which is agrees reasonably well with a device with roughly 50 % ballisticity and $C_{gg,p} = 10$ fF.

3.5 Scaling

Transcondcutance at $V_{GS} = V_{DS} = 0.5$V for a InAs MOSFET with $W_s = H_s = 10$ nm as a function of gate length is displayed in figure 3.7a. At large gate lengths the gate has good electrostatic control over the channel and g_m is slightly above $6mS/\mu m$. For short gate lengths the short-channel effects are not suppressed and the device has poor gate control and the g_m suffers as a consequence. To improve

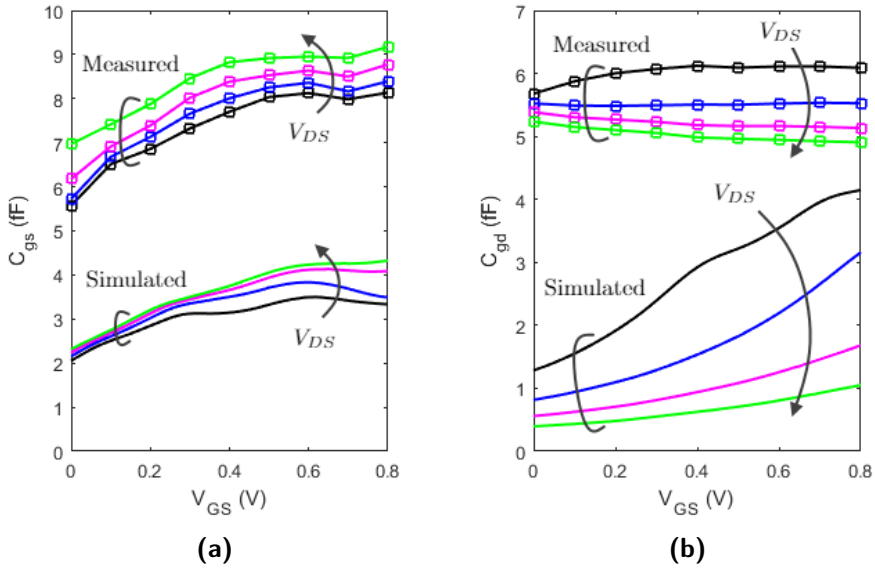

Figure 3.5: Measured C_{gs} and C_{gd} data from [12] together with simulated data for four different drain voltages, $V_{DS} = 0.25, 0.5, 0.75, 1$ V.

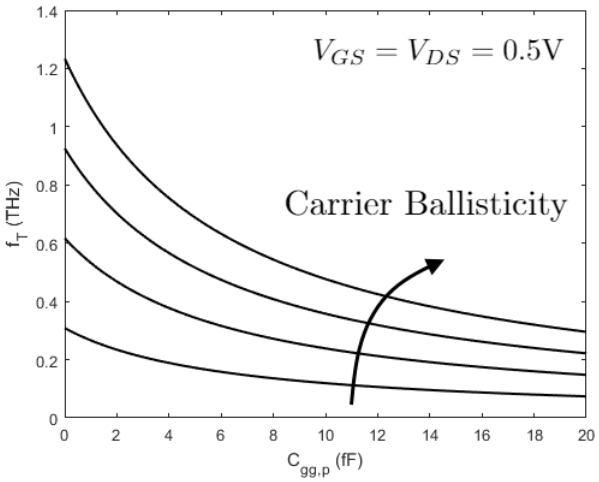

Figure 3.6: Computed f_T at $V_{GS} = V_{DS} = 0.5$ V versus different amounts of parasitic capacitance, $C_{gg,p}$. The carrier ballisticity is 100, 75, 50 and 25 %, with f_T being higher with a higher degree of ballisticity.

the electrostatics and hence the transconductance, the width and the height of the channel need to be reduced.

The capacitances C_{gs} and C_{gd} at $V_{GS} = V_{DS} = 0.5$ V for different gate lengths are shown in 3.7b. At these bias conditions nearly all of the injected charges in the channel with energy higher than E_0 originates from the source side and most of the charges from the source has energy higher than E_0. With a longer gate length more of these charges are situated inside the channel, hence the increase of C_{gs}. The majority of the charges injected from the drain has energy less than E_0, the amount of these charges are nearly independent of the gate length.

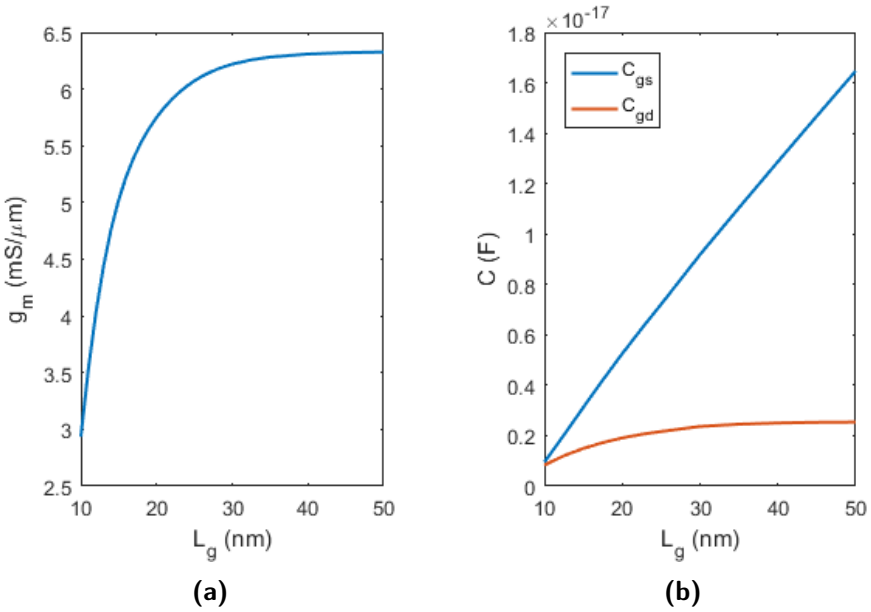

Figure 3.7: Transconductance (a) and C_{gs} & C_{gd} (b) at $V_{GS} = V_{DS} = 0.5$ V as a function of L_g for an InAs MOSFET with dimensions $W_s = H_s = 10$ nm.

The transition frequency, f_T is shown in figure 3.8 with different amounts of added parasitic capacitances $C_{gg,p}$ ranging from 0 to 3 fF for a device with 200 nanowires. With no additional parasitic capacitance the gain of a scaled device is clearly seen. With added parasitic capacitance the improvements from the gate length scaling are diminished. This happens because the intrinsic capacitances are approaching the same magnitude or even become smaller than the parasitic capacitances. For this reason f_T is strongly limited by parasitic capacitances at short gate lengths.

The output conductance g_d is presented in figure 3.9. Worse gate control lead to an increased output conductance at small gate lengths. The electrostatics can be improved by scaling the width and the height of the nanowire towards smaller dimensions. The intrinsic gain, g_m/g_d at $V_{GS} = V_{DS} = 0.5$ V is shown in figure 3.10. At small gate lengths the output conductance is approaching the same order

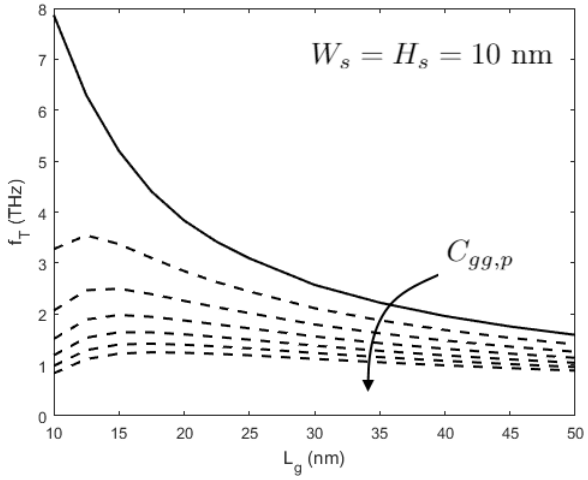

Figure 3.8: Simulated f_T versus L_g with different amounts of added parasitic capacitance, $C_{gg,p} = 0, 0.5, 1, 1.5, 2, 2.5, 3$ fF for a device with 200 nanowires. The InAs device dimensions are $W_s = H_s = 10$ nm.

as the mutual conductance and the intrinsic gain is small.

3.6 Evaluation of Fermi Level Adjustment Approximation

The gate charges of a InAs MOSFET with $W_s = H_s = 10$ nm and $L_g = 30$ nm has been simulated in to different ways. Data set a) in figure 3.11 is obtained by regular integration using equation 2.23 and 2.26. Data set b) is from equation 2.35 and the respective expression for the charges injected from the drain. Keep in mind that the gate charge here is not the total gate charge because the electrons with energy lower than E_0 are not included.

When each higher sub-band starts to be occupied by electrons the approximation shows some discrepancy. This is because of the singularity in the 1D DOS at the sub-band edge, this causes the adjusted Fermi level to be set a bit low, so at other positions z in the channel the amount of charges will be underestimated. When all charges at every z position is added, the total charge will then be lower than computed by the integration. This discrepancy is seen at every sub-band edge. When the source Fermi level is below E_0 (at low V_{GS}) and there is a very low charge in the channel, the error in the approximation is larger.

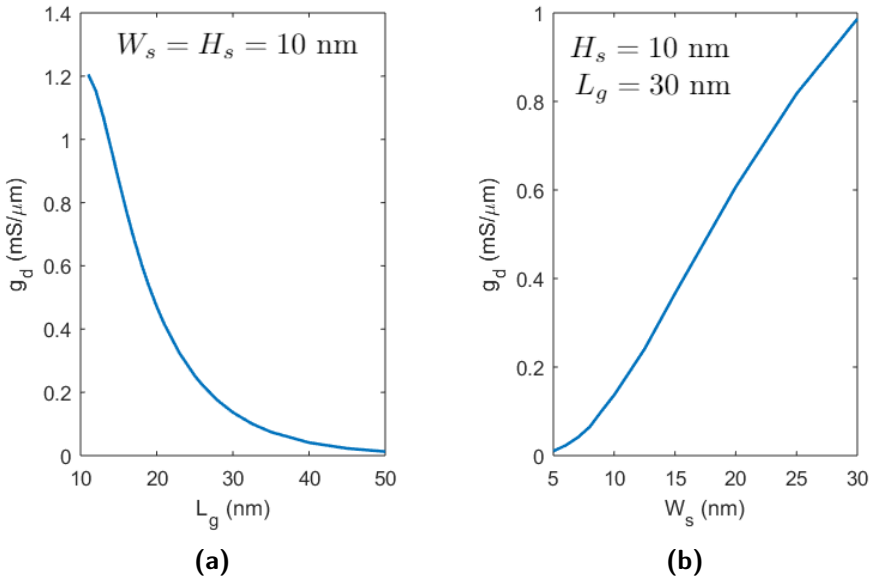

Figure 3.9: Output conductance as a function of L_g(a) and as a function of nanowire width W_s (b) at $V_{GS} = V_{DS} = 0.5$ V for an InAs MOSFET.

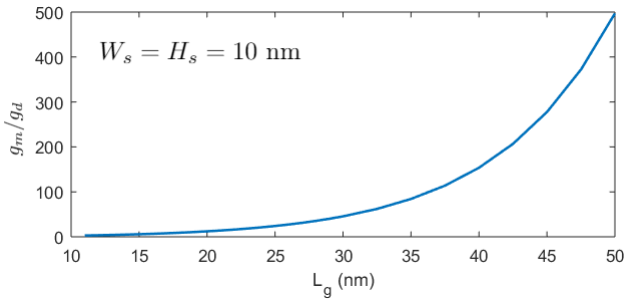

Figure 3.10: Intrinsic gain g_m/g_d at at $V_{GS} = V_{DS} = 0.5$ V as a function of gate length. The InAs device dimensions are $W_s = H_s = 10$ nm.

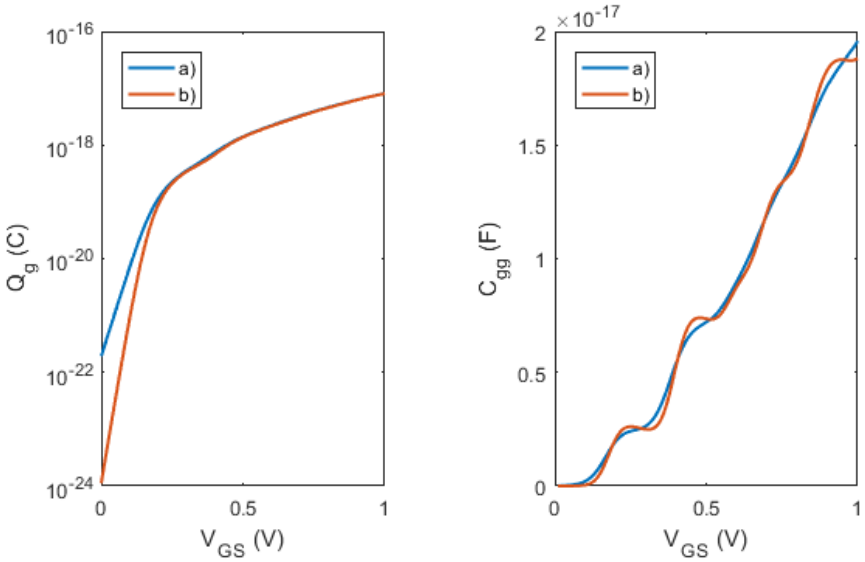

Figure 3.11: Gate charge Q_g and total gate capacitance C_{gg} with $V_{DS} = 0.5$ V computed by a) regular integration and b) Fermi level adjustment approximation.

Chapter 4
Conclusion & Outlook

In a ballistic top of the barrier model, even the charges which are not a part of the current has a large influence on the capacitance behavior of the device. Simplifying the potential along the channel to be constant works as a first approximation but does not suffice for more accurate modelling which predicts that C_{gs} and C_{gd} are non zero at zero gate bias. The importance to include these charges is further emphasized by the unrealistic f_T at low voltages if they are excluded.

Scaling the gate length and keeping short-channel effects under control by scaling the width and height of the nanowire is important for designing devices with large f_T. The RF-performance is strongly limited by parasitic capacitances for short gate lengths when the intrinsic capacitances are becoming very small. Transistors designed with minimized parasitic capacitances in mind is thus very important for high speed devices. The Fermi level adjustment approximation has a reasonable agreement , though it shows some discrepancy close to sub-band edges.

To further improve this model quantum mechanical effects such as tunneling and the effects of standing electron waves inside the channel should be considered. Some charges in the channel will be balanced from the source/drain areas and not by the gate potential, this effects should also be included. Degrading of the gate control duo traps in the oxide is another effect which has not been explored in this thesis. An analytic approximation for calculating the amount charges with energy less than E_0 similar to the approximation for charges above E_0 done in this thesis should be explored as well.

References

[1] J. A. del Alamo, "Nanometre-scale electronics with III–V compound semiconductors," *Nature*, vol. 479, pp. 317-323, Nov. 2011

[2] C. B. Zota, D. Lindgren, L.-E. Wernersson, and E. Lind, "Quantized conduction and high mobility in selectively grown $In_xGa_{1-x}As$ nanowires," *ACS Nano*, vol. 9, no. 10, pp. 9892–9897, 2015

[3] S. Takagi, T. Irisawa, T. Tezuka, T. Numata, S. Nakaharai, N. Hirashita, Y. Moriyama, K. Usuda, E. Toyoda, S. Dissanayake, M. Shichijo, R. Nakane, S. Sugahara, M. Takenaka and N. Sugiyama, "Carrier-Transport-Enhanced Channel CMOS for Improved Power Consumption and Performance," *IEEE Trans. Electron Devices*, vol. 55, no. 1, pp. 21-39, Jan. 2008

[4] C. B. Zota, L.-E. Wernersson and E. Lind, "High-Performance Lateral Nanowire InGaAs MOSFETs With Improved On-Current," *IEEE Electron Device Lett.*, vol. 37, no. 10, pp. 1264-1267, Oct. 2016

[5] M. Lundstrom and J. Guo, 2006, *Nanoscale Transistors: Device Physics, Modeling and Simulation*, New York: Springer Science & Business Media.

[6] M. Lundstrom, 2009, *Fundamentals of Carrier Transport*, Cambridge: Cambridge University Press

[7] E. Lind, "High frequency III–V nanowire MOSFETs," *Semicond. Sci. Technol.*, vol. 31, no. 9, 093005, Sept. 2016

[8] A. Godoy, Z. Yang, U. Ravaioli and F. Gámiz, "Effects of nonparabolic bands in quantum wires," *J. Appl. Phys.*, vol. 98, no. 1, 013702, July 2005

[9] F. J. G. Ruiz, I. M. Tienda-Luna, A. Godoy, L. Donetti and F. Gámiz, "Equivalent Oxide Thickness of Trigate SOI MOSFETs With High-κ Insulators," *IEEE Trans. Electron Devices*, vol. 56, no. 11, pp. 2711-2719, Nov. 2009

[10] B. Yu, L. Wang, Y. Yuan, P. M. Asbeck and Y. Taur, "Scaling of Nanowire Transistors," *IEEE Trans. Electron Devices*, vol. 55, no. 11, pp. 2846-2858, Nov. 2008

[11] M. D. Ko, C. W. Sohn, C. K. Baek and Y. H. Jeong, "Study on a Scaling Length Model for Tapered Tri-Gate FinFET Based on 3-D Simulation and Analytical Analysis," *IEEE Trans. Electron Devices*, vol. 60, no. 9, pp. 2721-2727, Sept. 2013

[12] C. B. Zota, G. Roll, L.-E. Wernersson and E. Lind, "Radio-Frequency Characterization of Selectively Regrown InGaAs Lateral Nanowire MOSFETs," *IEEE Trans. Electron Devices*, vol. 61, no. 12, pp. 4078-4083, Dec. 2014