University of Alberta

Transport Performance Projection of Emerging Nanoscale Devices

by

Sabbir Ahmed

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Solid State Electronics

Department of Electrical and Computer Engineering

© Sabbir Ahmed
Spring 2014
Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis and, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

To

My lovely wife, Emi

and

My wonderful parents, Dr. Golam Mowla Choudhury and Fatema Choudhury

Abstract

Emerging devices, such as those based on carbon (in the form of graphene or nanotubes) or III-V compound semiconductors, are constructed on an atomic scale, where the transport is governed by the Schrödinger equation of quantum mechanics or the Boltzmann transport equation (BTE) of semi-classical mechanics.

The solutions of the Schrödinger equation and the BTE offer an opportunity not only to explore and understand the rich physics of small-scale devices, but also to predict their performance potential. The Schrödinger equation can generally be tackled by the method of nonequilibrium Green's functions (NEGF), and the BTE can be solved with the aid of commercial numerical software, such as COMSOL. In this doctoral work, we utilize these state-of-the-art transport approaches to study the performance of emerging nanoscale transistors, namely, III-V high-electron-mobility transistors (HEMTs) and carbon-nanotube transistors (CNFETs).

In the first stage of work, we use the NEGF approach to show how quantum-mechanical transport impacts the cutoff frequencies of III-V HEMTs as the gate length is shrunk. We demonstrate that the cutoff frequencies tend to saturate as the gate length is scaled down, *i.e.*, that they attain a maximum value that ceases to improve with further scaling, and we tie this behavior to the low effective mass of electrons in III-V materials, which is a transport property.

In the second stage of work, we examine the impact of electron scattering on the performance of CNFETs via the BTE. We show that the collisions of electrons with substrate phonons (arising from lattice vibrations within the substrate on which the CNFET resides) is critical to their performance, and we thereby identify the best and worst choices of substrate for optimum performance.

For future work, we propose the creation of a tool that captures the transport of electrons in quantum-dot solar cells. The tool would utilize NEGF to account for quantum-mechanical transport in the presence of light, and its aim would be to facilitate the systematic understanding of cell operation and hence optimal cell design.

Acknowledgements

It is a great pleasure to express my deepest gratitude to my thesis supervisor, Dr. Mani Vaidyanathan, for his continuous support, invaluable advice, and critical and constructive evaluation towards this thesis. He has an ever-positive approach for training and an uncompromising attitude for quality, both of which always inspired me towards the utmost professionalism. His kindness and considerateness helped me to overcome some of the most difficult hurdles during my study. Very importantly, he has always guided me to thrive both as an individual and as a team member.

I am also deeply indebted to Dr. Diego Kienle of Universität Bayreuth, Germany for providing insightful suggestions and devoting a lot of his precious time to the work. Dr. Kienle has this remarkable ability to simplify a complex problem into manageable segments, which steered me through some of the very difficult challenges during my research.

I would like to sincerely thank the members of my supervisory and Ph.D. exam committees, Dr. Sandipan Pramanik, Dr. Xihua Wang, Dr. Ray DeCorby, and Dr. Masum Hossain for their valuable comments, questions, and suggestions to improve the quality of the thesis.

I am highly appreciative to my past and present colleagues of the Nanoelectronic Research Laboratory, Dr. Navid Paydavosi, Ahsan Ul Alam, and

Kyle Holland, for the healthy discussions and all the precious times we have spent together at the University of Alberta.

My heartiest thanks go to my wonderful parents, Dr. Golam Mowla Choudhury and Fatema Chowdhury, for their love, encouragement, and support throughout my life.

These acknowledgements would not be complete without my loving thanks and earnest gratitude to my beloved wife, Emi. This work would not have been possible without her sacrifice and tremendous support. During the course of my Ph.D., I have been immersed into my studies, while my wife quietly took care of everything and supported me selflessly.

Last, but not the least, I would like to thank the F.S. Chia Ph.D. Scholarship, Queen Elizabeth II Graduate Scholarship, and the Natural Sciences and Engineering Research Council (NSERC) of Canada for the wonderful financial support during this work.

Table of Contents

1. Introduction 1

1.1 Overview..1

1.2 Stages of Work..3

2. Impact of Effective Mass on the Scaling Behavior of the f_T and $f_{\max}$ of III-V High-Electron-Mobility Transistors 20

2.1 Introduction..20

2.2 Intrinsic f_T 23

2.2.1 Simulation Approach 23

2.2.2 Results..26

2.3 Complete Device..53

2.3.1 Small-Signal Equivalent Circuit 54

2.3.2 Modeling of Parasitics 55

2.3.3 Results..59

2.4 Conclusions..62

3. Impact of Substrate Material on the RF Performance of Carbon-Nanotube Transistors 65

3.1 Introduction..65

3.2 Intrinsic Device..68

3.2.1 BTE-Poisson Approach 68

3.2.2 Results for the Single Tube...72

3.3 Extrinsic Device...84

3.3.1 Approach...84

3.3.2 Results and Discussion 87

3.4 Conclusions...95

4. Conclusions and Future Work 99

4.1 Summary of Conclusions...99

4.1.1 Stage I (Chapter 2)..99

4.1.2 Stage II (Chapter 3) 101

4.1.3 Stage III (Future Work) 102

4.2 Future Work: Quantum Mechanical Transport in Nanoscale Quantum-Dot Solar Cells..102

4.2.1 Introduction...102

4.2.2 InAs/GaAs Quantum-Dot Solar Cell 104

4.2.3 Results...114

4.2.4 Summary...116

5. Bibliography 118

6. Appendix 131

List of Tables

Table 2.1. Element values for the circuit of Figure 2.13 when $L_g = 30$ nm.......... 55

Table 2.2. Parasitic element values of the HEMTs (for $W = 1$ μm)....................... 59

Table 3.1. Parameters for the SPP scattering in carbon nanotubes on various substrates [79]... 76

Table 3.2. Values for the parasitic resistances and capacitances used in the lumped element model of Figure 3.8 for the different substrates. In all cases, the total width of the array-based CNFET is 1 micron….. …87

Table 3.3. ITRS RF CMOS millimeter-wave technology requirements for the year 2015 [7] versus the array-based CNFET with SPP scattering included from various polar substrates. The tube pitch is 100 nm.. 95

List of Figures

Figure 1.1. Device structure of a III-V HEMT 6

Figure 1.2. (a) Extrinsic and intrinsic f_T of the HEMTs considered in this work, and the reported f_T of InGaAs HEMTs [3, Figure 10], versus L_g. The lines have been drawn as guides for the eye. (b) Conduction-band profiles along with the positions of the subband edges at peak f_T along the depth of the channel (z -direction), for the InGaAs HEMTs with L_g= 10 and 50 nm 8

Figure 1.3. (a) Normalized value of the transconductance g_m and its components versus gate length L_g for the InGaAs HEMT. (b) Reciprocal of total gate capacitance $1/C_{\mathrm{gg}}$ versus gate length L_g for the HEMTs 11

Figure 1.4. (a). Schematic of an array-based CNFET structure. (b) Intrinsic single-tube block.. ..13

Figure 1.5. (a) Dc drain current I_D and (b) intrinsic f_T vs. gate voltage V_G for a single-tube CNFET.. ..15

Figure 1.6. An InAs/GaAs quantum-dot solar cell 18

Figure 2.1. III-V HEMT structure used in this work. 24

Figure 2.2. Intrinsic unity-current-gain frequency f_T vs. dc gate voltage V_G of an InGaAs HEMT having the structure shown in Figure 2.1. The insulator thickness is t_{ins} = 3 nm and results are shown for three different gate lengths L_g. The dc drain voltage V_D is held at 0.5 V... 27

Figure 2.3. (a) Transconductance g_m and gate capacitance C_{gg} vs. gate bias V_G. (b) Conduction-band profile in the channel, *i.e.*, $E_C(x, z_{\mathrm{ch}})$ vs. x, at different gate-bias voltages

for the InGaAs HEMT with L_g= 30 nm considered in Figure 2.2. Dashed lines are used in part (a) for gate voltages above the value causing barrier collapse. 30

Figure 2.4. Conduction-band profile, *i.e.*, $E_C(x, z_{\text{ch}})$ vs. x, of the InGaAs HEMTs considered in Figure 2.2 at V_G= 0.3V, illustrating the impact of drain-induced barrier lowering. 31

Figure 2.5. Intrinsic *peak* f_T vs. gate length L_g for the III-V HEMTs considered in this work. The lines have been drawn as guides for the eye.. 32

Figure 2.6. Input equivalent circuit of the HEMTs. (a) Overall input circuit, which is just the input capacitance C_{gg}. (b) Separation of C_{gg} into the series combination of insulator and inversion-layer capacitances, C_{ins} and C_{inv}. (c) Further subdivision of C_{inv} into contributions arising from each subband, where $C_{\text{inv}\,i}$ is the contribution from the ith subband. 35

Figure 2.7. Total spectral function $A_S[x, z_{\text{ch}}, E(k_x, k_z)] + A_D[x, z_{\text{ch}}, E(k_x, k_z)]$ displayed as intensity vs. $E(k_x, k_z)$ and x, along with the positions of the Fermi levels μ_S and μ_D, for the InGaAs HEMT with (a) L_g= 50 nm and (b) L_g= 10 nm, showing that at the shorter gate length, the subband edges at peak f_T are lower in position with respect to the source Fermi level. 40

Figure 2.8. (a) The difference in the Fermi-Dirac integrals $\left\{F_{-\frac{1}{2}}[\mu_S - E(k_x, k_z)] - F_{-\frac{1}{2}}[\mu_D - E(k_x, k_z)]\right\}$, along with the positions of the occupied subband edges at the top of the barrier, $(x, z) = (x_{\text{top}}, z_{\text{ch}})$. (b) Source Fermi-Dirac integral $F_{-\frac{1}{2}}[\mu_S - E(k_x, k_z)]$ and source spectral function $A_S\left[x_{\text{top}}, z_{\text{ch}}, E(k_x, k_z)\right]$ vs. in-plane energy $E(k_x, k_z)$ for the InGaAs HEMT with gate lengths L_g= 10 and 50 nm; the source Fermi level is $\mu_S \equiv 0$, and the conduction-band edge E_C is marked. 45

Figure 2.9. Capacitance ratio $C_{\text{ins}}/(C_{\text{ins}}+C_{\text{inv}})$, total difference in Fermi-Dirac integrals ("difference in agenda"), *i.e.*, the result of the summation in (2.8), and the transconductance g_m (at peak f_T) versus gate length L_g for the InGaAs HEMT. The actual values have been normalized with respect to the value of each component at the gate length L_g= 50 nm in order to illustrate the scaling behavior, and the lines have been drawn as guides for the eye. 46

Figure 2.10. (a) Conduction-band profiles at peak f_T along the depth of the channel, *i.e.,* $E_C(x_{\text{top}}, z)$ vs. z, for the InGaAs HEMTs with L_g= 10 and 50 nm; also shown are the positions of the occupied subband edges at $(x, z) = (x_{\text{top}}, z_{\text{ch}})$, illustrating that the subbands are lower in position with respect to the source Fermi level for the shorter gate-length device. (b) Charge modulation along the transport direction (*i.e.*, the x-direction) for the InGaAs HEMTs with L_g= 10 and 50 nm; the results show a relatively greater charge modulation (under the gate) in the first three subbands of the shorter gate-length device 50

Figure 2.11. (a) Reciprocal of inversion-layer capacitance $1/C_{\text{inv}}$ and (b) reciprocal of total gate capacitance $1/C_{\text{gg}}$ versus gate length L_g for the HEMTs. The lines have been drawn as guides for the eye, and to emphasize the saturating behavior at short gate lengths, additional lines are shown to illustrate the expected behavior based on linear scaling with L_g, extrapolated to lower gate lengths from $L_g \sim 35$ nm 51

Figure 2.12. Conduction-band profiles at peak f_T along the depth of the channel, *i.e.,* $E_C(x_{\text{top}}, z)$ vs. z, for the InGaAs and GaN HEMTs with L_g= 10 nm; the subbands for the InGaAs device are lower in position with respect to the source Fermi level........ 53

Figure 2.13. (a) Small-signal equivalent circuit of an intrinsic III-V HEMT from [39, Figure 9.5]. (b) Circuit elements, to be evaluated at the dc operating point; the symbols bear their usual meanings. 55

Figure 2.14. Full device structure of the HEMTs. 57

Figure 2.15. Complete device model of the III-V HEMTs. The intrinsic device is represented by the small-signal circuit of Figure 2.13. 58

Figure 2.16. Current gain $|h_{21}|$, unilateral power gain U, and extrapolated f_T and $f_{\max}$ of the InGaAs HEMT at L_g= 30 nm. 59

Figure 2.17. Extrinsic and intrinsic f_T of the HEMTs considered in this work, and the reported f_T of InGaAs HEMTs [3, Figure 10], versus L_g. The lines have been drawn as guides for the eye. 60

Figure 2.18. Extrinsic $f_{\max}$ of the HEMTs versus L_g. The lines have been drawn as guides for the eye.. 62

Figure 3.1. Base geometry of the intrinsic single-tube CNFET structure used for the simulation. 69

Figure 3.2. Static (dc) drain current I_D vs. gate bias voltage V_G for a single-tube block (shown in Figure 3.1) with a tube pitch of 10 nm in the presence of SPP scattering on different substrates. The dc drain voltage V_D is held at 0.9 V. The curve labeled "Ballistic" is with all scattering turned off and the curve labeled "Native Phonons" includes scattering only from native acoustic and optical phonons; in each of these cases, an SiO_2 substrate is assumed solely for the purpose of solving Poisson's equation (and with no contribution to the scattering). The remaining curves are for the indicated substrates with *all* scattering (native *and* SPP) turned on. 73

Figure 3.3. Intrinsic unity-current-gain cutoff frequency f_T vs. gate voltage V_G for the single-tube CNFET block shown in Figure 3.1 and for block widths of (a) 10 nm and (b) 100 nm. The labeling convention for the curves is the same as that described in the caption of Figure 3.2. 78

Figure 3.4. (a) Transconductance g_m and (b) gate capacitance C_{gg} vs. dc gate voltage V_G of a single-tube CNFET block in Figure 3.1 with a width of 10 nm. The labeling convention for the curves is the same as that described in the caption of Figure 3.1. 81

Figure 3.5. (a) Energy-resolved dc distribution function $\bar{f}(E)$ at $x = 80$ nm (see Figure 3.1) and for $V_G = 0.4$ V, with SiO_2 used as the substrate and shown for three cases: (i) ballistic transport; (ii) only native acoustic and optical phonon scattering; (iii) all scattering, *i.e.*, additionally including substrate phonons. The inset to the figure shows a sketch of the conduction-band profile $E_c(x)$ vs. x, which is provided for reference. (b) Energy-resolved change in the distribution function $\Delta\bar{f}(E)$ at $x = 80$ nm and $V_G = 0.4$ V, where $\Delta\bar{f}(E)$ arises from an incremental change in gate voltage ΔV_G (equal to 10 mV). 82

Figure 3.6. Magnitude of the transistor's forward y-parameters, y_{11} and y_{21}, comparing the ballistic limit with the case of SPP scattering for the two substrates AlN and HfO_2, where the SPP scattering is the weakest and strongest, respectively. 84

Figure 3.7. Schematic of a top-gated, aligned, array-based CNFET structure similar to [49]. The gate length and width of the device are $L_g = 20$ nm and $W_g = 1$ μm, respectively. The yellow box shows the portion of the structure that is studied. The figure is not drawn to scale. 85

Figure 3.8. Extrinsic lumped-element model of the array-based CNFET structure [49]. The inner block represents the intrinsic multi-tube device described by the y-parameter matrix $[y_{\mathrm{m,int}}]_p$ specified by (3.5). 86

Figure 3.9. Magnitude of the common-source current gain $|h_{21}|$, unilateral power gain U, maximum stable gain (MSG), maximum available gain (MAG), and the stability factor **k** as function of frequency for the CNFET of Figure 3.7 with a tube pitch $p = 10$ nm and with (a) AlN and (b) HfO_2 as substrates. The extrapolated f_T and $f_{\max}$ are also shown. 89

Figure 3.10. Extrapolated extrinsic f_T of the array-based CNFET structure as a function of tube pitch in the presence of native and SPP scattering for different substrates. 91

Figure 3.11. $f_{\max}$ of the array-based CNFET structure as function of tube pitch in the presence of native and SPP scattering for different substrates... 93

Figure 3.12. Effective gate-drain capacitance $C_{\text{gd,eff}}$ for the array-based CNFET structure as a function of tube pitch in the presence of native and SPP scattering from different substrates.. 94

Figure 4.1. An InAs/GaAs quantum-dot solar cell... 105

Figure 4.2. Conduction band profile of an InAs/GaAs quantum-dot solar cell structure at a reverse bias voltage of 0.5 V.. 114

Figure 4.3. Current-voltage characteristic curves of an InAs/GaAs quantum-dot solar-cell structure. The lines have been drawn as guides for the eye. 115

Chapter 1

Introduction

1.1 Overview

In the context of electronic devices, the word "transport" loosely refers to the movement of charge carriers (electrons and holes) under the influence of applied voltages and (possibly) under illumination. For much of the past 40 years, transport within the key devices of electronics --- such as the silicon-based MOSFET and the simple *pn*-junction solar cell --- have been successfully described by the *classical* **drift-diffusion equation (DDE)**, written for electrons in the familiar form

$$J_n = q\,\mu_n n\,\mathcal{E} + q\,D_n\,\frac{dn}{dx} \tag{1.1}$$

where J_n is the electron current density, q is the magnitude of the electronic charge, μ_n is the electron mobility, n is the electron concentration, $\mathcal{E}$ is the electric field, D_n is the electron diffusivity, and x is position.

Modern devices, including modern transistors and solar cells, are constructed on an *atomic scale*, and often employ new materials, such as carbon in the form of graphene or nanotubes, and III-V compound semiconductors, rather than silicon. The transport in such *nanoscale* devices is generally *quantum-mechanical* or *semi-*

classical, governed by the **Schrödinger equation** of quantum mechanics or the **Boltzmann transport equation (BTE)** of semi-classical mechanics.

The solutions of the Schrödinger equation and the BTE offer an opportunity not only to explore and understand the rich physics of small-scale devices, but also to predict their performance potential, thereby helping to identify the most promising candidates for future technologies. The Schrödinger equation can generally be tackled by the *method of nonequilibrium Green's functions* (NEGF), and the BTE can be solved with the aid of commercial numerical software, such as COMSOL [1]. The specific cases involving the NEGF and the BTE considered in this thesis, and hence the connections between quantum-mechanical and semi-classical carrier transport and device performance, include the following:

1. Transport in III-V high-electron-mobility transistors (HEMTs): III-V HEMTs are candidates for terahertz applications, which require the highest possible *cutoff frequencies* for the transistor. We use the NEGF approach to show how quantum-mechanical transport impacts the cutoff frequencies as the gate length is shrunk. We demonstrate that the cutoff frequencies tend to *saturate* as the gate length is scaled down, *i.e.*, that they attain a maximum value that ceases to improve with further scaling, and we tie this behavior to the low effective mass of electrons in III-V materials, which is a transport property.
2. Transport in carbon-nanotube transistors (CNFETs): CNFETs are candidates to augment or even replace conventional MOSFETs. We

describe the transport of electrons in CNFETs semi-classically, via the BTE, and we examine the impact of *electron scattering* on the performance of CNFETs. We show that the collisions of electrons with *substrate phonons* (arising from lattice vibrations within the substrate on which the CNFET resides) is critical to their performance, and we thereby identify the best and worst choices of substrate for optimum performance.

3. Proposed future work: Transport in quantum-dot solar cells: Quantum-dot solar cells are among the latest candidates being considered as potentially cheap, highly efficient, green-energy alternatives to silicon-based solar cells. To extend the ideas from the work already completed for this thesis, we propose the creation of a tool that captures the transport of electrons in quantum-dot solar cells. The tool would utilize NEGF to account for quantum-mechanical transport effects in the presence of light, *i.e.*, it would account for the so-called *electron-photon interaction* within the framework of quantum transport, and its aim would be to facilitate the systematic understanding of cell operation and hence optimal cell design.

1.2 Stages of Work

To accomplish the tasks outlined above, the doctoral work can be categorized into two completed stages and a proposed future stage. For each completed stage, a summary and a description of the key points are provided below. Please note that the intention of these summaries and descriptions is merely to convey *the essence*

and *not every aspect* of each completed stage of work; interested readers can find far more information and all the supporting details in the subsequent chapters of this thesis. The proposed future work is also discussed below, with further details available in the concluding chapter of this thesis.

1. **Impact of Effective Mass on the Scaling Behavior of the f_T and f_{max} of III-V High-Electron-Mobility Transistors**

   Summary:

   The first stage involved research on III-V high-electron-mobility transistors (HEMTs) which, primarily from offering very high mobility, are among the contenders for applications at terahertz frequencies. The scaling behavior of these III-V HEMTs has not been extensively studied or explained, either experimentally or by quantum-mechanical simulations, with conflicting data and comments on the expected outcome of scaling. In our work, we report on a tendency for III-V devices with low effective-mass channel materials to exhibit a saturation in their unity-current-gain and unity-power-gain cutoff frequencies (f_T and f_{max}) with a down scaling of gate length. We focus on InGaAs and GaN HEMTs and examine gate lengths from 50 nm down to 10 nm. A self-consistent, quantum-mechanical solver based on the method of nonequilibrium Green's functions (NEGF) is used to quasi-statically extract the f_T for intrinsic III-V devices. This intrinsic model is then combined with the series resistances of the heterostructure stack and the parasitic resistances and

capacitances of the metal contacts to develop a complete extrinsic model, and to extract the extrinsic f_T and $f_{\max}$.

It is shown that the f_T and $f_{\max}$ of III-V devices will saturate, *i.e.*, attain a maximum value that ceases to increase as the gate length is scaled down, and that the saturation is caused by the low effective mass of III-V materials. It is also shown that the InGaAs HEMTs have faster f_T at long gate lengths, but as a consequence of their lower effective mass, they experience a more rapid f_T saturation than the GaN HEMTs, such that the two devices have a comparable f_T at very short gate lengths (~10 nm). On the other hand, due to favorable parasitics, it is shown that the InGaAs HEMTs have a higher $f_{\max}$ at all the gate lengths considered in our work.

Of three peer reviewers of our published results [2], the first commented that the work was *"very informative"* and *"commendable in the approach as well as the depth of the material,"* the second stated it was *"very nice work,"* and the third stated *"the results should be of great interest to the community."*

Key Points:

In a recent review of RF transistors for terahertz applications, and based on experimental data reported in the literature, Schwierz *et al.* [3, Figure 10] observed a tendency for the f_T of III-V HEMTs to saturate at very short gate lengths, *i.e.*, to attain a maximum value that ceases to improve with further scaling. The primary outcome of this stage of the work is a demonstration, with careful explanation, that III-V HEMTs constructed using low effective-mass

channel materials can indeed experience a saturation in their f_T and $f_{\max}$ with gate-length scaling. The device structure used in this work is shown in Figure 1.1.

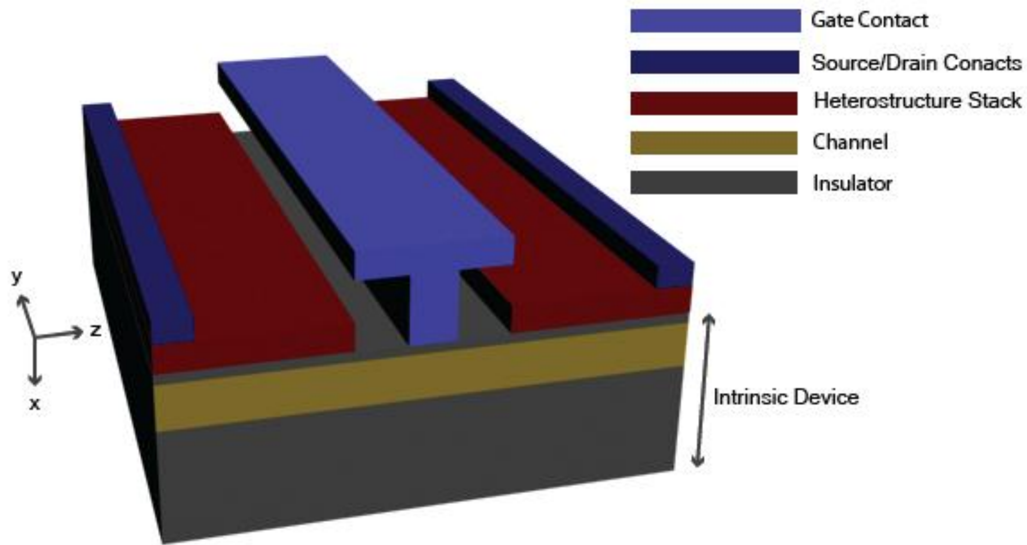


Figure 1.1. Device structure of a III-V HEMT.

We first shed insight into the signature saturation of the intrinsic peak f_T,[1] as displayed in Figure 1.2 (a), by analyzing the scaling behavior of the intrinsic transconductance g_m and the total intrinsic gate capacitance C_{gg}, since $f_T = g_m/(2\pi C_{\text{gg}})$. Towards this end, our work begins by establishing an important result: as L_g is scaled down, drain-induced barrier lowering (DIBL) causes the peak f_T to occur at successively lower values of the gate bias voltage V_G. As a consequence, the quantum well defining the channel (in the z-direction) at peak

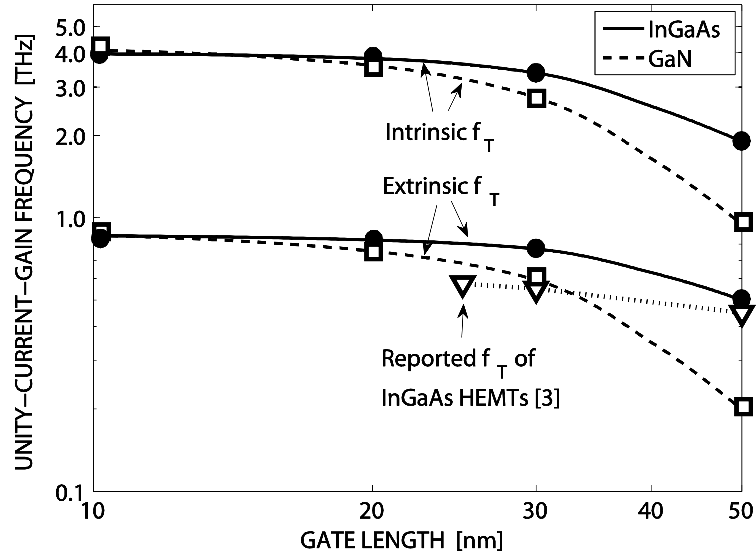
[1] By "peak f_T," we mean the highest f_T that can be attained at each L_G as the gate bias V_G is varied.

f_T is *less sharp i.e.*, there is reduced quantum confinement, and the channel subbands are hence lower in position with respect to the source Fermi level μ_S at shorter gate lengths; this scenario is depicted in Figure 1.2(b).

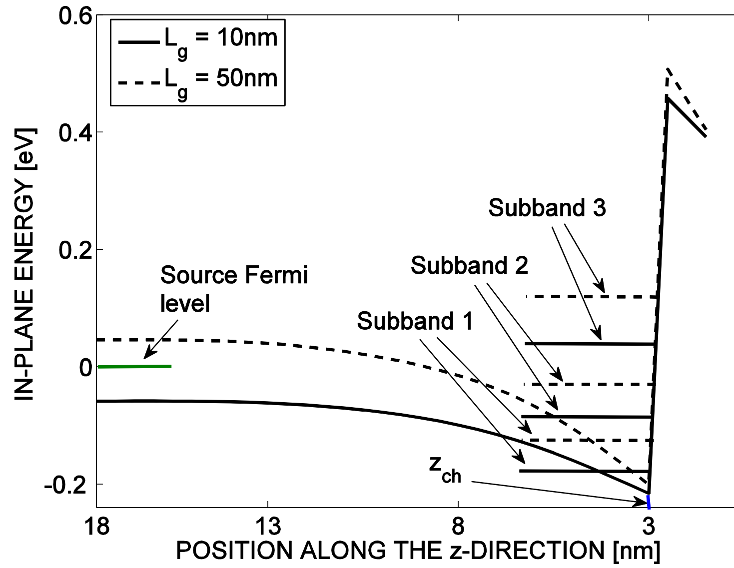
To understand the scaling behavior of g_m, we then appeal to the expression (2.8) for g_m, repeated below for convenience:

$$g_m = \frac{q^2}{\hbar^2}\sqrt{\frac{m^* k_B T}{2\pi^3}}\frac{C_{\text{ins}}}{C_{\text{ins}} + C_{\text{inv}}}\sum_i \left\{F_{-\frac{1}{2}}[\mu_S - E_{C\,\text{top}} - \Delta_i] - F_{-\frac{1}{2}}[\mu_D - E_{C\,\text{top}} - \Delta_i]\right\}. \quad (1.2)$$

The details of this expression are unimportant for the present discussion; the reader need only note that the scaling behaviors of two factors impact the scaling behavior of g_m. (i) The capacitance ratio $C_{\text{ins}}/(C_{\text{ins}} + C_{\text{inv}})$, with C_{ins}



(a)



(b)

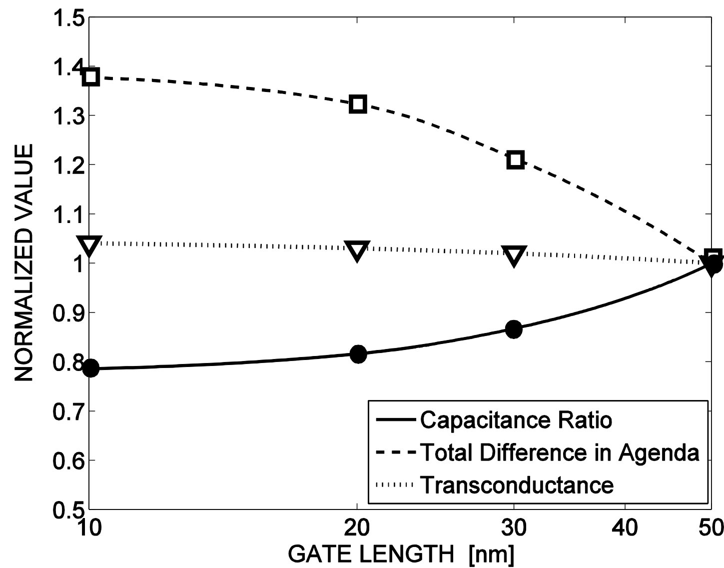
Figure 1.2. (a) Extrinsic and intrinsic f_T of the HEMTs considered in this work, and the reported f_T of InGaAs HEMTs [3, Figure 10], versus L_g. The lines have been drawn as guides for the eye. (b) Conduction-band profiles along with the positions of the subband edges at peak f_T along the depth of the channel (z -direction), for the InGaAs HEMTs with L_g= 10 and 50 nm.

and C_{inv} being the gate-insulator and inversion-layer capacitance, respectively; physically, this factor describes the ability of the gate to modulate the source-to-drain channel barrier. (ii) The difference in Fermi-Dirac integrals (within curly brackets) evaluated at each subband edge i, which describes the "difference in agenda" [4, Ch. 1] of the source and drain contacts for each subband. The downward shift in the positions of the subbands with down scaling causes an enhancement to the difference in the Fermi-Dirac integrals (and hence to the "difference in agenda") for each subband, and this will tend to *increase* g_m. On the other hand, the same shift in the subband edges causes the capacitance ratio $C_{\text{ins}}/(C_{\text{ins}} + C_{\text{inv}})$ to decrease, which will tend to *decrease* g_m. The decrease in the capacitance ratio occurs because of the relative behavior of C_{ins} and C_{inv}. C_{ins} is primarily determined by the insulator thickness, which is fixed in this work, causing C_{ins} to scale linearly with gate length. On the other hand, the downward shift in the subband edges causes C_{inv} to be larger at each gate length than would be expected from a purely linear dependence on L_g, which we show occurs because of a greater-than-expected charge modulation in the shorter gate-length devices. Hence, the factors governing g_m act in opposition to each other, such that the corresponding g_m remains relatively insensitive to scaling, as depicted in Figure 1.3(a).

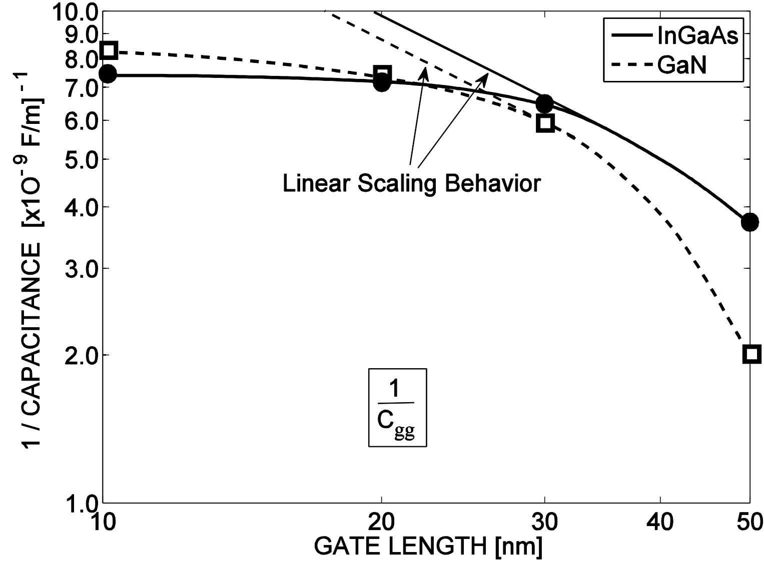
Regarding the gate capacitance C_{gg}, the low effective mass implies $C_{\text{ins}} \gg C_{\text{inv}}$, and hence that $C_{\text{gg}} = \frac{C_{\text{ins}} C_{\text{inv}}}{C_{\text{ins}} + C_{\text{inv}}}$ is roughly equal to the inversion

capacitance C_{inv}. As a result, C_{gg} scales according to C_{inv}; as previously mentioned, the downward shift in the subband edges depicted in Figure 1.2(b) causes C_{inv} to be larger at each gate length than would be expected from a purely linear dependence on L_g, *i.e.*, the scaling to occur at a slower rate than would be expected from a purely linear dependence on L_g. The resulting trend in $1/C_{\text{gg}}$ is plotted in Figure 1.3(b), where the saturating behavior of $1/C_{\text{gg}}$ at short gate lengths is of particular importance.

Since g_m remains relatively insensitive to scaling, and since $1/C_{\text{gg}}$ scales more slowly than predicted by a linear dependence on the gate length and saturates, then the intrinsic $f_T = g_m/(2\pi C_{\text{gg}})$ also scales more slowly than predicted by a linear dependence on the gate length and also saturates, consistent with the experimental data reported by Schwierz *et al.* [3, Figure 10].



(a)



(b)

Figure 1.3. (a) Normalized value of the transconductance g_m and its components versus gate length L_g for the InGaAs HEMT. (b) Reciprocal of total gate capacitance $1/C_{\text{gg}}$ versus gate length L_g for the HEMTs.

2. Impact of Substrate Material on the RF Performance of Carbon-Nanotube Transistors

Summary:

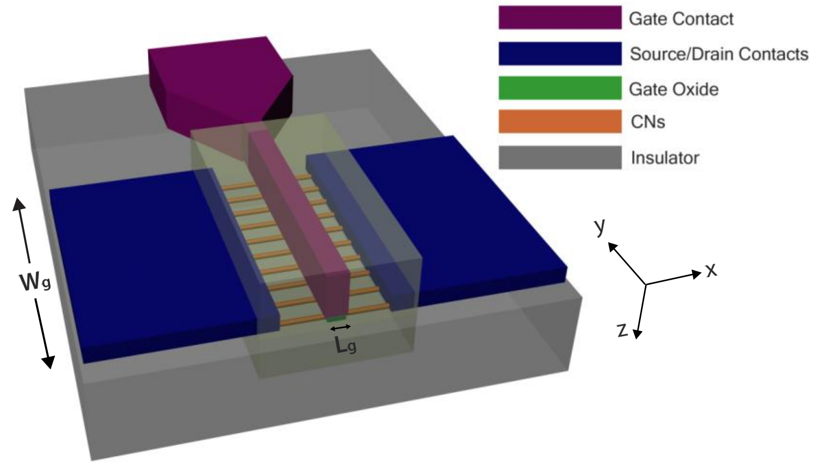
Carbon-nanotube, field-effect transistors are being widely considered for future high-performance radio-frequency (RF) electronic devices [3]. In the second stage, we examine the effect of the substrate material on the RF behavior of these devices by considering the impact of *substrate polar phonons*. We consider substrate polar phonon (SPP) scattering from AlN, SiO_2, HfO_2, and ZrO_2 substrates within a semi-classical approach by solving the time-dependent Boltzmann transport equation self-consistently with the Poisson equation. A

semi-classical approach is employed because there is no tunneling between the source and drain terminals in these devices, and because scattering can then be modeled without much complexity. Various RF figures of merit, such as the unity-current-gain frequency f_T, the unity-power-gain frequency $f_{\max}$, the transconductance g_m, and the two-port y-parameters, are determined in order to characterize the impact of SPP scattering. We first consider the impact of SPP scattering on the RF behavior of an intrinsic single-tube CNFET. These single-tube results are then combined with the external parasitic elements to analyze the pitch-dependent, RF behavior of an extrinsic array-based CNFET. It is shown that AlN substrates have the least impact in degrading the RF performance of a CNFET, while the more polar substrates (HfO_2 or ZrO_2) have a greater impact. This varying behavior can be attributed to the SPP energies, which are higher in AlN compared to the other materials, making CNFETs with AlN substrates less susceptible to SPP scattering even at room temperature. Our results suggest that substrate engineering will become an important component in the design process of emerging devices to achieve an optimized RF performance. The results of this stage have been published in the *IEEE Transactions on Nanotechnology* [5].

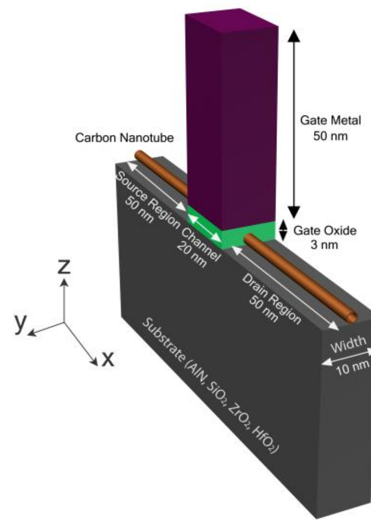
Key Points:

We investigate the impact of substrate polar phonon scattering on the RF characteristics of array-based, carbon-nanotube transistors at varying tube pitches; a schematic of an array-based CNFET is shown in Figure 1.4(a). We

demonstrate and explain that SPP scattering substantially affects the device's RF performance and, most importantly, displays a strong dependence on the



(a)

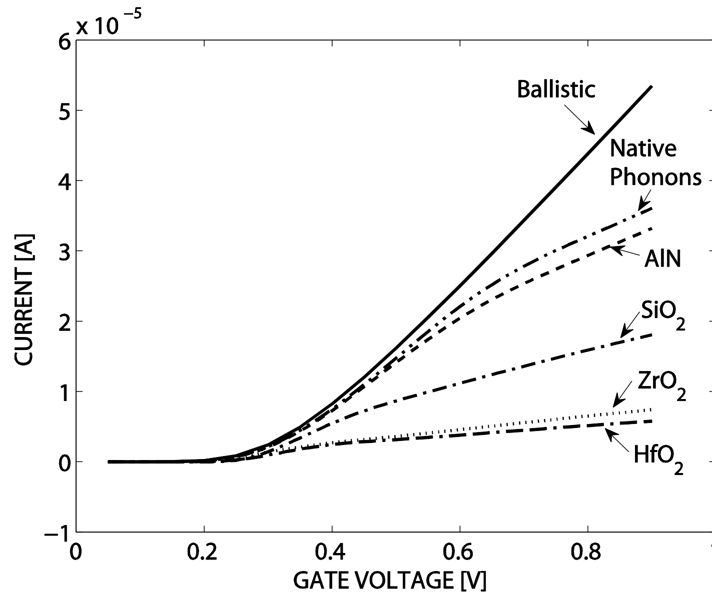


(b)

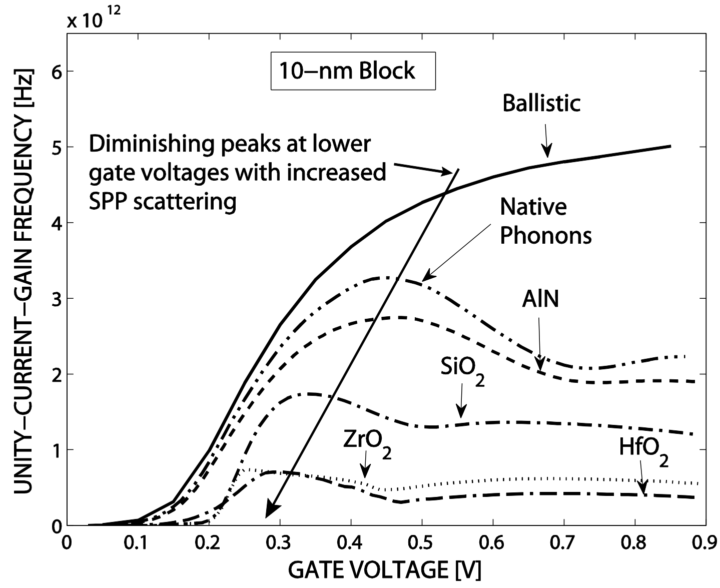
Figure 1.4(a). Schematic of an array-based CNFET structure. (b) Intrinsic single-tube block.

choice of substrate. This result is in accord with the qualitative expectation of Perebeinos *et al*. [6], who commented on the possible impact of different polar substrates on the dc mobility.

The impact of SPP scattering can best be understood by considering the RF behavior of an intrinsic single-tube CNFET block having a width of 10 nm, illustrated in Figure 1.4(b). It can be discerned from Figure 1.5 --- which plots the dc drain current I_D and the intrinsic f_T as a function of the gate bias V_G for such a CNFET block with different substrates --- that SPP scattering causes a significant reduction of the transistor's performance compared to a ballistic device, and that the magnitude of this reduction depends on the substrate.



(a)



(b)

Figure 1.5. (a) Drain current I_D and (b) intrinsic f_T vs. gate voltage V_G for a single-tube CNFET.

The strong variation in the device performance with a change in substrate can be explained based on the substrate phonon modes in two parts. (i) For the first phonon mode, the relative strengths of the SPP scattering can be discerned by the prefactor of the SPP collision integral specified in (A.4) of the Appendix. Substrate materials with low phonon energies will lead to a strong collision-integral prefactor through its two components, the strength of the polarization field and the phonon number, where both exhibit a strong inverse dependence on the phonon energy. Thus, the low-energy SPP phonons in ZrO_2 and HfO_2, and to a lesser extent SiO_2, cause increased SPP scattering and hence a more severe degradation in performance. (ii) For the second phonon mode, it is required to consider more than the prefactor in determining the relative

strengths of the scattering. The higher phonon energies of the second mode suggest that the corresponding SPP scattering will become important only at higher gate biases, when the channel band profile has been "pushed down" sufficiently below the source Fermi level for carriers injected from the source to have available states to occupy in the channel and in the channel-to-drain barrier region upon emitting the higher energy SPP phonons; mathematically, the requirement for available states is taken into account by the presence of the density of states and distribution function in the collision integral (A.4). In this case, ZrO_2 and HfO_2 can thus be expected to suffer from more severe SPP scattering that becomes important at a lower gate voltage, since the phonon energy is significantly lower and hence requires less gate voltage to push the bands down a sufficient amount for a phonon emission. The curves in Figure 1.5 are indeed consistent with this expectation, with AlN leading to the weakest impact of SPP scattering on the performance degradation, *i.e.*. with AlN leading to higher I_D and f_T values as a function of applied gate voltage V_G than for HfO_2 and ZrO_2.

Despite the presence of SPP scattering, and except for the transconductance g_m, we ultimately show [Table 3.3 in Chapter 3] that array-based CNFETs built on AlN and SiO_2 substrates continue to offer performance that meets and exceeds the ITRS specifications for RF CMOS [7].

3. Future Work: Quantum-Mechanical Transport in Nanoscale Quantum-Dot Solar Cells

<u>Summary:</u>

For future work, we propose the extension of our research on the impact of transport on the performance of emerging devices by focusing on the field of renewable energy sources, *i.e.*, solar cells. Solar cells are devices that convert energy from the sun into electricity.

Among the latest contenders in solar cells are those utilizing quantum dots, in which an array of nanoparticles ("dots") can cheaply be fabricated and engineered for an optimum cost-efficiency tradeoff. Critical to the success of quantum-dot, solar-cell technology is the ability to systematically investigate, design, and engineer the parameters of each cell. In turn, this requires a computer model to gain a proper understanding of the physics of cell operation, which is fundamentally quantum-mechanical. However, to date, no computer-based, quantum-mechanical tool for cell design exists that can comment on the impact of useful design parameters, such as the size of the quantum dots, the types of materials used to make the dots, the distance between the dots, and irregularities in the array. In this stage of work, we propose the development of a quantum-mechanical tool that has the potential to shed useful insight into the device physics and to facilitate systematic cell design, thereby providing a means to speed the development of quantum-dot, solar-cell technology.

Key Points:

The major advantage of quantum dots is that they have bandgaps that are tunable across a wide range by changing the size and types of dots [8]. Such tunability facilitates the ability to fabricate solar cells consisting of a number of different bandgap materials, each of which can be optimized to convert solar energy within a portion of the sun's spectrum, thus maximizing overall efficiency. Among the emerging quantum-dot solar cells are devices characterized by the existence of InAs quantum dots in a GaAs channel region [9, 10], and we assume this type of structure as a starting point for our work. A sample InAs/GaAs quantum-dot solar cell structure is shown in Figure 1.6.

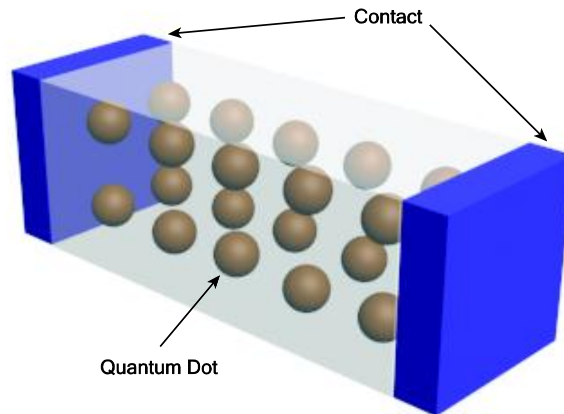


Figure 1.6. An InAs/GaAs quantum-dot solar cell.

A 1D NEGF-Poisson solver, which is capable of describing *both* the optical *and* transport properties, including quantum effects, is proposed to model the InAs/GaAs quantum-dot solar cell. Results from the solver can be utilized to extract and explain the important solar-cell figures of merit, such as

the short-circuit current density J_{sc}, open-circuit voltage V_{oc}, fill factor FF, and efficiency η. The developed solver can also be extended to quantum-dot solar cells fabricated with other materials.

Chapter 2

Impact of Effective Mass on the Scaling Behavior of the f_T and $f_{\max}$ of III-V High-Electron-Mobility Transistors[2]

2.1 Introduction

High-electron-mobility transistors (HEMTs) fabricated with III-V materials are among the leading candidates for terahertz radio-frequency (RF) applications [11]. Of principle interest in assessing the ultimate potential of these devices is the behavior of their unity-current-gain and unity-power-gain cutoff frequencies (f_T and $f_{\max}$) as the device gate length is scaled down. To date, the scaling behavior has not been extensively studied or explained, with conflicting data and comments on the expected behavior of the f_T and $f_{\max}$ with scaling, and on the relative performance of the different types of HEMTs.

Regarding the gate-length scaling, in a recent review of RF transistors for terahertz applications, and based on experimental data reported in the literature, Schwierz *et al.* [3, Figure 10] observed a tendency for the f_T of III-V HEMTs to saturate at very short gate lengths, *i.e.*, to show no further increase with decreasing gate length once the gate length is sufficiently small; they also commented on the

[2] A version of this chapter has been published [2].

importance of considering the impact of a low density of states near the bottom of the conduction band in assessing the potential of III-V transistors, as originally discussed in [12, 13] and as recently discussed in [14, 15], but they did not connect the low density of states to the gate-length scaling behavior. On the other hand, in contrast to the observation of Schwierz *et al.*, the simulations of Ayubi-Moak *et al*. [16] and Akis *et al*. [17] displayed no saturation in f_T versus gate length; their work was based on a semi-classical Monte Carlo approach that included electron scattering and which scaled InGaAs HEMTs down to 10 nm.

Regarding the relative performance of different III-V HEMTs, according to the 2009 ITRS [7], InGaAs and GaN devices are among the most important for RF and analog/mixed-signal technology. The InGaAs HEMTs have very high f_T and $f_{\max}$ [11, 18-20], an outcome of the very high electron mobility in the channel [11, 18, 19]. The GaN HEMTs have emerged as interesting candidates, because they offer not only high f_T and $f_{\max}$ [21-23] but also the ability to operate at high voltages and high powers, owing to the large bandgap and breakdown field in nitride-based materials [24, 25]. However, it is not clear how these HEMTs will perform in comparison with each other as the gate length is scaled down. Some researchers are confident that the InGaAs HEMTs are faster [11, 18, 19, 23] whereas the 2009 ITRS requires the GaN HEMTs to be as fast [7].

In this work, we examine the gate-length scaling behavior of the f_T and $f_{\max}$ of III-V HEMTs, with the primary aim of clarifying the connection between the effective mass (and hence density of states) of the channel material and the scaling

behavior. We focus on InGaAs and GaN HEMTs, and we use a fully quantum-mechanical approach based on the method of nonequilibrium Green's functions (NEGF). The primary outcome of our work is a demonstration that devices constructed using low effective-mass channel materials can indeed be expected to experience a saturation in their f_T and $f_{\max}$ with gate-length scaling, as observed by Schwierz *et al.* [3, Figure 10], along with a careful explanation of why the saturation occurs. Hence, we show that there is an inherent tradeoff in III-V devices: a low effective mass yields high electron velocity and mobility, but also leads to diminishing speed improvements with scaling at small gate lengths.

To date, the simulation work carried out for predicting the f_T and $f_{\max}$ in III-V HEMTs has utilized a semi-classical framework [16, 17, 23]. While semi-classical approaches are amenable to the inclusion of electron scattering, they require the use of an effective potential profile to model the quantum well defining the channel [26, 27]. However, as will be shown in this work, a proper prediction of the scaling behavior requires a precise modeling of the quantum well and the associated subbands, which can only be accomplished by a fully quantum-mechanical approach; electron scattering is less important at short gate lengths (< 50 nm), as shown in [28] by the good agreement between ballistic simulations and experiment, such that a ballistic quantum-mechanical approach is sufficient for a first-order study.

This chapter is organized as follows. Section 2.2 briefly explains the NEGF simulation approach and then presents results for the intrinsic f_T of III-V HEMTs.

The bias dependence of the intrinsic f_T at various gate lengths is first examined, and these results are then used to help explain the gate-length scaling behavior of the *peak* f_T by investigating the scaling behavior of the corresponding transconductance g_m and the gate capacitance C_{gg}, where $f_T = g_m/(2\pi C_{\mathrm{gg}})$. Short gate lengths of 10, 20, 30, and 50 nm are considered and the insulator thickness is fixed at 3 nm. The insulator is not scaled with the gate length since it was observed in [28] that scaling the insulator from 3.8 to 3 nm has a negligible effect on device performance. Section 2.3 then discusses the construction of a complete device model by way of an equivalent circuit that adds external parasitics to the intrinsic model; the parasitics include the series resistances arising from the heterostructure stack and metal contacts, and the external pad capacitances, which are computed from COMSOL [1] using an open-device model. The complete circuit is used to extract the extrinsic f_T and $f_{\max}$ and to examine their behavior versus gate-length scaling. Section 2.4 summarizes the conclusions.

2.2 Intrinsic f_T

2.2.1 Simulation Approach

To extract the intrinsic f_T of III-V HEMTs, we employ a two-dimensional NEGF-Poisson solver and use the same simulation structure as in [28], shown in Figure 2.1. The HEMT has a channel-layer thickness of 15 nm along the z-direction, and the upper and lower insulators (wide-bandgap layers) have thicknesses of 3 and 500

nm, respectively. As in [18, 21, 22, 28, 29], InAlAs is the insulator for the InGaAs HEMTs, while AlGaN is used for the GaN HEMT. The doping densities in the δ-doped layer, which electrostatically dope the underlying n^+ and n^{++} regions, are consistent with those in [28], equal to 2×10^{12}cm^{-2} and 1×10^{13}cm^{-2}, respectively. The value for the n^{++} regions is higher than that realized in practice, but it aids in numerical stability and otherwise does not affect the device [28].

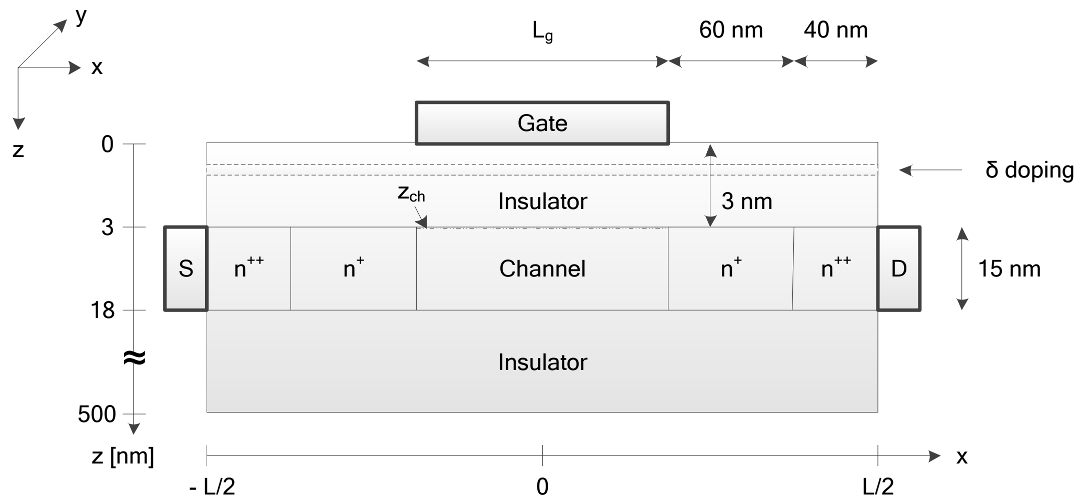


Figure 2.1. III-V HEMT structure used in this work.

The two-dimensional ballistic quantum transport in the channel, n^+, n^{++}, and upper insulator regions is described by the NEGF approach within the effective-mass approximation. The width W is assumed to be sufficiently large for the potential to be translationally invariant along the y-direction in Figure 2.1. The effective-mass Hamiltonian of the device $h(x, z)$ is discretized in the two-dimensional (x, z) space, as outlined in [30], and the quantum open boundary conditions for transport at the source and drain terminals are modeled through the

self-energies Σ_S and Σ_D, respectively [31]. The retarded Green's function [30] can then be written as $G[E(k_x,k_z)] = [E(k_x,k_z)I - h(x,z) - \Sigma_S - \Sigma_D]^{-1}$, where $E(k_x,k_z) \equiv E - E_{k_y}$ is the in-plane energy, E_{k_y} is the plane-wave energy along the width direction y, and E is the total electron energy. The Green's function $G[E(k_x,k_z)]$ can be used to evaluate the electron density $n(x,z)$ and ballistic current following the usual approach [30-33].

The electrostatic potential $V(x,z)$ is obtained by solving the two-dimensional Poisson equation $\nabla \cdot [\epsilon_r(x,z)\nabla V(x,z)] = -\rho_V(x,z)/\epsilon_0$, where ϵ_0 is the permittivity of free space, $\epsilon_r(x,z)$ is the relative permittivity, and $\rho_V(x,z) = q[N_D(x,z) - n(x,z)]$ is the volume charge density, with $N_D(x,z)$ being the doping density and q being the magnitude of the electronic charge. Holes are presumed to be negligible in comparison to $N_D(x,z)$ and $n(x,z)$ and are hence ignored. The potential $V(x,z)$ is taken to be the vacuum potential $V(x,z) = -E_{\mathrm{vac}}(x,z)/q$, and the conduction-band edge is defined by $E_C(x,z) = E_{\mathrm{vac}}(x,z) - q\chi_{\mathrm{aff}}$, where χ_{aff} is the electron affinity. Dirichlet boundary conditions are used at the gate, whereas homogeneous Neumann boundary conditions are applied at all other boundary points.

The NEGF-Poisson system was solved self-consistently using well-established methods [30-33], and the solution was validated by comparing the current-voltage characteristics of an InGaAs HEMT having a gate length L_g= 60 nm with the results of [28].

2.2.2 Results

Results for the intrinsic f_T are discussed below. *For the convenience of the reader, where appropriate, we have summarized the key results from the detailed discussions; the reader may find the italicized statements near the ends of subsections A.1, D.5, E, and F to be particularly useful.*

A Bias Dependence of Intrinsic f_T

The intrinsic f_T can be obtained from the NEGF-Poisson results using the well-known expression

$$f_T = \frac{1}{2\pi}\frac{dI_D}{dQ_G} \tag{2.1}$$

where dI_D and dQ_G are the changes in the current and the magnitude of gate-electrode (or channel) charge, respectively, that result from a small change dV_G in gate voltage while the drain voltage is held fixed and the source is taken as the reference.

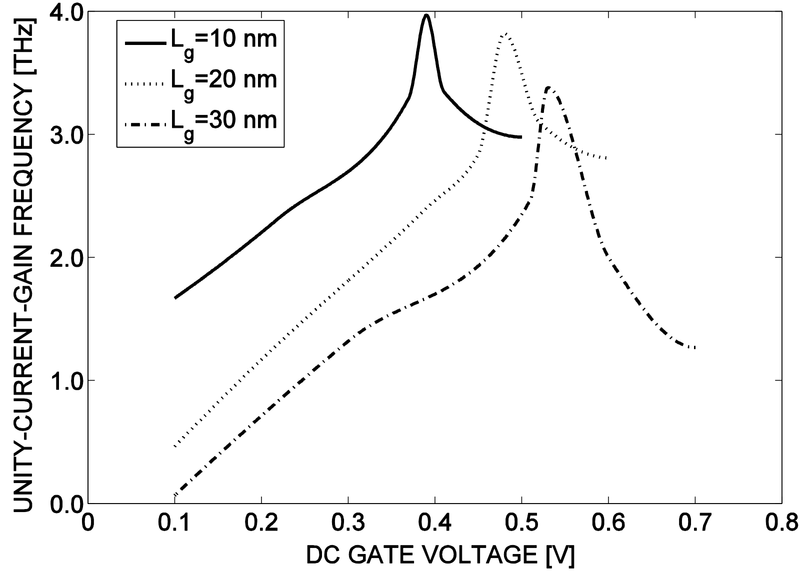


Figure 2.2. Intrinsic unity-current-gain frequency f_T vs. dc gate voltage V_G of an InGaAs HEMT having the structure shown in Figure 2.1. The insulator thickness is t_{ins} = 3 nm and results are shown for three different gate lengths L_g. The dc drain voltage V_D is held at 0.5 V.

Figure 2.2 shows the intrinsic f_T as a function of dc gate voltage V_G for an InGaAs HEMT at three different gate lengths equal to 10, 20, and 30 nm, and with the insulator thickness fixed at t_{ins} = 3 nm. The drain bias V_D is held constant at 0.5 V, which is a typical bias voltage for HEMTs [34], [35, Ch. 3], [36, 37]; HEMTs are required to operate at voltages substantially below 1 V in order to compete with Si CMOS technology [38] and to reduce the active power dissipation of the device [35, Ch. 10].

Two important features of the bias dependence of the intrinsic f_T can be discerned from Figure 2.2. First, for a fixed gate length L_g, *i.e.*, for a given curve in Figure 2.2, the f_T shows a significant variation with V_G, including a well-defined

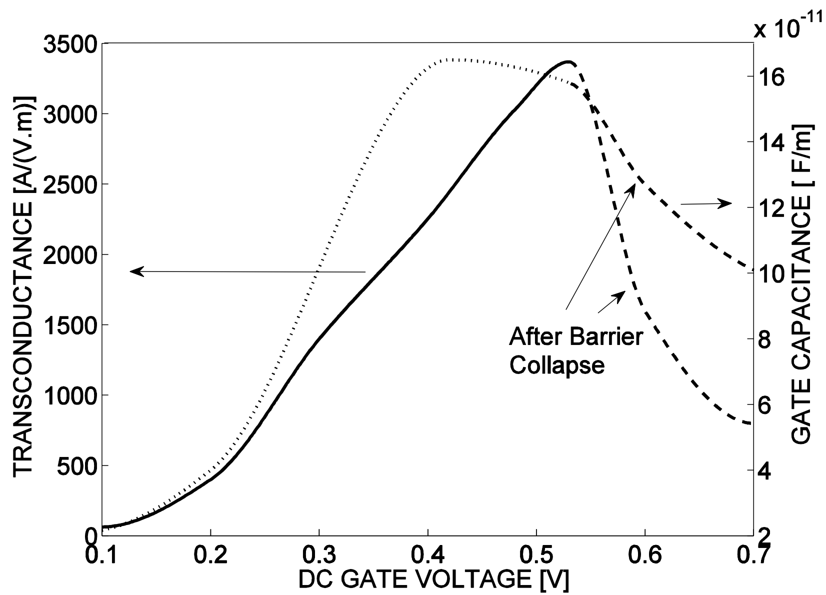
peak. Second, a comparison of the three curves in Figure 2.2 reveals that the gate bias V_G at which the f_T peaks depends on the gate length L_g. We will now discuss each of these features in turn, and later refer back to the discussion (in subsections B - E below) to help explain the scaling behavior of the *peak* f_T.

A.1 Variation of f_T with V_G

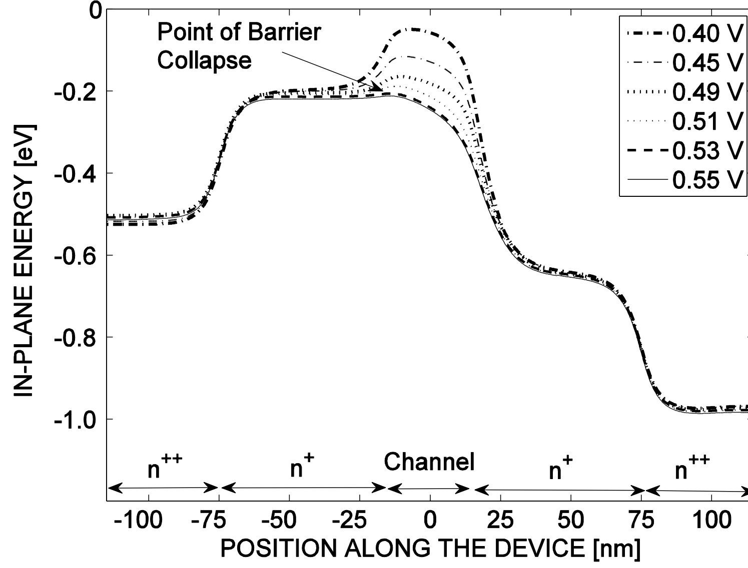
The variation of f_T with V_G (at a fixed L_g, *i.e.*, for a given curve in Figure 2.2) can be understood by first writing $f_T = g_m/(2\pi C_{\text{gg}})$, where $g_m = dI_D/dV_G$ is the transconductance and $C_{\text{gg}} = dQ_G/dV_G$ is the total intrinsic gate capacitance. Figure 2.3(a) plots g_m and C_{gg} for the L_g= 30 nm case from Figure 2.2. As shown, after reaching a maximum, both C_{gg} and g_m decrease with increasing V_G, but g_m degrades more rapidly, such that g_m controls the peaking in f_T. The trends in g_m and C_{gg} can be explained by analyzing the effect of gate voltage on the conduction-band edge in the channel, as depicted in Figure 2.3(b) for the L_g= 30 nm case.

Figure 2.3(b) shows the simulated conduction-band profile in the channel at different gate-bias voltages; here and elsewhere, the term “conduction-band profile in the channel” refers to $E_C(x, z_{\text{ch}})$ vs. x, where z_{ch} is a depth just below the insulator-channel interface, near the tip of the quantum well defining the channel, as marked in Figure 2.1 [and Figs. 2.10(a) and 2.12 further below]. Initially, changes in the gate bias V_G are effective in pushing down the barrier at the n^+-

channel junction, and correspondingly, incremental changes dV_G in gate voltage are effective in introducing more electrons into the channel. However, once the gate bias has pushed the conduction-band edge to the point of "barrier collapse," *i.e.*, to the point where the band edge in the channel reaches the same level as that in the n^+region near the source, as shown by the dashed curve ($V_G = 0.53$ V) in Figure 2.3(b), an incremental change in gate voltage dV_G can only weakly modulate the n^{++} - n^+ junction [28]; the incremental change dV_G thus loses the ability to introduce new electrons into the channel. The resulting increments in channel charge dQ_G and channel current dI_D arising from dV_G are hence diminished, leading to values of $g_m = dI_D/\ dV_G$ and $C_{\text{gg}} =\ dQ_G/\ dV_G$ that decrease beyond the point of barrier collapse.



(a)



(b)

Figure 2.3. (a) Transconductance g_m and gate capacitance C_{gg} vs. gate bias V_G. (b) Conduction-band profile in the channel, *i.e.*, $E_C(x, z_{\mathrm{ch}})$ vs. x, at different gate-bias voltages for the InGaAs HEMT with L_g= 30 nm considered in Figure 2.2. Dashed lines are used in part (a) for gate voltages above the value causing barrier collapse.

Overall, the results in Figs. 2.3(a) and 2.3(b) establish that, for a fixed gate length L_g, *peak f_T occurs at the gate bias corresponding to the onset of barrier collapse.*

A.2 Variation in V_G for Peak f_T

The variation in the gate bias V_G at which the f_T peaks (as L_g changes, *i.e.*, between curves in Figure 2.2) can be understood to be a result of drain-induced barrier lowering (DIBL). As the gate length is scaled down, it is well-known that the effect of the drain potential on the barrier gets stronger, acting as an additional source of

barrier lowering [34], [35, Ch. 10], [39, Ch. 6]. Thus, for the same gate bias V_G, the devices with shorter gate lengths will have lower barriers, as shown by the simulation results in Figure 2.4. Therefore, the scenario of barrier collapse leading to peak f_T is achieved with smaller gate bias voltages at shorter gate lengths, which explains why the locations of the peaks in Figure 2.2 shift to the left as L_g is reduced.

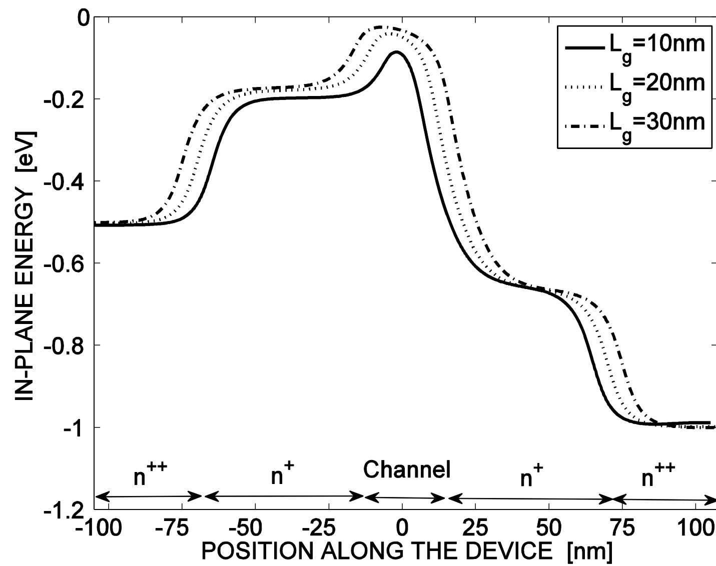


Figure 2.4. Conduction-band profile, *i.e.*, $E_C(x, z_{\text{ch}})$ vs. x, of the InGaAs HEMTs considered in Figure 2.2 at V_G= 0.3V, illustrating the impact of drain-induced barrier lowering.

B Gate-Length Scaling of Intrinsic f_T

To study the scaling behavior, the intrinsic *peak* f_T is plotted as a function of the gate length in Figure 2.5. It is evident that the peak f_T of III-V HEMTs shows a signature saturation as the gate length L_g is scaled down, an outcome that can also be discerned from the experimental results collected by Schwierz *et al.* [3, Figure

10]. The saturation can be explained by analyzing the scaling behavior of the intrinsic g_m and C_{gg} [since $f_T = g_m/(2\pi C_{\mathrm{gg}})$], where the relevant values of g_m and C_{gg} are those at the gate bias corresponding to the onset of barrier collapse, *i.e.*, those leading to the peak f_T at each L_g, as discussed in the previous subsection. In what follows, we first establish the capacitive input equivalent circuit seen looking into the gate, and then examine the scaling behavior of g_m and C_{gg}, referring to the circuit as an aid when appropriate; the observations are then used to explain the relative scaling behavior of InGaAs and GaN HEMTs, based on the difference in the effective mass of these materials.

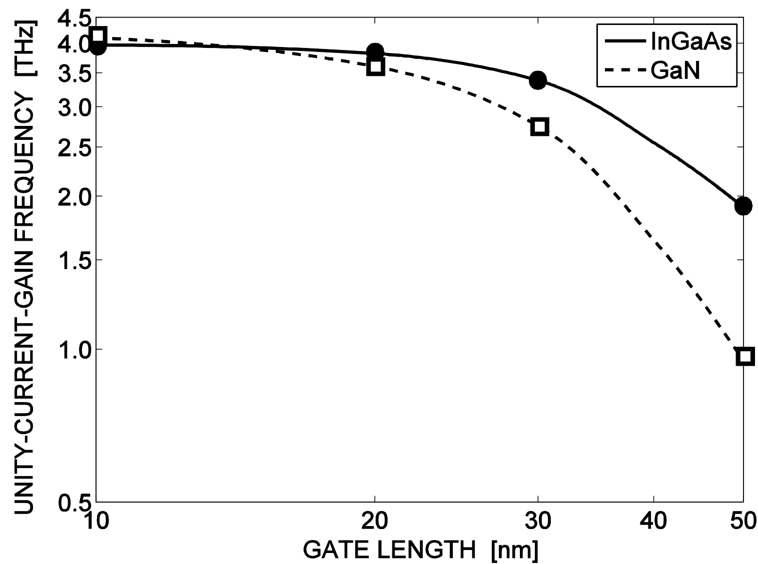


Figure 2.5. Intrinsic *peak* f_T vs. gate length L_g for the III-V HEMTs considered in this work. The lines have been drawn as guides for the eye.

C Input Equivalent Circuit

Consider first the input equivalent circuit seen from the gate under the conditions

of a perturbation in gate voltage dV_G, as sketched in the different parts of Figure 2.6, where the top terminal is the gate and the bottom terminal is the shorted source and drain combination, taken here as the reference, since both terminals are at ac ground under the conditions needed for an f_T extraction. In its simplest form, the input circuit is just the total input capacitance $C_{\mathrm{gg}} = dQ_G/ dV_G$, as shown in Figure 2.6(a). However, it is well-known that C_{gg} can be modeled as a series combination of insulator capacitance C_{ins} and a so-called inversion-layer capacitance[3] C_{inv} [35, Ch. 3], [40], with C_{ins} and C_{inv} each being found as an integrated value of a change in charge with respect to potential along the channel:

$$C_{\mathrm{ins}} = \int_{-L/2}^{L/2} \frac{dQ_G(x)}{[dV_G + (1/q)dE_C(x, z_{\mathrm{ch}})]}\, dx \tag{2.2}$$

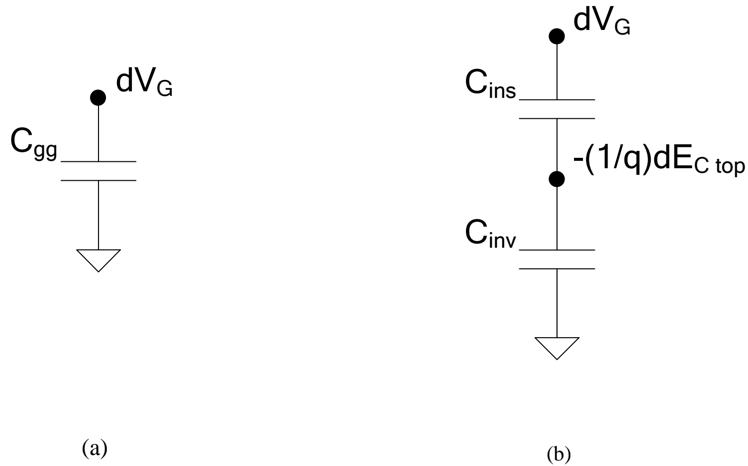
and

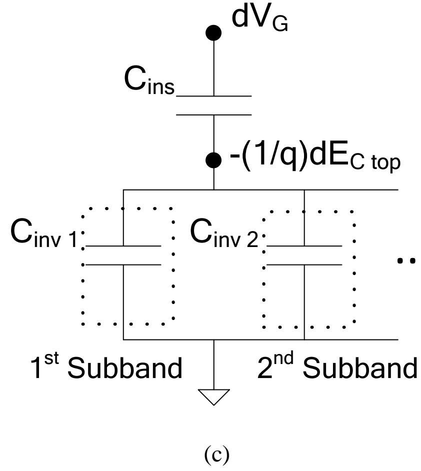
$$C_{\mathrm{inv}} = \int_{-L/2}^{L/2} \frac{dQ_G(x)}{-(1/q)dE_C(x, z_{\mathrm{ch}})}\, dx \tag{2.3}$$

where $dQ_G(x)$ and $dE_C(x, z_{\mathrm{ch}})$ represent the changes (due to dV_G) in gate-

[3] It is worth noting that the term "inversion-layer capacitance" used in the context of HEMTs is equivalent to the term "quantum capacitance" used in the context of emerging transistors [4, Ch. 7], [80, 98]. In the context of HEMTs, the term "quantum capacitance" is often used in a different context [40].

electrode charge and conduction-band edge, respectively, at each point x, and L is the total length of the device from the source to the drain, as depicted in Figure 2.1. A simple representation of C_{gg} is then given by the circuit in Figure 2.6(b), where $dE_{C\,\mathrm{top}}$ is the change in the conduction-band edge at the *top* of the n^{+}-channel barrier [41], located at a point $(x, z) = (x_{\mathrm{top}}, z_{\mathrm{ch}})$; this circuit applies in a lumped model of a ballistic device even when the conduction-band edge is not flat across the channel [42]. An alternative representation is provided in Figure 2.6(c),





(c)

Figure 2.6. Input equivalent circuit of the HEMTs. (a) Overall input circuit, which is just the input capacitance C_{gg}. (b) Separation of C_{gg} into the series combination of insulator and inversion-layer capacitances, C_{ins} and C_{inv}. (c) Further subdivision of C_{inv} into contributions arising from each subband, where $C_{\text{inv}\,i}$ is the contribution from the ith subband.

where C_{inv} is further subdivided into a parallel combination of capacitances arising from the occupied subbands, with $C_{\text{inv}\,i}$ representing the inversion-layer capacitance from the ith subband, and where the subbands themselves arise due to vertical confinement within the channel (*i.e.*, in the z-direction of Figure 2.1).

D Gate-Length Scaling of g_m

D.1 Expression for g_m

Consider now the scaling behavior of the transconductance g_m at peak f_T. From its definition, and utilizing the expression for current within the NEGF formalism [30], the intrinsic g_m can be written as

$$g_m = \frac{dI_D}{dV_G} = \frac{d}{dV_G}\left(\frac{q}{\hbar^2}\sqrt{\frac{m^* k_B T}{2\pi^3}} \int_{-\infty}^{\infty} T[E(k_x, k_z)]\left\{F_{-\frac{1}{2}}[\mu_S - E(k_x, k_z)] - F_{-\frac{1}{2}}[\mu_D - E(k_x, k_z)]\right\} dE(k_x, k_z)\right) \tag{2.4}$$

where m^* is the electron effective mass, $T[E(k_x, k_z)]$ is the total transmission function at an in-plane energy $E(k_x, k_z)$, μ_S is the source Fermi level, μ_D is the drain Fermi level, $F_{-\frac{1}{2}}$ is the Fermi-Dirac integral of order $-1/2$,

$$F_{-\frac{1}{2}}(\theta) = \int_{0}^{\infty} \frac{\eta^{-1/2}}{1 + \exp[\eta - (\theta/k_B T)]} d\eta \tag{2.5}$$

and it is to be understood that the right side of (2.4) and all subsequent expressions in this discussion should be evaluated at the gate bias corresponding to peak f_T.

For the purpose of examining the scaling behavior, the circuit in Figure 2.6(b) can be exploited to substitute $dV_G = [(C_{\text{ins}} + C_{\text{inv}})/C_{\text{ins}}](-1/q)dE_{C\,\text{top}}$ into (2.4), yielding

$$g_m = \frac{C_{\text{ins}}}{C_{\text{ins}} + C_{\text{inv}}} \times \frac{d}{(-1/q)dE_{C\text{ top}}}\left(\frac{q}{\hbar^2}\sqrt{\frac{m^* k_B T}{2\pi^3}}\int_{-\infty}^{\infty} T[E(k_x, k_z)]\left\{F_{-\frac{1}{2}}[\mu_S - E(k_x, k_z)] - F_{-\frac{1}{2}}[\mu_D - E(k_x, k_z)]\right\} dE(k_x, k_z)\right). \tag{2.6}$$

Equation (2.6) then reveals that the scaling behavior of the transconductance depends on the scaling behavior of two quantities. The first is the capacitance ratio $C_{\text{ins}}/(C_{\text{ins}} + C_{\text{inv}})$, which reflects (through voltage division) the ability of a perturbation in gate voltage dV_G to move the conduction-band edge at the top of the barrier by an amount $dE_{C\text{ top}}$. The second is the change in the integrated electron current for a given change $dE_{C\text{ top}}$, as specified by the remaining factor, *i.e.*, the derivative in (2.6); the key elements in this derivative are the transmission function $T[E(k_x, k_z)]$, which reflects the likelihood that an electron incident from the source with an in-plane energy $E(k_x, k_z)$ will be able to reach the drain, and the difference in Fermi-Dirac integrals, which reflects the "difference in agenda" [4, Ch. 1] between the source and drain contacts at each in-plane energy $E(k_x, k_z)$.

To gain further insight from (2.6), we note that under ballistic transport, the transmission function can be approximated as a sum of unit-step functions, with the steps occurring at those energies corresponding to the subband edges at the top of

the barrier:

$$T[E(k_x,k_z)] = \sum_i u\big[E(k_x,k_z) - E_{C\,\text{top}} - \Delta_i\big] \tag{2.7}$$

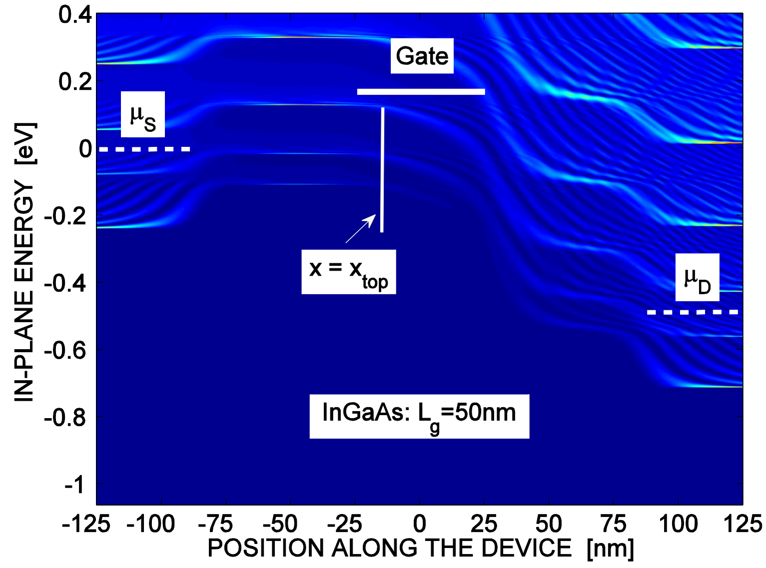
where Δ_i is the bottom edge of subband i with respect to $E_{C\,\text{top}}$. The use of (2.7) in (2.6) then leads to the following result for g_m:

$$g_m = \frac{q^2}{\hbar^2}\sqrt{\frac{m^* k_B T}{2\pi^3}}\frac{C_{\text{ins}}}{C_{\text{ins}} + C_{\text{inv}}}\sum_i \Big\{F_{-\frac{1}{2}}[\mu_S - E_{C\,\text{top}} - \Delta_i] - F_{-\frac{1}{2}}[\mu_D - E_{C\,\text{top}} - \Delta_i]\Big\}. \tag{2.8}$$

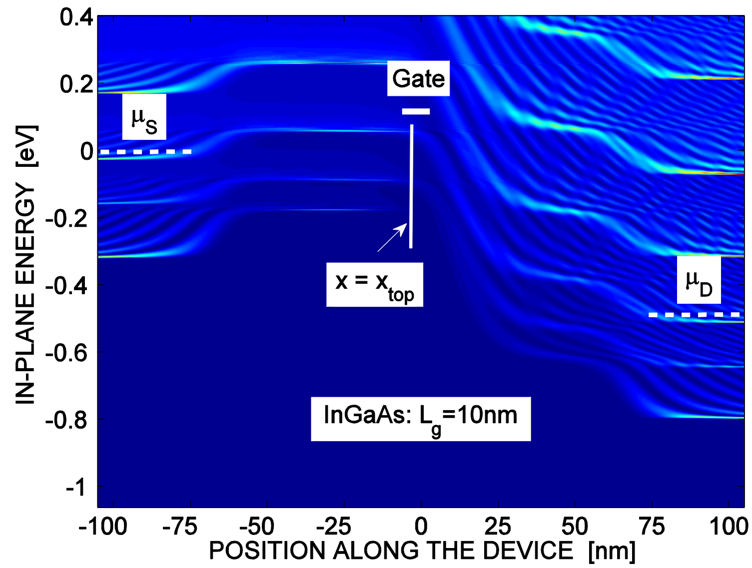
According to (2.8), the scaling behavior of g_m thus ultimately depends on the scaling behavior of the capacitance ratio $C_{\text{ins}}/(C_{\text{ins}} + C_{\text{inv}})$, which describes the ability of the gate to modulate the top of the barrier, and on the difference in Fermi-Dirac integrals *evaluated at each subband edge* $\varepsilon_i \equiv E_{C\,\text{top}} + \Delta_i$, which describes the "difference in agenda" of the source and drain contacts. As we now discuss, the capacitance ratio and the difference in Fermi-Dirac integrals at peak f_T change only weakly with scaling and in opposition to each other, such that the corresponding g_m remains relatively insensitive to scaling.

D.2 Position of Subband Edges

To describe this outcome, it is necessary to follow the relative positions of the subband edges at peak f_T as the gate length is scaled down. While a detailed explanation of the phenomenon will be provided in subsection E, for the present discussion, it suffices to note that the edges of the first few subbands at peak f_T will be located *further below* the source Fermi level as the gate length is scaled down. This result can be discerned from the plots of the spectral functions in Figure 2.7. The figure shows the total spectral function in the channel vs. in-plane energy and position, *i.e.*, $A_S[x, z_{\mathrm{ch}}, E(k_x, k_z)] + A_D[x, z_{\mathrm{ch}}, E(k_x, k_z)]$ displayed as an intensity vs. $E(k_x, k_z)$ and x, for the two extreme gate-length InGaAs devices, *i.e.*, the 10- and 50-nm devices, at the gate bias corresponding to peak f_T. An inspection of the plots reveals that the subband edges (indicated by the brightly shaded regions) under the gate do indeed move down with respect to μ_S as the gate length is scaled from 50 to 10 nm.



(a)



(b)

Figure 2.7. Total spectral function $A_S[x, z_{\mathrm{ch}}, E(k_x, k_z)] + A_D[x, z_{\mathrm{ch}}, E(k_x, k_z)]$ displayed as intensity vs. $E(k_x, k_z)$ and x, along with the positions of the Fermi levels μ_S and μ_D, for the InGaAs HEMT with (a) L_g= 50 nm and (b) L_g= 10 nm, showing that at the shorter gate length, the subband edges at peak f_T are lower in position with respect to the source Fermi level.

D.3 Impact on Fermi-Dirac Integrals

The impact of a change in the relative positions of the subband edges on the Fermi-Dirac integrals in (2.8) is shown in Figure 2.8(a), where the subband edges at the top of the barrier and at the gate bias corresponding to peak f_T are superimposed on a sketch of the difference $\left\{F_{-\frac{1}{2}}[\mu_S - E(k_x, k_z)] - F_{-\frac{1}{2}}[\mu_D - E(k_x, k_z)]\right\}$. The shift in the positions of the subbands with down scaling causes an enhanced contribution to the difference in the Fermi-Dirac integrals evaluated at each subband edge, *i.e.*, there is a greater difference in agenda between the source and drain contacts at peak f_T for each subband as the gate length is scaled down, and this will tend to increase g_m.

D.4 Impact on Capacitance Ratio

By contrast, as the gate length is scaled down, the lower position of the subband edges causes the capacitance ratio $C_{\text{ins}}/(C_{\text{ins}} + C_{\text{inv}})$ to *decrease*, which will tend to *decrease* g_m. The decrease in the capacitance ratio can be understood by noting that C_{ins} is primarily determined by the insulator thickness, which is fixed in this work, causing C_{ins} to scale linearly with gate length, while C_{inv} is larger at each gate length than would be expected from a purely linear dependence on L_g.

To understand the behavior of C_{inv}, we note that the charge at node (x, z) from the NEGF formalism [30] can be represented as an appropriate integral (over energy) of $F_{-\frac{1}{2}}$ times the source and drain spectral functions (local densities of

states):

$$q \times n(x,z) = \frac{q}{ab}\sqrt{\frac{m^* k_B T}{2\pi^3 \hbar^2}} \int_{-\infty}^{\infty} \Big\{ F_{-\frac{1}{2}}[\mu_S - E(k_x,k_z)] A_S[x,z,E(k_x,k_z)] + F_{-\frac{1}{2}}[\mu_D - E(k_x,k_z)] A_D[x,z,E(k_x,k_z)] \Big\} dE(k_x,k_z) \tag{2.9a}$$

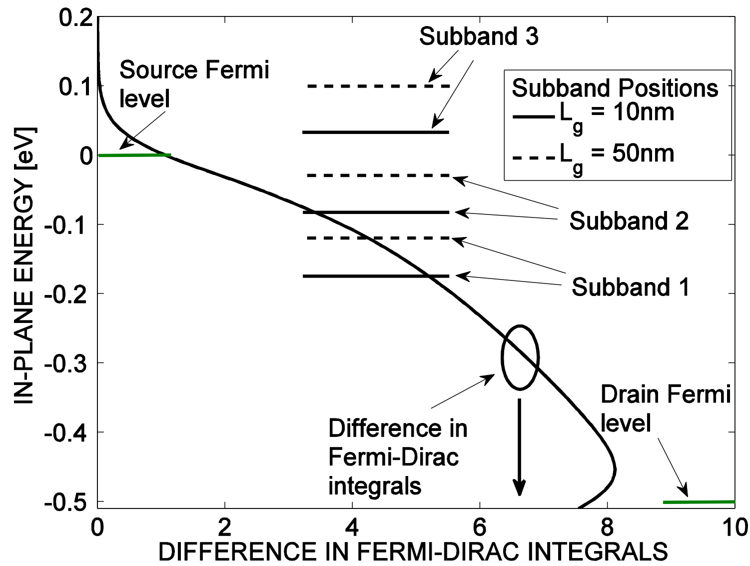
$$\approx \frac{q}{ab}\sqrt{\frac{m^* k_B T}{2\pi^3 \hbar^2}} \int_{-\infty}^{\infty} F_{-\frac{1}{2}}[\mu_S - E(k_x,k_z)] A_S[x,z,E(k_x,k_z)] dE(k_x,k_z) \tag{2.9b}$$

where a and b are the grid sizes along the x- and z-directions, respectively, A_S and A_D are the spectral functions due to the source and drain contacts, respectively, and μ_D is assumed to lie sufficiently below μ_S for the second term in the integrand of (2.9a) to be neglected in comparison with the first at all points x within the channel to yield (2.9b) (which will be true at typical drain bias voltages). The inversion capacitance C_{inv} is then given by the prescription in (2.3), where $dQ_G(x)$ is equal to the right side of (2.9b) integrated over z in the channel. Rather than integrate over z, it is more instructive to examine the trends in the charge located at the single point corresponding to the top of the barrier, *i.e.*, the point $(x,z) = (x_{\text{top}}, z_{\text{ch}})$,

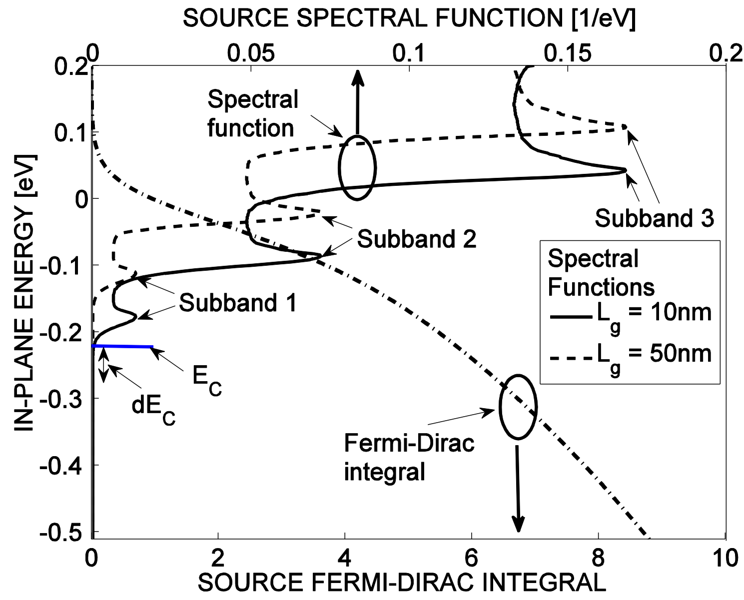
which will be representative of the trends in $dQ_G(x)$. Figure 2.8(b) shows plots of the key factors of (2.9b) found in this way, at peak f_T and for the same InGaAs device and gate lengths considered in Figure 2.7.

Plotted in Figure 2.8(b) are hence the Fermi-Dirac integral $F_{-\frac{1}{2}}[\mu_S - E(k_x, k_z)]$ and the spectral function $A_S[x_{\text{top}}, z_{\text{ch}}, E(k_x, k_z)]$; the Fermi-Dirac integral looks the same at both gate lengths, with $\mu_S \equiv 0$ taken as the reference, so that the only difference is in the spectral function. As expected, the spectral function shows a sequence of $1/\sqrt{E(k_x, k_z)}$ dependencies that are consistent with the existence of a sequence of one-dimensional subbands arising from confinement in the z-direction, *i.e.*, consistent with a dispersion relation of the form $E(k_x, k_z) = \varepsilon_n + \frac{\hbar^2}{2m^*} k_x^2$; the subband edges are marked in the figure. In each case, the charge $q \times n(x_{\text{top}}, z_{\text{ch}})$ is given by the area under the *overlap* of the Fermi-Dirac integral and the spectral function; more importantly, under dynamic conditions, *i.e.*, under a modulation dE_C in E_C, which can be visualized as shifting the spectral functions, the *incremental* overlap and hence the *incremental* charge is greater in the shorter gate-length device. This result is not immediately obvious, but it can be discerned from the figure for two reasons: (i) the incremental charge will be greater in the first two subbands of the shorter gate-length device, since they are further below the source Fermi level and hence experience a greater population modulation (since the Fermi-Dirac integral is greater); (ii) the third subband will participate in the shorter gate-length device, whereas in the longer gate-length device, it is too far above the source Fermi level to hold any charge or to participate in charge

modulation, *i.e.*, the associated states are always empty. As a result of the greater charge modulation with a modulation dE_C in E_C, the capacitance C_{inv} is larger in the smaller device than would be expected from a purely linear dependence on gate length, and the capacitance ratio $C_{\mathrm{ins}}/(C_{\mathrm{ins}}+C_{\mathrm{inv}})$ thus *decreases* with a down scaling of gate length.



(a)



(b)

Figure 2.8. (a) The difference in the Fermi-Dirac integrals $\left\{F_{-\frac{1}{2}}[\mu_S - E(k_x,k_z)] - F_{-\frac{1}{2}}[\mu_D - E(k_x,k_z)]\right\}$, along with the positions of the occupied subband edges at the top of the barrier, $(x,z) = (x_{\text{top}}, z_{\text{ch}})$. (b) Source Fermi-Dirac integral $F_{-\frac{1}{2}}[\mu_S - E(k_x,k_z)]$ and source spectral function $A_S[x_{\text{top}}, z_{\text{ch}}, E(k_x,k_z)]$ vs. in-plane energy $E(k_x,k_z)$ for the InGaAs HEMT with gate lengths L_g= 10 and 50 nm; the source Fermi level is $\mu_S \equiv 0$, and the conduction-band edge E_C is marked.

D.5 Overall Scaling of g_m

Overall, the capacitance ratio and the difference in Fermi-Dirac integrals in (2.8) thus act in opposition to each other as the gate length is scaled down, as shown in Figure 2.9, leading to a g_m (at peak f_T) that does not scale significantly with gate length, as also shown in Figure 2.9. *In essence, as the gate length is scaled down, the subband positions at peak f_T move lower in energy with respect to the source Fermi level, causing the difference in agenda between the source and drain*

contacts to increase for each subband (reflected by the greater difference in Fermi-Dirac integrals) but with this increase being offset by a weaker control of the channel barrier by the gate (reflected by the decreased capacitance ratio). It is worth mentioning that a similar trend in g_m was experimentally observed in [43, Figure 6] for an InAs HEMT with t_{ins} = 4 nm, where scaling the gate length below 90 nm resulted in an insignificant improvement (by only a few percent) in g_m.

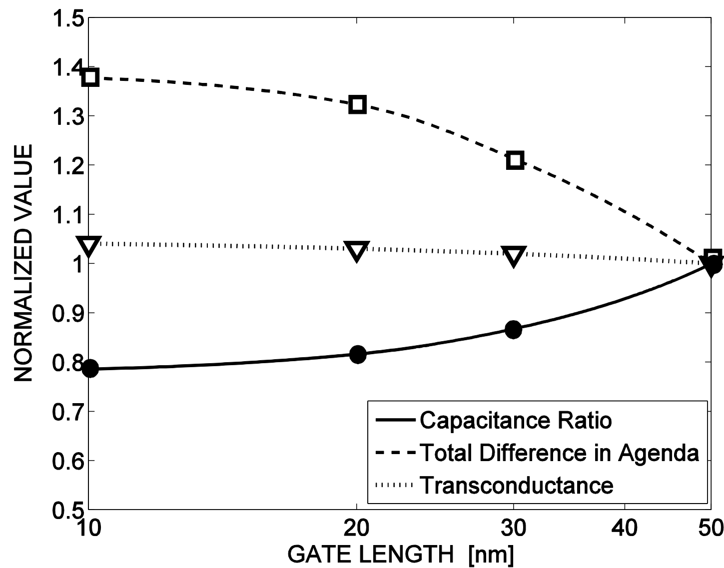


Figure 2.9. Capacitance ratio $C_{\text{ins}}/(C_{\text{ins}} + C_{\text{inv}})$, total difference in Fermi-Dirac integrals ("difference in agenda"), *i.e.*, the result of the summation in (2.8), and the transconductance g_m (at peak f_T) versus gate length L_g for the InGaAs HEMT. The actual values have been normalized with respect to the value of each component at the gate length L_g= 50 nm in order to illustrate the scaling behavior, and the lines have been drawn as guides for the eye.

Two points are worth adding regarding the insensitivity of g_m to gate-length scaling. First, we found the same outcome even when the insulator thickness was scaled (from 3 nm down to 2 nm) with gate length, where the g_m improved only

slightly (by a few percent) and where the same tradeoffs occurred. Second, while we have used the InGaAs devices to illustrate the result, a similar trend in g_m can be expected irrespective of the material, *i.e.*, irrespective of the precise value of the effective mass; this follows because the compensating increase and decrease leading to an insensitivity of g_m to gate length will always occur, with only the extent of the increase and decrease varying between materials.

Since the g_m is relatively constant with gate length, the scaling behavior of $f_T = g_m/(2\pi C_{\mathrm{gg}})$ is determined by the scaling behavior of C_{gg}.

E Gate-Length Scaling of C_{gg}

In contrast to the behavior of g_m, the total gate capacitance C_{gg} at peak f_T is significantly affected by the scaling of gate length.

As previously illustrated in Figure 2.6(b), C_{gg} can be modeled as a series combination of insulator and inversion capacitance:

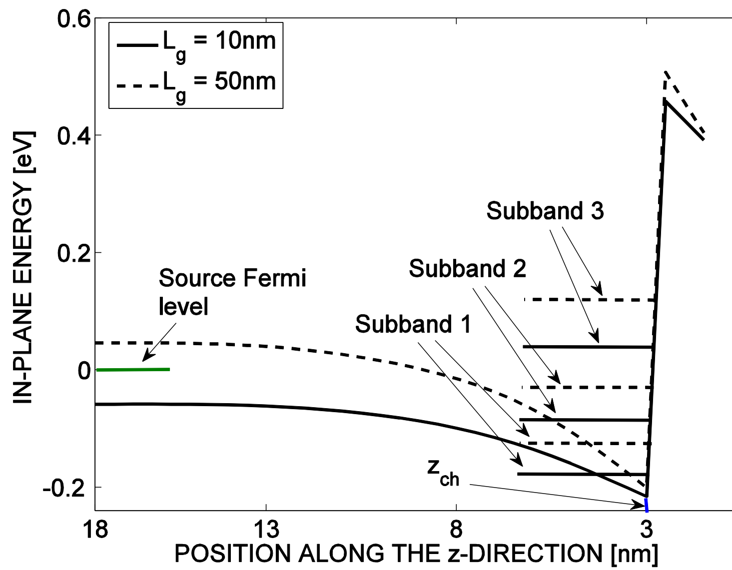
$$C_{\mathrm{gg}} = \frac{C_{\mathrm{ins}}C_{\mathrm{inv}}}{C_{\mathrm{ins}}+C_{\mathrm{inv}}}. \tag{2.10}$$

Since III-V materials have a relatively low density of states (due to a relatively small electron effective mass) and are fabricated with very thin, high-k insulators, then C_{inv} is significantly smaller than C_{ins}; for example, the ratio $C_{\mathrm{inv}}/C_{\mathrm{ins}}$ ranged from 0.05 to 0.26 for the InGaAs and 0.25 to 0.39 for the GaN device when the gate

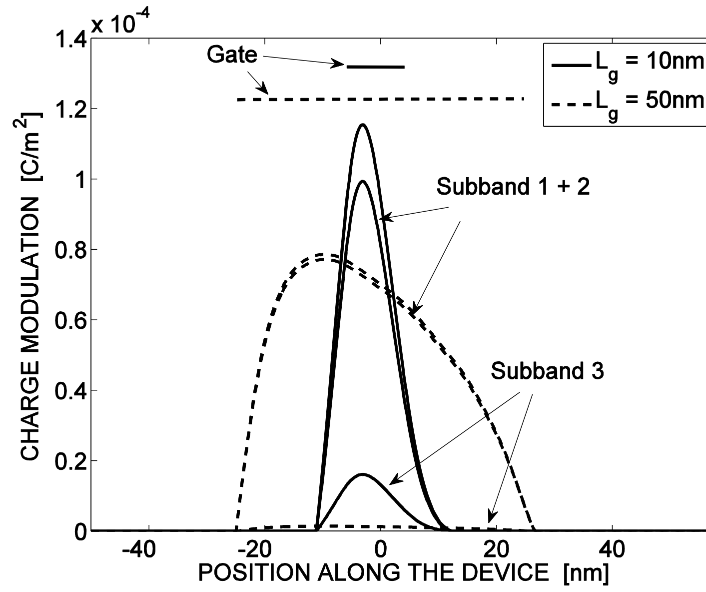
length was scaled from 50 to 10 nm. Hence, to a first approximation, $C_{\text{gg}} \sim C_{\text{inv}}$, and C_{inv} controls the scaling behavior of C_{gg}.

As explained earlier, DIBL causes the peak f_T to occur at a smaller gate bias V_G for the shorter gate-length devices. As a consequence, the channel charge is less tightly held near the insulator interface by V_G, or equivalently, the quantum well defining the channel is *less sharp*; this can be observed from the plot of conduction-band profiles along the depth of the channel (along the z-direction) at $x = x_{\text{top}}$, *i.e.*, from a plot of $E_C(x_{\text{top}}, z)$ vs. z, as illustrated in Figure 2.10(a), which shows results for the same InGaAs device and gate lengths considered in Figure 2.7. For the shorter gate length, the triangular shape of the well is less pronounced, *i.e.*, there is reduced quantum confinement, and hence the subbands lie closer to the conduction-band edge at the tip of the well [44, Ch. 1], *i.e.*, closer to $E_C(x_{\text{top}}, z_{\text{ch}})$. Since the position of $E_C(x_{\text{top}}, z_{\text{ch}})$ relative to μ_S at peak f_T is fixed, which follows because the latter always occurs at the onset of barrier collapse, *i.e.*, when E_C in the channel aligns with that in the n^+ region near the source (Figure 2.3(b)), it then follows that the subbands at peak f_T are lower in position with respect to μ_S at shorter gate lengths. This result, already mentioned above in subsection D, is readily seen from the data in Figure 2.10(a). Since the subband positions are lower with respect to μ_S at shorter gate lengths, then as already discussed in conjunction with Figure 2.8(b), a modulation dE_C in E_C leads to a relatively larger charge modulation (at each point x under the gate) in the first three subbands. This outcome is illustrated in Figure 2.10(b) for the same InGaAs devices considered in Figure 2.7.

Therefore, at shorter gate lengths, there is a relatively larger overall charge modulation than would be expected from a purely linear dependence on gate length, causing C_{inv} to be larger than would be expected from a purely linear dependence on gate length. This behavior of C_{inv} with scaling is illustrated in Figure 2.11(a), where we have plotted $1/C_{\text{inv}}$; as shown, the larger-than-expected values of C_{inv} cause the behavior of $1/C_{\text{inv}}$ to saturate at short gate lengths. In turn, $1/C_{\text{gg}} \sim 1/C_{\text{inv}}$ behaves in a similar manner, as shown in Figure 2.11(b).



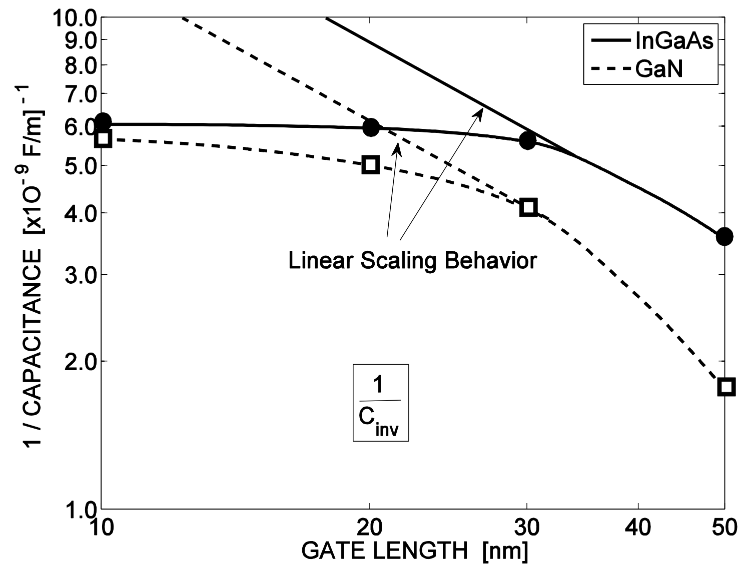
(a)



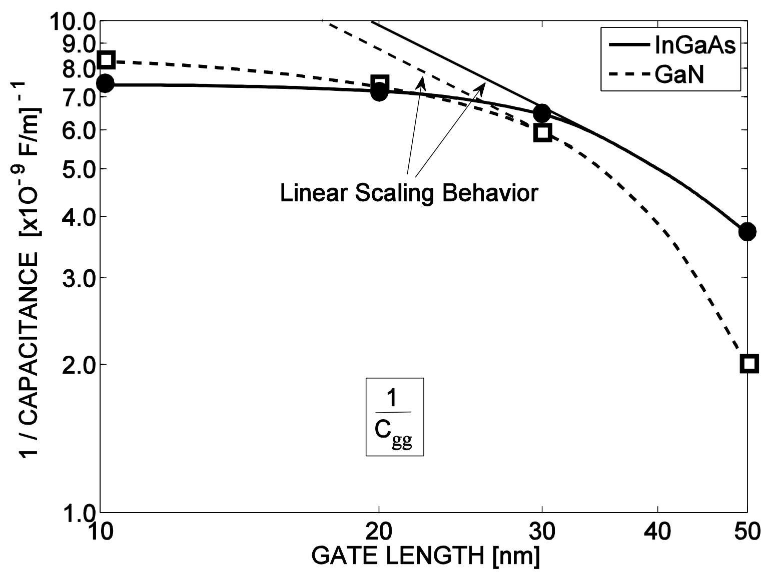
(b)

Figure 2.10. (a) Conduction-band profiles at peak f_T along the depth of the channel, *i.e.*, $E_C(x_{\mathrm{top}}, z)$ vs. z, for the InGaAs HEMTs with L_g= 10 and 50 nm; also shown are the positions of the occupied subband edges at $(x, z) = (x_{\mathrm{top}}, z_{\mathrm{ch}})$, illustrating that the subbands are lower in position with respect to the source Fermi level for the shorter gate-length device. (b) Charge modulation along the transport direction (*i.e.*, the x-direction) for the InGaAs HEMTs with L_g= 10 and 50 nm; the results show a relatively greater charge modulation (under the gate) in the first three subbands of the shorter gate-length device.

Since $1/C_{\mathrm{gg}}$ saturates, while the corresponding g_m remains almost constant (subsection D), *the peak* $f_T = g_m/(2\pi C_{\mathrm{gg}})$ *in III-V materials will exhibit a signature saturation with a down scaling of gate length, as we depicted in Figure 2.5 and as originally observed by Schwierz et al.* [3, Figure 10].



(a)



(b)

Figure 2.11. (a) Reciprocal of inversion-layer capacitance $1/C_{\mathrm{inv}}$ and (b) reciprocal of total gate capacitance $1/C_{\mathrm{gg}}$ versus gate length L_g for the HEMTs. The lines have been drawn as guides for the eye, and to emphasize the saturating behavior at short gate lengths, additional lines are shown to illustrate the expected behavior based on linear scaling with L_g, extrapolated to lower gate lengths from $L_g \sim 35$ nm.

F Impact of Channel-Material Effective Mass

Among the HEMT channel materials considered in this work, InGaAs has the smaller effective mass in comparison to GaN ($0.048 \times m_0$ [28] vs. $0.2 \times m_0$ [45], where m_0 is the electron rest mass). While many factors can impact the relative positions of the subbands at the point of barrier collapse and hence at peak f_T, we note that *when comparing materials*, the most important consideration is the requirement that E_C in the channel be aligned with that in the n^+ region near the source; this requirement then implies that the charge at the top of the barrier (integrated over the extent of the channel in the z-direction) be equal (to first order) to that in the n^+ region near the source [28], where the latter is determined by the doping concentration and is hence the same for the two HEMTs under consideration.

To accommodate the required charge at the top of the barrier, the subband edges in InGaAs — which has the lower effective mass and hence the lower density of states — must move further below the source Fermi level than those in GaN; this outcome is illustrated in the plot of conduction-band profiles along the z-direction at $x = x_{\text{top}}$ in Figure 2.12 for the InGaAs and GaN HEMTs having a gate length L_g= 10 nm. For reasons already discussed in subsections D and E above, the inversion capacitance C_{inv} in InGaAs, and hence the total gate capacitance $C_{\text{gg}}{\sim}C_{\text{inv}}$ in InGaAs, will thus exhibit a greater deviation from purely linear gate-length scaling than in GaN; equivalently, the reciprocal capacitances $1/C_{\text{inv}}$ and

$1/C_{\mathrm{gg}}$ will experience a more pronounced saturation at shorter gate lengths in InGaAs vs. GaN, as depicted in the two parts of Figure 2.11. Since the saturation of $1/C_{\mathrm{gg}}$ is more pronounced in the lighter mass material, the peak $f_T = g_m/(2\pi C_{\mathrm{gg}})$ experiences a more pronounced saturation, as shown in Figure 2.5. More generally, these results illustrate that *devices with a small effective mass will suffer from a more rapid saturation in peak f_T with a down scaling of gate length in comparison with devices having a heavy effective mass*.

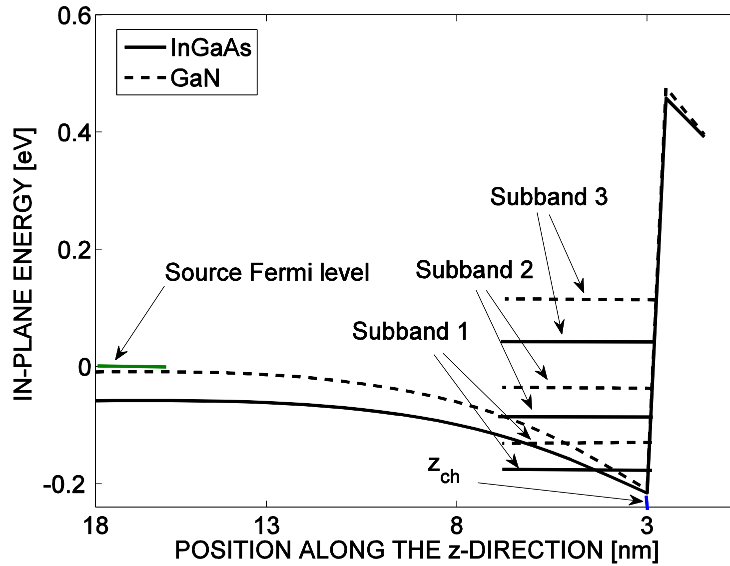


Figure 2.12. Conduction-band profiles at peak f_T along the depth of the channel, *i.e.*, $E_C(x_{\mathrm{top}}, z)$ vs. z, for the InGaAs and GaN HEMTs with L_g= 10 nm; the subbands for the InGaAs device are lower in position with respect to the source Fermi level.

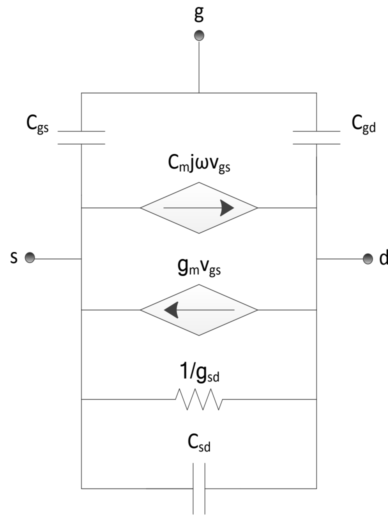
2.3 Complete Device

To examine the scaling behavior of the *extrinsic* RF metrics, the NEGF-Poisson solver was utilized to find the components of a small-signal equivalent circuit for

the intrinsic device, which was then combined with the parasitic elements to develop a complete (extrinsic) device model. This model was then used to extract the device y-parameters and subsequently the extrinsic f_T and $f_{\max}$.

2.3.1 Small-Signal Equivalent Circuit

The small-signal equivalent circuit for an intrinsic III-V HEMT is the classical circuit usually used for field-effect transistors [39, Figure 9.5], shown here in Figure 2.13(a), with the definitions of the elements (which have their usual meanings) provided in Figure 2.13(b). The source/drain charge partitioning factor χ is not critical to the results of interest in this work; for completeness, we have chosen $\chi = 0.4$, which is the value suggested in [46] and a value that is also commonly used for MOSFETs. The values of all the other parameters, for both the InGaAs and GaN HEMTs, are provided in Table 2.1; values are listed for the L_g= 30 nm case as a representative example.



(a)

$$C_{gs} = \frac{dQ_G}{dV_S}$$
$$C_{gd} = \frac{dQ_G}{dV_D}$$
$$C_{dg} = \chi \frac{dQ_G}{dV_G}$$
$$C_m = C_{dg} - C_{gd}$$
$$C_{sd} = (1-\chi)\frac{dQ_G}{dV_D}$$
$$g_m = \frac{dI_D}{dV_G}$$
$$g_{sd} = \frac{dI_D}{dV_D}$$

(b)

Figure 2.13. (a) Small-signal equivalent circuit of an intrinsic III-V HEMT from [39, Figure 9.5]. (b) Circuit elements, to be evaluated at the dc operating point; the symbols bear their usual meanings.

TABLE 2.1. ELEMENT VALUES FOR THE CIRCUIT OF FIGURE 2.13 WHEN $L_g = 30\ \mathbf{nm}$

	C_{gs} [F/m]	C_{gd} [F/m]	C_m [F/m]	C_{sd} [F/m]	g_m [S/m]	g_{sd} [S/m]
InGaAs HEMT	1.08×10^{-10}	0.46×10^{-10}	0.15×10^{-10}	0.27×10^{-10}	3.30×10^{3}	1.20×10^{2}
GaN HEMT	1.14×10^{-10}	0.47×10^{-10}	0.17×10^{-10}	0.28×10^{-10}	2.78×10^{3}	1.25×10^{2}

2.3.2 Modeling of Parasitics

Figure 2.14 shows the device structure used to extract the parasitic resistances and capacitances; the structure is consistent with those reported in [28, 29, 47, 48]. The source and drain metal contacts are 50 nm in both length and height, and they are placed 1 µm apart on InGaAs/InAlAs and GaN/InGaN heterostructure stacks for

the InGaAs and GaN HEMTs, respectively, with each stack having a thickness of 15 nm. The height of the gate metal is 150 nm, with the lower and upper parts having lengths of L_g and L_g+300 nm, respectively, where L_g varies from 50 to 10 nm with scaling. The extent of the intrinsic device is indicated in the figure, and the width W in the y-direction is considered to be 1 μm in this work.

The gate resistance R_g arising from the gate metal is considered to be distributed in nature, similar to MOSFETs, and it is modeled by a single lumped and effective resistance in series with the gate lead, as in [49]. Ni/Au and Ti/Pt/Au are assumed to be the gate metals for the InGaAs and GaN HEMTs, as in [50] and [51], respectively. Although very small, the source and drain metal resistances, R_s and R_d, respectively, are also considered in the complete device; Ni/Ge/Au and Ti/Al are used as the source and drain contact metals, as in [29] and [51], for the InGaAs and the GaN HEMTs, respectively. In each case, instead of modeling the multilayer nature of the contact metals, only the lowermost metal layer is considered to represent the entire contact; this assumption can be justified by the fact that the lowermost metal layer is the material that sets the work function and thus controls the overall behavior of the contact.

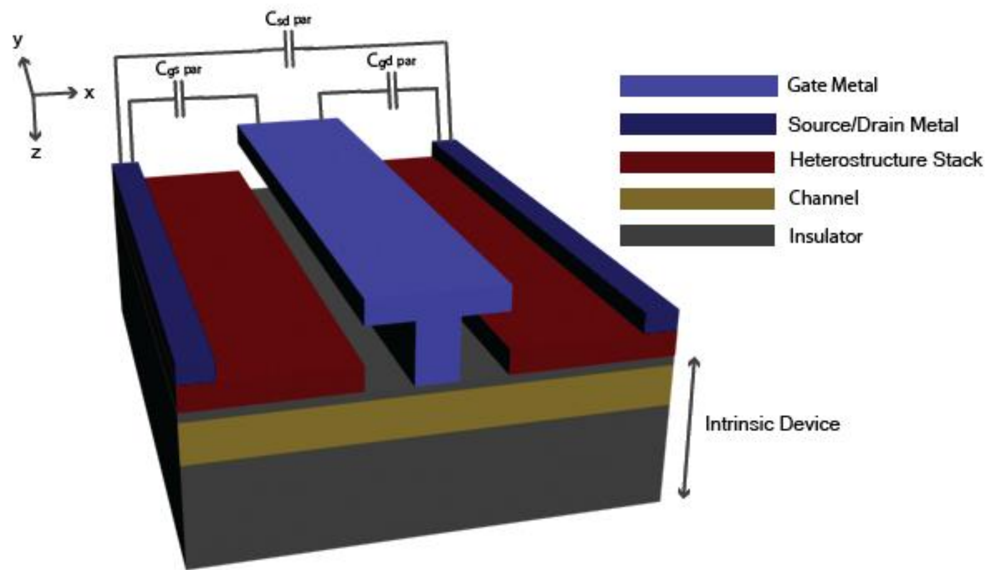


Figure 2.14. Full device structure of the HEMTs.

The parasitic capacitances between the contact metals are extracted by designing an “open-device” model [52] for the HEMTs, where the structure is exactly like the actual device but with zero charge in the intrinsic portion, and then using COMSOL as outlined in [49]. These parasitic capacitances are represented as $C_{\mathrm{gs\ par}}$, $C_{\mathrm{gd\ par}}$, and $C_{\mathrm{sd\ par}}$ in the complete device model, the topology of which is based on [39, Figure 8.30] and which is shown in Figure 2.15.

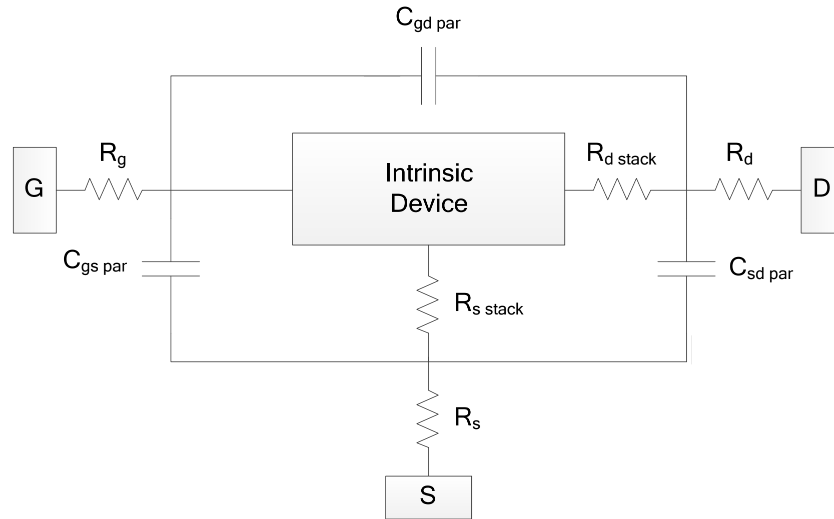


Figure 2.15. Complete device model of the III-V HEMTs. The intrinsic device is represented by the small-signal circuit of Figure 2.13.

Apart from the above parasitic elements, the heterostructure stacks at the source and drain can also impact the performance of the HEMTs. This effect can be modeled simply by two series resistances, $R_{s\ \text{stack}}$ and $R_{d\ \text{stack}}$, connected at the two ends of the intrinsic model [28, 34], [35, Ch. 3]. These are considered to be $200\ \Omega\ \mu\text{m}/W$ for the InGaAs HEMTs [28], [35, Ch. 3] and $400\ \Omega\ \mu\text{m}/W$ for the GaN HEMTs [53].

The values of the parasitic elements are listed in Table 2.2 for both the InGaAs and GaN HEMTs; the values do not scale with gate length except for R_g, which is listed for the L_g= 30 nm case as a representative example. With these values, the model of Figure 2.15 was utilized to generate the overall y-parameters and hence to find the extrinsic f_T and f_{max}.

2.3.3 Results

We focus on the scaling behavior of the extrinsic f_T and $f_{\max}$, where for each gate length, the bias point was chosen to be that for peak f_T of the intrinsic device. The current gain $|h_{21}| = |y_{21}/y_{11}|$ [54] and the unilateral power gain U [55], calculated from the overall y-parameters, are extrapolated to obtain the extrinsic f_T and $f_{\max}$, respectively; sample plots of $|h_{21}|$ and U are provided in Figure 2.16.

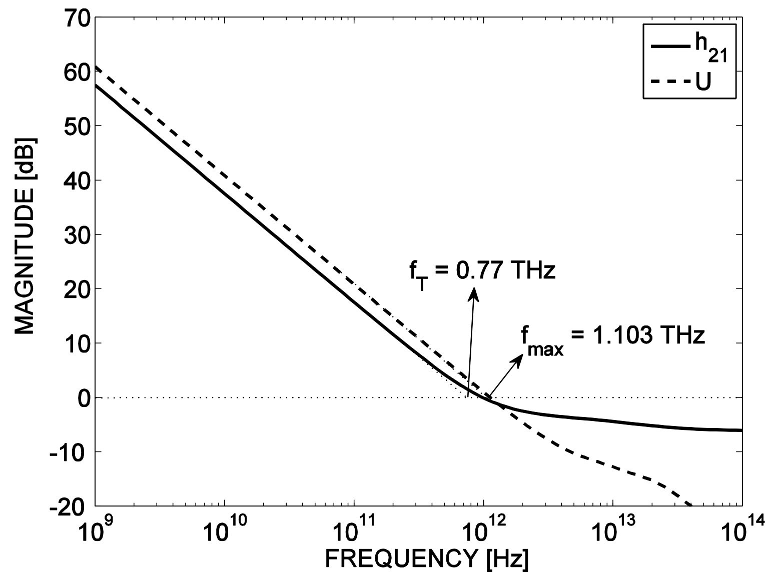


Figure 2.16. Current gain $|h_{21}|$, unilateral power gain U, and extrapolated f_T and $f_{\max}$ of the InGaAs HEMT at L_g= 30 nm.

TABLE 2.2. PARASITIC ELEMENT VALUES OF THE HEMTS (FOR $W = 1$ **μm**)

	$C_{\text{gs par}}$ [F]	$C_{\text{gd par}}$ [F]	$C_{\text{sd par}}$ [F]	R_g [for $L_g =$ 30 nm] [Ω]	R_s [Ω]	R_d [Ω]	$R_{s\text{ stack}}$ [Ω]	$R_{d\text{ stack}}$ [Ω]
InGaAs HEMT	3.41×10^{-17}	3.37×10^{-17}	9.29×10^{-18}	3.69	0.072	0.072	200	200
GaN HEMT	2.56×10^{-17}	2.53×10^{-17}	6.27×10^{-18}	20.51	0.4	0.4	400	400

Figure 2.17, which plots the extrinsic f_T versus gate length L_g, shows that the extrinsic f_T exhibits the same saturation at short gate lengths that was discussed for the intrinsic f_T. Moreover, as expected, the InGaAs HEMT exhibits a more pronounced saturation, given its lower effective mass. As a result of the rapid saturation in the f_T of the InGaAs HEMTs, the GaN HEMTs have comparable values of extrinsic f_T at short gate lengths. We have also shown experimental data for the f_T of InGaAs HEMTs in Figure 2.17, as collected in [3, Figure 10], and they exhibit the same trend as our simulations.

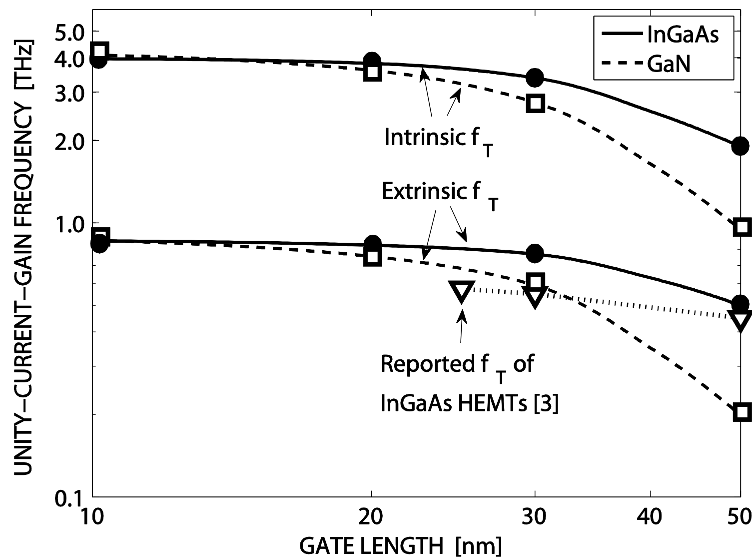


Figure 2.17. Extrinsic and intrinsic f_T of the HEMTs considered in this work, and the reported f_T of InGaAs HEMTs [3, Figure 10], versus L_g. The lines have been drawn as guides for the eye.

Regarding the $f_{\max}$, it is well known that, to a first approximation, $f_{\max} \propto \sqrt{f_T/(RC)_{\text{eff}}}$ [56-58] for RF transistors, where $(RC)_{\text{eff}}$ refers to an effective charging time. While the scaling behavior of $f_{\max}$ based on such a relationship can be involved due to the involved nature of $(RC)_{\text{eff}}$, it is clear that the saturating behavior of the f_T at short gate lengths will also tend to saturate the $f_{\max}$, as depicted in Figure 2.18. In addition, among the HEMTs under consideration, the GaN devices suffer from substantially larger parasitic resistances in the gate, source, and drain leads, as indicated by the higher values of the associated resistances in Table 2.2, and this will tend to degrade the $f_{\max}$ through a larger value of $(RC)_{\text{eff}}$; thus, although the two HEMTs have comparable extrinsic f_T at short gate lengths, the GaN HEMTs have a lower $f_{\max}$ at all gate lengths, as shown in Figure 2.18.

It should be mentioned that apart from the effective mass, other effects can contribute to the diminishing enhancement of the extrinsic f_T and $f_{\max}$ at short gate lengths. For example, while not an issue for the devices studied in this work, parasitic resistances and capacitances that do not scale with gate length at the same rate as the elements of the intrinsic device can exacerbate the saturating behavior of the extrinsic figures of merit.

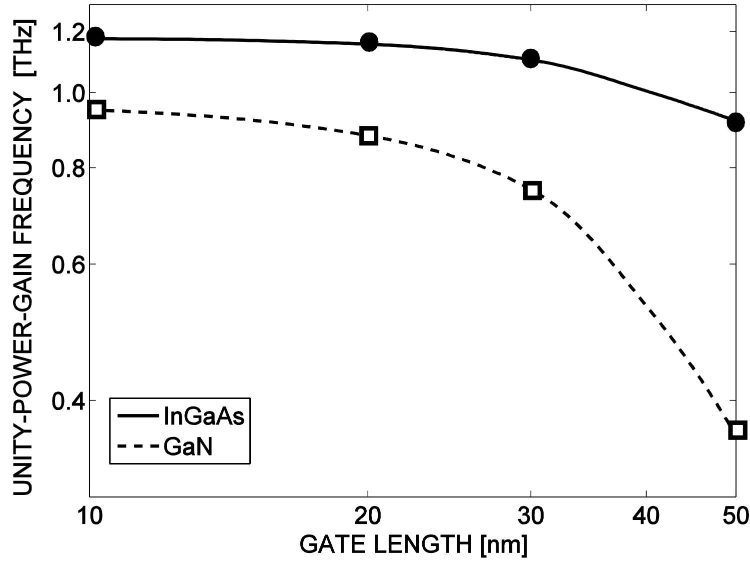


Figure 2.18. Extrinsic $f_{\max}$ of the HEMTs versus L_g. The lines have been drawn as guides for the eye.

2.4 Conclusions

The following conclusions can be drawn from this study of the impact of effective mass on the gate-length scaling behavior of the f_T and $f_{\max}$ of III-V HEMTs:

1. The intrinsic peak f_T occurs at gate voltages corresponding to the point of barrier collapse, beyond which the f_T degrades significantly.
2. At shorter gate lengths, drain-induced barrier lowering (DIBL) causes the barrier to collapse at lower gate bias voltages, such that the intrinsic peak f_T occurs at lower gate voltages.
3. For a given channel material, with a down scaling of gate length, the transconductance g_m at the gate bias corresponding to peak f_T remains relatively insensitive to scaling. On the other hand, the low effective mass causes the intrinsic gate capacitance C_{gg} to roughly equal the inversion

capacitance C_{inv}, which scales slower than would be expected from a purely linear dependence on gate length; the slower scaling is an outcome of the peak f_T occurring at lower gate biases at shorter gate lengths (due to DIBL), which reduces the sharpness of the potential well defining the channel and thereby lowers the position of the conduction subbands with respect to the source Fermi level, leading to a larger-than-expected charge modulation and hence a larger-than-expected capacitance. The intrinsic peak $f_T = g_m/(2\pi C_{\mathrm{gg}})$ thus exhibits a saturating behavior when the gate length is scaled down, *i.e.*, it shows no further increase with decreasing gate length once the gate length is sufficiently small; our results (Figs. 2.11 and 2.17) show this to occur for gate lengths below approximately 30 nm.

4. In comparing channel materials, the material with lower effective mass will exhibit a more pronounced saturation in its peak f_T as the gate length is scaled down; this occurs because the subbands in the lighter mass material must move further below the source Fermi level to accommodate the required charge at the top of the barrier at peak f_T (as set by the doping in the n^+ region), and the lower positioning of the subbands accentuates the larger-than-expected gate capacitance as the gate length is scaled down.
5. The extrinsic peak f_T of HEMTs reflect the saturating behavior of the intrinsic f_T. In comparing HEMTs, the InGaAs HEMTs have a more pronounced saturation in their f_T due to their lower effective mass, such that the f_T of the two HEMTs are comparable at short gate lengths.

6. The saturating behavior of the f_T of III-V HEMTs at short gate lengths will contribute to the saturating behavior of the $f_{\max}$. In comparing HEMTs, the larger parasitic resistances of GaN HEMTs causes them to have a lower $f_{\max}$ than the InGaAs HEMTs at all gate lengths.

Overall, the most important outcome of this work is the connection between the effective mass and the scaling behavior of RF performance. While a low effective mass is desirable for high mobility and potentially high-speed operation, it leads to diminishing improvements in the peak f_T and $f_{\max}$ as the gate length is scaled down.

Chapter 3

Impact of Substrate Material on the RF Performance of Carbon-Nanotube Transistors[4]

3.1 Introduction

Since their discovery in 1991, carbon nanotubes (CNs) have stimulated a plethora of experimental and theoretical research to better understand the charge transport in (quasi) 1D-material systems, and to assess whether CN-based channel materials have the potential to replace or augment the well-matured, silicon-based CMOS technology [3].

One of the major concerns has been to understand how scattering by lattice vibrations *native* to the nanotube (native phonons) affects the terminal behavior of CN field-effect transistors (CNFETs). So far, most studies on native phonon scattering have been conducted for CNFETs under steady-state (dc) operating conditions [59-66]; both measurements and simulations of the dc mobility and output currents have revealed that native acoustic and optical phonons strongly contribute to limiting the transport in metallic and semiconducting tubes, including a saturation of the output current at high bias [59, 63, 67].

[4] A version of this chapter has been published [5].

Another source of scattering (over and above that from native phonons) originates from phonons within the substrate on which the tube resides, and this has received detailed attention only recently, when it was realized that transport in nanoscale channels may be quite sensitive to the substrate type [68-70]. Many of these substrates are *polar* in nature, so that the lattice vibrations of the substrate form oscillating bond dipoles that generate polarization fields that die out exponentially from the surface [71]. In carbon-based materials, such as CNs and graphene, the distance between the atomically thin channel and the substrate is within 0.35 nm (van-der Waals distance), and the polarization fields can interact quite strongly with the channel electrons, leading to the so-called "substrate polar phonon (SPP) scattering" [68]. It has been argued that beyond a temperature of T = 100 K, SPP scattering is activated and dominates the electronic transport, causing the dc mobility to degrade by a factor of 10 [6].

Regarding the radio-frequency (RF) behavior, recent studies have shown that CNFETs have the potential to outperform other high-frequency transistors, even in the presence of significant scattering from native acoustic and optical phonons [49]. However, this outcome should be revisited, given that SPP scattering has been flagged as an additional major contributor to device degradation at ambient temperatures [6]. As far as we know, a detailed investigation of the importance of SPP scattering on the RF performance of CNFETs has yet to be performed, either theoretically or experimentally.

In this work, we investigate the impact of substrate polar phonon scattering on the high-frequency characteristics of array-based, carbon-nanotube transistors at

varying tube pitches. The RF performance is compared for four substrate materials: AlN, SiO_2, HfO_2, and ZrO_2. We show that SPP scattering substantially affects the device's RF performance and, most importantly, displays a strong dependence on the choice of substrate. Among the materials studied, it is demonstrated that AlN degrades the device's RF performance the least, which can be attributed to the high energy of the phonons in AlN in comparison to the other substrates. This result is in accord with the qualitative expectation of Perebeinos *et al*. [6], who commented on the possible impact of different polar substrates on the dc mobility; in this work, we focus on the RF performance, and we provide detailed quantitative results.

The simulation approach that allows us to extract the relevant RF figures of merit in the presence of various scattering sources consists of two steps. First, a solver for the time-dependent Boltzmann transport equation (BTE) self-consistently coupled to the Poisson equation, developed in [67] using COMSOL [1], is employed to simulate the RF behavior of an *intrinsic* single-tube device at various widths (pitches) of the tube block. The collision integrals for the SPP scattering, based on the formalism of [6], are developed and included in the BTE along with the collision integrals for the native acoustic and optical phonons. Second, the external parasitic resistances and capacitances are combined with the results of the intrinsic single-tube block as described in [49] to establish the *extrinsic* *y*-parameters of the array-based structure; these allow the extraction of key extrinsic high-frequency figures of merit, including the unity-current-gain frequency f_T, the unity-power-gain frequency $f_{\max}$, the unilateral power gain U, the maximum stable gain (MSG), and the maximum available gain (MAG) [72].

This chapter is organized as follows. Section 3.2 briefly summarizes the BTE-Poisson approach used to determine the RF transport characteristics of a single-tube block, including the scattering due to native acoustic, native optical, and substrate phonons; the details on the respective collision integrals are provided in the Appendix. We then proceed to determine the intrinsic high-frequency characteristics of a single-tube CNFET and to examine how these characteristics are impacted by the choice of substrate. In Section 3.3, a complete lumped circuit model is developed for an overall array-based CNFET, including all parasitic elements, and we discuss how SPP scattering from different substrates affects the transistor's extrinsic RF figures of merit at various tube densities. The results of our investigation are summarized in Section 3.4.

3.2 Intrinsic Device

3.2.1 BTE-Poisson Approach

Figure 3.1 shows the base geometry of the single-tube transistor used in the simulations, which consists of a semi-conducting (16, 0) zigzag carbon nanotube with a tube diameter $d = 1.25$ nm and a bandgap of $E_g = 0.55$ eV. The 50-nm-long source and drain contact regions are doped n-type, each at a level of 10^9 m^{-1}. The nanotube channel region underneath the gate is undoped and has a length of 20 nm in correspondence to the ITRS requirements for the year 2015 for RF CMOS millimeter-wave (10-100 GHz) technology [7]; HfO_2 is used as the gate oxide with an effective thickness (from the gate metal to the surface of the tube) of ~2 nm; the

gate metal has a thickness of 50 nm and a work function of 4.5 eV, a typical value for W or Cr; the tube resides on a 100-nm thick substrate consisting of different polar materials, as detailed further below. Varying the width of the base structure in Figure 3.1 is equivalent to varying the tube pitch of an array containing the block, as explained in Section 3.3.1. Block widths and hence array pitches of 10 and 100 nm correspond to 100 and 10 tubes, respectively, across a 1-μm-wide, array-based device, which we use for demonstration purposes in this work.

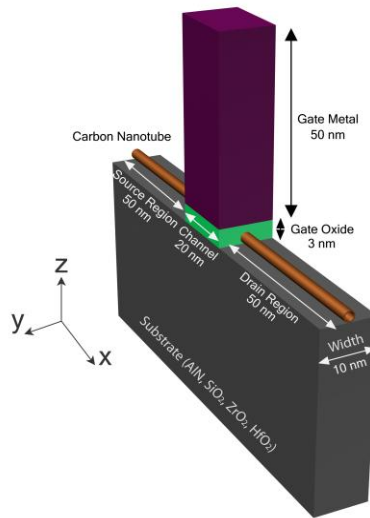


Figure 3.1. Base geometry of the intrinsic single-tube CNFET structure used for the simulation

The electron dynamics within the tube is described by a one-dimensional (1D) BTE for the time-dependent distribution function $f = f(x, k, t)$ for electrons within the lowest conduction subband

$$\frac{\partial f}{\partial t} + v(k)\frac{\partial f}{\partial x} - \frac{q\mathcal{E}_x}{\hbar}\frac{\partial f}{\partial k} = S_o f \tag{3.1}$$

where the symbols are as follows: t is the time; k is the electron wave vector; $v(k)$ is the band velocity for an electron in a state k, obtained from $v(k) = (1/\hbar)[dE(k)/dk]$, with $E(k)$ being the band dispersion, given by [4, Eq. (6.1.12)]; $\mathcal{E}_x = \mathcal{E}_x(x,t)$ is the time-dependent, circumferentially averaged x-component of the electric field along the tube surface; q is the magnitude of the electronic charge; $\hbar$ is the reduced Planck constant; and $S_o f$ denotes the total collision integral comprising all electronic scattering contributions, including acoustic $(S_o f)_{\mathrm{ac}}$, optical $(S_o f)_{\mathrm{op}}$, and surface polar phonons $(S_o f)_{\mathrm{spp},\nu}$, as specified in the Appendix.

As detailed in the Appendix, we follow the "unscreened" approach of Perebeinos *et al.* [6] in formulating the SPP collision integrals, *i.e.*, we ignore the impact of "screening" as modeled by a Thomas-Fermi screening length [73, 74] or as determined by the coupling of the channel charge to the substrate phonons, where the latter leads to so-called "hybridized, interfacial plasmon-phonon (IPP) modes" [75-77]. While such screening may play an important role in 2D/3D materials, its significance in quasi-1D systems such as CNs appears less severe [78]. We also note that our results with regard to RF performance degradation in tubes (*e.g.*, see Figs. 2 and 3) are generally consistent in terms of the ordering of substrates (*e.g.*, AlN and SiO_2 vs. HfO_2 and ZrO_2) with those found from a Thomas-Fermi approach for the dc mobility in 2D graphene [74, Figure 5], and that a Thomas-Fermi

approach is generally in agreement with a fully coupled IPP approach [76, Figure 3]. Thus, while it must be noted that the precise quantitative nature of our results may be impacted by a more detailed consideration of screening, for a first assessment of the relative importance of various substrates on the RF performance, we use the unscreened approach of [6].

To proceed, we make the usual assumption that a sinusoidal perturbation at a radian frequency ω is applied at one of the terminals, thus inducing a perturbation to the distribution function f and the electric field $\mathcal{E}_x$ along the transport direction. For small signal amplitudes, both f and $\mathcal{E}_x$ can be split into their dc and ac components via $f = \bar{f} + \tilde{f}e^{j\omega t}$ and $\mathcal{E}_x = \overline{\mathcal{E}}_x + \widetilde{\mathcal{E}}_x e^{j\omega t}$, respectively, where here and elsewhere, we use the "$\bar{\ }$" and "$\tilde{\ }$" notation to distinguish between the dc and ac parts. Inserting this ansatz into (3.1), a BTE for each component can be derived following the approach of [67]:

$$v\frac{\partial \bar{f}}{\partial x} - \frac{q\overline{\mathcal{E}}_x}{\hbar}\frac{\partial \bar{f}}{\partial k} = S_o\bar{f} \tag{3.2}$$

$$j\omega\tilde{f} + v\frac{\partial \tilde{f}}{\partial x} - \frac{q\overline{\mathcal{E}}_x}{\hbar}\frac{\partial \tilde{f}}{\partial k} - \frac{q\widetilde{\mathcal{E}}_x}{\hbar}\frac{\partial \bar{f}}{\partial k} = S_o\{\tilde{f}, \bar{f}\} \tag{3.3}$$

where $S_o\bar{f}$ and $S_o\{\tilde{f}, \bar{f}\}$ refer to the total collision integrals applicable to the dc and ac equations, given by (A.2) – (A.4) with $f \rightarrow \bar{f}$ and by (B.1) – (B.3) in the Appendix, respectively.

The BTEs in (3.2) and (3.3) are solved self-consistently along with their corresponding Poisson equation, as described in [67]. All simulations are performed at room temperature $T = 300$ K.

3.2.2 Results for the Single Tube

A DC Output Current

We begin our analysis by studying the effect of SPP scattering on the dc output characteristics of a single-tube block (Figure 3.1) with a width of 10 nm and a supply voltage of $V_D = V_{\mathrm{DD}} = 0.9$ V, where the latter is consistent with the ITRS requirements for the year 2015 for RF CMOS millimeter-wave (10-100 GHz) technology [7].

Figure 3.2 shows the dc drain current as a function of the gate bias V_G for a CNFET situated on different substrates. The curve labeled "Ballistic" is with all scattering turned off and the curve labeled "Native Phonons" includes scattering only from native acoustic and optical phonons; in each of these cases, an SiO_2 substrate is assumed solely for the purpose of solving Poisson's equation (but with no contribution to the scattering). The remaining curves in Figure 3.2 are for the indicated substrates with *all* scattering, *i.e.*, native *and* SPP, turned on. Starting with the ballistic limit, one observes a monotonic increase of the drain current with increasing gate bias. The inclusion of native acoustic and optical phonons leads to the expected reduction of current, which becomes pronounced at large V_G. The additional inclusion of SPP scattering causes a further reduction of the current, as

depicted in Figure 3.2, which for SiO_2 is about a factor of two at $V_G = 0.9$ V. What is striking is the strong dependence of the current on the type of substrate, as Figure 3.2 demonstrates. Among all the substrates considered, AlN seems to play a distinguished role, since SPP scattering with AlN does not affect the device performance as severely as with the other substrates; for AlN, the drain current I_D is only slightly below that in the presence of only native phonon scattering.

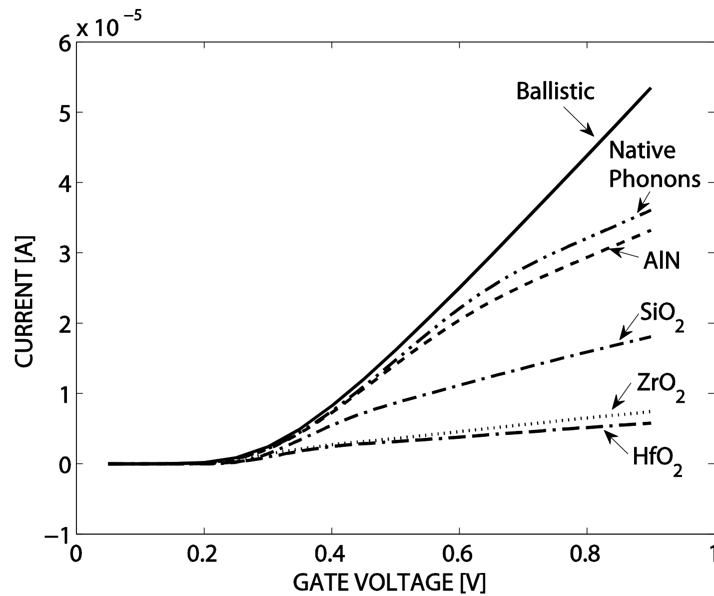


Figure 3.2. Static (dc) drain current I_D vs. gate bias voltage V_G for a single-tube block (shown in Figure 3.1) with a tube pitch of 10 nm in the presence of SPP scattering on different substrates. The dc drain voltage V_D is held at 0.9 V. The curve labeled "Ballistic" is with all scattering turned off and the curve labeled "Native Phonons" includes scattering only from native acoustic and optical phonons; in each of these cases, an SiO_2 substrate is assumed solely for the purpose of solving Poisson's equation (and with no contribution to the scattering). The remaining curves are for the indicated substrates with *all* scattering (native *and* SPP) turned on.

The strong variation of the drain current with a change in substrate can be explained *in part* by the relative strengths of SPP scattering for the first ($\nu = 1$)

substrate phonon mode, as indicated by the prefactor of the SPP collision integral specified in (A.4) of the Appendix; the role of the second ($\nu = 2$) substrate phonon mode, which corresponds to relatively higher values of phonon energy, requires a more careful consideration, as will be discussed shortly.

The prefactor $\mathrm{PF}_{\mathrm{spp},\nu}$ of the SPP collision integral in (A.4) can itself be written as a product of three factors:

$$\mathrm{PF}_{\mathrm{spp},\nu} = F^2_{\mathrm{spp},\nu} \times P_{\mathrm{spp},\nu} \times \left(N_{\mathrm{spp},\nu} + 1\right) \tag{3.4}$$

where the symbols are as follows: $F^2_{\mathrm{spp},\nu}$ (known as the Fröhlich coupling) specifies the coupling strength between the electrons and the substrate phonons; $P_{\mathrm{spp},\nu}$ is indicative of the strength of the polarization field due to the substrate phonons at a distance s away from the substrate, with a strong inverse dependence on the energy $\hbar\omega_{\mathrm{spp},\nu}$ of the phonon mode, as indicated by an inspection of (A.5) and (A.6) in the Appendix; and $N_{\mathrm{spp},\nu} = 1/(e^{\hbar\omega_{\mathrm{spp},\nu}/k_B T} - 1)$ is the number of substrate phonons, which also exhibits a strong inverse dependence on the energy $\hbar\omega_{\mathrm{spp},\nu}$ of the phonon mode. As Table 3.1 reveals, the $\nu = 1$ modes a have a similar electron-phonon coupling factor, $F^2_{\mathrm{spp},1} \sim 0.5 - 0.75$ V^2m, while the phonon energies $\hbar\omega_{\mathrm{spp},1}$ are well-ordered, from about 90 meV for AlN, 60 meV for SiO_2, and 28 and 22 meV for ZrO_2 and HfO_2, respectively. As the values in Table 3.1 show, the differences in these energies dominates the collision-integral prefactor $PF_{\mathrm{spp},1}$, through the impact on the polarization factor $P_{\mathrm{spp},1}$ and phonon

number $\left(N_{\text{spp},1} + 1\right)$, and this is especially true for ZrO_2 and HfO_2. Overall, and at least for the first phonon mode, the low-energy SPP phonons in ZrO_2 and HfO_2, and to a lesser extent SiO_2, can thus be expected to cause increased SPP scattering and hence a more severe degradation in the current, as demonstrated in Figure 3.2.

For the second phonon mode, the situation is more involved, as one needs to consider more than simply the prefactor in the collision integral (A.4) in determining the relative strengths of the scattering. The higher phonon energies of the second mode suggest that the corresponding SPP scattering will become important only at higher gate biases, when the channel band profile has been "pushed down" sufficiently below the source Fermi level for carriers injected from the source to have available states to occupy in the channel and in the channel-to-drain barrier region upon emitting the higher ($\nu = 2$) energy SPP phonons; mathematically, the requirement for available states is taken into account by the presence of the density of states and distribution function in the collision integral (A.4). Again, in this case, ZrO_2 and HfO_2 can be expected to suffer from more severe SPP scattering, *i.e.*, scattering that becomes important at a lower gate voltage, since $\hbar\omega_{\text{spp},2}$ is significantly lower for these materials, thus requiring less gate voltage to push the bands down a sufficient amount. The curves in Figure 3.2 are indeed consistent with this expectation.

Overall, considering both substrate phonon modes, the important conclusion is that the lower substrate phonon energies ($\hbar\omega_{\text{spp},1}$ and $\hbar\omega_{\text{spp},2}$ in Table 3.1) in ZrO_2 and HfO_2, and to a lesser extent SiO_2, lead to a greater degradation of the current than in AlN.

TABLE 3.1. PARAMETERS FOR THE SPP SCATTERING IN CARBON NANOTUBES ON VARIOUS SUBSTRATES [79]

	AlN	**SiO_2**	**ZrO_2**	**HfO_2**
$\boldsymbol{\epsilon^0}$	9.14	3.90	24.0	22.0
$\boldsymbol{\epsilon^\infty}$	4.80	2.50	4.00	5.03
$\boldsymbol{\hbar\omega_{\mathrm{spp},1}}$ **[meV]**	90	61	28	22
$\boldsymbol{F^2_{\mathrm{spp},1}} \times 10^{-11}$ **[V²m]**	0.55	0.75	0.60	0.55
$\boldsymbol{P_{\mathrm{spp},1}} \times 10^{9}$ **[eV/V² s]**	0.67	0.97	1.95	2.44
$\boldsymbol{N_{\mathrm{spp},1}} + 1$	1.03	1.09	1.48	1.73
$\mathbf{PF_{spp,1}} \times 10^{6}$ **[eVm/s]**	0.21	0.44	0.96	1.29
$\boldsymbol{\hbar\omega_{\mathrm{spp},2}}$ **[meV]**	106	149	76	54
$\boldsymbol{P_{\mathrm{spp},2}} \times 10^{9}$ **[eV/V² s]**	0.57	0.40	0.79	1.09
$\boldsymbol{F^2_{\mathrm{spp},2}} \times 10^{-11}$ **[V²m]**	1.61	1.66	1.88	0.53
$\boldsymbol{N_{\mathrm{spp},2}} + 1$	1.01	1.00	1.05	1.13
$\mathbf{PF_{spp,2}} \times 10^{6}$ **[eVm/s]**	0.51	0.36	0.86	0.36
Peak f_T [THz] 10 nm - 100 nm	2.75 - 1.90	1.74 - 1.27	0.73 - 0.57	0.71 - 0.53
V_G for peak f_T [V] 10 nm - 100 nm	0.46 - 0.46	0.34 - 0.35	0.25 - 0.28	0.29- 0.31

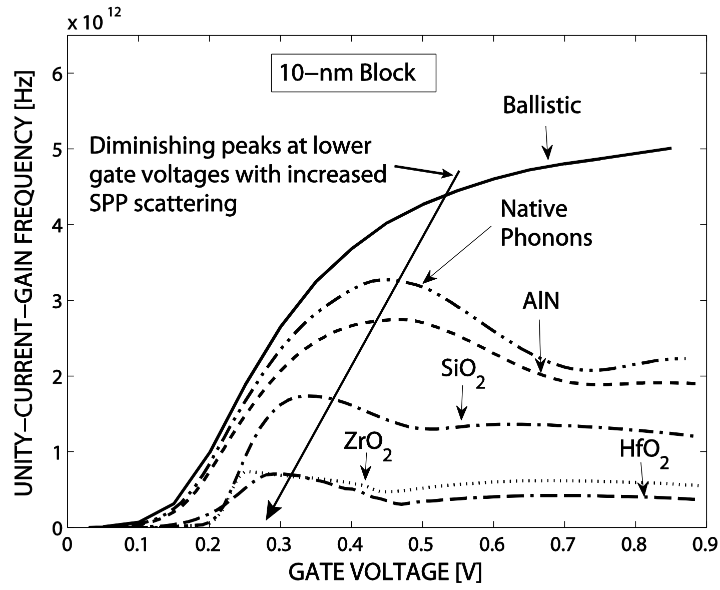
B Cutoff Frequency

B.1 Result

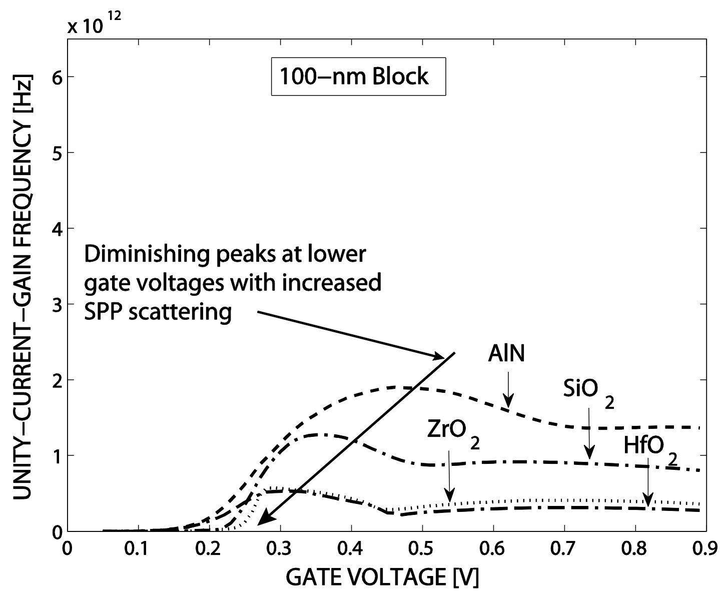
Over and above the dc current-voltage behavior, SPP scattering also has a strong impact on the device's RF figures of merit, such as the unity-current-gain, cut-off frequency $f_T = g_m/(2\pi C_{\mathrm{gg}})$, where g_m is the transconductance and C_{gg} is the total gate capacitance.

Figure 3.3 shows the f_T vs. gate bias V_G for the same cases (and with the same labeling convention) as in Figure 3.2, but for two different single-tube block widths of 10 nm and 100 nm. First focusing on the curves for a 10-nm block width [part (a) of Figure 3.3], a distinct pattern in the behavior of f_T vs. V_G is evident. In the

ballistic case, the f_T exhibits a monotonic increase with V_G, before saturating at high values of V_G. In the presence of native acoustic and optical phonon scattering, the f_T still increases monotonically with V_G at low gate bias, but then exhibits a clear peak before falling off and saturating at high gate bias. In the additional presence of SPP scattering, the peaking behavior is still present, but the peaks diminish and shift to lower gate voltages in an order corresponding to the expected strength of the SPP scattering as already discussed in conjunction with Figure 3.2, i.e., starting from AlN (having high SPP phonon energies), followed by SiO_2 (with intermediate SPP phonon energies), and finally by ZrO_2 and HfO_2 (having low SPP phonon energies). In the case of the 100-nm block width [part (b) of Figure 3.3], the behavior is similar, but the f_T values are lower, a result that will be explained in Section 3.3.2. For reference, the peak f_T values and the gate bias V_G at which they occur for the different substrates are listed in Table 3.1.



(a)



(b)

Figure 3.3. Intrinsic unity-current-gain cutoff frequency f_T vs. gate voltage V_G for the single-tube CNFET block shown in Figure 3.1 and for block widths of (a) 10 nm and (b) 100 nm. The labeling convention for the curves is the same as that described in the caption of Figure 3.2.

B.2 Discussion

One can understand the behavior of the f_T in Figure 3.3 by considering separately the behavior of the g_m and C_{gg}, both of which are plotted vs. gate bias V_G in Figure 3.4; we restrict our attention in this discussion to the case of a 10-nm block width.

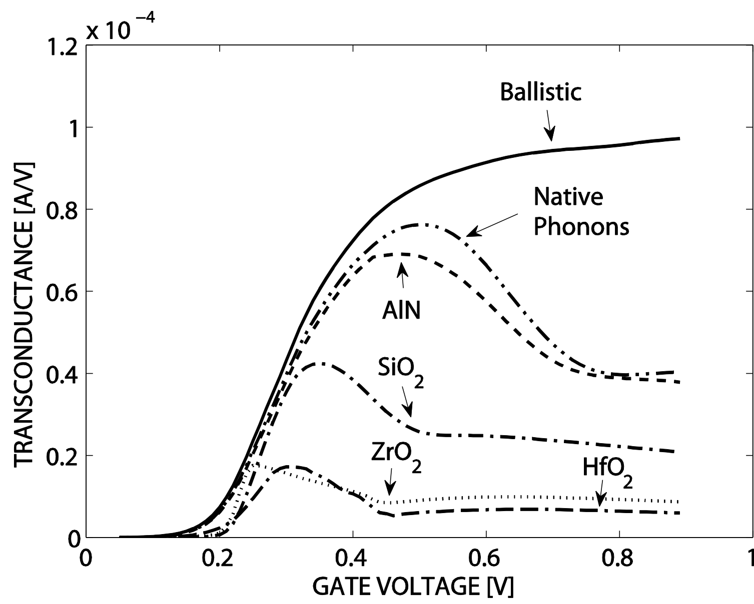
The transconductance $g_m = \partial I_D/\partial V_G$ in Figure 3.4(a) represents the local slope of the I_D-V_G characteristics, and hence peaks at the point of inflection of the corresponding I_D-V_G curve in Figure 3.2. Based on the earlier discussion of the I_D-V_G curves, the peaks in g_m are hence diminished and occur at successively lower gate voltages with increased scattering, and this behavior is primarily responsible for the peaking pattern of the f_T in Figure 3.3. However, C_{gg} also plays a role, accentuating the peaking in f_T by itself peaking, as shown in Figure 3.4(b). These peaks in C_{gg} are worth a more careful consideration.

Figure 3.5(a) shows a plot of the energy-resolved dc charge distribution $\bar{f}(E)$ for the case of SiO_2 as the substrate; we choose a bias point of $V_G = 0.4$ V for illustration, and we choose to plot the distribution at the point $x = 80$ nm (see Figure 3.1), which is just past the channel region and where the effects of scattering (when present) are clearly visible. The plot shows $\bar{f}(E)$ for three cases: (i) ballistic transport, (ii) with only native acoustic and optical phonon scattering, and (iii) with all scattering, *i.e.*, by additionally including SPP phonons. In the ballistic case, $\bar{f}(E)$ shows the expected peak for energies above the top of the barrier and below the source Fermi level ($E_{\text{top}} \lesssim E \lesssim \mu_S$), representing source-injected electrons that can cross the channel barrier; $\bar{f}(E)$ is otherwise zero except for energies around the

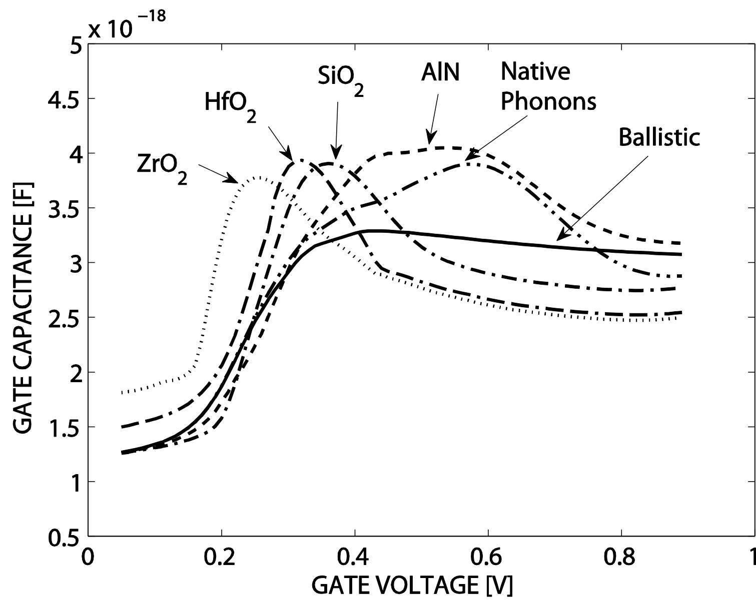
drain Fermi level ($E \sim \mu_D$), representing drain-injected electrons. In the presence of only native acoustic and optical phonon scattering, the distribution function takes a similar shape, except now we see "sidebands" of occupancy facilitated by optical phonon emissions; carriers are moved down from the ballistic peak in bands separated by the optical phonon energy of $\hbar\omega_{\mathrm{op}} = 200$ meV [59]. With the addition of substrate polar phonons, the signature of the native optical phonon sidebands is "washed out" because the substrate polar phonons allow transitions over additional energies (specified by the values of $\hbar\omega_{\mathrm{spp,1}}$ and $\hbar\omega_{\mathrm{spp,2}}$ in Table 3.1). The important point to note is that *scattering moves carriers from the ballistic peak down to lower energies*. As a result, with scattering, a perturbation in gate voltage ΔV_G leads to a modulation of the distribution function $\Delta\bar{f}(E)$ at energies below the ballistic peak, as shown in Figure 3.5(b); overall, the additional modulation at lower energies more than compensates the decreased modulation in the ballistic peak, such that the integrated charge modulation $\Delta Q_G \approx q\int_{\mu_D}^{\mu_S}\left[\Delta\bar{f}(E) \times D(E)\right] dE$ is higher in the presence of increased scattering,[5] which will tend to increase the capacitance $C_{\mathrm{gg}} \sim \Delta Q_G/\Delta V_G$ seen looking into the gate. The effect is most pronounced at the gate bias corresponding to the onset of scattering, since the lower energy states fill up at higher gate biases, thereby causing increased Pauli blocking and hence a reduced modulation; this is why the capacitance peaks at the onset of scattering in Figure 3.4(b).

---

[5] Here, we have used $\Delta\bar{f}(E) \approx 0$ outside the interval of integration and the fact that $D(E)$ [given by (A.1) in the Appendix] decreases monotonically with energy.

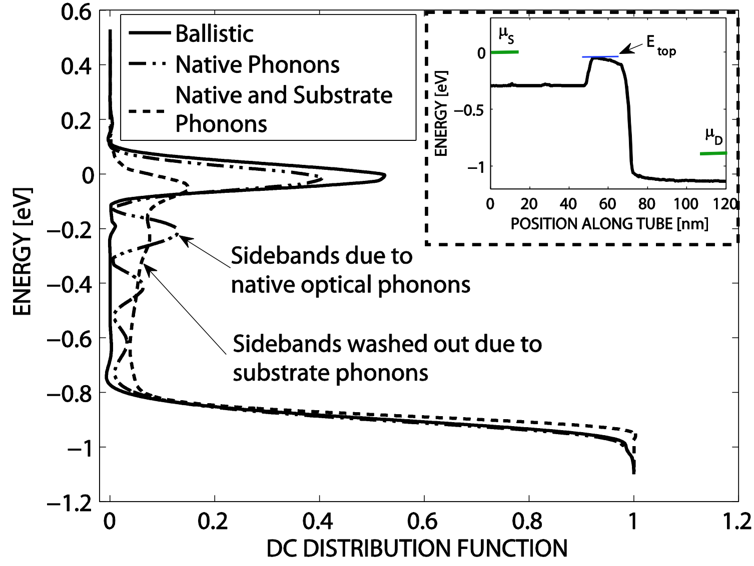


(a)

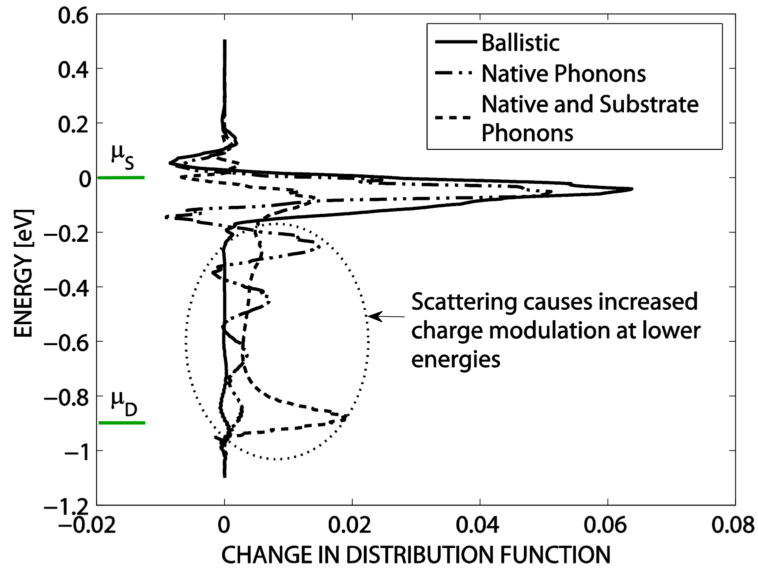


(b)

Figure 3.4. (a) Transconductance g_m and (b) gate capacitance C_{gg} vs. dc gate voltage V_G of a single-tube CNFET block in Figure 3.1 with a width of 10 nm. The labeling convention for the curves is the same as that described in the caption of Figure 3.1.



(a)



(b)

Figure 3.5. (a) Energy-resolved dc distribution function $\bar{f}(E)$ at $x = 80$ nm (see Figure 3.1) and for $V_G = 0.4$ V, with SiO_2 used as the substrate and shown for three cases: (i) ballistic transport; (ii) only native acoustic and optical phonon scattering; (iii) all scattering, *i.e.*, additionally including substrate phonons. The inset to the figure shows a sketch of the conduction-band profile $E_c(x)$ vs. x, which is provided for reference. (b) Energy-resolved change in the distribution function $\Delta\bar{f}(E)$ at $x = 80$ nm and $V_G = 0.4$ V, where $\Delta\bar{f}(E)$ arises from an incremental change in gate voltage ΔV_G (equal to 10 mV).

C y-parameters

We conclude our analysis of the RF behavior of the single-tube block by extracting its frequency-dependent y-parameters from our self-consistent BTE-Poisson solution, as described in [67]. Figure 3.6 shows the magnitude of the transistor's intrinsic forward y-parameters, y_{11} and y_{21}, for the single-tube block of Figure 3.1 having a width of 10 nm; results are shown for a ballistic device and a device with all scattering, *i.e.*, due to native acoustic and optical phonons as well as substrate phonons. The ballistic device is again assumed to lie on a SiO_2 substrate, solely for the purpose of solving Poisson's equation. For the devices with all scattering included, the focus is on the choice of AlN and HfO_2 as substrates, *i.e.*, the materials where the signatures of scattering are the weakest and strongest, respectively. For the ballistic case, the operating bias is chosen to be equal to $V_G = V_{\mathrm{DD}}/2$ as in [67], and for the cases with scattering, V_G is chosen to be the gate voltage corresponding to peak f_T, which is 0.48 V for AlN and 0.28 V for HfO_2; in all cases, the drain voltage is held fixed at 0.9 V.

As Figure 3.6 reveals, scattering lowers the forward y-parameters for both AlN and HfO_2 over the entire frequency range. Despite the strong impact of the SPP phonons, one still observes a resonant behavior at about the same frequency of approximately 18 THz. This resonant behavior has been discussed in [54, 67, 80-82]. Compared to the ballistic case, the resonances are not only lowered but also strongly broadened due to scattering, which is equivalent to a reduction in the quality (Q) factor. As Figure 3.6 shows, the degradation in the y-parameters is

more severe for HfO_2 than for AlN since, as discussed earlier, SPP scattering has a stronger impact (due to the lower phonon energies) in HfO_2 than AlN.

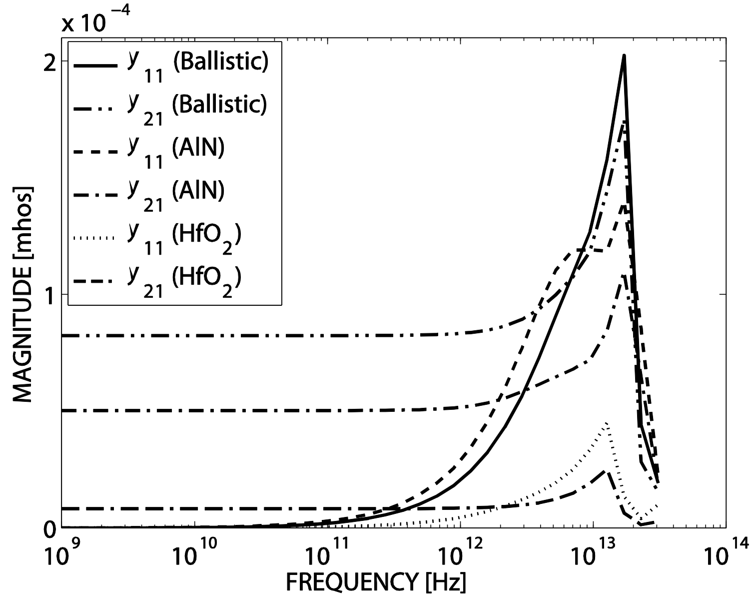


Figure 3.6. Magnitude of the transistor's forward y-parameters, y_{11} and y_{21}, comparing the ballistic limit with the case of SPP scattering for the two substrates AlN and HfO_2, where the SPP scattering is the weakest and strongest, respectively.

3.3 Extrinsic Device

3.3.1 Approach

A Array-based Structure

A schematic of the multi-tube CNFET structure used in this work is shown in Figure 3.7, and it consists of a parallel arrangement of the single-tube blocks from Figure 3.1; the total width of the device in the y-direction, *i.e.*, the gate width, is fixed to be $W_g = 1$ μm, and the distance between the tubes, *i.e.*, the tube pitch p, is varied between 10 and 100 nm, corresponding to a total of 100 and 10 tubes, respectively.

Here, it is important to note that the pitch p of the array in Figure 3.7 is equivalent to the width of a single-tube block in Figure 3.1; this follows since the block of Figure 3.1 can represent a portion of an array when the appropriate boundary conditions are applied at the edges (along the tube length) of the block [83], which is what we have done throughout this work. The *intrinsic* y-parameters of the multi-tube array, *i.e.*, *excluding parasitics*, can be found by a scaling of the intrinsic y-parameters of a single-tube block; for our 1-μm wide transistor, we can write [49]

$$\left[y_{\mathrm{m,int}}\right]_p = \left[y_{\mathrm{s,int}}\right]_p \times \left(\frac{1000\ [\mathrm{nm}]}{p\ [\mathrm{nm}]}\right) \tag{3.5}$$

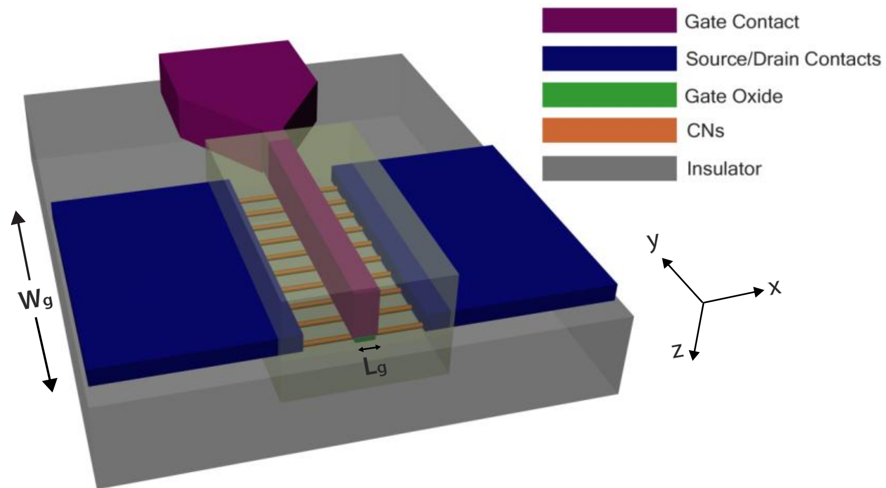


Figure 3.7. Schematic of a top-gated, aligned, array-based CNFET structure similar to [49]. The gate length and width of the device are $L_g = 20$ nm and $W_g = 1$ μm, respectively. The yellow box shows the portion of the structure that is studied. The figure is not drawn to scale.

where $[y_{\mathrm{m,int}}]_p$ and $[y_{\mathrm{s,int}}]_p$ denote the y-parameter matrices of the multi- and single-tube structures, respectively; both matrices are functions of the tube pitch p.

B Modeling of Parasitics

A major cause for the overall performance degradation of the device is the impact of the extrinsic capacitances and resistances originating from the contacts, which makes it necessary to incorporate these into a complete device model, as shown in Figure 3.8. To determine the values of these external elements, we follow the procedure explained in [49], utilizing COMSOL [1]; for the details, including the choice of materials and dimensions of the contacts, the reader is referred to [49], and here we simply list the extracted values in Table 3.2.

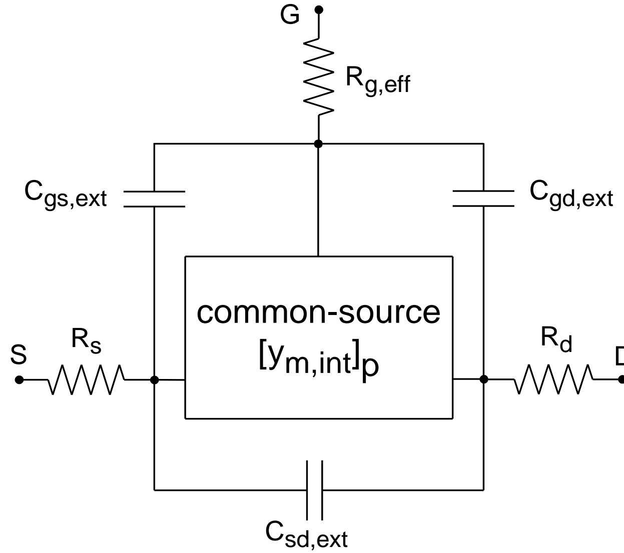


Figure 3.8. Extrinsic lumped-element model of the array-based CNFET structure [49]. The inner block represents the intrinsic multi-tube device described by the y-parameter matrix $[y_{\mathrm{m,int}}]_p$ specified by (3.5).

TABLE 3.2. VALUES FOR THE PARASITIC RESISTANCES AND CAPACITANCES USED IN THE LUMPED ELEMENT MODEL OF FIGURE 3.8 FOR THE DIFFERENT SUBSTRATES. IN ALL CASES, THE TOTAL WIDTH OF THE ARRAY-BASED CNFET IS 1 MICRON.

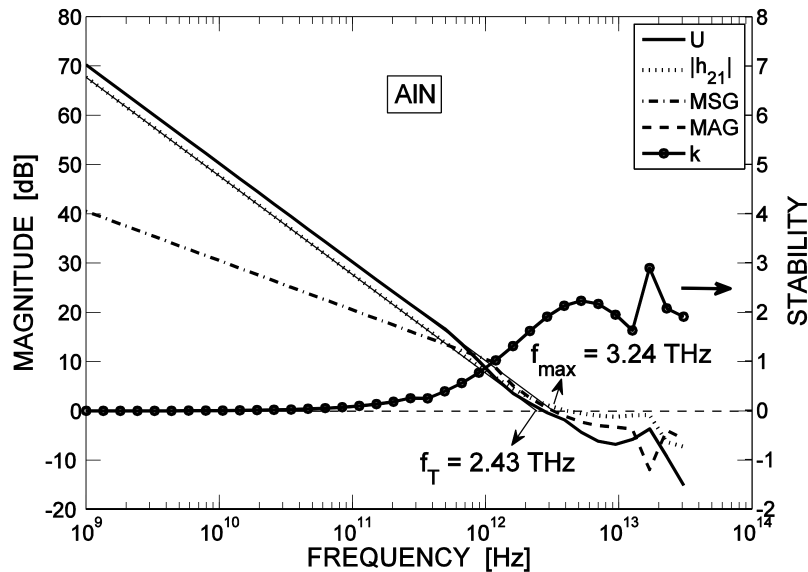
	AlN	**SiO_2**	**ZrO_2**	**HfO_2**
$R_{g,eff}$ (Ω) (Tungsten)	67	67	67	67
$R_{s/d}$ (Ω) (Yttrium)	1.77	1.77	1.77	1.77
$C_{gs,ext}$ (aF)	34	22	70	51
$C_{gd,ext}$ (aF)	32	21	66	48
$C_{sd,ext}$ (aF)	9.7	4.7	29	18

3.3.2 Results and Discussion

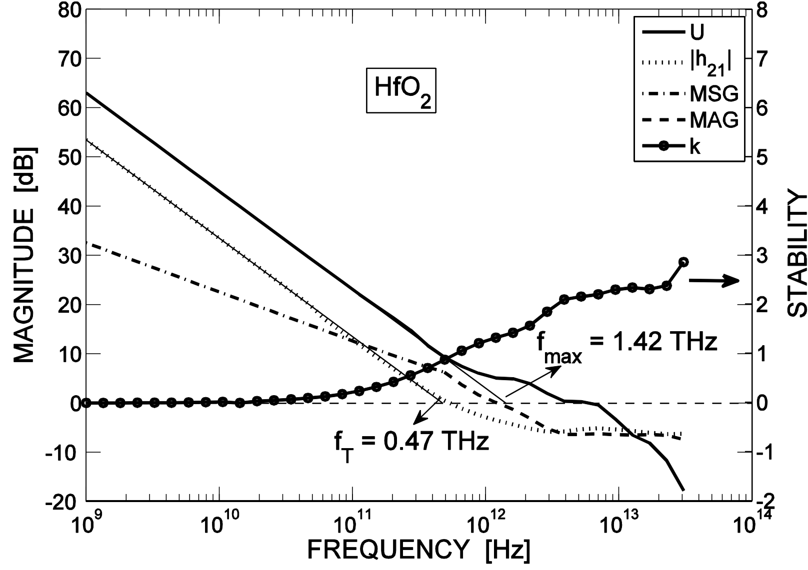
A RF Figures of Merit

Figure 3.9 shows the extrinsic common-source current gain $|h_{21}| = |y_{21}/y_{11}|$ vs. frequency and the extrinsic cutoff frequency f_T for a 1-μm-wide, array-based structure with a pitch $p = 10$ nm; here and for all the extrinsic results, the operating bias is chosen to be the gate voltage corresponding to peak intrinsic f_T with the drain voltage held fixed at 0.9 V. Notably, despite the presence of SPP scattering, CNFETs with AlN substrates still offer the potential for an extrinsic f_T of about 2.43 THz, and thus remain ahead of the present records for RF transistors reported in the literature, including graphene transistors with 300 GHz [84], silicon MOSFETs with 485 GHz [85], III-V high-electron-mobility transistors with 644 GHz [86], and heterojunction bipolar transistors with 710 GHz [87]. However, this is no longer the case for a substrate leading to strong SPP scattering, such as HfO_2, where the extrinsic f_T falls to 470 GHz, comparable to silicon MOSFETs.

Other RF measures to characterize the device are also shown in Figure 3.9, such as the unilateral power gain U, the maximum stable gain (MSG), the maximum available gain (MAG), and the Kurokawa stability factor **k** [55]; all figures of merit are calculated based on well-known expressions [55] using the *y*-parameters extracted from the extrinsic device model shown in Figure 3.8. Similar to [67], the MSG is plotted for those frequencies where the transistor is conditionally stable ($\mathbf{k} < 1$) and is replaced by the MAG when the device becomes unconditionally stable ($\mathbf{k} \geq 1$) [55]. Extrapolating the power-gain U to unity, the $f_{\max}$ can be extracted to be 3.24 THz for AlN and a lower value of 1.42 THz for HfO_2, where the SPP scattering is stronger.



(a)



(b)

Figure 3.9. Magnitude of the common-source current gain $|h_{21}|$, unilateral power gain U, maximum stable gain (MSG), maximum available gain (MAG), and the stability factor **k** as function of frequency for the CNFET of Figure 3.7 with a tube pitch $p = 10$ nm and with (a) AlN and (b) HfO_2 as substrates. The extrapolated f_T and $f_{\max}$ are also shown.

B Effect of Pitch on f_T and f_{max}

B.1 Unity-Current-Gain Frequency f_T

Figure 3.10 shows the extrinsic f_T as a function of the tube pitch in the presence of native and SPP scattering from the various substrates. As shown, the f_T can be improved by a factor of about 2.5 to 3.5 by decreasing the pitch from 100 nm to 10 nm. One can understand the improvement by considering that the pitch dependency of the extrinsic f_T is determined by the pitch dependency of the intrinsic transconductance and gate capacitance of the multi-tube system (neglecting parasitics), both of which are given by the single-tube values multiplied by the

number of tubes N: $g_{\mathrm{m,int}}^{m}\big|_p = N \times g_{\mathrm{m,int}}^{s}\big|_p$ and $C_{\mathrm{gg,int}}^{m}\big|_p = N \times C_{\mathrm{gg,int}}^{s}\big|_p$, where the superscripts m and s refer to "multi-tube" and "single-tube," respectively, and where the notation " $|_p$" is used to emphasize that the corresponding quantities depend on the tube pitch p. To first order, we can estimate the pitch dependency by writing the extrinsic $f_T \sim g_{\mathrm{m,int}}^{m}\big|_p / C_{\mathrm{gg,int}}^{m}\big|_p = g_{\mathrm{m,int}}^{s}\big|_p / C_{\mathrm{gg,int}}^{s}\big|_p$. With a decrease in pitch (higher tube density), both $g_{\mathrm{m,int}}^{s}\big|_p$ and $C_{\mathrm{gg,int}}^{s}\big|_p$ decrease due to tube screening [83]; the important point is that $C_{\mathrm{gg,int}}^{s}\big|_p$ decreases by 30-35% (depending on the substrate) whereas $g_{\mathrm{m,int}}^{s}\big|_p$ reduces by only 3-7% when the pitch is scaled down from 100 nm to 10 nm (not shown). This difference in reduction results in an effective increase of the f_T as reported in [83] and as depicted in Figure 3.10; it also explains why the f_T results for the 10-nm block presented earlier in Figure 3.3(a) exceed those for the 100-nm block in Figure 3.3(b). More importantly, Figure 3.10 shows that the RF performance advantage of an AlN substrate continues to exist for the *extrinsic* cutoff frequency f_T, a consequence of AlN's high-energy SPP phonons causing weaker SPP scattering as compared to HfO_2 and ZrO_2 substrates, as discussed earlier in Section 3.2.

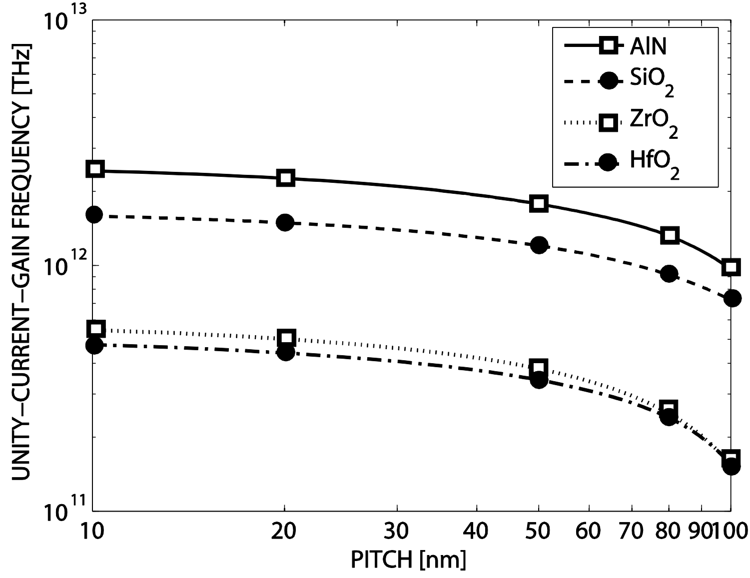


Figure 3.10. Extrapolated extrinsic f_T of the array-based CNFET structure as a function of tube pitch in the presence of native and SPP scattering for different substrates.

B.2 Unity-Power-Gain Frequency $f_{\max}$

Contrary to the f_T, the transistor's $f_{\max}$ is much less susceptible to variations of the tube pitch; for example, it grows by a factor of about 1.2 to 1.5 for the HfO_2 and ZrO_2 substrates and not at all for SiO_2 and AlN when the pitch is decreased from 100 to 10 nm, as can be discerned from Figure 3.11. This behavior of the $f_{\max}$ can be explained with the aid of the approximate expression [56, 57]

$$f_{\max} \approx \sqrt{\frac{f_T}{8\pi R_{\mathrm{g,eff}} C_{\mathrm{gd,eff}}}} \tag{3.6}$$

which has been shown to reproduce the pitch-dependency of the numerically

obtained $f_{\max}$ quite well [49]. Here, $R_{\mathrm{g,eff}}$ is the effective gate resistance (shown in Figure 3.8) and $C_{\mathrm{gd,eff}}$ is the total effective gate-drain capacitance, which can be expressed as [49]

$$C_{\mathrm{gd,eff}} = N \times C^{s}_{\mathrm{gd,int}}\Big|_{p} + C_{\mathrm{gd,ext}} \tag{3.7}$$

where $C^{s}_{\mathrm{gd,int}}\Big|_{p}$ refers to the gate-drain capacitance of the intrinsic single-tube CNFET block of Figure 3.1, and $C_{\mathrm{gd,ext}}$ is the external gate-drain capacitance of the array-based CNFET shown in Figure 3.6 and listed in Table 3.2. Decreasing the pitch corresponds to a larger N, causing the term $N \times C^{s}_{\mathrm{gd,int}}\Big|_{p}$ and thus $C_{\mathrm{gd,eff}}$ to grow, as Figure 3.12 shows; this increase of $C_{\mathrm{gd,eff}}$ compensates the increase of the f_T discussed in conjunction with Figure 3.10, causing the $f_{\max}$ to remain almost unaffected from a variation of the pitch [49].

Another interesting feature of Figure 3.11 is that the $f_{\max}$ for SiO_2 is slightly higher than for AlN at all tube pitches, despite the lower f_T for SiO_2 in Figure 3.10. The reason for the higher $f_{\max}$ of SiO_2-based CNFETs is due to the relative permittivity of SiO_2, which at 3.9 is the lowest among all substrates, as seen from Table 3.1. The low permittivity causes the external parasitic capacitance $C_{\mathrm{gd,ext}}$ for SiO_2 to be the smallest, which translates into the smallest overall $C_{\mathrm{gd,eff}}$, as shown in Figure 3.12; owing to the inverse dependence of the $f_{\max}$ on $C_{\mathrm{gd,eff}}$ according to (3.6), CNFETs with SiO_2 rather than AlN substrates thus have a slightly higher $f_{\max}$

at all tube pitches. In the case of HfO_2 and ZrO_2, which have already been flagged as the worst for RF performance due to strong SPP scattering, the relatively higher values of the dielectric permittivity (see Table 3.1) add to the degradation of $f_{\max}$ due to a larger $C_{\mathrm{gd,eff}}$.

It should be noted that the results here (and throughout this paper) assume room temperature operation. Self-heating has been experimentally observed at small gate lengths, which could potentially increase the operating temperature and impact the device behavior, especially at high drive currents [59, 88]. For simplicity, and for a first assessment, we omit self-heating effects in our study; this assumes good heat removal and is consistent with our assumption [see the Appendix] of including only phonon emission processes.

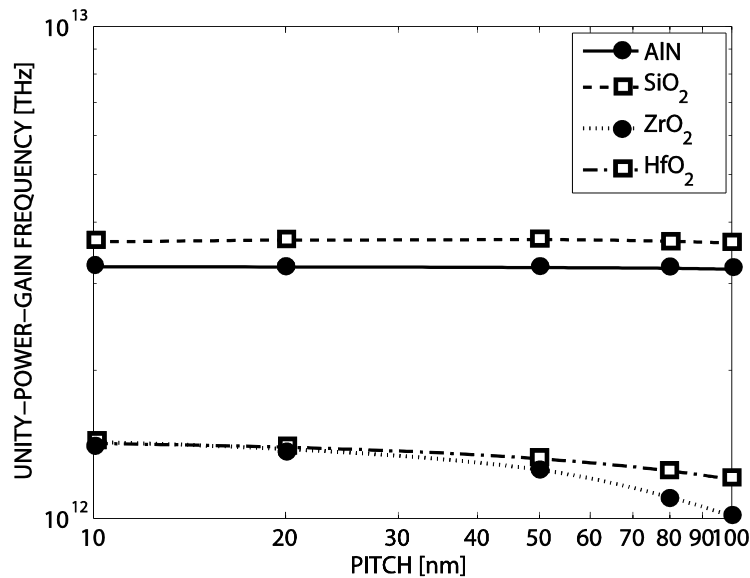


Figure 3.11. $f_{\max}$ of the array-based CNFET structure as function of tube pitch in the presence of native and SPP scattering for different substrates.

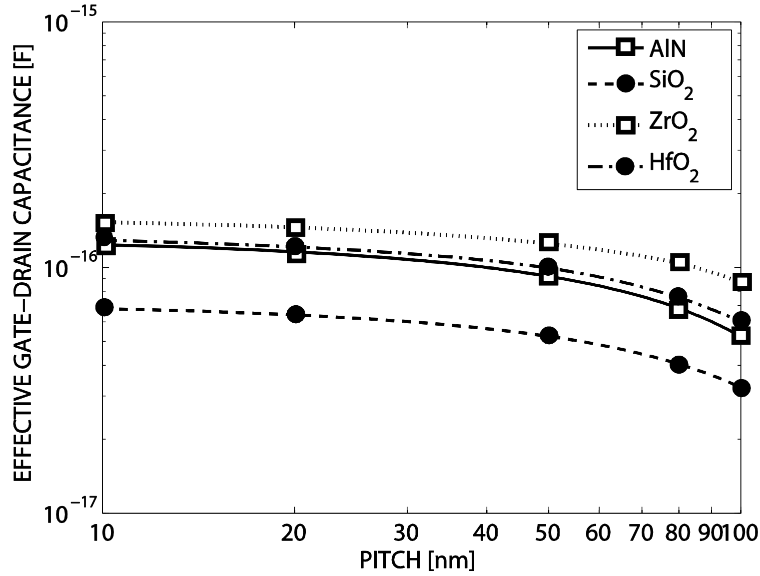


Figure 3.12. Effective gate-drain capacitance $C_{\mathrm{gd,eff}}$ for the array-based CNFET structure as a function of tube pitch in the presence of native and SPP scattering from different substrates.

B.3 Summary

In Table 3.3, we summarize and compare the RF figures of merit for the array-based CNFET with the ITRS RF CMOS millimeter-wave technology requirements for the year 2015 [7]. Note that the reported values for the array-based CNFET are for a tube pitch of 100 nm, and can be improved by increasing the tube density as discussed in the previous sections; included also are the transistor's gain at frequencies of 24, 60, and 94 GHz, since these are of commercial interest [7]. As shown in Table 3.3, despite the presence of SPP scattering (over and above native acoustic and optical phonons), and except for the transconductance g_m (which could be improved with a lower tube pitch than the assumed 100 nm), array-based

CNFETs with AlN and SiO_2 substrates continue to offer performance that meets and exceeds the ITRS specifications for RF CMOS.

TABLE 3.3. ITRS RF CMOS MILLIMETER-WAVE TECHNOLOGY REQUIREMENTS FOR THE YEAR 2015 [7] VERSUS THE ARRAY-BASED CNFET WITH SPP SCATTERING INCLUDED FROM VARIOUS POLAR SUBSTRATES. THE TUBE PITCH IS 100 **nm**.

		Power Supply Voltage V_{DD} [V]	**Gate Length L_g [nm]**	**Peak f_T [THz]**	**Peak f_{max} [THz]**	**Peak g_m at $V_{DS}=V_{DD}$ [S/mm]**	**MSG/MAG [dB] at 24 GHz**	**MSG/MAG [dB] at 60 GHz**	**MSG/MAG [dB] at 94 GHz**
Array-Based CNFET Gate Width 1μm [67]		0.9	20	1.29	4.77	0.78	22	18	16
Array-Based CNFET with SPP scattering (This Study)	AlN	0.9	20	0.96	3.21	0.74	19	16	13
	SiO_2	0.9	20	0.72	3.63	0.44	19	15	12
	ZrO_2	0.9	20	0.15	1.01	0.20	11	8	5
	HfO_2	0.9	20	0.16	1.20	0.18	12	9	7
RF CMOS (ITRS)		0.9	20	0.44	0.56	2.24	17.4	13.4	11.5

3.4 Conclusions

The following conclusions can be drawn from this study of the impact of the substrate material via SPP scattering on the RF performance of array-based, carbon-nanotube transistors:

1. Scattering due to surface polar phonons, modeled according to the approach of [6], causes a significant reduction of the transistor's RF performance (Figs. 3.3, 3.10, and 3.11) compared to a ballistic device;

the magnitude of this reduction depends strongly on the substrate.

2. Among the materials considered in this work, AlN leads to the weakest impact of SPP scattering on the performance degradation; this can be attributed to the high energies of the excited SPP modes of AlN. On the other hand, the substrate phonon energies of HfO_2 and ZrO_2 are rather low, leading to strong SPP scattering and thus to a severe RF performance degradation, which becomes evident in a much lower cutoff frequency f_T (Figure 3.3). Another consequence of the disparity in the SPP phonon energies is that the f_T adopts its maximum at a larger gate bias V_G for AlN than for HfO_2 and ZrO_2 (Figure 3.3).

3. SPP scattering lowers the forward y-parameters over the entire frequency range, with the decrease becoming worse as the scattering gets stronger (Figure 3.6). In addition, SPP scattering causes the resonant behavior in the y-parameters [54, 67, 80-82] to become smeared out (lower Q factor).

4. The extrinsic f_T increases by a factor of about two as the pitch is decreased from 100 nm down to 10 nm (Figure 3.10). The extrinsic f_T for AlN remains the largest for all the pitches since SPP scattering with AlN has the least impact on the RF degradation. Correspondingly, HfO_2 and ZrO_2 lead to the lowest extrinsic f_T as a result of their strong susceptibility to SPP scattering.

5. The extrinsic f_{max} remains quite insensitive to pitch variations (Figure 3.11), which is a result of the increase in the effective gate-drain

capacitance $C_{\mathrm{gd,\,eff}}$ with decreasing pitch (Figure 3.12), thus counteracting the improvement of the extrinsic f_T. Due to the low dielectric permittivity (ϵ^0 in Table 3.1) of SiO_2 substrates, the external parasitic capacitance $C_{\mathrm{gd,ext}}$ and thus the $C_{\mathrm{gd,eff}}$ is the smallest; as a consequence, the $f_{\max}$ with an SiO_2 substrate becomes the largest for all pitches. On the other hand, ZrO_2 and HfO_2 suffer from a high dielectric permittivity leading to a large $C_{\mathrm{gd,ext}}$, which tends to degrade the already poor $f_{\max}$ arising from a low f_T with these substrates due to pronounced SPP scattering.

6. Despite the presence of SPP scattering (over and above native acoustic and optical phonons), and except for the transconductance g_m (which could be improved with a lower tube pitch), the results in Table 3.3 show that an array-based CNFET with a gate width of 1 μm, a gate length of 20 nm, and a tube pitch of 100 nm built on AlN and SiO_2 substrates continue to offer performance that meets and exceeds the ITRS specifications for RF CMOS [7]. The same is not true for HfO_2 and ZrO_2 substrates.

In summary, this work shows that SPP scattering can significantly impact the RF performance of array-based CNFETs. Our reported findings may be experimentally probed by performing an RF characterization of array-based CNFETs fabricated on different substrates. More generally, our results suggest that substrate effects should be carefully considered to improve the performance of all

future high-frequency devices; for example, similar results on the RF degradation could be expected for graphene-based FETs, since the distance between the substrate and channel material is comparable.

Chapter 4

Conclusions and Future Work

4.1 Summary of Conclusions

The conclusions from each stage of the work are summarized in this chapter. The full details of the work conducted in each stage leading to these conclusions are discussed in the preceding chapters of this thesis. Here we list the specific conclusions and then indicate the overall (collective) contribution of each stage.

4.1.1 Stage I (Chapter 2)

The specific conclusions from the first stage, entitled, "Impact of Effective Mass on the Scaling Behavior of the f_T and f_{max} of III-V High-Electron-Mobility Transistors," are as follows:

1. The intrinsic peak f_T occurs at gate voltages corresponding to the point of source-to-drain barrier collapse, beyond which the f_T degrades significantly.
2. At shorter gate lengths, drain-induced barrier lowering (DIBL) causes the barrier to collapse at lower gate bias voltages, such that the intrinsic peak f_T occurs at lower gate voltages.

3. The intrinsic peak f_T exhibits a saturating behavior when the gate length is scaled down; *i.e.*, it shows no further increase with decreasing gate length once the gate length is sufficiently small.
4. In comparing channel materials, the material with lower effective mass exhibits a more pronounced saturation in its peak f_T as the gate length is scaled down.
5. The extrinsic peak f_T of HEMTs reflect the saturating behavior of the intrinsic f_T.
6. The saturating behavior of the f_T of III-V HEMTs at short gate lengths contributes to the saturating behavior of the $f_{\max}$. Among the HEMTs considered, the larger parasitic resistances of GaN HEMTs causes them to have a lower $f_{\max}$ than the InGaAs HEMTs at all gate lengths.

Overall, the most important outcome of this work is the connection between the effective mass and the scaling behavior of RF performance. While a low effective mass is desirable for high mobility and potentially high-speed operation, it leads to diminishing improvements in the peak f_T and $f_{\max}$ as the gate length is scaled down.

4.1.2 Stage II (Chapter 3)

The specific conclusions from the second stage, entitled, "Impact of Substrate Material on the RF Performance of Carbon-Nanotube Transistors," are as follows:

1. Scattering due to surface polar phonons (SPP) causes a significant reduction in the transistor's RF performance compared to a ballistic device; the magnitude of this reduction depends strongly on the substrate.
2. Among the materials considered in our work, AlN leads to the weakest impact of SPP scattering on the performance degradation. On the other hand, the substrate phonon energies of HfO_2 and ZrO_2 lead to strong SPP scattering and thus to a severe RF performance degradation, which becomes evident in a much lower cutoff frequency f_T.
3. SPP scattering lowers the forward y-parameters over the entire frequency range, with the decrease becoming worse as the scattering gets stronger.
4. The extrinsic f_T increases by a factor of about two as the pitch is decreased from 100 nm down to 10 nm. The extrinsic f_T for AlN remains the largest for all the pitches since SPP scattering with AlN has the least impact on the RF degradation. Correspondingly, HfO_2 and ZrO_2 lead to the lowest extrinsic f_T.
5. The extrinsic $f_{\max}$ remains quite insensitive to pitch variations. Due to the low dielectric permittivity, the $f_{\max}$ with an SiO_2 substrate becomes the largest for all pitches. On the other hand, ZrO_2 and HfO_2 suffer from a high dielectric permittivity which tends to degrade the already poor $f_{\max}$.

6. Despite the presence of SPP scattering, an array-based CNFET with a gate width of 1μm, a gate length of 20 nm, and a tube pitch of 100 nm built on AlN and SiO_2 substrates continue to offer performance that meets and exceeds the ITRS specifications for RF CMOS [7].

Collectively, this work shows that SPP scattering can significantly impact the RF performance of array-based CNFETs. This outcome suggests that substrate effects should be carefully considered to improve the performance of all future high-frequency devices.

4.1.3 Stage III (Future Work)

The anticipated contribution of the third stage, entitled, "Quantum-Mechanical Transport in Nanoscale Quantum-Dot Solar Cells," has already been outlined in Section 1.2. Further details are provided in the following section.

4.2 Future Work: Quantum-Mechanical Transport in Nanoscale Quantum-Dot Solar Cells

4.2.1 Introduction

In recent years, there has been a significant and growing interest in renewable energy sources. The majority of the earth's total usable energy comes from fossil fuels. One of the main concerns of the widespread use of fossil fuels is the threat to the climate, since fossil fuels emit a large volume of greenhouse gases into the

atmosphere. These fuels are also in limited supply, making the industrialized economy dependent on fluctuating fuel prices and supply. On the other hand, renewable energy sources have the potential to be the key to all these problems and hence are being considered to complement fossil fuels and perhaps even replace them as a primary energy source. Solar-cell technology, which utilizes the sun's energy, is perhaps the most tempting because of the versatile, inexhaustible, and environmentally friendly features of the source.

The main concern in solar-cell technology is the tradeoff between efficiency and production cost. Traditional technologies, such as those utilizing crystalline silicon, offer high efficiency but only at the expense of a high production cost [89]; the hope for emerging technologies is to offer *cheap* but *sufficient* efficiency at the device level such that the installed cost of a system (often cited in $/Watt-peak) drives the technology to being a cost-effective, green alternative to fossil fuels.

Quantum-dot solar cells are among the most popular of third-generation photovoltaic devices being pursued to achieve an optimal cost-efficiency tradeoff [90]. The major advantage of quantum dots is that they have bandgaps that are tunable across a wide range by changing the size and types of dots [8]. Such tunability facilitates the ability to fabricate solar cells consisting of a number of different bandgap materials, each of which can be optimized to convert solar energy within a portion of the sun's spectrum, thus maximizing overall efficiency. In fact, the use of multiple bandgap materials makes it theoretically possible to *exceed* the efficiency of single-junction cells [91], the latter being limited to the well-known value of ~33%.

Achieving the promise of bandgap engineering in quantum-dot cells, and hence achieving a cost-effective alternative to silicon, requires a proper description of the electron and hole transport within the structure, which is fundamentally quantum-mechanical. To date, no computer-based tool accounting for quantum-mechanical transport exists to explore the design parameters, such as the size of the dots, the types of the dots, the distance between the dots, the background material containing the dots, and irregularities in the structure. The development of the technology so far has been based on forcibly applying rudimentary classical concepts such as "electron mobility" [92] and the drift-diffusion equation [(1.1) of Chapter 1], and on utilizing *trial-and-error* cell design. Toward this end, in this stage of the work, a simulation tool that captures the quantum-mechanical transport in these cells, and which hence captures their essential physics and facilitates their systematic design, is proposed. Such a tool has the potential to substantially speed the movement of the technology from the laboratory to the market and to substantially reduce development costs.

4.2.2 InAs/GaAs Quantum-Dot Solar Cell

A Overview of the Project

Among the emerging quantum-dot solar cells are devices characterized by the existence of InAs quantum dots in a GaAs channel region [9, 10], and we assume this type of structure as a starting point for our work. A sample InAs/GaAs quantum-dot solar cell structure is shown in Figure 4.1. The GaAs channel region,

with a bandgap of 1.43 eV, is considered to have a length of 120 nm, and InAs quantum dots, each with a bandgap of 0.36 eV and a diameter of 3 nm, are placed 3 nm apart, in a regular array, inside the channel region. Contacts at the two sides of the channel region are each considered to be 15 nm long, with one being n^+ doped and the other being p^+ doped, where both the doping densities are 1×10^{18} cm^{-3}.

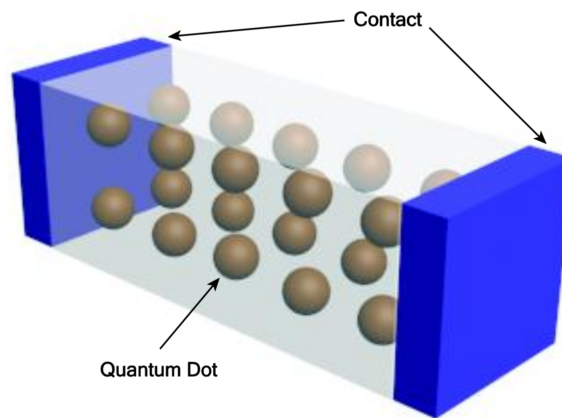


Figure 4.1. An InAs/GaAs quantum-dot solar cell.

The cross-sectional area of the device in Figure 4.1 can be considered to be quite large with respect to the individual atoms, and hence a one-dimensional (1D) transport simulation is assumed adequate for a first-order representation of device behavior. Toward this end, a 1D NEGF-Poisson solver, which is capable of describing *both* the optical *and* transport properties, including quantum effects, is proposed to model the InAs/GaAs quantum-dot solar cell. A tight-binding Hamiltonian under the effective-mass approximation can be used to model the hopping of electrons and holes between neighboring energy states. This technique

will yield a powerful solver in the sense that it will be capable of precisely modeling (i) the locations and sizes of the quantum dots within the active region of the device, (ii) the effective mass of the charge carriers, and (iii) the bandgaps of different regions in space. Results from the solver can be utilized to extract and explain the important solar-cell figures of merit, such as the short-circuit current density J_{sc}, open-circuit voltage V_{oc}, fill factor FF, and efficiency η. The final outcome of this project would hence be a solver for InAs/GaAs quantum-dot solar cells and the insight into their operation provided by the solver. The developed solver can then be extended to quantum-dot solar cells fabricated with other materials.

B Simulation Approach

Initial work towards development of the solver has already been completed. *It must be emphasized that the approach described here is preliminary in nature and only intended to provide a starting point.* While the approach does yield the curves in Figure 4.3, it should be revisited, scrutinized, and improved, with the aim of moving the work towards publishable results; this will be done by the candidate and other group members in the near future. The reader should hence view the description below only as a sketch of work "under construction."

The quantum-dot solar cell is modeled using the NEGF formalism following the approach described in [93, 94], where the interaction between the electrons and photons are taken into the NEGF equations through the self-energy matrices. The device Green's function can then be represented as

$$G(E) = [EI - h - \Sigma_S - \Sigma_D - \Sigma_{\text{ph}}]^{-1} \tag{4.1}$$

where E is the total electron energy, h is the device Hamiltonian, I is the identity matrix, Σ_S and Σ_D are the source and drain self-energies, and Σ_{ph} is the self-energy due to the electron-photon coupling. The electron-photon self-energy Σ_{ph} is further represented in terms of the in-scattering and out-scattering self-energies, Σ^{in} and Σ^{out}, respectively, as [93]

$$\Sigma_{\text{ph}}(E) = -\frac{i}{2}\left[\Sigma^{\text{in}}(E) + \Sigma^{\text{out}}(E)\right] \tag{4.2}$$

B.1 Incident Light Source

Solar radiation considered in the work is the ASTM G173-03 reference spectra derived from SMARTS v. 2.9.2 for AM 1.5 [95], which provides data of spectral irradiance as a function of photon wavelength, and which facilitates the determination of illumination intensities at particular wavelengths. The effect of this solar radiation on a solar cell is modeled through the incoming photon flux $I_\omega = \frac{I_{\text{int}}(\lambda)}{\hbar\omega}$, where $I_{\text{int}}(\lambda)$ is the intensity of illumination and $\hbar\omega$ is the energy of the photon, calculated from its wavelength λ. A range of photon energies which are

higher than the bandgap of the quantum-dot material and the corresponding photon intensities, calculated[6] from [95], are considered in the work.

The strength of the interaction between carriers and photons is described by a mathematical quantity M, given at a lattice site (l, m) by [94]

$$M_{\text{lm}} = \frac{q\hbar}{i2am_m{}^*}\sqrt{\frac{\hbar\sqrt{\mu\epsilon}}{2N\omega\epsilon_0 c}I_\omega}\, P_{\text{lm}} \tag{4.3}$$

where a is the grid spacing, $m_m{}^*$ is the effective mass of the electron at site m, μ and ϵ are the magnetic permeability and the dielectric constant of the material, respectively, ϵ_0 is the dielectric constant in free space, c is the velocity of light, N is the number of photons with energy $\hbar\omega$, and

$$P_{\text{lm}} = \begin{cases} +1, & m = l + 1 \\ -1, & m = l - 1 \\ 0\,, & \text{otherwise} \end{cases} \tag{4.4}$$

B.2 Generation and Recombination

Generation is the process by which mobile charge carriers are created. Under illumination, the absorbed photons excite the electrons from the valence band to the conduction band and an electron-hole pair is generated. The separated charge

[6] The intensity at a particular wavelength λ_i can be obtained by multiplying the irradiance at λ_i by $\Delta\lambda_i$, where $\Delta\lambda_i$ is the difference between two consecutive wavelengths λ_i and λ_{i+1}.

carriers get transported to their respective electrodes, *i.e.*, electrons toward the n^+ region and holes toward the p^+ region giving rise to an electric current corresponding to the incoming photons. During the transport, some of the carriers recombine with each other, *i.e.* an electron in the conduction band can move into an empty state in the valence band, which is equivalent to an electron-hole annihilation (recombination) event.

The effect of generation and recombination is modeled in the simulation through the in-scattering and out-scattering self-energies, Σ^{in} and Σ^{out}, which are described, under the illumination of a monochromatic light with a photon energy of $\hbar\omega$, at each lattice site (l,m) as

$$\Sigma_{\text{lm}}{}^{\text{in}}(E) = \sum_{\text{pq}} M_{\text{lp}} M_{\text{qm}} [N G_{\text{pq}}{}^{n}(E - \hbar\omega) + (N+1) G_{\text{pq}}{}^{n}(E + \hbar\omega)] \qquad (4.5\text{a})$$

and

$$\Sigma_{\text{lm}}{}^{\text{out}}(E) = \sum_{\text{pq}} M_{\text{lp}} M_{\text{qm}} [N G_{\text{pq}}{}^{p}(E + \hbar\omega) + (N+1) G_{\text{pq}}{}^{p}(E - \hbar\omega)] \qquad (4.5\text{b})$$

where the symbols are as follows: l, m, p and q are real-space indices; $G_{\text{pq}}{}^{n}$ and $G_{\text{pq}}{}^{p}$ are the electron and hole correlation functions at lattice site (p,q), respectively, and are expressed in (4.6a) and (4.6b) below.

In (4.5a) and (4.5b), the first term defines the absorption process, where an electron absorbs the incoming photon and jumps from a state in the valence band to a state in the conduction band, with the two states separated by the photon energy. This process thus creates an electron-hole pair and hence describes the generation process. At the same time, the second term defines the emission process, where an electron moves from a state in the conduction band to a state in the valence band, with the two states separated by the photon energy, and where a photon is emitted. This process thus destroys an electron-hole pair and hence describes a type of recombination process. A major advantage of the NEGF approach is that it does not require any approximate expression to model the generation and recombination process as utilized in a drift-diffusion approach [96]; rather, it precisely models the movement of charge carriers going from one state to the other, separated by the photon energy, as just described.

A term requiring definition in (4.5a) and (4.5b) is the photon number N. It can be assumed that, when there are electron-photon interactions, the participating photons are in equilibrium inside the solar cell and that they are at the same temperature as the charge carriers. So, the photon number inside the device can be described by the Bose-Einstein distribution and this is what is considered in this work.

For the case of irradiation by the solar spectrum, several photon frequencies will be present in the illumination (vs. monochromatic illumination). The effect of these multiple photon frequencies is taken into account in the simulation by modifying (4.5a) and (4.5b), where the photons at each incoming frequency are

considered to impact the device independently, *i.e.*, not coupled to each other. The contribution of each illuminating frequency (*i.e.*, of photons with energies of $\hbar\omega_1$, $\hbar\omega_2$, and so on) to the absorption and emission terms in (4.5a) and (4.5b) are summed together so that the in-scattering and out-scattering functions include the effect of all the photons. While this approach is a major simplification, it suffices as a first step to demonstrate the viability of the overall simulation technique.

B.3 Correlation Functions and Current

The electron and hole correlation functions describe how the states are filled by the charge carriers. These correlation functions, including the electron-photon coupling, can be expressed as [93]

$$G^n(E) = G(E)\left[\Gamma_S(E)f_S(E) + \Gamma_D(E)f_D(E) + \Sigma^{\text{in}}(E)\right]G(E)^+ \tag{4.6a}$$

and

$$\begin{aligned} G^p(E) = G(E)[\Gamma_S(E)[1-f_S(E)] + \Gamma_D(E)[1-f_D(E)] \\ + \Sigma^{\text{out}}(E)]G(E)^+ \end{aligned} \tag{4.6b}$$

where $\Gamma_{S/D}(E) = i\left[\Sigma_{S/D}(E) - \Sigma_{S/D}(E)^+\right]$ is the broadening function of the source/drain contact, and $f_{S/D}(E)$ is the source/drain Fermi distribution.

Once the correlation functions are obtained, the energy-resolved current spectrum at a lattice site l can be calculated as

$$J_l(E) = \frac{4qi}{h} \mathrm{Tr}\left[h_{l,l+1}(E) G^n_{l,l+1}(E) - h_{l+1,l}(E) G^n_{l+1,l}(E)\right] \tag{4.7}$$

Finally, the device current, which is the fundamental parameter of interest for a solar cell, can be obtained by integrating (4.7) over the energy E.

B.4 Simulation Steps

The overall simulation is performed in two steps. In the first step, the simulation is performed for the case of no illumination. The NEGF transport equations in (4.1), (4.6a) and (4.6b), without the electron-photon interaction part, are solved self-consistently with the Poisson equation at each bias point; the generated self-consistent potential profile is used as an initial guess for the second step.

In the second step, the electron-photon interaction terms are taken into account in the NEGF equations. At first, the in-scattering and out-scattering self-energies are calculated according to (4.5a) and (4.5b), where the electron and hole correlation functions are taken to be the ones for the case of no illumination. These self-energies are then used to calculate the correlation functions using (4.6a) and (4.6b), which also make use of the Green's function of the device under no illumination. The new correlation functions are inserted back in (4.5a) and (4.5b) to again calculate the self-energies, which are then utilized to calculate the Green's function from (4.1).

This process of iteration between (4.5), (4.6) and (4.1) is continued until convergence is achieved. Typically, to reduce computational cost, the NEGF equations are *not* generally solved self-consistently with each other [93] ; however, without complete self-consistency, the results are expected to significantly lack in precisely depicting the interaction between the carriers and the photons, and hence the additional self-consistency has been performed in this work. Finally, the self-consistent NEGF equations are solved with Poisson's equation at each bias point for all the important photon energies and corresponding illumination intensities, thus incorporating the usual self-consistency with the electrostatics. Upon achieving *overall* self-consistency, the solver can be utilized to extract and explain the important solar-cell figures of merit.

B.5 Inclusion of Traps

NEGF is a very powerful tool that can include the effect of trap states. The most common type of traps are defined by the presence of energy states in the bandgap region between the conduction and valence bands, where such states are primarily created by impurities in the lattice. A charge carrier can jump or fall into these energy states and result in generation or recombination, respectively and thus affect the overall device behavior.

In the drift-diffusion approach, these trap-assisted generation/recombination events are described by approximate expressions, whereas, in the case of NEGF, these traps can be handled more precisely; in NEGF, the trap states can be modeled

by introducing a localized potential well, *i.e.*, a potential that dips at the location of the traps. Once the potential profile is modified, the NEGF transport equations will automatically capture the charge confinement in these precise trap locations and hence model the presence of trap states.

As a first-order study, in this work, the effect of trap states has not been considered. However, work is under way on the investigation of the effect of trap states on device performance.

4.2.3 Results

A Initial Work

Initial work has been done to illustrate the viability of the solution approach; a sample conduction-band energy diagram, under the condition of no illumination, is shown in Figure 4.2 for an InAs/GaAs quantum-dot cell structure, as depicted Figure 4.1, at a reverse bias voltage of 0.5 V.

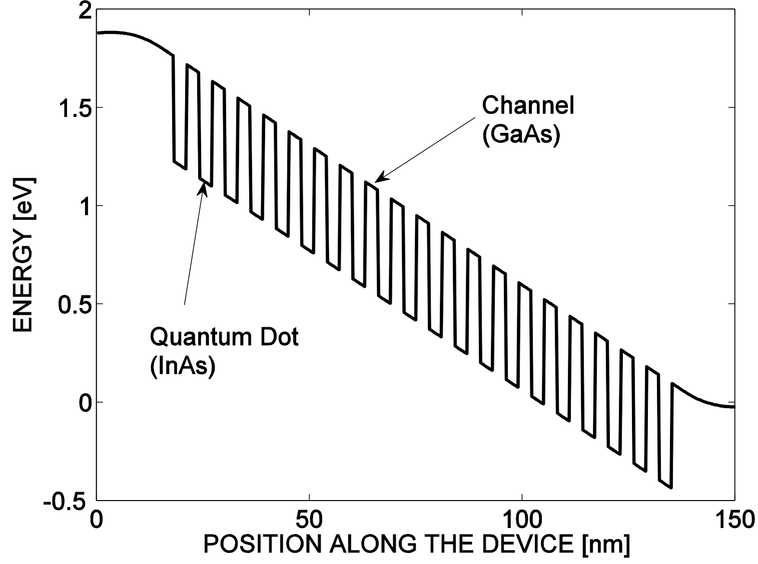


Figure 4.2. Conduction-band profile of an InAs/GaAs quantum-dot solar-cell structure at a reverse bias voltage of 0.5 V.

B Current-Voltage Characteristics

The solar cell structure as shown in Figure 4.1 has been simulated for a range of reverse bias voltages from 0 V to 1 V at all the important photon energies and corresponding illumination intensities. The resulting current-voltage characteristic curves for the InAs/GaAs quantum-dot solar cell are shown in Figure 4.3 where we have plotted the curves for the cases of illumination and no illumination; the plots have been flipped with respect to the horizontal axis to represent the current densities of the solar cell in a conventional way. The current density curve under illumination also facilitates the extraction of the short-circuit current density J_{sc} and open-circuit voltage V_{oc} which are observed to be 187 A/m^2 and 0.85 V, respectively.

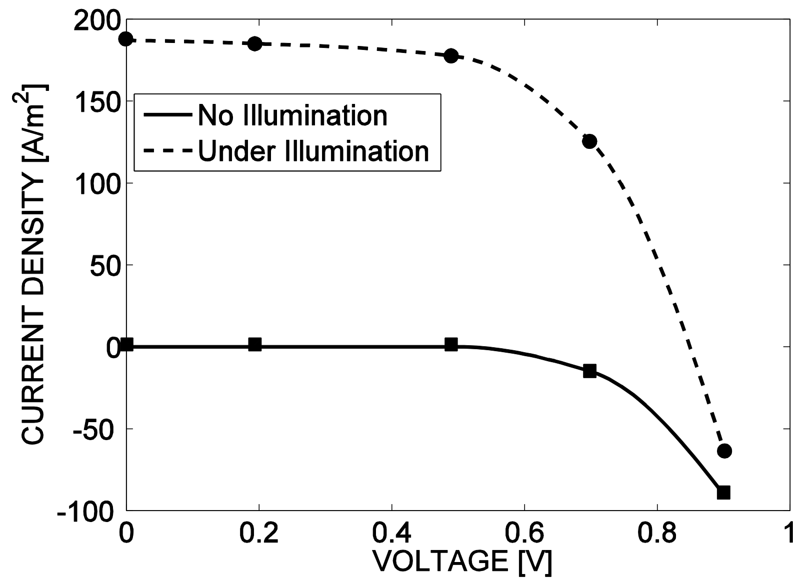


Figure 4.3. Current-voltage characteristic curves of an InAs/GaAs quantum-dot solar-cell structure. The lines have been drawn as guides for the eye.

C Work in Progress

The proposed simulation tool has the potential to be very powerful and will be capable of investigating several design parameters of the quantum-dot solar cell, such as the size of the dots, the types of the dots, the distance between the dots, the background material containing the dots, and irregularities in the structure. Work is under way towards this direction and the ultimate outcome will be the comparison of performances of devices with different structures, which will act as a guide towards optimization of quantum-dot solar cells.

4.2.4 Summary

This study of *Quantum-Mechanical Transport in Nanoscale Quantum-Dot Solar Cells* utilizes the expertise developed while working on the previous projects, outlined in Chapters 2 and 3, to look into the impact of carrier transport in the field of renewable energy sources and provide important insight into the device behavior.

The methodology for the prediction of the performance of InAs/GaAs quantum-dot solars cell is general and can be applied to all kinds of nanoscale solar cells. The quantum-mechanical simulation tool proposed in this work will not only be useful to experimentalists, but also to software firms (such as Lumerical Solutions Inc. in Canada) involved in optoelectronic software and who have already expressed interest. While quantum transport is nontrivial, it can be made tractable using the NEGF approach, which is capable of describing both the optical and transport properties of solar cells, including all relevant quantum effects.

This concludes the Ph.D. thesis, with a title of *Transport Performance Projection of Emerging Nanoscale Devices*.

Bibliography

[1] COMSOL Multiphysics Version 3.5a. COMSOL, Inc. Stockholm, Sweden, 2008.

[2] S. Ahmed, K. D. Holland, N. Paydavosi, C. M. S. Rogers, A. U. Alam, N. Neophytou, D. Kienle, and M. Vaidyanathan, "Impact of effective mass on the scaling behavior of the f_T and f_{max} of III-V high-electron-mobility transistors," *IEEE Trans. Nanotechnol.,* vol. 11, no. 6, pp. 1160-1173, Nov. 2012.

[3] C. Rutherglen, D. Jain, and P. Burke, "Nanotube electronics for radiofrequency applications," *Nat. Nano,* vol. 4, no. 12, pp. 811-819, Nov. 2009.

[4] S. Datta, *Quantum Transport : Atom to Transistor.* Cambridge, U.K.: Cambridge University Press, 2005.

[5] S. Ahmed, N. Paydavosi, A. U. Alam , K. D. Holland, D. Kienle, and M. Vaidyanathan, "Impact of Substrate Material on the RF Performance of Carbon-Nanotube Transistors," *IEEE Trans. Nanotechnol.,* vol. 13, no. 1, pp. 123-135, Jan. 2014.

[6] V. Perebeinos, S. V. Rotkin, A. G. Petrov, and P. Avouris, "The effects of substrate phonon mode scattering on transport in carbon nanotubes," *Nano Lett.,* vol. 9, no. 1, pp. 312-316, Jan. 2009.

[7] "Radio frequency and analog/mixed-signal technologies for wireless communications," International Technology Roadmap for Semiconductors, 2009 edition. Available:

http://www.itrs.net/Links/2009ITRS/2009Chapters_2009Tables/2009_Wireless.pdf

[8] P. Ardalan, T. P. Brennan, H. Lee, J. R. Bakke, I. Ding, M. D. McGehee, and S. F. Bent, "Effects of self-assembled monolayers on solid-state CdS quantum dot sensitized solar cells," *ACS Nano,* vol. 5, no. 2, pp. 1495-1504, Feb. 2011.

[9] A. Marti, N. Lopez, E. Antolin, E. Canovas, C. Stanley, C. Farmer, L. Cuadra, and A. Luque, "Novel semiconductor solar cell structures: The quantum dot intermediate band solar cell," *Thin Solid Films,* vol. 511-512, pp. 638-644, Jul. 2006.

[10] V. Aroutiounian, S. Petrosyan, A. Khachatryan, and K. Touryan, "Quantum dot solar cells," *J. Appl. Phys.,* vol. 89, no. 4, pp. 2268-2271, Feb. 2001.

[11] F. Schwierz and J. J. Liou, "RF transistors: recent developments and roadmap toward terahertz applications," *Solid State Electronics,* vol. 51, no. 8, pp. 1079-1091, Aug. 2007.

[12] M. V. Fischetti and S. E. Laux, "Monte Carlo simulation of transport in technologically significant semiconductors of the diamond and zinc-blende structures. II. Submicrometer MOSFET's," *IEEE Trans. Electron Devices,* vol. 38, no. 3, pp. 650-660, Mar. 1991.

[13] P. M. Solomon and S. E. Laux, "The ballistic FET: Design, capacitance and speed limit," in *Proc. IEEE IEDM,* Dec. 2001, pp. 95-98.

[14] A. Rahman, G. Klimeck, and M. Lundstrom, "Novel channel materials for ballistic nanoscale MOSFETs-bandstructure effects," in *Proc. IEEE IEDM,* Dec. 2005, pp. 601-604.

[15] K. C. Saraswat, C. O. Chui, D. Kim, T. Krishnamohan, and A. Pethe, "High mobility materials and novel device structures for high performance nanoscale MOSFETs," in *Proc. IEEE IEDM,* Dec. 2006, pp. 1-4.

[16] J. Ayubi-Moak, R. Akis, D. K. Ferry, S. M. Goodnick, N. Faralli, and M. Saraniti, "Towards the global modeling of InGaAs-based pseudomorphic HEMTs," *J. Comput. Electron.,* vol. 7, no. 3, pp. 187-191, Sep. 2008.

[17] R. Akis, J. Ayubi-Moak, D. K. Ferry, S. M. Goodnick, N. Faralli, and M. Saraniti, "Full-band cellular Monte Carlo simulations of terahertz high electron mobility transistors," *J. Phys.: Condens. Matter,* vol. 20, no. 38, pp. 384201-1-384201-9, Aug. 2008.

[18] Y. Yamashita, A. Endoh, K. Shinohara, K. Hikosaka, T. Matsui, S. Hiyamizu, and T. Mimura, "Pseudomorphic $In_{0.52}Al_{0.48}As/In_{0.7}Ga_{0.3}As$ HEMTs with an ultrahigh f_T of 562 GHz," *IEEE Electron Device Lett.,* vol. 23, no. 10, pp. 573-575, Oct. 2002.

[19] B. O. Lim, K. L. Mun, J. B. Tae, M. Han, C. K. Sung, and J. K. Rhee, "50-nm T-gate InAlAs/InGaAs metamorphic HEMTs with low noise and high f_T characteristics," *IEEE Electron Device Lett.,* vol. 28, no. 7, pp. 546-548, Jun. 2007.

[20] W. R. Deal, "Solid-state amplifiers for terahertz electronics," in *Proc. IEEE MTT,* May 2010, pp. 1122-1125.

[21] T. Palacios, Y. Dora, A. Chakraborty, C. Sanabria, S. Keller, S. P. DenBaars, and U. K. Mishra, "Optimization of AlGaN/GaN HEMTs for high frequency operation," *Phys. Stat. Sol. (a),* vol. 203, no. 7, pp. 1845-1850, May 2006.

[22] A. Endoh, Y. Yamashita, K. Ikeda, M. Higashiwaki, K. Hikosaka, T. Matsui, S. Hiyamizu, and T. Mimura, "Fabrication of sub-50-nm-gate i-AlGaN/GaN HEMTs on sapphire," *Phys. Stat. Sol. (c),* vol. 0, no. 7, pp. 2368-2371, Oct. 2003.

[23] D. Guerra, R. Akis, F. A. Marino, D. K. Ferry, S. M. Goodnick, and M. Saraniti, "Aspect ratio impact on RF and DC performance of state-of-the-art short-channel GaN and InGaAs HEMTs," *IEEE Electron Device Lett.,* vol. 31, no. 11, pp. 1217-1219, Oct. 2010.

[24] R. Tulek, A. Ilgaz, S. Gokden, A. Teke, M. K. Ozturk, M. Kasap, S. Ozgelik, E. Arslan, and E. Ozbay, "Comparison of the transport properties of high quality AlGaN/AlN/GaN and AlInN/AlN/GaN two-dimensional electron gas heterostructures," *J. Appl. Phys.,* vol. 105, no. 1, pp. 013707-1- 013707-6, Nov. 2009.

[25] Y. Wang, L. Ma, Z. Yu, and L. Tian, "Optimization of two-dimensional electron gases and I-V characteristics for AlGaN/GaN HEMT devices," *Superlattice. Microst.,* vol. 36, no. 4-6, pp. 869-875, Oct.-Dec. 2004.

[26] D. K. Ferry, "The onset of quantization in ultra-submicron semiconductor devices," *Superlattice. Microst.,* vol. 27, no. 2-3, pp. 61-66, Feb. 2000.

[27] R. Akis, L. Shifren, D. K. Ferry, and D. Vasileska, "The effective potential and its use in simulation," *Phys. Stat. Sol. (b),* vol. 226, no. 1, pp. 1-8, Jul. 2001.

[28] N. Neophytou, T. Rakshit, and M. S. Lundstrom, "Performance analysis of 60-nm gate-length III-V InGaAs HEMTs: simulations versus experiments," *IEEE Trans. Electron Devices,* vol. 56, no. 7, pp. 1377-1387, Jul. 2009.

[29] D. Kim and J. A. del Alamo, "Lateral and vertical scaling of $In_{0.7}Ga_{0.3}As$ HEMTs for post-Si-CMOS logic applications," *IEEE Trans. Electron Devices,* vol. 55, no. 10, pp. 2546-2553, Sep. 2008.

[30] R. Venugopal, Z. Ren, S. Datta, M. S. Lundstrom, and D. Jovanovic, "Simulating quantum transport in nanoscale transistors: real versus mode-space approaches," *J. Appl. Phys.,* vol. 92, no. 7, pp. 3730-3739, Oct. 2002.

[31] S. Datta, "Nanoscale device modeling: the Green's function method," *Superlattice. Microst.,* vol. 28, no. 4, pp. 253-278, May 2000.

[32] M. P. L. Sancho, J. M. L. Sancho, and J. Rubio, "Highly convergent schemes for the calculation of bulk and surface Green functions," *J. Phys. F: Met. Phys.,* vol. 15, no. 4, pp. 851-858, Apr. 1985.

[33] A. Svizhenko, M. P. Anantram, T. R. Govindan, B. Biegel, and R. Venugopal, "Two-dimensional quantum mechanical modeling of nanotransistors," *J. Appl. Phys.,* vol. 91, no. 4, pp. 2343-2354, Feb. 2002.

[34] Y. Liu and M. S. Lundstrom, "Simulation of III-V HEMTs for high-speed low-power logic applications," in *ECS Transactions,* vol. 19, pp. 331-342, May 2009.

[35] S. Oktyabrsky and P. D. Ye, Eds., *Fundamentals of III-V Semiconductor MOSFETs*, 1st ed. New York: Springer, 2010.

[36] S. P. Kumar, A. Agrawal, R. Chaujar, S. Kabra, M. Gupta, and R. S. Gupta, "3-dimensional analytical modeling and simulation of fully depleted AlGaN/GaN modulation doped field effect transistor," in *Proc. IWSPD,* Dec. 2007, pp. 373-376.

[37] G. Steinhoff, B. Baur, G. Wrobel, S. Ingebrandt, A. Offenhausser, A. Dadgar, A. Krost, M. Stutzmann, and M. Eickhoff, "Recording of cell action potentials with AlGaN/GaN field-effect transistors," *Appl. Phys. Lett.,* vol. 86, no. 3, pp. 33901-1-33901-3, Jan. 2005.

[38] J. A. del Alamo, "Is nanometer-scale lll-V CMOS cool enough to rejuvenate Moore's Law?" *Compound Semiconductor,* vol. 16, no. 5, pp. 19-22, Jul. 2010.

[39] Y. Tsividis, *Operation and Modeling of the MOS Transistor.* 2nd ed. New York: McGraw-Hill, 1999.

[40] D. Jin, D. Kim, T. Kim, and J. A. del Alamo, "Quantum capacitance in scaled down III-V FETs," in *Proc. IEEE IEDM,* Dec. 2009, pp. 1-4.

[41] A. Rahman, J. Guo, S. Datta, and M. S. Lundstrom, "Theory of ballistic nanotransistors," *IEEE Trans. Electron Devices,* vol. 50, no. 9, pp. 1853-1864, Sep. 2003.

[42] S. Hasan, S. Salahuddin, M. Vaidyanathan, and M. A. Alam, "High-frequency performance projections for ballistic carbon-nanotube transistors," *IEEE Trans. Nanotechnol.,* vol. 5, no. 1, pp. 14-22, Jan. 2006.

[43] D. Kim and J. A. del Alamo, "Scalability of sub-100 nm InAs HEMTs on InP substrate for future logic applications," *IEEE Trans. Electron Devices,* vol. 57, no. 7, pp. 1504-1511, Jul. 2010.

[44] B. V. V. Zeghbroeck, *Principles of Semiconductor Devices and Heterojunctions*. 1st ed. Upper Saddle River, N.J.: Prentice Hall, 2009.

[45] B. K. Meyer, D. Volm, A. Graber, H. C. Alt, T. Detchprohm, K. Amano, and I. Akasaki, "Shallow donors in GaN - The binding energy and the electron effective mass," *Solid State Commun.,* vol. 95, no. 9, pp. 597-600, Sep. 1995.

[46] D. L. John, F. Allerstam, T. Rodle, S. K. Murad, and G. D. J. Smit, "A surface-potential based model for GaN HEMTs in RF power amplifier applications," in *Proc. IEEE IEDM,* Dec. 2010, pp. 8.3.1-8.3.4.

[47] J. M. Lopez, T. Gonzalez, D. Pardo, S. Bollaert, T. Parenty, and A. Cappy, "Design optimization of AlInAs-GaInAs HEMTs for high-frequency applications," *IEEE Trans. Electron Devices,* vol. 51, no. 4, pp. 521-528, Apr. 2004.

[48] W. D. Hu, X. S. Chen, Z. J. Quan, X. M. Zhang, Y. Huang, C. S. Xia, W. Lu, and P. D. Ye, "Simulation and optimization of GaN-based metal-oxide-semiconductor high-electron-mobility transistor using field-dependent drift velocity model," *J. Appl. Phys.,* vol. 102, no. 3, pp. 034502-1-034502-7, Aug. 2007.

[49] N. Paydavosi, J. P. Rebstock, K. D. Holland, S. Ahmed, A. U. Alam, and M. Vaidyanathan, "RF performance potential of array-based carbon-nanotube transistors—part II: extrinsic results," *IEEE Trans. Electron Devices,* vol. 58, no. 7, pp. 1941-1951, Jun. 2011.

[50] Y. Q. Wu, W. K. Wang, O. Koybasi, D. N. Zakharov, E. A. Stach, S. Nakahara, J. C. M. Hwang, and P. D. Ye, "0.8-V supply voltage deep-submicrometer inversion-mode $In_{0.75}Ga_{0.25}As$ MOSFET," *IEEE Electron Device Lett.,* vol. 30, no. 7, pp. 700-702, Jun. 2009.

[51] Y. Yamashita, A. Endoh, K. Ikeda, K. Hikosaka, T. Mimura, M. Higashiwaki, T. Matsui, and S. Hiyamizu, "Effect of thermal annealing on 120-nm -T-shaped- TiPtAu -gate AlGaNGaN high electron mobility transistors," *J. Vac. Sci. Technol. B,* vol. 23, no. 3, pp. 895-899, May 2005.

[52] C. Kocabas, S. Dunham, Q. Cao, K. Cimino, X. Ho, H. Kim, D. Dawson, J. Payne, M. Stuenkel, H. Zhang, T. Banks, M. Feng, S. V. Rotkin, and J. A. Rogers, "High-frequency performance of submicrometer transistors that use aligned arrays of single-walled carbon nanotubes," *Nano Lett.,* vol. 9, no. 5, pp. 1937-1943, Apr. 2009.

[53] H. Yu, L. McCarthy, S. Rajan, S. Keller, S. Denbaars, J. Speck, and U. Mishra, "Ion implanted AlGaN-GaN HEMTs with nonalloyed ohmic contacts," *IEEE Electron Device Lett.,* vol. 26, no. 5, pp. 283-285, May 2005.

[54] N. Paydavosi, K. D. Holland, M. M. Zargham, and M. Vaidyanathan, "Understanding the frequency- and time-dependent behavior of ballistic carbon-nanotube transistors," *IEEE Trans. Nanotechnol.,* vol. 8, no. 2, pp. 234-244, Mar. 2009.

[55] M. S. Gupta, "Power gain in feedback amplifiers, a classic revisited," *IEEE Trans. Microwave Theory Tech.,* vol. 40, no. 5, pp. 864-879, May 1992.

[56] L. C. Castro and D. L. Pulfrey, "Extrapolated f_{max} for carbon nanotube field-effect transistors," *Nanotechnology,* vol. 17, no. 1, pp. 300-304, Jan. 2006.

[57] M. Vaidyanathan and D. L. Pulfrey, "Extrapolated f_{max} of heterojunction bipolar transistors," *IEEE Trans. Electron Devices,* vol. 46, no. 2, pp. 301-309, Feb. 1999.

[58] C. L. Tao and G. A. Armstrong, "The impact of the intrinsic and extrinsic resistances of double gate SOI on RF performance," *Solid State Electron.,* vol. 50, no. 5, pp. 774-783, May 2006.

[59] J. Y. Park, S. Rosenblatt, Y. Yaish, V. Sazonova, H. Ustunel, S. Braig, T. A. Arias, P. W. Brouwer, and P. L. McEuen, "Electron-phonon scattering in metallic single-walled carbon nanotubes," *Nano Lett.,* vol. 4, no. 3, pp. 517-520, Feb. 2004.

[60] T. Dürkop, S. A. Getty, E. Cobas, and M. S. Fuhrer, "Extraordinary mobility in semiconducting carbon nanotubes," *Nano Lett.,* vol. 4, no. 1, pp. 35-39, Jan. 2004.

[61] X. Zhou, J. Y. Park, S. Huang, J. Liu, and P. L. McEuen, "Band structure, phonon scattering, and the performance limit of single-walled carbon nanotube transistors," *Phys. Rev. Lett.,* vol. 95, no. 14, pp. 146805-1-146805-4, Sep. 2005.

[62] V. Perebeinos, J. Tersoff, and P. Avouris, "Electron-phonon interaction and transport in semiconducting carbon nanotubes," *Phys. Rev. Lett.,* vol. 94, no. 8, pp. 086802-1-086802-4, Mar. 2005.

[63] J. Guo and M. Lundstrom, "Role of phonon scattering in carbon nanotube field-effect transistors," *Appl. Phys. Lett.,* vol. 86, no. 19, pp. 193103-1-193103-3, May 2005.

[64] A. Svizhenko and M. P. Anantram, "Effect of scattering and contacts on current and electrostatics in carbon nanotubes," *Phys. Rev. B,* vol. 72, no. 8, pp. 085430-1-085430-10, Aug. 2005.

[65] S. O. Koswatta, S. Hasan, M. S. Lundstrom, M. P. Anantram, and D. E. Nikonov, "Ballisticity of nanotube field-effect transistors: role of phonon energy and gate bias," *Appl. Phys. Lett.,* vol. 89, no. 2, pp. 23125-1-23125-3, Jul. 2006.

[66] G. Pennington, N. Goldsman, A. Akturk, and A. E. Wickenden, "Deformation potential carrier-phonon scattering in semiconducting carbon nanotube transistors," *Appl. Phys. Lett.,* vol. 90, no. 6, pp. 62110-1-62110-3, Feb. 2007.

[67] N. Paydavosi, A. U. Alam, S. Ahmed, K. D. Holland, J. P. Rebstock, and M. Vaidyanathan, "RF performance potential of array-based carbon-nanotube transistors—part I: intrinsic results," *IEEE Trans. Electron Devices,* vol. 58, no. 7, pp. 1928-1940, Jun. 2011.

[68] A. G. Petrov and S. V. Rotkin, "Energy relaxation of hot carriers in single-wall carbon nanotubes by surface optical phonons of the substrate," *JETP Lett.,* vol. 84, no. 3, pp. 156-160, Oct. 2006.

[69] V. Perebeinos, J. Tersoff, and P. Avouris, "Mobility in semiconducting carbon nanotubes at finite carrier density," *Nano Lett.,* vol. 6, no. 2, pp. 205-208, Jan. 2006.

[70] B. Chandra, V. Perebeinos, S. Berciaud, J. Katoch, M. Ishigami, P. Kim, T. F. Heinz, and J. Hone, "Low Bias Electron Scattering in Structure-Identified Single Wall Carbon Nanotubes: Role of Substrate Polar Phonons," *Phys. Rev. Lett.,* vol. 107, no. 14, pp. 146601-1-146601-4, Sep. 2011.

[71] G. D. Mahan, "Surface polaritons and the theory of image charges," *Phys. Rev. B,* vol. 5, no. 2, pp. 739-744, Jan. 1972.

[72] R. Ludwig and G. Bogdanov, *RF Circuit Design: Theory and Applications.* 2nd ed. Upper Saddle River, New Jersey: Prentice-Hall, 2009.

[73] S. Fratini and F. Guinea, "Substrate-limited electron dynamics in graphene," *Phys. Rev. B,* vol. 77, no. 19, pp. 195415-1-195415-6, May 2008.

[74] A. Konar, T. Fang, and D. Jena, "Effect of high-k gate dielectrics on charge transport in graphene-based field effect transistors," *Phys. Rev. B,* vol. 82, no. 11, pp. 115452-1-115452-7, Sep. 2010.

[75] Z. Y. Ong and M. V. Fischetti, "Theory of interfacial plasmon-phonon scattering in supported graphene," *Phys. Rev. B,* vol. 86, no. 16, pp. 165422-1-165422-15, Jun. 2012.

[76] Z. Y. Ong and M. V. Fischetti, "Erratum: Theory of interfacial plasmon-phonon scattering in supported graphene," *Phys. Rev. B*, vol. 86, no. 19, pp. 199904-1-199904-2, Nov. 2012.

[77] Z. Y. Ong and M. V. Fischetti, "Top oxide thickness dependence of remote phonon and charged impurity scattering in top-gated graphene," *Appl. Phys. Lett.,* vol. 102, no. 18, pp. 183506-1-183506-4, May 2013.

[78] F. Wang, D. J. Cho, B. Kessler, J. Deslippe, P. J. Schuck, S. G. Louie, A. Zettl, T. F. Heinz, and Y. R. Shen, "Observation of excitons in one-dimensional metallic single-walled carbon nanotubes," *Phys. Rev. Lett.,* vol. 99, no. 22, pp. 227401-1-227401-4, Nov. 2007.

[79] M. V. Fischetti, D. A. Neumayer, and E. A. Cartier, "Effective electron mobility in Si inversion layers in metal-oxide-semiconductor systems with a high-k insulator: The role of remote phonon scattering," *J. Appl. Phys.,* vol. 90, no. 9, pp. 4587-4608, Nov. 2001.

[80] N. Paydavosi, M. M. Zargham, K. D. Holland, C. M. Dublanko, and M. Vaidyanathan, "Non-quasi-static effects and the role of kinetic inductance in ballistic carbon-nanotube transistors," *IEEE Trans. Nanotechnol.,* vol. 9, no. 4, pp. 449-463, Jul. 2010.

[81] D. Kienle, M. Vaidyanathan, and F. Léonard, "Self-consistent ac quantum transport using nonequilibrium Green functions," *Phys. Rev. B,* vol. 81, no. 11, pp. 115455-1-115455-10, Mar. 2010.

[82] D. Kienle and F. Léonard, "Terahertz response of carbon nanotube transistors," *Phys. Rev. Lett.,* vol. 103, no. 2, pp. 026601-1-026601-4, Jul. 2009.

[83] K. Holland, N. Paydavosi, and M. Vaidyanathan, "Self-consistent simulation of array-based CNFETs: Impact of tube pitch on RF performance," in *Proc. 14th IWCE,* Oct. 2010, pp. 291-294.

[84] L. Liao, Y. C. Lin, M. Bao, R. Cheng, J. Bai, Y. Liu, Y. Qu, K. L. Wang, Y. Huang, and X. Duan, "High-speed graphene transistors with a self-aligned nanowire gate," *Nature,* vol. 467, pp. 305-308, Sep. 2010.

[85] S. Lee, B. Jagannathan, S. Narasimha, A. Chou, N. Zamdmer, J. Johnson, R. Williams, L. Wagner, J. Kim, J. -. Plouchart, J. Pekarik, S. Springer, and G. Freeman, "Record RF performance of 45-nm SOI CMOS technology," in *Proc. IEEE IEDM,* Dec. 2007, pp. 255-258.

[86] D. H. Kim and J. A. del Alamo, "30-nm InAs PHEMTs with f_T = 644 GHz and $f_{\max}$ = 681 GHz," *IEEE Electron Device Lett.,* vol. 31, no. 8, pp. 806-808, Aug. 2010.

[87] W. Hafez, W. Snodgrass, and M. Feng, "12.5 nm base pseudomorphic heterojunction bipolar transistors achieving f_T = 710 GHz and f_{MAX} = 340 GHz," *Appl. Phys. Lett.,* vol. 87, no. 25, pp. 252109-1-252109-3, Dec. 2005.

[88] H. Dai, A. Javey, E. Pop, D. Mann, W. Kim, and Y. Lu, "Electrical transport properties and field effect transistors of carbon nanotubes," *Nano,* vol. 1, no. 1, pp. 1-13, Jul. 2006.

[89] D. M. Bagnall and M. Boreland, "Photovoltaic technologies," *Energy Policy,* vol. 36, no. 12, pp. 4390-4396, Dec. 2008.

[90] A. J. Nozik, "Quantum dot solar cells," in *Proc. Workshop on Nanostructures in Photovoltaics,* Apr. 2002, pp. 115-120.

[91] L. Tsakalakos, "Nanostructures for photovoltaics," *Mat. Sci. Eng. R: Reports,* vol. 62, no. 6, pp. 175-189, Aug. 2008.

[92] I. Hwang, C. R. McNeill, and N. C. Greenham, “Drift-diffusion modeling of photocurrent transients in bulk heterojunction solar cells,” *J. Appl. Phys.*, vol. 106, no. 9, pp. 094506-1-094506-10, Nov. 2009.

[93] J. Guo, M. A. Alam and Y. Yoon, "Theoretical investigation on photoconductivity of single intrinsic carbon nanotubes," *Appl. Phys. Lett.,* vol. 88, no. 13, pp. 133111-1-133111-3, Mar. 2006.

[94] L. E. Henrickson, "Nonequilibrium photocurrent modeling in resonant tunneling photodetectors," *J. Appl. Phys.,* vol. 91, no. 10, pp. 6273-6281, May 2002.

[95] ASTM Standard G173-03, "Standard Tables for Reference Solar Spectral Irradiances: Direct Normal and Hemispherical on 37° Tilted Surface," ASTM International, West Conshohocken, PA, 2012.

[96] W. Tress, K. Leo, and M. Riede, "Optimum mobility, contact properties, and open-circuit voltage of organic solar cells: A drift-diffusion simulation study," *Phys. Rev. B,* vol. 85, no. 15, pp. 155201-1- 155201-11, Apr. 2012.

[97] M. Lundstrom, *Fundamentals of Carrier Transport.* New York: Cambridge University Press, 2000.

[98] S. Luryi, "Quantum capacitance devices," *Appl. Phys. Lett.,* vol. 52, no. 6, pp. 501-503, Feb. 1988.

Appendix

A Full Collision Integrals

In the following, we provide expressions for the full collision integrals appearing in the BTE transport equations (3.1) – (3.3) due to the scattering of electrons with native acoustic, native optical, and the νth surface polar phonon modes, all of which can be calculated following the standard approach described in [97] and [67]. Similar to [67], only the process due to phonon emission is considered, in which case the collision integrals contain factors for the effective phonon occupancy of the form $(N_\gamma + 1)$, where $\gamma = \text{ac}, \text{ op}, \text{ or spp}, \nu$, corresponding to native acoustic, native optical, and the νth surface polar phonon modes. At a given temperature T and phonon energy $\hbar\omega_\gamma$, this occupancy is given through the Bose-Einstein distribution, $N_\gamma = 1/(e^{\hbar\omega_\gamma/k_B T} - 1)$, with k_B being Boltzmann's constant. Another factor that determines the magnitude of the collision integral for each scattering process γ is how many states are available at a specific energy $E(k)$ for an electron in a state k; this information is contained in the density of states of the semiconducting nanotube,

$$D[E(k)] = \frac{8}{3\pi a_0 t_E} \frac{[E(k) + E_0]}{\sqrt{[E(k) + E_0]^2 - E_0^2}} \tag{A.1}$$

where $E_0 = E_g/2$ is half of the tube's band gap.

A.1 Acoustic Phonons

The collision integral for acoustic phonons can be represented as

$$(S_o f)_{\mathrm{ac}} = \frac{\pi {D_A}^2}{2\rho v_s} D[E(k)]|2k|(N_{\mathrm{ac}} + 1) [f(-k) - f(k)] \tag{A.2}$$

where D_A = 5 eV is the deformation potential for acoustic phonons [59] [7], $\rho = 3 \times 10^{-15}$ kg m^{-1} is the mass density of the carbon nanotube [66], and $v_s = 21.1 \times 10^3$m s^{-1} is the velocity of sound [65]. Acoustic phonons have an energy $\hbar\omega_{\mathrm{ac}} = \hbar v_s \beta$, where $\beta = |k - k'|$ is the phonon momentum [97].

A.2 Optical Phonons

The collision integral for optical phonons can be expressed as

[7] The values of the acoustic and optical deformation potentials D_A and D_{op} appearing in (A.2) and (A.3) have been taken from [59], because the low- and high-bias mean-free paths computed from them were shown to be consistent with the experiments in the same paper; specifically, the experimental mean-free paths were found to lie within the upper bounds determined by D_A and D_{op}, which is to be expected since the experiments would include additional (above and beyond native) sources of scattering.

$$(S_o f)_{\mathrm{op}} = \frac{\pi {D_{\mathrm{op}}}^2}{2\rho\omega_{\mathrm{op}}} (N_{\mathrm{op}} \tag{A.3}$$

$$+ 1)\left\{D[E^+(k)][1 - f(k)] \sum_{k^+} f(k^+)\right.$$

$$\left. - D[E^-(k)]\, f(k) \sum_{k^-} [1 - f(k^-)]\right\}$$

where $D_{\mathrm{op}} = 12.8 \times 10^{10}$ eV m^{-1} is the deformation potential for optical phonons, which have an energy of $\hbar\omega_{\mathrm{op}} = 200$ meV [59]. The values $f(k^{\pm})$ refer to the electron distribution function for those states with an energy $\hbar\omega_{\mathrm{op}}$ higher and lower than the state k, with $E^{\pm} = E(k) \pm \hbar\omega_{\mathrm{op}}$. The summations in (A.3) are taken over all such allowed states $k^{\pm}$ [67].

A.3 Substrate Surface Polar Phonons (SPP)

The collision integral arising from the νth phonon mode of the substrate can be written as

$$(S_o f)_{\mathrm{spp},\nu} = F^2_{\mathrm{spp},\nu} \times P_{\mathrm{spp},\nu} \times \left(N_{\mathrm{spp},\nu} + 1\right) \tag{A.4}$$

$$\times \left\{D[E^+(k)]\,[1 - f(k)] \sum_{k^+} f(k^+) - D[E^-(k)] f(k) \sum_{k^-} [1 - f(k^-)]\right\}$$

where the symbols are as follows: $F_{\mathrm{spp},\nu}^2$ is the Fröhlich coupling of the νth surface phonon mode, given in [6] as $F_{\mathrm{spp},\nu}^2 = \frac{\hbar\omega_{\mathrm{spp},\nu}}{2\pi\varepsilon_0}\left(\frac{1}{\epsilon^\infty+1} - \frac{1}{\epsilon^0+1}\right)$, where ϵ^0 and ϵ^∞ are the low- and high-frequency dielectric constants of the polar substrate, and ε_0 is the vacuum permittivity; d is the tube diameter; $P_{\mathrm{spp},\nu}$ is indicative of the strength of the polarization field due to the substrate phonons at a distance s away from the substrate, given by

$$P_{\mathrm{spp},\nu} = \frac{4\pi^2 q^2 e^{-2sq_x}}{\hbar d q_x \sqrt{2\pi(d+2s)q_x}} \tag{A.5}$$

where s is the separation of the surface of the tube from the substrate and is set to 0.35 nm, as given by the van der Waals interaction [6], and $\hbar q_x$ is the electron momentum transfer along the tube axis, determined from [6, 68]

$$\hbar v_F q_x = \left\{\hbar\omega_{\mathrm{spp},\nu}\left(2\Delta + \hbar\omega_{\mathrm{spp},\nu}\right)\right\}^{1/2} \tag{A.6}$$

with $v_F \approx 10^8$ cm s^{-1} being the Fermi velocity and $2\Delta \approx 0.9/d$ eV. The values for the phonon energies and the dielectric constants of the polar substrates are taken from [79] and are listed in Table 3.1.

B Linearized Collision Integrals

The small-signal form of the collision integrals can be obtained in a manner similar to the linearization of the BTE outlined in Section 3.2.1 by inserting the ansatz $f = \bar{f} + \tilde{f}e^{j\omega t}$ into the full expressions (A.2) – (A.4) for the collision integrals and retaining terms to linear order in $\tilde{f}$. For each type of scattering, the results are as follows:

$$S_o\{\tilde{f},\bar{f}\}_{\mathrm{ac}} = \frac{\pi {D_A}^2}{2\rho v_s} D[E(k)]|2k|(N_{\mathrm{ac}}+1)[\tilde{f}(-k)-\tilde{f}(k)] \tag{B.1}$$

$$S_o\{\tilde{f},\bar{f}\}_{\mathrm{op}} = \frac{\pi {D_{\mathrm{op}}}^2}{2\rho\omega_{\mathrm{op}}}(N_{\mathrm{op}}+1) \tag{B.2}$$

$$\times\left\{D[E^+(k)]\sum_{k^+}[1-\bar{f}(k)]\tilde{f}(k^+)-\bar{f}(k^+)\tilde{f}(k)\right.$$

$$\left.-D[E^-(k)]\sum_{k^-}[1-\bar{f}(k^-)]\tilde{f}(k)-\bar{f}(k)\tilde{f}(k^-)\right\}$$

and

$$S_o\{\tilde{f},\bar{f}\}_{\mathrm{spp},\nu} = F_{\mathrm{spp},\nu}^2 \times P_{\mathrm{spp},\nu} \times \left(N_{\mathrm{spp},\nu} + 1\right) \tag{B.3}$$

$$\times \Bigg\{ D[E^+(k)] \sum_{k^+} \big[1 - \bar{f}(k)\big] \tilde{f}(k^+) - \bar{f}(k^+) \tilde{f}(k)$$

$$- D[E^-(k)] \sum_{k^-} \big[1 - \bar{f}(k^-)\big] \tilde{f}(k) - \bar{f}(k) \tilde{f}(k^-) \Bigg\}.$$