InAs Nanowire
Devices and Circuits

Kristofer Jansson

LUND UNIVERSITY

Doctoral Thesis
Electrical Engineering
Lund, August 2015

Academic thesis which, by due permission of the Faculty of Engineering at Lund University, will be
publicly defended on Friday, September 11, 2015, at 10^{15} a.m. in lecture hall E:1406, Department of
Electrical and Information Technology, Ole Römers väg 3, 223 63 Lund, Sweden, for the degree of Doctor
of Philosophy in Electrical Engineering. The academic thesis will be defended in English.
Abstract:
Since the introduction of the transistor and the integrated circuit, the semiconductor industry has developed at a remarkable pace. By continuously fabricating smaller and faster transistors, it has been possible to maintain an exponential increase in performance, a phenomenon famously described by Moore’s Law. Today, billions of transistors are integrated on a single chip and the size of a transistor is on the scale of tens of nanometres. Until recently, the improvements in performance and integration density have been mostly driven by scaling down the transistor size. However, as the length scale is rapidly approaching that of only a few atoms, this scaling paradigm may not continue forever. Instead, the research community, as well as the industry, is investigating alternative structures and materials in order to further increase the performance.

One emerging technology for use in future electronic circuits is transistors based on nanowires. The nanowire transistor structure investigated in this work combines a number of key technologies to achieve a higher performance than traditional Si-based transistors. Epitaxially grown nanowires are naturally oriented in the vertical direction, which means that the devices may be fabricated from the bottom and up. This three-dimensional structure allows a higher integration density and enables the gate to completely surround the channel in a gate-all-around configuration. Combined with a high-$\varepsilon$ dielectric, this results in an excellent electrostatic gate control. Furthermore, nanowires have the unique ability to combine semiconductor materials with significantly different lattice constants. By introducing InAs as a channel material, a much higher electron mobility than for Si is achieved.

In this work, simulations of nanowire-based devices are performed and the ultimate performance is predicted. A nanowire transistor architecture with a realistic footprint is proposed and a roadmap is established for the scaling of the device structure, based on a set of technology nodes. Benchmarking is performed against competing technologies, both from a device and circuit perspective. The physical properties of nanowire transistors, and the corresponding capacitor structure, are investigated by band-structure simulations. Based on these simulations, a ballistic transport model is used to derive the intrinsic transistor characteristics. This is combined with an extensive evaluation and optimization of the parasitic elements in the transistor structure for each technology node. It is demonstrated that an optimized nanowire transistor has the potential to operate at terahertz frequencies, while maintaining a low power consumption. A high quality factor and extremely high integration density is predicted for the nanowire capacitor structure. It is concluded that InAs nanowire devices show great potential for use in future electronic circuits, both in digital and analogue applications.

Keywords:
Nanowire, MOSFET, Metal-oxide-semiconductor field-effect transistor, InAs, III-V semiconductor, Band structure, RF, Transistor, Capacitor, Ballistic, Simulation, Modelling, Circuit, Amplifier.

Classification system and/or index terms (if any):

Language: English

Supplementary bibliographical information (if any):

Key title and ISSN:
Series of licentiate and doctoral theses, 1654-790X; No. 72

Recipient’s notes:
Number of pages: 178
Price:

Security classification: Unclassified
InAs Nanowire Devices and Circuits

Kristofer Jansson

Lund University

Doctoral Thesis
Electrical Engineering
Lund, August 2015
Since the introduction of the transistor and the integrated circuit, the semiconductor industry has developed at a remarkable pace. By continuously fabricating smaller and faster transistors, it has been possible to maintain an exponential increase in performance, a phenomenon famously described by Moore’s Law. Today, billions of transistors are integrated on a single chip and the size of a transistor is on the scale of tens of nanometres. Until recently, the improvements in performance and integration density have been mostly driven by scaling down the transistor size. However, as the length scale is rapidly approaching that of only a few atoms, this scaling paradigm may not continue forever. Instead, the research community, as well as the industry, is investigating alternative structures and materials in order to further increase the performance.

One emerging technology for use in future electronic circuits is transistors based on nanowires. The nanowire transistor structure investigated in this work combines a number of key technologies to achieve a higher performance than traditional Si-based transistors. Epitaxially grown nanowires are naturally oriented in the vertical direction, which means that the devices may be fabricated from the bottom and up. This three-dimensional structure allows a higher integration density and enables the gate to completely surround the channel in a gate-all-around configuration. Combined with a high-κ dielectric, this results in an excellent electrostatic gate control. Furthermore, nanowires have the unique ability to combine semiconductor materials with significantly different lattice constants. By introducing InAs as a channel material, a much higher electron mobility than for Si is achieved.

In this work, simulations of nanowire-based devices are performed and the ultimate performance is predicted. A nanowire transistor architecture with a realistic footprint is proposed and a roadmap is established for the scaling of the device structure, based on a set of technology nodes. Benchmarking is performed against competing technologies, both from a device and circuit perspective. The physical properties of nanowire transistors, and the corresponding capacitor structure, are investigated by band-structure simulations. Based on these simulations, a ballistic transport model is used to derive the intrinsic transistor characteristics. This is combined with an extensive evaluation and optimization of the parasitic elements in the transistor structure for
each technology node.

It is demonstrated that an optimized nanowire transistor has the potential to operate at terahertz frequencies, while maintaining a low power consumption. A high quality factor and extremely high integration density is predicted for the nanowire capacitor structure. It is concluded that InAs nanowire devices show great potential for use in future electronic circuits, both in digital and analogue applications.
Sedan introduktionen av transistorn och den integrerade kretsen har teknologin inom halvledarindustrin utvecklats i en mycket hög takt. Genom att bygga allt mindre och snabbare transistorer har en exponentiell förbättring av prestanda bibehållits sedan 1960-talet, ett fenomen känt som Moores Lag. Fram tills nyligen har utvecklingen mot högre prestanda och fler transistorer per chip framförallt drivits genom att minska storleken på transistorerna. I dagsläget är det inte ovanligt med miljardtals transistorer per chip och varje transistor är i storleksordningen av tiotals nanometer, motsvarande ett par hundra atomer eller en tiotusendel av tjockleken av ett hårråt.


I detta arbete presenteras simuleringar av nanotrådstransistorer där strukturen för transistorn optimeras och den förväntade prestandan uppskattas. Prestandan jämförs med konkurrerande teknologier och undersöks även från ett kretsperspektiv.
Acknowledgments

"It is the struggle itself that is most important. We must strive to be more than we are. It does not matter that we will not reach our ultimate goal. The effort itself yields its own reward."

– Gene Roddenberry, Creator of Star Trek

My deepest gratitude is extended to Lars-Erik Wernersson for giving me the opportunity to pursue nanoelectronics in a highly accomplished research group and ultimately earn a PhD degree in the field. For an interesting and fruitful collaboration, I honour Erik Lind, together with all former and present members of the nanoelectronics group. I would also like to take the opportunity to thank all colleagues, lecturers, and students encountered over the years for providing a friendly and creative environment.

With all my heart I thank my family for your undying and unconditional love and support. Although I am not always capable of communicating it, I am immensely thankful for everything you have done for me. Your virtuous compassion and integrity has made me into the man I am today. Although our academic tradition is next to none, you have always encouraged me to pursue my dreams and provided me with the tools to do it. It is in your honour that I proudly carry your legacy into uncharted territory.

Dear friends, I dare not mention your names in fear of forgetting someone, but you know who you are. Always supporting, you are my tether to this world and the source of my mental sanity. You have given me so much enjoyment and I regret not being able to spend as much time with you as I would like. May you all find happiness and success.

My final words I dedicate to my very good friend and colleague Martin Berg. You have been my greatest support over the last ten years and we have shared not only our education and scientific endeavours, but also many other great moments together. I am honoured to have met such a true humanist and visionary, dedicated to the progress of mankind. You are destined for greatness my friend and I wish you all luck. I thank you for these years together and hope that our paths will cross again in the future.

Live long and prosper,
Kristofer Jansson

Lund, August 2015
Contents

Abstract iii
Populärvetenskaplig Sammanfattning v
Acknowledgements vii
Contents ix
Preface xiii
  Structure of the Thesis ........................................ xiii
  Included Papers ........................................... xv
Abbreviations and Symbols xvii
  Abbreviations .............................................. xvii
  Greek Symbols ............................................. xviii
  Latin Symbols ............................................... xix
  Functions and Operators ................................... xxii

INTRODUCTION 1

1: Background 3
  1.1: The MOSFET Transistor ................................ 5
  1.2: Transistor Materials .................................. 6
  1.3: Advanced Transistor Structures ....................... 8

2: Nanowire Transistors 9
  2.1: Transistor Structure .................................. 10
  2.2: Nanowire Growth ..................................... 11
  2.3: Device Fabrication .................................... 13
  2.4: Current Status ....................................... 15

3: Figures of Merit 17
3.1: DC Transistor Metrics ........................................... 17
  3.1.1: Drive Current .............................................. 17
  3.1.2: Threshold Voltage .......................................... 18
  3.1.3: Transconductance .......................................... 18
  3.1.4: Subthreshold Slope ........................................ 19
  3.1.5: Output Conductance ....................................... 20
  3.1.6: Voltage gain ............................................... 20
3.2: RF Transistor Metrics ............................................ 20
  3.2.1: Cutoff Frequency ........................................... 20
  3.2.2: Maximum Oscillation Frequency ......................... 21
3.3: Capacitor Metrics ................................................ 22
  3.3.1: Quality Factor ............................................. 22
3.4: Digital Circuit Metrics .......................................... 22
  3.4.1: Propagation delay ......................................... 22
  3.4.2: Power Dissipation ......................................... 23
  3.4.3: Power-Delay Product ...................................... 23
  3.4.4: Energy-Delay Product .................................... 24
3.5: Analogue Circuit Metrics ....................................... 24
  3.5.1: Maximum Power Gain .................................... 24
  3.5.2: IIP$_3$ ..................................................... 25
  3.5.3: 1 dB Compression Point .................................. 25

4: Band Structure Simulations ........................................ 27
  4.1: Semiconductor Physics ....................................... 27
    4.1.1: Nanowires ............................................... 28
  4.2: Band Structure Simulations .................................. 28
    4.2.1: Carrier Concentration ................................ 30
    4.2.2: Non-Parabolic Bands ................................... 32
    4.2.3: Electrostatics ........................................... 32
    4.2.4: Self-Consistency ....................................... 33
  4.3: Ballistic Transistor Modelling ............................... 33
    4.3.1: Gate Capacitance ....................................... 35
    4.3.2: Gate Barrier ............................................ 36
    4.3.3: Transistor Characteristics ............................ 37
    4.3.4: Short-Channel Effects ................................ 38
This thesis is the culmination of more than five years of work in the nanoelectronics group at Lund University under the supervision of Professor Lars-Erik Wernersson. The focus of this work is on the performance evaluation and optimization of a vertically aligned InAs nanowire transistor architecture, as well as the corresponding capacitor structure.

STRUCTURE OF THE THESIS

The thesis is structured to introduce two comprehensive papers published in Transactions on Electron Devices, encompassing simulation, optimization and benchmarking of InAs nanowire transistors and capacitors. This is supported by a submitted manuscript to Solid-State Electronics, regarding ballistic transistor simulations. Further manuscripts include a circuit simulation paper regarding amplifier design using these transistors and a paper describing fabrication, measurements, and modelling of an InAs nanowire capacitor. Finally, a manuscript concerning a fabricated nanowire transistor structure is included.

- **Chapter 1 - Background**
  This chapter gives a short introduction to the field with a historical perspective and motivation for the research.

- **Chapter 2 - Nanowire Transistors**
  This chapter introduces the transistor structure, discusses fabrication, and gives an overview of the current state of the art devices.

- **Chapter 3 - Figures of Merit**
  This chapter defines relevant figures of merit used for the performance benchmarking and optimization.

- **Chapter 4 - Band Structure Simulations**
  This chapter discusses simulations of the intrinsic band structure of nanowire transistors and capacitors. A customized Schrödinger-Poisson solver is presented to simulate the band structure within the nanowire.
From this, the intrinsic performance of the transistor and capacitor structure is calculated. This chapter is related to Paper II and III.

- **Chapter 5 - Parasitic Elements**
  This chapter discusses the parasitic elements in the vertical nanowire transistor structure. In order to quantify the parasitic capacitances, the device is modelled using a field solver known as FastCap. A MATLAB interface to automatically generate 3D structures using a set of structural parameters is developed in order to enable optimization of the structure. The validity of the parasitic capacitances within the structure is verified by measurements on fabricated devices. This chapter is related to Paper I and VI.

- **Chapter 6 - Benchmarking**
  In this chapter, the transistor architecture is optimized and benchmarked against competing technologies. The intrinsic transistor data is combined with the calculated parasitic elements in order to evaluate the RF performance. A standardized scaling of the device architecture is introduced in the form of a roadmap, establishing a set of scaling rules based on fabrication experience. For each of these technology nodes, ranging from nanowire diameters from 50 nm down to 8 nm, a complete optimization of the structural parameters is performed and the ultimate performance is predicted. This chapter is related to Paper I and III.

- **Chapter 7 - Design variations**
  In this chapter, alternative transistor structures are investigated and the impact on the performance is quantified. This chapter is related to Paper I.

- **Chapter 8 - Nanowire Capacitors**
  This chapter discusses fabrication, measurements, and simulations of a nanowire capacitor structure. The influence of the resistance within the channel of a long nanowire capacitor is modelled using a distributed RC-network. This allows the frequency response and quality factor to be calculated. The model is also used to fit the measurement data of a fabricated nanowire capacitor in order to extract the nanowire resistivity. This chapter is related to Paper II and V.

- **Chapter 9 - Circuit Simulations**
  In this chapter, a number of benchmarking circuits are simulated using the developed nanowire device models. This chapter is related to Paper I-V.
Chapter 10 - Outlook
This chapter summarizes the conclusions from the thesis and discusses the future of the research field.

INCLUDED PAPERS

The following papers are included in this thesis and the respective published or draft versions are appended at the back of this thesis.

► Performed almost all of the work.

► Performed almost all of the work.

► Performed almost all of the work.

► Performed almost all of the work.

► Performed modelling and simulations. Assisted with process development, mask design, and measurements.

► Performed simulations of the parasitic capacitances in the transistor structure.
Abbreviations and Symbols

ABBREVIATIONS

1D One-dimensional
2D Two-dimensional
3D Three-dimensional

AC Alternating current
Al₂O₃ Aluminium trioxide
ALD Atomic layer deposition
Au Gold

BCB Benzocyclobutene

CF Capacitance-frequency
CG Common-gate
CMOS Complementary metal-oxide-semiconductor
COMSOL A computer software package for simulations
CPU Central processing unit
CS Common-source
CV Capacitance-voltage

D Drain
DC Direct current
DOS Density of States

EBL Electron beam lithography
EMA Effective mass approximation

FastCap A computer software package for calculations of capacitances
FinFET Fin field-effect transistor

G Gate
GaAs Gallium arsenide
GaP Gallium phosphide
GaSb Gallium antimonide
Ge Germanium
GPU Graphics processing unit
HEMT High electron mobility transistor
HfO$_2$ Hafnium dioxide
HSQ Hydrogen silsesquioxane

III-V Group three and five in the periodic system of elements
InAs Indium arsenide
InP Indium phosphide
InSb Indium antimonide
ITRS International Technology Roadmap for Semiconductors

MATLAB A computer software package for calculations
MOS Metal-oxide-semiconductor
MOSFET Metal-oxide-semiconductor field-effect transistor

NMOS N-type metal-oxide-semiconductor
NP Non-parabolic
NW Nanowire
NWFET Nanowire field-effect transistor

PDN Pull-down network
PMOS P-type metal-oxide-semiconductor
PUN Pull-up network

RC Resistance and capacitance
RF Radio frequency

S Source
S1813 An organic photoresist
SCAM A computer software package for nodal analysis in MATLAB
SEM Scanning electron microscopy
Si Silicon
Si$_3$O$_4$ Silicon nitride
SiO$_2$ Silicon dioxide
SP Schrödinger-Poisson

TB Tight-binding
Ti Titanium

UV Ultraviolet

VLS Vapour-liquid-solid, a technique for nanowire growth.

W Tungsten
GREEK SYMBOLS

\( \alpha \) Nonparabolicity factor
\( \bar{\alpha} \) Modified nonparabolicity factor

\( \varepsilon \) Permittivity
\( \varepsilon_0 \) Permittivity of free space
\( \varepsilon_{\text{ox}} \) Relative permittivity of the gate dielectric
\( \varepsilon_s \) Relative permittivity of the semiconductor
\( \varepsilon_{\text{spc}} \) Relative permittivity of the spacer layer

\( \kappa \) Relative permittivity

\( \lambda \) Natural Length for the nanowire geometry

\( \nu \) Azimuthal quantum number

\( \phi_s \) (V) Surface potential

\( \rho \) (Ωm) Resistivity
\( \rho_{\text{D,e}} \) Drain electrode resistivity
\( \rho_G \) Gate electrode resistivity
\( \rho_m \) Electrode resistivity
\( \rho_{\text{S,e}} \) Source electrode resistivity
\( \rho_w \) Ungated nanowire resistivity
\( \rho_c \) (Ωm\(^2\)) Contact resistivity

\( \sigma \) Standard deviation

\( \mu \) (m\(^2\)/Vs) Carrier mobility

\( \omega \) (Hz) Angular frequency

LATIN SYMBOLS

\( A_v \) (dB) Voltage gain

\( C \) (F) Capacitance
\( C_{\text{ds}} \) Drain-source capacitance
\( C_g \) Total gate capacitance
\( C_{\text{gd}} \) Gate-drain capacitance
\( C_{\text{gd,e}} \) Gate-drain electrode capacitance
\( C_{\text{gd,we}} \) External gate-drain nanowire capacitance
\( C_{\text{gd,wi}} \) Internal gate-drain nanowire capacitance
Intrinsic gate capacitance  \( C_{g,i} \)
Gate-source capacitance  \( C_{gs} \)
Gate-source electrode capacitance  \( C_{gs,e} \)
External gate-source nanowire capacitance  \( C_{gs,we} \)
Internal gate-source nanowire capacitance  \( C_{gs,wi} \)
Geometric gate capacitance  \( C_{ox} \)
Quantum capacitance  \( C_q \)
Semiconductor capacitance  \( C_s \)

1 dB compression point  \( CP_{1dB} \)

Nanowire diameter  \( d_{nw} \)

Energy  \( E \)
  Fermi level  \( E_f \)
  Drain Fermi level  \( E_{f,d} \)
  Source Fermi level  \( E_{f,s} \)
  Subband energy levels  \( E_{n,v} \)

Energy-delay product  \( EDP \)

Frequency  \( f \)
  Maximum oscillation frequency  \( f_{\text{max}} \)
  in the reverse direction  \( f_{\text{max},i} \)
  Cutoff frequency  \( f_T \)
  in the reverse direction  \( f_{T,i} \)
  Geometric mean of  \( f_T \) and  \( f_{\text{max}} \)  \( f_{T,max} \)
  in the reverse direction  \( f_{T,max,i} \)

Transconductance  \( g \)
  Output conductance  \( g_d \)
  Transconductance  \( g_m \)

Power gain  \( G \)
  Maximum power gain  \( G_{\text{max}} \)

Planck’s constant  \( h \)
  Reduced Planck’s constant  \( \hbar \)

Current  \( I \)
  Drain current  \( I_D \)
  Source current  \( I_S \)

Third-order intercept point  \( \text{IIP}_3 \)

Current density  \( J \)
  Current density from source to drain  \( J^- \)
\( J^+ \) Current density from drain to source

\( k \) (m\(^{-1}\)) Wave vector
\( k_B \) (J/K) The Boltzmann constant

\( L \) (m) Length
- \( L_e \) Electrode length
- \( L_{ext} \) Electrode length extension
- \( L_g \) Gate length
- \( L_T \) Transfer length

\( m_e \) (kg) Electron mass
\( m^* \) Effective mass

\( n \) Radial quantum number
\( n^- \) (m\(^{-3}\)) Mobile carriers in the source to drain direction
\( n^+ \) (m\(^{-3}\)) Mobile carriers in the drain to source direction
\( n_x \) Number of nanowires in the array in the x direction
\( n_y \) Number of nanowires in the array in the y direction
\( N_D \) Doping concentration

\( P_A \) (W) Average power dissipation
\( PDP \) (J) Power-delay product

\( q \) (C) Elementary charge
\( Q \) Quality factor

\( R \) (\( \Omega \)) Resistance
- \( R_C \) Contact resistance
- \( R_C \) Capacitor series resistance
- \( R_D \) Drain resistance
- \( R_{D,e} \) Drain contact resistance
- \( R_{D,e} \) Drain electrode resistance
- \( R_{D,w} \) Drain nanowire resistance
- \( R_e \) Electrode resistance
- \( R_G \) Gate resistance
- \( R_S \) Source resistance
- \( R_{S,e} \) Source contact resistance
- \( R_{S,e} \) Source electrode resistance
- \( R_{S,w} \) Source nanowire resistance
- \( R_w \) Ungated nanowire resistance
- \( R_{w,tot} \) Total ungated nanowire resistance
- \( R_{NW} \) (m) Nanowire radius
$S_{nw}$ (m) Nanowire array spacing
$S$ (dec/V) Subthreshold slope

$t$ (m) Thickness
  $t_D$ Drain electrode thickness
  $t_{gd}$ Drain spacer thickness
  $t_{gs}$ Source spacer thickness
  $t_{ox}$ Gate dielectric thickness
  $t_S$ Source electrode thickness

$t_p$ (s) Propagation delay
$T$ (K) Temperature

$U_B$ (V) Gate barrier height

$V$ (V) Voltage
  $V_{DD}$ Supply voltage
  $V_{ds}$ Drain-source voltage
  $V_{gd}$ Gate-drain voltage
  $V_{gs}$ Gate-source voltage
  $V_{gs,i}$ Intrinsic gate-source voltage
  $V_{in}$ Input voltage
  $V_{od}$ Gate overdrive voltage
  $V_{out}$ Output voltage
  $V_t$ Threshold voltage

$W$ (m) Gate width
$W_e$ (m) Electrode width

$X_C$ (Ω) Capacitor reactance

FUNCTIONS AND OPERATORS

$E_{n,v}(K)$ Subband energy dispersion relation

$f(E)$ Fermi-Dirac distribution

$g(E)$ Density of states

$n(r)$ Radial carrier concentration

$v(E)$ Carrier velocity
$V(r)$ Radial electrostatic potential

$\psi_{n,v}(r)$ Radial eigenfunctions
INTRODUCTION
Background

“If the auto industry advanced as rapidly as the semiconductor industry, a Rolls Royce would get half a million miles per gallon, and it would be cheaper to throw it away than to park it.”

– Gordon Moore, Cofounder of Intel Corporation

Since the introduction of the first integrated circuits in the early 1960s, the evolution of modern electronics has experienced an exponential improvement in integration density and performance. This phenomenon was famously observed by Gordon E. Moore in 1965, who formulated what is now known as Moore’s law, stating that the number of transistors in an integrated circuit roughly doubles every year, later revised to every 24 months [1,2]. This statement has been transformed into somewhat of a self-fulfilling prophecy, as the semiconductor industry is struggling to keep up with the expectations from both consumers and investors. Remarkably, this trend has been more or less maintained over the last 50 years, and will probably be sustained at least in the near future. The evolution of the number of transistors in high-performance desktop CPUs is presented in Figure 1.1, for a set of milestone Intel products.

Although many technological hurdles have been overcome, the basis for this improvement has mainly been the continuous scaling of the transistors to even smaller dimensions. A reduction in the size of each transistor is also an economic incentive, as it reduces the die area required in the fabrication. The evolution of the size of a typical CMOS transistor is presented in Figure 1.2, where the size is defined as the commonly referred technology node. Until the early 2000’s, this traditional scaling has been the main focus of the industry, with only minor changes to the actual transistor structure and materials. However, as the size of the transistors are rapidly approaching a length scale of only a few atoms, it is apparent that this scaling can not continue indefinitely. There have been different approaches to address this problem, focusing on various aspects of the transistor architecture.

Since around 2005, there has been a paradigm shift in the overall architecture of integrated circuits, as the development has shifted from essentially a large single processing unit to multiple parallel cores. Although suitable for highly parallel problems, such as massive graphical calculations in a
Figure 1.1: Evolution of the number of transistors in consumer-grade desktop CPUs.

Figure 1.2: Evolution of the transistor size in commercial CMOS production. The referred size is the semiconductor manufacturing process node.
GPU, the performance gain is much smaller than ideal for most applications. This limitation was famously described by Gene Amdahl in 1967, henceforth known as Amdahl's law [3]. The performance gain from increasing the number of processing cores is illustrated in Figure 1.3. It is clearly observed that a high level of parallelization is required to take advantage of a large number of processing cores. This limitation in parallel computing is a driving force for the continued improvement of the performance of each processing unit instead of just increasing the number of cores. Ultimately, this requires improvements in the speed of the most fundamental part of the processing unit, the metal-oxide-semiconductor field-effect transistor (MOSFET).

![Figure 1.3: Amdahl's law illustrates the effective performance gain from increasing the number of processing threads.](image)

### 1.1 THE MOSFET TRANSISTOR

The MOSFET is the most common type of transistor in modern electronic circuits and has been completely dominating in microprocessors since the 1970s. A MOSFET is a three terminal device, where the conductivity between the drain and source terminal is regulated by the voltage applied at the gate terminal. It is represented by the symbol in Figure 1.4(a).

The fundamental operation of the MOSFET is based on a metal-oxide-semiconductor (MOS) interface, which forms a capacitor between the gate and the semiconductor substrate, separated by the gate dielectric. By applying a voltage on the gate terminal, opposite charges are accumulated near the surface of the semiconductor, which form a channel of high conductivity. This allows a current to flow when a voltage is applied between the source and drain terminals. A schematic illustration of a planar MOSFET structure is presented in Figure 1.4(b).
For low voltages between the drain and source terminals, the MOSFET operates like a voltage-controlled resistor. This behaviour is similar to a simple potentiometer, although the resistance is set by the applied gate voltage instead of turning a knob. When the drain-source voltage is further increased, the MOSFET eventually enters saturation, where the transistor acts as a voltage-controlled current source, dependent only on the gate-source voltage. As a small gate-source voltage may be amplified to a large current, this is utilized in the design of amplifier circuits. Furthermore, the ability to turn the current through the MOSFET on and off by applying either a high or low voltage on the gate is the basis for digital logic, where the high and low states represent 0 and 1, respectively.

1.2 TRANSISTOR MATERIALS

As the current through the transistor is basically a flow of electric charge, it is proportional to the number of charge carriers and their velocities. The amount of charge carriers is dependent on the gate capacitance, which means that the performance for a fixed transistor size may be improved by increasing the gate capacitance per area. This has traditionally been performed by decreasing the thickness of the gate dielectric. However, leakage due to quantum mechanical tunnelling becomes a problem for very thin dielectric films.

Another approach is to increase the relative permittivity of the gate dielectric. A recent development has been the replacement of SiO₂, traditionally used as the gate dielectric, with so-called high-κ dielectric films, such as HfO₂ or Al₂O₃. As a high-κ dielectric has a much higher permittivity than SiO₂, the gate capacitance may be increased without a further reduction of the thickness of the dielectric, thus avoiding an increased gate leakage current. Commercially, a high-κ dielectric was introduced into production by Intel in 2007 for the 45 nm technology node.
Table 1.1: Mobilities of common semiconductor materials.

<table>
<thead>
<tr>
<th></th>
<th>(cm²/Vs)</th>
<th>Si</th>
<th>Ge</th>
<th>GaSb</th>
<th>GaAs</th>
<th>InP</th>
<th>InAs</th>
<th>InSb</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Electron Mobility</strong></td>
<td></td>
<td>1400</td>
<td>3900</td>
<td>3000</td>
<td>8500</td>
<td>5400</td>
<td>40000</td>
<td>77000</td>
</tr>
<tr>
<td><strong>Hole Mobility</strong></td>
<td></td>
<td>450</td>
<td>1900</td>
<td>1000</td>
<td>400</td>
<td>200</td>
<td>500</td>
<td>850</td>
</tr>
</tbody>
</table>

The electrical mobility is a material property that describes how the carrier velocity is related to an applied electric field. A potentially ground-breaking development would be to change the channel material from Si to a high-mobility III-V material, such as InAs. The improved mobility of InAs allows the electrons to travel at a much higher velocity, thereby potentially increasing the performance. An overview of the mobility for a number of common semiconductor materials is presented in Table 1.1. It is evident that the electron mobility is much higher in most III-V semiconductors than in Si.

Due to the difference in electron and hole mobility, different materials should ideally be used for n- and p-type transistors. In p-type transistors, the current is conducted by holes and thus the important metric is the hole mobility. A good alternative for a p-type transistor would be to replace the channel material with Ge, which has a higher hole mobility than Si. Among III-Vs, the most promising p-type material is GaSb.

Introduction of new channel materials into a CMOS architecture has been discussed for many years. The current ITRS roadmap predicts a transition to III-V MOSFETs in CMOS technology in 2018–2026, with an expected performance increase in the switching speed of about 50%, while reducing the power dissipation by 40% [4]. There are, however, significant technological challenges that must be overcome, such as the lattice mismatch between the Si substrate and the introduced material.

![Figure 1.5: Different gate configurations, illustrating the improved electrostatic gate control in a multi-gate structure.](image-url)
1.3 ADVANCED TRANSISTOR STRUCTURES

Another development has been the transition from a planar transistor architecture to more intricate 3D structures with a multi-gate design. This was introduced commercially into Si CMOS by Intel in 2012 for the 22 nm node, with a transistor technology known as FinFETs or tri-gate transistors. In this technology, the channel is extended in the vertical dimension and is surrounded by the gate, not only from the top, but also from the sides. This allows for an improved electrostatic gate control, which increases the performance. A transition from a planar architecture tri-gates is estimated by Intel to give a performance increase of 37%, with a reduction of the power dissipation by 50% [5]. A further improvement would be a gate that completely surrounds the channel, in a so-called gate-all-around configuration. An overview of different gate configurations is presented in Figure 1.5.
In this work, a transistor architecture utilizing nanowires is investigated. The transistor structure is fundamentally a III-V MOSFET, which offers the prospect of extending the transistor roadmap by combining a number of key technologies [4, 6–10]. The cylindrical geometry of the nanowires enables the gate to circumvent the channel in a gate-all-around configuration, offering excellent electrostatic control, exceeding that of both traditional MOSFET structures and the more recent tri-gate architecture [11–14]. Together with the introduction of a high mobility III-V channel material and a high-κ gate dielectric, this transistor structure has the potential to achieve a performance beyond that of traditional Si-based devices, while at the same time reducing the power dissipation. [15–17]. According to the ITRS roadmap and estimations by Intel, a device like has the potential to improve the performance by more than 100% in comparison to a planar Si CMOS architecture, while reducing the power consumption by 80% [4, 5].

**Figure 2.1:** Vertical nanowire transistor structure in comparison to a standard planar transistor technology.

In contrast to a traditional MOSFET, the proposed InAs nanowire transistor architecture is aligned in the vertical dimension, as illustrated in Figure 2.1. Interestingly, the nanowire transistor architecture is inherently vertical as
the nanowires are grown using seed particles on a substrate. The vertical alignment offers the potential for a higher integration density and even the prospect to stack components on top of each other.

The utilization of III-V nanowire transistors in electronic circuits is attractive, with the potential of achieving a high performance while maintaining a low power dissipation [6]. The gate-all-around implementation combined with the advantageous transport properties of III-V materials promise a high transconductance, combined with good sub-threshold characteristics [11,14,18]. Compared to a high performance high-electron-mobility transistor (HEMT), the footprint of a nanowire MOSFET structure is much smaller, ultimately approaching the size of a conventional MOSFET architecture. Furthermore, the unique ability of nanowires to combine materials with significant differences in lattice constant is a major selling point [19].

2.1 TRANSISTOR STRUCTURE

An 3D rendering of the studied nanowire transistor structure is presented in Figure 2.2. This structure is similar in design to previously fabricated devices [20–23]. It is based on a hexagonal array of vertically arranged InAs nanowires, typically ranging from a single nanowire up to a few hundred nanowires. Each nanowire has a diameter of a few tens of nanometers and is a few hundred nanometers long.

![Figure 2.2: A 3D rendering of the nanowire transistor array, produced by Martin Berg.](image)

The nanowires are covered by a high-κ gate dielectric, such as HfO₂ or Al₂O₃, and is surrounded by the gate in a gate-all-around configuration. For a single nanowire, this means that the gate width is defined as the circumference of the nanowire. However, scaling the diameter is not a good
method to increase the gate width, as this would drastically change the transport properties within the nanowire. Instead, the drive current is scaled by changing the number of nanowires in the array. The total gate width of the transistor is thus defined as the number of parallel nanowires, multiplied with the nanowire circumference. Unless stated otherwise, the performance is normalized to this gate width.

In this architecture, vertically arranged electrodes are used to contact the nanowires. The source, gate, and drain electrodes are stacked on top of each other, separated by spacer materials. The electrodes are usually metal based, although highly doped semiconductors may also be used. In the simulations presented in this work, a gate electrode of W is used, whereas Au is used for the source and drain contacts. Spacer layers are used to separate the electrodes. Unless otherwise stated, a spacer layer with a relative permittivity of three is considered, roughly corresponding to the commonly used photoresist from the S18xx family.

A schematic of the nanowire transistor architecture is presented in Figure 2.3, illustrating the electrode layout as well as the location of the parasitic capacitances and resistances in the structure. The series resistance is caused mainly by the resistance from the ungated parts of the nanowires, combined with the contact resistance between the nanowires and electrodes. There are also large overlap capacitances in the structure, caused by the direct overlap between the vertically aligned electrodes. These parasitic capacitances, together with the series resistance, are the main parameters which cause degradation of the extrinsic device performance.

2.2 NANOWIRE GROWTH

Nanowires are grown on a substrate using seed particles, which are used as nucleation sites for the epitaxial growth. A commonly used seed particle material is Au, although other metals may be used. An alternative is to utilize seed crystals of a semiconductor material. Seed particles may be deposited using aerosols, although in this case the placement of the particles is random. In order to fabricate a device of the nanowires, a method to place the nanowires in pre-defined positions is necessary. This may be achieved by using a number of different lithography techniques. One method is to use electron beam lithography (EBL) to write dots on the substrate where the nanowires should be placed. An alternative would be to use nanoimprint lithography [24].

In an EBL-based patterning process, the focused electron beam is used to expose dots in a resist covering the substrate, which results in holes after the development of the resist. Depending on the dose, as well as the focus of the
Figure 2.3: (a) A cross-sectional layout of the transistor structure, showing the location of the parasitic elements. (b) Top view of the structure, illustrating the electrode configuration, as well as the arrangement of the nanowire array.

electron beam, the diameter of the seed particles may be controlled. Due to the requirements on the focus and limitations in the equipment, it is much easier to write large particles. EBL-defined seed particles down to 15 nm have been demonstrated, although diameters in the range of 20-40 nm are most commonly used [20, 21, 25].

The EBL patterning is followed by the deposition of a thin layer of the seed particle material using evaporation. Dissolving the remaining resist leads to lift-off of the metal film on top resulting in seed particles of the chosen material on the substrate. From the seed particles, the nanowires are grown epitaxially using the vapour-liquid-solid (VLS) method. This is performed in a reactor dedicated to metalorganic vapour phase epitaxy (MOVPE). The length of the grown nanowires are dependent not only on the growth conditions and growth time, but also on the diameter of the seed particles and the distance between the particles. In general, it is more difficult to reliably grow thin and tightly spaced nanowires. Nanowire growth has been demonstrated on various semiconductor substrates, including Si. In order to simplify growth of InAs nanowires on a Si-substrate, a thin planar InAs layer may be used [26]. This layer has the additional benefit of enabling device isolation by mesa etching.

During the growth, it is possible to change the material in order to produce heterostructure nanowires. Furthermore, doping of varying concentrations may be introduced into the nanowire. However, it is very difficult to achieve
sharp doping interfaces, due to the memory effect of residual dopants still present in the gold particle, as well as diffusion.

2.3 DEVICE FABRICATION

After growth, the nanowires are covered by a few nanometre thick film, consisting of a high-\(\kappa\) dielectric, such as HfO\(_2\) or Al\(_2\)O\(_3\). This is performed using atomic layer deposition (ALD) in a dedicated reactor. Using ALD, the thickness of the film may be controlled down to a single atomic layer using a self-limiting cycle-based approach. Multiple layers of dielectric films may be produced, composed of different materials.

![Figure 2.4: Cross-section (a) and top-down view (b) of a fabricated nanowire transistor structure after gate definition. The images are taken in a scanning electron microscope (SEM) and coloured by hand. The images are reproduced courtesy of Elvedin Memisevic and Sofia Johansson [27], respectively.](image)

The nanowires are contacted by vertically arranged electrodes, typically fabricated by evaporating or sputtering metals, such as Ti, W and Au. Alternatively, the bottom electrode may be replaced by a highly doped InAs epitaxial layer [20, 26]. Patterning of the electrodes may be performed using UV-lithography or EBL. EBL has a much higher resolution, which is essential in order to reduce the size of the metal electrodes and thus the parasitic capacitances. Etching of the electrodes may be performed either using etch-back or lift-off. Both dry and wet etches are typically used during the fabrication. SEM images of a nanowire array after patterning of the gate electrode is presented in Figure 2.4 and an optical image of a finished nanowire transistor is presented in Figure 2.5.

The electrodes are separated by spacer layers, which ideally should have a low relative permittivity in order to minimize the parasitic capacitances. The
material should also be mechanically stable and it is important to be able to accurately control the thickness of the deposited film. Typical spacer layers include organic photoresists, such as S1813 or BCB, or inorganic films, such as Si₃N₄, SiO₂, or HSQ [28,29]. An alternative is to completely remove the spacer layer, although that complicates the fabrication and may cause a mechanically unstable structure.

Photoresists have the advantage of being easy to deposit on the sample by spinning and baking. However, the thickness of the resist in a spinning process is difficult to control, which means that the layer generally needs to be etched down to the desired thickness. This requires time-consuming manual work by moving the sample between the etching equipment and an SEM. Ultimately, the accuracy of this method is limited to about 50 nm. Furthermore, the thickness of the resist is highly inconsistent over the sample. Another disadvantage is that organic photoresists are sensitive to high temperatures, which is a problem if any post-deposition annealing of the sample is performed.

Inorganic materials are usually deposited layer by layer, which means that the thickness may be accurately controlled down to the nanometre scale. However, this requires dedicated equipment for the deposition. Furthermore, the permittivity of most alternative materials is higher than for S1813. There are also considerations based on etching properties and selectivity, which may influence the choice of material.
2.4 CURRENT STATUS

The best fabricated vertical nanowire transistors currently demonstrate a performance exceeding 100 GHz. In the current state of the art devices, the nanowires usually have a diameter of about 30 nm, with a 6 nm thick gate dielectric consisting of a dual-layer of Al$_2$O$_3$/HfO$_2$. The nanowires are arranged in a hexagonally close-packed array of about 100 nanowires. In order to decrease the parasitic capacitances, the nanowires are packed closely together and the gate electrode is patterned by EBL.

The bottom contact consists of a highly doped InAs layer, whereas the top contact is fabricated from a stack of Ti/W/Au and the gate of W. An inorganic spacer of SiO$_2$ is used at the source side, which may be as thin as a few tens of nanometres. For the top spacer, an organic photoresist, typically S1813, is used. Other spacer layer materials, such as HSQ are being investigated as an alternative. The series resistance is reduced by a high doping in the ungated top and bottom segments of the nanowire. In order to further reduce the resistance, the diameter of the ungated parts of the nanowires may be increased by regrowth. Alternatively, a thicker nanowire may be used, where the channel diameter is reduced by digital etching.
In order to quantify the performance of electronic devices, well-defined figures of merit are required. Preferably, these metrics should be universally accepted, in order to simplify benchmarking against competing technologies. In this chapter, a number of important figures of merit are defined, separated into four categories: DC transistor metrics, RF transistor metrics, capacitor metrics, and circuit metrics.

3.1 DC TRANSISTOR METRICS

The main DC transistor metrics are simply derived from the transfer and output characteristics of the transistor. Of main importance are the drive current, transconductance, output conductance, and subthreshold slope.

3.1.1 DRIVE CURRENT

The most fundamental function of a MOSFET is to deliver a current based on the applied voltages on the gate and drain terminals. The dependence of the current on the drain-source voltage is called the output characteristics for the transistor, whereas the same for the gate-source voltage is called the transfer characteristics. Typical output and transfer characteristics for a MOSFET are presented in Figure 3.1.

The current level is an important metric for both analogue and digital applications. In digital applications, the current in both the on- and off-state is important. A high on current allows the load capacitances in logic gates to be more quickly charged, which decreases the propagation delay and thus allows for a faster switching speed. However, a low off-current is required in order to reduce the leakage and thus the static power dissipation. A common figure of merit for digital applications is the ratio between the on- and off-current.

In order to be useful for benchmarking, a comparison should be made under the same bias conditions. In this work, the current is given for a gate overdrive and drain-source voltage of 0.5 V, unless stated otherwise. Furthermore, as the drive current scales with the width of the transistor, it needs to be normalized. In the case of a nanowire transistor in a gate-all-around configuration, the total gate width is defined as the circumference of
the nanowire, multiplied with the number of nanowires in the array.

### 3.1.2 THRESHOLD VOLTAGE

The threshold voltage, $V_t$, is ideally defined as the voltage where the transistor switches state from off to on. In experimental data it is usually derived as an extrapolation from the linear slope of the transfer characteristics, as shown in Figure 3.1(b). When modelling the device physics it may be derived from the flat-band voltage. In this work it is defined as the applied gate-source voltage, for which the Fermi level is aligned with the bottom of the lowest subband. In this case, no current is flowing at a temperature of 0 K.

![Figure 3.1](image_url)

**Figure 3.1**: (a) Output and (b) transfer characteristics for a MOSFET.

### 3.1.3 TRANSCONDUCTANCE

A metric related to the drive current is the transconductance, $g_m$, which states how much current, $I_D$, is gained from a corresponding change in the gate-source voltage, $V_{gs}$. This is illustrated in Figure 3.1(b), with the resulting transconductance in 3.2(a):

$$g_m = \frac{\partial I_D}{\partial V_{gs}}.$$  \hspace{1cm} (3.1)

This metric is of interest in analogue circuit design as it essentially limits the potential amplification for the transistor. For benchmarking, either the maximum peak transconductance is given or a value at a specific bias point. Unless otherwise stated, the transconductance in the work will be given for a gate overdrive and drain-source voltage of 0.5 V.

It is important to distinguish between the intrinsic transconductance and the extrinsic transconductance, which includes the degradation from the parasitic resistances. When measuring on a fabricated device, the extrinsic
transconductance is obtained, whereas a band structure simulation models only the intrinsic transconductance. The extrinsic performance of the transistor may be calculated by adding the parasitic elements to the intrinsic transconductance and vice versa.

![Figure 3.2](image-url)

**Figure 3.2:** (a) Transconductance, (b) logarithmic transfer characteristics, and (c) output conductance for a MOSFET.

### 3.1.4 SUBTHRESHOLD SLOPE

The subthreshold slope, $S$, is defined similarly to the transconductance, although in the subthreshold region and on a logarithmic scale, as illustrated in Figure 3.2(b):

$$S = \frac{\partial}{\partial V_{gs}} \log_{10}(I_D).$$ (3.2)

This metric is important for digital applications, as it basically sets the ratio between the current in the on- and off-state for a fixed voltage span. A steep slope means that the voltage level between the on- and off-state may be reduced while retaining the same on/off ratio, thereby increasing the switching speed or reducing the power dissipation. Most commonly, the subthreshold swing is given instead, which is defined as the inverse of the subthreshold slope. The minimum subthreshold swing achievable in a MOSFET design at room temperature is 60 mV/dec, as given by:

$$S^{-1} = \frac{k_B T}{q} \ln(10) \approx 60 \text{ mV/dec.}$$ (3.3)

Here, $T$ is the temperature, assumed to be 300 K, $k_B$ the Boltzmann constant and $q$ the elementary charge. Note that this limit applies to MOSFET devices and may be circumvented by e.g. transistors utilizing quantum mechanical tunnelling.
3.1.5 OUTPUT CONDUCTANCE

The output conductance, $g_d$, of the transistor is derived from the output characteristics, as illustrated in Figure 3.1(a), and is defined as the change in current from the drain-source voltage, $V_{ds}$:

$$g_d = \frac{\partial I_D}{\partial V_{ds}}. \tag{3.4}$$

A typical output conductance is presented in Figure 3.2(c), where the linear and saturation regions of the transistor operation are clearly visible. Ideally, the output conductance should be zero in the saturation region, which is true for an ideal MOSFET. However, effects such as channel length modulation and drain induced barrier lowering introduce additional output conductance. A low output conductance is important in amplifier design in order to improve the gain and ensure linearity.

3.1.6 VOLTAGE GAIN

The voltage gain, $A_v$, also called the self gain, of the transistor is given by the transconductance per output conductance and is an important metric in amplifier design:

$$A_v = \frac{g_m}{g_d}. \tag{3.5}$$

A high voltage gain is typically achieved in the saturation region, where the output conductance is low, and for high gate-source voltages, where the transconductance is high.

3.2 RF TRANSISTOR METRICS

The high frequency performance of the transistor is limited not only by the transconductance, but also by the capacitances in the structure. The most commonly stated metrics for the RF-performance are the cutoff frequency, $f_T$, and the maximum oscillation frequency, $f_{\text{max}}$.

3.2.1 CUTOFF FREQUENCY

The cutoff frequency, $f_T$, is defined as the frequency for which the current gain is unity, as illustrated in Figure 3.3(a). A simple approximation of $f_T$, assuming no series resistance, is given by the transconductance together with the gate-source, $C_{gs}$, and gate-drain capacitances, $C_{gd}$:

$$f_T \approx \frac{g_m}{2\pi(C_{gs} + C_{gd})}. \tag{3.6}$$
The key to obtain a high performance is to maximize the transconductance, while minimizing the gate capacitance. However, the analytical expression gets more complicated when series resistances at the source, $R_S$, and drain, $R_D$, are included:

$$f_T \approx \frac{1}{2\pi} \left( \frac{C_{gs} + C_{gd}}{g_m} + C_{gd}(R_S + R_D) \right)^{-1}. \quad (3.7)$$

It is evident that it is important to also reduce the series resistance. As a further complication, the capacitances and resistances are distributed in the structure. In this case an analytical expression is unwieldy and the cutoff frequency is instead calculated numerically using modified nodal analysis, applied on a small-signal model.

![Figure 3.3](image)

**Figure 3.3:** (a) Current gain and definition of $f_T$. (b) Unilateral power gain and definition of $f_{max}$.

### 3.2.2 MAXIMUM OSCILLATION FREQUENCY

The maximum oscillation frequency, $f_{max}$, is defined as the frequency for which the unilateral power gain is unity, as illustrated in Figure 3.3(b). It is approximately given by:

$$f_{max} \approx \sqrt{\frac{f_T}{4\pi(R_G C_{gd})}}. \quad (3.8)$$

The maximum oscillation frequency is sensitive to the gate resistance, $R_G$, and it is more important to decrease the gate-drain than the gate-source capacitance. As with $f_T$, deriving an accurate analytical expression gets extremely complex if all distributed parasitic elements are included, and it
is thus calculated numerically. For optimization, the geometrical mean of \( f_T \) and \( f_{\text{max}} \) is often used, in this work denoted by \( f_{T,\text{max}} \):

\[
f_{T,\text{max}} = \sqrt{f_T \cdot f_{\text{max}}} \tag{3.9}
\]

### 3.3 CAPACITOR METRICS

Real capacitors are not ideal components with a pure capacitive reactance. The performance of capacitors, especially at high frequencies, is dependent on the parasitic elements in the device structure.

#### 3.3.1 QUALITY FACTOR

An important figure of merit for capacitors is the quality factor, or Q-factor. It is essentially a metric describing the ideality of the capacitor and is of importance in circuit design, where a high quality factor is desired. It is defined as the ratio between the reactive, \( X_C \), and resistive, \( R_C \), parts of the impedance of the capacitor:

\[
Q = \frac{X_C}{R_C} = \frac{1}{\omega C R_C} \tag{3.10}
\]

It is evident that the quality factor is degraded as the series resistance of the capacitor is increased and that a capacitor without any series resistance would have an infinite quality factor. As the reactance is dependent on the frequency, the quality factor degrades at higher frequencies. For benchmarking, the quality factor at 1 MHz or 1 GHz is often stated.

### 3.4 DIGITAL CIRCUIT METRICS

The performance of a logic gate for digital applications is primarily evaluated based on the maximum switching speed as well as the power dissipation.

#### 3.4.1 PROPAGATION DELAY

The propagation delay, \( t_p \), is defined as the time it takes for a digital gate to switch from one state to another. This is measured as the time between the 50\% transition of the input signal to the same for the output signal, as illustrated in Figure 3.4. The inverse of the propagation delay sets the absolute maximum possible switching speed for the logic gate. The propagation delay is related to the amount of delivered current in comparison to the load capacitance. A high current means that the load capacitances can be charged faster, resulting in a faster switching speed.
3. Figures of Merit

![Figure 3.4](image.png)

**Figure 3.4:** Definition of the propagation delay for a digital gate.

As the propagation delay is dependent on the load capacitance, and thus the fan-out, it is typically given for an inverter with another inverter as load. The propagation delay is usually measured using a ring oscillator, which is a loop of an odd number of inverters. In this case, the propagation delay may be simply calculated from the oscillation frequency. In this work, the propagation delay is given for a single inverter with another inverter both at the input and output, corresponding to a conventional ring oscillator measurement.

### 3.4.2 POWER DISSIPATION

A low power dissipation is desirable in order to improve battery life for portable devices and reduce the need for cooling. The power dissipation for a digital gate consists of both a static part and a dynamic part, which is dependent on the switching speed. The static power dissipation is primarily due to off-current leakage through the transistors in the logic gate.

The dynamic power dissipation is due to charging and discharging the load capacitances and is thus dependent on the switching speed. In a CMOS technology the dynamic power consumption is usually larger than the static power. However, in e.g. a ratioed logic architecture, the static power consumption is dominating. For benchmarking, the average power dissipation is used, calculated at the maximum switching speed as determined by the propagation delay.

### 3.4.3 POWER-DELAY PRODUCT

The power-delay product, PDP, is defined as the product between the propagation delay, \( t_p \), and the average power dissipation, \( P_A \):

\[
PDP = t_p P_A.
\]  (3.11)
InAs Nanowire Devices and Circuits

This figure of merit is useful in order to optimize a transistor technology for low-power digital applications.

### 3.4.4 ENERGY-DELAY PRODUCT

In the energy-delay product, EDP, more weight is attributed to the propagation delay:

$$\text{EDP} = t_p^2 P_A. \quad (3.12)$$

This figure of merit is useful in order to optimize a transistor technology for high performance digital applications.

### 3.5 ANALOGUE CIRCUIT METRICS

In analogue circuit design, the maximum power gain is an important metric to benchmark the performance of an amplifier. Furthermore, it is important to consider the linearity and stability of the amplifier, together with the power consumption.

#### 3.5.1 MAXIMUM POWER GAIN

The maximum power gain is obtained when both the input and output are conjugately matched, which means that there is no power reflection at the source or load. The maximum available gain for an unconditionally stable amplifier may be derived from the S-parameters:

$$G_{\text{max}} = \left| \frac{S_{21}}{S_{12}} \right| \left( K - \sqrt{K^2 - 1} \right). \quad (3.13)$$

Here, $K$ is the stability factor defined as:

$$K = \frac{1 - |S_{11}|^2 - |S_{22}|^2 + |\Delta|^2}{2|S_{11}S_{12}|} \quad (3.14)$$

$$\Delta = S_{11}S_{22} - S_{12}S_{21}. \quad (3.15)$$

The maximum available gain is only defined for an unconditionally stable device, i.e. when the stability factor, $K$, is greater than one. If the amplifier is only conditionally stable, i.e. $K < 1$, the maximum stable gain, $G_{\text{MSG}}$, is given by:

$$G_{\text{MSG}} = \frac{S_{21}}{S_{12}}. \quad (3.16)$$

In this case, the amplifier must be carefully designed in order to ensure stability in the targeted operation region. In this work, the maximum power gain is calculated using the circuit simulation software Cadence.
3.5.2 IIP$_3$

A source of non-linearity in an amplifier is the amplification of higher-order harmonics of the input signal. This linearity is conventionally evaluated by studying the third order intermodulation intercept point, IIP$_3$. This metric is defined as the input power level, for which the output power of the third-order harmonic is equal to that of the fundamental tone. For an ideal common-source amplifier with no output conductance, it may be derived from the transistor characteristics as [30, 31]:

\[
\text{IIP}_3 = \frac{4 \left( \frac{\delta I_D}{\delta V_{gs}} \right)}{R_S \left( \frac{\delta^3 I_D}{\delta V_{gs}^3} \right)}. \tag{3.17}
\]

Here, $R_S$ is the source resistance, usually specified to 50 Ω. This figure of merit is usually given in dBm.

3.5.3 1 DB COMPRESSION POINT

The 1 dB compression point, $CP_{1\text{dB}}$, is a metric of linearity of the amplifier in regard to the amplitude of the input signal. The compression point essentially limits the highest input power that can be used in order to maintain a linear gain. As the amplitude of the input signal increases, the operation point of the transistor is no longer constant, which means that the small-signal approximation is no longer valid. The result is a gain that is dependent on the power of the input signal. This non-linear behaviour is quantified by the 1 dB compression point, which is defined as the input power that causes a gain degradation of 1 dB. If only third order distortion is considered, it may be estimated based on the related to the IIP$_3$ level [32]:

\[
CP_{1\text{db}} = \text{IIP}_3 - 9.6 \text{ dB}. \tag{3.18}
\]

The input power for the 1 dB compression point is usually given in dBm.
In order to evaluate the performance of nanowire devices, reliable and accurate models must be developed. This requires an understanding of the underlying semiconductor physics and its application to the nanowire structure. The device models also need to be sufficiently fast in order to perform an optimization of various transistor structures, which often requires extensive parameter sweeps. In this section, the development of a model for determining the band structure within a nanowire is presented together with an analysis of the derived simulation data. This model is used to calculate the capacitance-voltage characteristics of a nanowire capacitor, as well as for ballistic transport simulations in a nanowire MOSFET structure.

4.1 SEMICONDUCTOR PHYSICS

Before delving into the complicated band structure of semiconductor nanowires, a short introduction to the underlying physics is presented. Semiconductors are contrasted to metals and insulators based on the electrical conductivity of the material. As is well known, metals have a high conductivity, whereas insulators have a very low conductivity. Semiconductors, such as InAs, fall somewhere in between. This phenomenon is most commonly explained by the band structure model, which describes the available states for charge carriers, i.e. electrons, in the material. In a bulk material, these states form relatively continuous bands, separated by band gaps. The available states are filled up according to the Fermi-Dirac distribution, \( f(E) \), which describes the number of occupied states as a function of the energy, \( E \), above the Fermi level, \( E_F \):

\[
f(E) = \frac{1}{1 + e^{(E-E_F)/(kT)}}.
\]

In a metal, the Fermi level is located within one of the bands, where a lot of free states are available. This means that the charge carriers are free to move around and the conductivity is high. In an insulator, on the other hand, the Fermi level is located within a band gap, which means that any charge transport is effectively blocked. This concept is illustrated in Figure 4.1.

A semiconductor is reminiscent to an insulator, with the Fermi level located between bands, which means that the conductivity of a semiconductor is very
low under standard conditions. In comparison to an insulator, however, the band gap is relatively small. This means that only a low energy is required for carriers to enter the band and allow a current to flow. Semiconductors thus have the unique ability to increase the conductivity relatively easily by providing energy from an external source. This may be done by heating the material or by radiating it with light. In electronics, the conductivity is increased by introducing impurities into the crystal lattice, in a process known as doping. In a MOSFET, the bands are populated by carriers by applying a voltage on the gate terminal.

### 4.1.1 NANOWIRES

As a semiconductor structure is scaled down, the conductivity is no longer dependent only on the material, but also on the dimensions. In comparison to a bulk semiconductor, the conduction band in a semiconductor nanowire gets increasingly quantized as the diameter is scaled down. As in a quantum well, the subbands are pushed up to higher energies due to the stricter confinement in the thin nanowire. This means that fewer subbands are contributing to the conduction. By calculating the energy levels and contribution of these subbands, the carrier concentration and velocities within the nanowire may be determined.

### 4.2 BAND STRUCTURE SIMULATIONS

In order to determine the number of carriers within the nanowire, the band structure must be calculated. The band structure in thin nanowires of various materials have been extensively studied using various simulation approaches
Common techniques include tight-binding simulations and \( k \cdot p \) perturbation theory \([13,42–46]\). In order to calculate the band structure in InAs nanowires, it is important to consider the highly non-parabolic conduction band. In many cases this effect is neglected by using the simple effective mass approximation \([47–49]\). This is a relatively rough approximation, especially under high bias conditions. Furthermore, self-consistency between the carrier structure and the electrostatic potential should be ensured. This effect is especially important in thick nanowires, where the potential in the centre of the nanowire may differ a lot from the surface potential.

In this work, the band structure simulations are performed by solving a modified version of the Schrödinger equation in the electrostatic potential within the nanowire \([50–52]\). The Schrödinger equation is a partial differential equation, where the eigenvalues correspond to the different subband energy levels. The square of the eigenfunctions describe the probability to find carriers at a specific location within the nanowire. In this work, this eigenfunction problem is solved numerically using a customized solver in MATLAB. This model considers both self-consistency and includes the effect from non-parabolic conduction bands.

A thin nanowire may be modelled as a one-dimensional system, where conduction is only considered along the nanowire axis. As the nanowire is a cylindrical structure, the Schrödinger equation is formulated in polar coordinates, with the assumption that the band structure is constant along the length of the nanowire. This is performed for a specific electrostatic potential along the radial dimension, \( V(r) \). This results in an equation system, which has solutions in the form of eigenvalues, \( E_{n,v} \), with two quantum numbers, one radial, \( n \), and one azimuthal, \( v \). Due to symmetry, the azimuthal quantum number must be an integer. For each azimuthal quantum number, the eigenfunctions, \( \psi_{n,v}(r) \), along the radial dimension are calculated:

\[
E_{n,v} |\psi_{n,v}(r)\rangle = \left[ -\frac{\hbar^2}{2m^*} \left( \frac{d^2}{dr^2} + \frac{1}{r} \frac{d}{dr} - \frac{\nu^2}{r^2} \right) + V(r) \right] |\psi_{n,v}(r)\rangle. \tag{4.2}
\]

Here, \( m^* \) is the effective mass of the semiconductor material and \( \hbar \) the reduced Planck constant. In the effective mass approximation, the effective mass of the electrons is assumed to be constant. The electrostatic potential is not known at this stage and will later be calculated by Poisson’s equation. As a simple approximation it may be assumed to be constant inside the nanowire, equal to the surface potential. Another approach is to estimate the carrier concentration using Fermi-Dirac statistics and calculate the corresponding electrostatic potential. As the surface potential is increased by an applied gate-source voltage, more and more subbands are populated as they descend below the Fermi level.
A perfect dielectric interface is assumed with no influence from traps, which may degrade the performance [53–57]. In a real device, the wave function extends slightly into the gate dielectric. However, in these simulations, no penetration of the wave function inside the dielectric is allowed. This means that as a boundary condition, the eigenfunctions must be zero at the surface of the nanowire. In the centre of the nanowire, the eigenfunction must be continuous due to symmetry. The different eigenfunctions have the form of Bessel functions, as illustrated in Figure 4.2.

![Figure 4.2: Eigenfunctions for the lowest subbands in a nanowire with a diameter of (a) 12 nm and (b) 22 nm.](image)

4.2.1 CARRIER CONCENTRATION

From the calculated energy levels of the subbands, as well as the eigenfunctions, it is possible to calculated how the carrier concentration, \( n(r) \), varies within the nanowire. This is done by integrating the density of states, \( g(E) \), and the Fermi distribution, \( f(E) \), over the energy within each subband:

\[
n(r) = \sum_{n,\nu} |\psi_{n,\nu}(r)|^2 \int_{E_{n,\nu}(0)}^{\infty} f(E)g(E) \, dE. \tag{4.3}
\]

The resulting carrier concentration for different surface potentials is presented in Figure 4.3. The number of available states at a given energy level is described by the density of states, as shown in Figure 4.4. For a one-dimensional quantum wire, the density of states is proportional to the inverse slope of the energy dispersion relation and is given by:

\[
g(E) = \frac{2}{\pi} \left( \frac{dE}{dk} \right)^{-1}. \tag{4.4}
\]
Figure 4.3: Carrier concentration within a nanowire with a diameter of (a) 12 nm and (b) 22 nm. This is shown for different surface potentials together with a comparison between the effective mass approximation and the non-parabolic model.

Figure 4.4: Density of states for a nanowire with a diameter of (a) 12 nm and (b) 22 nm.

Figure 4.5: Energy dispersion relation for a nanowire with a diameter of (a) 12 nm and (b) 22 nm.
In the effective mass approximation, the effective mass of the charge carriers is assumed to be constant, independent on the energy. This means that the energy dispersion relation for each subband is completely parabolic in relation to the wave vector, $k$:

$$E(k) = E_0 + \frac{\hbar^2 k^2}{2m^*}. \tag{4.5}$$

### 4.2.2 NON-PARABOLIC BANDS

The effective mass approximation, however, is a rough approximation only valid for low energies in the band structure. At higher energies, the dispersion relation is no longer parabolic and instead approaches a linear slope that is material dependent [12, 47–49]. This effect is relatively strong in InAs, where the bands are highly non-parabolic and approach a linear slope of $\alpha = 1.4 \text{eV}^{-1}$, as illustrated in Figure 4.5 [58, 59].

In order to account for the non-parabolicity, the solution method needs to be modified to consider the change in effective mass with the energy. This effect may be modelled by solving the following Schrödinger-like equation for each wave vector, while accounting for the non-parabolicity factor, $\alpha$:

$$\left(1 + \tilde{\alpha}\right) E_{n,\nu}(k) |\psi_{n,\nu}(r)\rangle = \left[ -\frac{\hbar^2}{2m^*} \left( \frac{d^2}{dr^2} + \frac{1}{r} \frac{d}{dr} - \frac{\nu^2}{r^2} - k^2 \right) + (1 + 2\tilde{\alpha}) V(r) - \langle \psi_{n,\nu}(r)|\alpha V(r)|^2 \psi_{n,\nu}(r)\rangle \right] |\psi_{n,\nu}(r)\rangle \right]. \tag{4.6}$$

$$[\tilde{\alpha} = \langle \psi_{n,\nu}(r)|\alpha \psi_{n,\nu}(r)\rangle E_{n,\nu}(k)] \tag{4.7}$$

This equation is much more complicated to solve, as it introduces an additional dependence on the wave vector, $k$. This means that the equation must be solved for each value of $k$, which vastly increases the simulation time. It is important to remember that the linearisation of the non-parabolicity also is an approximation and that the band bending is even more complex at higher energies. This phenomenon may be simulated by the use of a complete tight-binding model. In order to verify the results, they are compared to a tight-binding simulation in Figure 4.6 [13]. For a nanowire with a diameter of 25 nm, it is evident that the two models correspond well, at least for the lowest subbands. It is clearly observed that the effective mass approximation gets increasingly worse at higher energies.

### 4.2.3 ELECTROSTATICS

As a rough estimation, the electrostatic potential may be assumed to be constant within the nanowire, set directly by the surface potential. However,
as the states in the conduction band are populated, the charge within the nanowire increases. This causes a deformation of the electrostatic potential, as described by Poisson’s equation:

\[
\left( \frac{d^2}{dr^2} + \frac{1}{r} \frac{d}{dr} \right) V(r) = -\frac{q (N_D - n(r))}{\varepsilon_s \varepsilon_0}.
\]  

Here, \( N_D \) is the doping level and \( \varepsilon_s \) the relative permittivity of the semiconductor. The resulting electrostatic potential within the nanowire for different surface potentials is shown in Figure 4.7.

4.2.4 SELF-CONSISTENCY

The carrier concentration calculated by the Schrödinger-like equation is used as input to Poisson’s equation and the result is a radius-dependent electrostatic potential. This electrostatic potential is then used as new input to the Schrödinger-like equation. By iterating between the two equations until convergence, a self-consistent solution is achieved. An overview of the simulation method is presented in Figure 4.8.

4.3 BALLISTIC TRANSISTOR MODELLING

In order to derive the transistor characteristics, a ballistic transistor model is used based on the work in [60–62]. When calculating the carrier concentration, the dispersion relation for each band is split in two, with carriers travelling in the positive and negative directions, respectively. The states in the positive
Figure 4.7: The electrostatic potential within a nanowire with a diameter of (a) 12 nm and (b) 22 nm. This is shown for different surface potentials together with a comparison between the effective mass approximation and the non-parabolic model.

Figure 4.8: Overview of the complete simulation method, showing the iterations between the solution of the Schrödinger-like equation and Poisson’s equation until convergence is achieved.
direction are filled with carriers up to the source Fermi level, whereas the states in the negative direction are filled up to the drain Fermi level. The principle of the ballistic conduction is illustrated in Figure 4.9.

![Cross-section and layout of a generic nanowire transistor structure](image)

**Figure 4.9:** Cross-section (a) and side view (b) of a generic nanowire transistor structure. (c) Illustration of ballistic transport within the nanowire.

The transistor current is regulated by applying a gate-source voltage, which controls the surface potential under the gate and thus the electrostatic potential within the nanowire. The change in electrostatic potential shifts the band structure, thereby allowing more or less carriers to flow under the gate. The resulting band structure within a nanowire MOSFET structure is presented in Figure 4.10.

![Band structures with different diameters](image)

**Figure 4.10:** Band structure within a nanowire MOSFET with a diameter of (a) 12 nm and (b) 22 nm.

### 4.3.1 GATE CAPACITANCE

The total intrinsic gate capacitance, $C_g$, consists of the geometric gate capacitance, $C_{ox}$ in series with the quantum capacitance, $C_q$ [63]:

$$ C_g = \left( \frac{1}{C_{ox}} + \frac{1}{C_q} \right)^{-1}. \quad (4.9) $$
The geometric gate capacitance is dependent on the permittivity, \( \varepsilon_{\text{ox}} \), and thickness, \( t_{\text{ox}} \) of the gate dielectric. The structure is in a cylindrical geometry, which is similar to a coaxial cable. Thus the geometric capacitance may be calculated as:

\[
C_{\text{ox}} = 2\pi \varepsilon_{\text{ox}} \varepsilon_0 \ln \left( \frac{R + t_{\text{ox}}}{R} \right)^{-1} \cdot L_g.
\]  

(4.10)

If the gate dielectric thickness is scaled as a one tenth of the nanowire diameter, the geometric gate capacitance per unit length is a constant value of about 4.57 aF/nm. The quantum capacitance is defined as the change in the charge under the gate, \( n \), from a change in the surface potential, \( \Phi_s \):

\[
C_q = \frac{\delta(qn)}{\delta\Phi_s}.
\]  

(4.11)

The quantum capacitance for a 12 nm and 22 nm transistor is presented in Figure 4.11. It is clearly observed that the quantum capacitance is increased with the gate-source voltage, as more carriers occupy the bands. The population of each new band is seen as steps, most prominently observed in the thinner nanowire. Operation close to the quantum capacitance limit is predicted in both nanowires, as the normalized quantum capacitance is much smaller than the geometric gate capacitance of 4.57 aF/nm [30, 64, 65].

![Figure 4.11: Quantum capacitance in a nanowire transistor with a diameter of (a) 12 nm and (b) 22 nm.](image)

**4.3.2 GATE BARRIER**

The gate control of the surface potential may be modelled as a potential barrier between the source and drain, which pushes the subbands to higher
energies. The barrier height is indirectly controlled by the gate-source voltage, depending on the relative size between the geometric gate capacitance and the quantum capacitance [61]:

\[ U_B = -qV_{GS} + \frac{q^2}{C_{ox}}(n^+ + n^-). \] (4.12)

Here \( n^+ \) and \( n^- \) are the concentrations of mobile carriers travelling in the positive and negative direction, respectively. If the geometric gate capacitance is much larger than the quantum capacitance, the second term is close to zero and the barrier height is set directly by the source-gate voltage. On the other hand, if the geometric gate capacitance is much smaller than the quantum capacitance, the barrier height will be dependent on the charge under the gate, which degrades the gate control. As the barrier height is dependent on the carrier concentration under the gate, this equation must be solved self-consistently with the calculation of the carrier concentration from the band structure. This is performed by an iterative method until the solution converges.

4.3.3 TRANSISTOR CHARACTERISTICS

In the ballistic limit, the transistor current is basically proportional to the number of carriers multiplied with their velocities. From the calculated band structure, both the carrier concentration and the carrier velocities may be calculated. The carrier velocity is dependent on the kinetic energy of the carriers and is derived from the slope of the energy dispersion relation:

\[ v(E) = \frac{1}{\hbar} \left( \frac{dE}{dk} \right). \] (4.13)

In the effective mass approximation, the bands are parabolic, which means that the velocity is increasing linearly with the energy. However, if non-parabolicity is considered, the slope of the energy dispersion relation approaches a constant value, which means that the velocity saturates for high energies. The mean carrier velocity inside the nanowire is shown in Figure 4.12.

Depending on the barrier height, the resulting current density, \( J^{+/−} \), in the positive and negative direction is given by:

\[ J^{+/−}(r) = \frac{1}{2} \sum_{n,\nu} |\psi_{n,\nu}(r)|^2. \]

\[ \int_{E_{n,\nu}(0)}^{\infty} f(E - E_{F,s/d}) \cdot g(E - U_B) \cdot v(E - U_B) \, dE. \] (4.14)
Figure 4.12: Mean velocity in a nanowire transistor with a diameter of (a) 12 nm and (b) 22 nm.

Here, \( E_{F,s/d} \) is the Fermi level at the source and drain side, with the difference set by the applied drain-source voltage. By integrating over the area and subtracting the carriers travelling in the negative direction, the transistor characteristics in Figure 4.13 are derived. A current level of about 2 A/mm and a transconductance of about 4 S/mm is predicted at a gate overdrive of 0.5 V.

Figure 4.13: (a) Output characteristics, (b) transfer characteristics, and (c) transconductance for nanowire transistors with a diameter of 12 nm and 22 nm.

4.3.4 SHORT-CHANNEL EFFECTS

In order to allow ballistic conduction and reduce the gate capacitance, the gate length should be as short as possible. However, a short gate length introduces short channel effects, such as drain-induced barrier lowering. This results in a
higher output conductance, as well as a degraded subthreshold slope. Thus, the channel should preferably be as short as possible, while at the same time minimizing these effects. The shortest recommended gate length, as described by [13], is \( L_g > 4.6 \lambda \), where \( \lambda \) is the natural length for an InAs nanowire transistor [66]:

\[
\lambda = d_{nw} \sqrt{\frac{2 \varepsilon_{\text{InAs}} \ln(1 + 2t_{ox}d_{nw}^{-1})}{16\varepsilon_{ox}}}.
\] (4.15)

For a scaling of the gate dielectric thickness of one tenth of the nanowire diameter, this correspond to a gate length approximately 30\% larger than the nanowire diameter. Unless otherwise stated, this gate length is assumed in the following simulations.

\[
L_g > 4.6 \cdot \lambda \approx 1.30 \cdot d_{nw}
\] (4.16)

Thus far, the effect of the source and drain potential on the barrier has been neglected in the simulations. However, as the gate length is scaled down, this influence gets increasingly important. The influence on the barrier from the source and drain may be modelled as:

\[
U_B = -q(\alpha_G V_G + \alpha_D V_D + \alpha_S V_S) + \frac{q^2}{C_G + C_D + C_S}(n^+ + n^-).
\] (4.17)

Here, the \( \alpha \) factors represent the relative control of the barrier height from the gate, drain, and source respectively:

\[
\alpha_G = \frac{C_G}{C_G + C_D + C_S}, \quad \alpha_D = \frac{C_D}{C_G + C_D + C_S}, \quad \alpha_S = \frac{C_S}{C_G + C_D + C_S}.
\] (4.18)

In order to quantify this effect, a full 3D simulation of the electrostatic potential has been performed in COMSOL, by solving the Laplace equation in cylindrical coordinates. This model does not, however, take into account the carrier concentration within the nanowire. The resulting degradation of the gate control of the barrier height is presented in Figure 4.14(a) for a nanowire with a diameter of 12 nm. It is concluded that the barrier height is controlled by the gate to an efficiency of about \( \alpha_G \approx 87 \% \), in comparison to the source and drain contribution. This is for a gate length 30\% longer than the nanowire diameter. Similar results are found for other nanowire diameters, as the gate length is scaled linearly with the diameter. As the gate length is reduced down to a few nanometers, tunnelling through the gate barrier introduces further degradation of the device operation, although this effect is not considered in this work.

The effect on the output characteristics of the transistor is presented in 4.14(b), where it is seen that the influence from the drain potential on the
barrier, $\alpha_D$, introduces drain induced barrier lowering. This results in an increased output conductance, most prominent as a higher slope in the saturation region. Furthermore, the subthreshold slope is degraded from the ideal 60 mV/dec to about 68 mV/dec, as illustrated in Figure 4.15. The transconductance is slightly lower and the threshold voltage is shifted to a lower value.

**Figure 4.14:** (a) Effective gate control of the barrier height for different gate lengths. (b) Output characteristics for a 12 nm nanowire transistor, with and without influence from the source and drain potentials.

**Figure 4.15:** Influence on the barrier height from the source and drain potentials on (a) the transfer characteristics, (b) transconductance, and (c) subthreshold slope.
Parasitic Elements

The performance of a real transistor architecture is often far from the expected intrinsic performance. This is due to significant parasitic capacitances and resistances in the device structure that degrade the performance. This means that the parasitic elements must be evaluated in order to predict the practical performance of a device architecture. Furthermore, an understanding of the parasitics is important in order to identify bottlenecks in the design and allows for an optimization of the structural parameters.

Compared to a conventional, planar MOSFET structure, the parasitic elements are aligned differently in a vertical nanowire transistor. The most prominent difference is the large overlap between the vertically aligned electrodes. This introduces large gate-source and gate-drain capacitances over the dielectric spacer layers that are not present in a planar architecture. The presence of spacer layers also introduces large series resistances in the ungated segments of the nanowire. Both of these elements introduce significant performance degradation. An overview of the location of the parasitic elements is presented in Figure 5.1.

5.1 PARASITIC CAPACITANCES

Parasitic capacitances in the structure degrade the performance of the device, due to the time it takes for the capacitances to be charged and discharged. In the ideal case, where the parasitic elements are non-existent, the high frequency performance is limited only by the intrinsic gate capacitance. As the parasitic capacitances and resistances are distributed within the nanowire, they are modelled as multiple individual elements in the form of an RC-network. In a simple approximation, the contribution from the nanowire surfaces consists of two parts in a first order pi-network, one internal, $C_{g,wi}$, and one external, $C_{g,we}$. The external part is combined with the parasitic capacitances from the corresponding source and drain electrode, whereas the internal part is combined with the intrinsic gate capacitance. This simple model is suitable both for tabulation of the parasitics and for circuit simulations.
5.1.1 FASTCAP

As the structure is fairly complex, analytical expressions for the parasitic capacitances quickly get unwieldy. In order to solve this problem, the capacitances have been calculated using the FastCap algorithm, which is a numerical fast field solver based on the boundary-element method [67, 68]. To facilitate the simulations, a MATLAB interface has been developed using the COM automation interface to control the FastCap software. A MATLAB script is used to automatically generate 3D models of the structure based on a set of structural parameters. This allows an extensive parameter analysis to be performed for all structural parameters within the structure. An example of a generated 3D structure is presented in Figure 5.2.

The generated 3D model is used as input to the FastCap algorithm and all capacitances in the transistor structure are calculated. By tagging the different surfaces in the structure, it is possible to separate the gate-source, gate-drain, and drain-source capacitances. The surfaces are also tagged depending on the segment they represent, such as a metal electrode or a nanowire. Furthermore, the nanowire capacitances are distributed in an RC network by tagging parts of the nanowire as different surfaces. The separation of the surfaces allows the different parasitic capacitances to be represented at their correct locations in a small-signal model of the nanowire transistor, as shown in Figure 5.3.
Figure 5.2: Three-dimensional transistor structure generated by the MATLAB interface, which is used as input to the FastCap algorithm.

Figure 5.3: The small-signal model of the transistor used in the simulations, including both intrinsic and extrinsic circuit elements.

In order to calculate the capacitances accurately, the surfaces in the structure must be discretized into smaller sub-surfaces. As the discretization level is increased, the accuracy of the simulation is improved. However, the simulation time increases exponentially with the number of sub-surfaces, which means that the chosen discretization level is a compromise between speed and accuracy. In the generation of the 3D model, the discretization level is varying within the structure. Close to the nanowire array, where a high accuracy is necessary, the discretization level is high, whereas it is lower further away on the electrodes. Simulations with different levels of discretization have been performed in order to ensure that the accuracy is sufficiently high, while maintaining a reasonable simulation time.

The simulation time increases exponentially with the number of nanowires in the structure. Due to the extremely long simulation time for more than a few nanowires, a method based on linear extrapolation has been developed.
Each parasitic element is divided into one constant part and one part dependent on the number of nanowires. The simulation is then run for a structure without nanowires and then sequentially with one, two and three nanowires. The result is a near linear dependence on the number of nanowires, offset by the capacitances present in the electrode structure. It is verified that this method is highly accurate by comparing it with a brute-force method, as illustrated in Figure 5.4. Another benefit of the linear representation is that it is possible to calculate the parasitic capacitances for any number of nanowires without rerunning the simulation. This is highly useful in the optimization of the architecture as well as for circuit simulations.

![Figure 5.4: Comparison between the linear extrapolation and a brute-force simulation of the parasitic capacitances.](image)

### 5.2 Parasitic Resistances

There are also parasitic resistances in the structure that limit the performance. The most significant part is the series resistance within the ungated segments of the nanowires. The resistivity of the nanowire, $\rho_w$, is calculated from the doping level, $N_D$, and the predicted mobility, $\mu$, using data from [69]. The resistance, $R_w$, is then calculated conventionally from the nanowire geometry, where $t_{gs, gd}$ is the length of the ungated segment, corresponding to the thickness of the spacer layer:

$$R_w = \frac{4t_{gs, gd}\rho_w}{d_{nw}^2\pi}, \quad \left[ \rho_w = \frac{1}{N_D\mu} \right]. \quad (5.1)$$

Between the nanowire and the metal electrode, there is also a contact resistance that must be considered. The contact resistance, $R_c$, is modelled as a distributed model as described in [70]. In order to minimize the resistance, the thickness of the source and drain electrodes, $t_{S, D}$, should be significantly
longer than the transfer length, $L_T$. For the contact resistivity, $\rho_{cw}$, a value of 1.0 $\Omega \mu m^2$ is employed from [71]. The series and contact resistance form a resistive network with an equivalent resistance described by:

$$R_c = \frac{4\rho_w L_T}{\pi d_{nw}^2} \coth \left( \frac{l_{SD}}{L_T} \right), \quad L_T = \sqrt{\frac{d_{nw} \rho_c}{4\rho_w}}.$$  

(5.2)

Additional series resistances are found within the source and drain electrodes. There is also a resistance within the gate electrode. These resistances, $R_e$, are calculated analytically from the length, $L$, width, $W$, and thickness, $h$ of the electrodes, together with the bulk resistivity of the metal, $\rho_m$:

$$R_e = \rho_m \frac{L}{Wh}.$$  

(5.3)

### 5.2.1 EXPERIMENTAL VERIFICATION

In order to verify the validity of the capacitance calculations, the model is compared to measurements on the fabricated device in [27]. A schematic illustration of the device structure is presented in Figure 5.5. The fabricated device has narrow EBL-defined electrodes, as illustrated in 5.6, which means that the parallel plate model is no longer accurate due to the large fringe capacitances from the electrode side walls. Reference samples using large UV-defined electrodes are also available. The parasitic capacitances in the transistor structure are calculated using FastCap and the simple parallel plate model is used as a reference.

The result is presented in Figure 5.7, where it is seen that both models correspond well to the experimental data for large UV defined electrodes. This is due to the very small influence from fringe capacitances relative to the large overlap capacitances. The gate-drain capacitance is much smaller than the gate-source capacitance, due to a thicker spacer layer with a lower permittivity. The source-drain capacitance is close to zero as it is shielded by the gate electrode.

When the drain electrode is patterned by EBL, the gate-drain capacitance is cut almost in half. However, the reduction is smaller than predicted by the parallel plate model, due to the large contribution from fringe capacitances. If instead the gate electrode is patterned by EBL, both the gate-source and gate-drain capacitances are reduced by about 50%. However, this introduces a large source-drain capacitance, due to a direct overlap as the shielding effect of the gate is reduced. This may be reduced by patterning both the gate and the drain electrode, after which only the fringe capacitances between source and drain remains. Note that no measurement data for the drain-source capacitance is available, as the relatively small size made it very difficult to reliably extract from the small-signal model.
Figure 5.5: Schematic illustration of the nanowire transistor structure fabricated in [27].

Figure 5.6: Electrode configuration for the transistor structure in [27], showing the use of finger gates.
Figure 5.7: Summary of the parasitic capacitances in the fabricated nanowire transistor structure. A comparison is presented between the measured capacitances and those calculated by the FastCap simulations. The parallel plate model is included as a reference.
In Figure 5.8, the effect of removing the spacer layer material is investigated. This is performed for the design with an EBL-defined gate electrode. As air has a relative permittivity of one, this would greatly reduce the parasitic capacitances. A potential problem with removing the spacer material is that the structure may collapse without support. A compromise would be to remove the spacer layer only outside of the electrodes, here denoted as a partial spacer. This ensures structural integrity, while at the same time eliminating most of the fringe capacitances. However, the overlap capacitances are unaffected and removing the spacer layer also within the array would of course reduce the parasitic capacitances even more.
Benchmarking is the art of comparing the performance and advantages of one technology to another. In this chapter, the focus is on the RF performance of the proposed nanowire transistor architecture, evaluated by the metrics $f_T$ and $f_{\text{max}}$. In order to simplify the optimization, a single figure of merit is often desired. In this case the geometrical mean of $f_T$ and $f_{\text{max}}$ is used. The optimization is performed for maximum performance, with no regard to the power consumption. The complete transistor architecture is simulated by combining the predicted intrinsic properties with the calculated parasitic elements in the transistor structure. All of these elements are combined into the small-signal model in Fig. 5.3. From this model, the $y$-parameters are calculated using nodal analysis in the MATLAB tool SCAM [72,73]. From the $y$-parameters, $f_T$ and $f_{\text{max}}$ are calculated by deriving the current gain, $h_{21}$, and the unilateral power gain, $U$.

6.1 SCALING ROADMAP

A roadmap for the scaling of the nanowire transistor structure is established in order to determine the highest practical performance and how it scales with the dimensions of the device structure. The proposed roadmap is based on a set of technology nodes, corresponding to the nanowire diameter, which is scaled from 50 nm down to 8 nm. For each node, the structural parameters in the device are optimized and the performance is calculated. According to the band structure simulations, a thinner nanowire demonstrates a higher intrinsic transconductance and the parasitic capacitances are lower in a smaller transistor structure. This should translate into an improved performance as the device structure is scaled down. However, the series resistance is higher in a thin nanowire, which causes a degradation of the extrinsic transconductance.

6.1.1 INTRINSIC TRANSISTOR PERFORMANCE

In this chapter, the intrinsic transistor parameters are based on the ballistic transistor simulations in [13]. These simulations were performed using a tight-binding method, although without considering self-consistency. Fur-
thermore, the output conductance is simply assumed to be 10% of the transconductance, corresponding to a voltage gain of 10. The intrinsic gate capacitance is derived from band structure calculations in [13]. Operation in the saturation region is assumed, which means that the intrinsic gate-source capacitance is 2/3 of the total intrinsic gate capacitance, whereas the intrinsic gate-drain capacitance is zero [74]. A summary of the intrinsic transistor properties for each technology node is presented in Table 6.1.

**Table 6.1:** Summary of the intrinsic transistor parameters for each technology node.

<table>
<thead>
<tr>
<th>Node (Diameter)</th>
<th>$L_g$ (nm)</th>
<th>$C_{gg,i}$ (aF/NW)</th>
<th>$g_m$ (S/mm)</th>
<th>$g_d$ (S/mm)</th>
<th>$\mu_i$ (cm$^2$/Vs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 nm</td>
<td>67.3</td>
<td>71.4</td>
<td>1.71</td>
<td>0.171</td>
<td>9370</td>
</tr>
<tr>
<td>35 nm</td>
<td>47.1</td>
<td>40.0</td>
<td>2.88</td>
<td>0.288</td>
<td>6200</td>
</tr>
<tr>
<td>22 nm</td>
<td>29.6</td>
<td>19.7</td>
<td>4.86</td>
<td>0.486</td>
<td>3460</td>
</tr>
<tr>
<td>16 nm</td>
<td>21.5</td>
<td>11.8</td>
<td>5.77</td>
<td>0.577</td>
<td>2200</td>
</tr>
<tr>
<td>12 nm</td>
<td>16.1</td>
<td>8.20</td>
<td>6.12</td>
<td>0.612</td>
<td>1350</td>
</tr>
<tr>
<td>8 nm</td>
<td>11.1</td>
<td>3.20</td>
<td>7.14</td>
<td>0.714</td>
<td>508</td>
</tr>
</tbody>
</table>

The intrinsic parameters used in this chapter are not the same as those calculated in Chapter 4. The reason is that the benchmarking in this chapter was performed before the band structure simulations. This means that the calculated transconductance and output conductance were not available for this analysis. The main difference is the neglection of the self-consistency in [13]. According to the simulations in Chapter 4, the transconductance is about 2/3 of the values used in this analysis, although with an output conductance approximately half as high. This deviation is relatively small and the conclusions regarding the optimization of the structure should still be valid. However, the ultimate performance may be a few tens of percent lower than predicted. The resulting performance from a change in transconductance or output conductance is presented in Figure 6.1.

6.1.2 MATERIAL PARAMETERS

The nanowire transistor is based on the standard materials presented in Chapter 2, as summarized in Table 6.2. Here $\varepsilon_{ox}$, $\varepsilon_{spc}$, and $\varepsilon_{sub}$ are the relative permittivities of the gate dielectric, spacer layers, and substrate, respectively. The source and drain electrodes consist of Au, whereas W is used for the gate, with resistivities of $\rho_{S,D,e}$ and $\rho_G$, respectively. The nanowire contact
Figure 6.1: Change in the RF performance from a change in the transconductance (a), and output conductance (b).

Table 6.2: Summary of the material parameters in the nanowire transistor structure.

<table>
<thead>
<tr>
<th>(\varepsilon_{\text{ox}})</th>
<th>(\varepsilon_{\text{spc}})</th>
<th>(\varepsilon_{\text{sub}})</th>
<th>(\rho_{G} (\Omega \mu \text{m}))</th>
<th>(\rho_{S,D,e} (\Omega \mu \text{m}))</th>
<th>(\rho_{c} (\Omega \mu \text{m}^{2}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>3.0</td>
<td>11.7</td>
<td>0.056</td>
<td>0.024</td>
<td>1.0</td>
</tr>
</tbody>
</table>

resistivity is denoted by \(\rho_{c}\). In order to decrease the series resistance, the doping level is increased for each technology node as described by [15].

6.1.3 STRUCTURAL PARAMETERS

The dimensions of most structural parameters in the design are relatively straightforward to optimize, as a smaller scale results in lower parasitic capacitances. However, these parameters are limited by lithography capabilities and other fabrication challenges. In this case, the scaling is based on a reasonable transistor footprint estimated from practical experience. More aggressive scaling is assumed as the technology is scaled down, with the size related to the technology node by a fixed multiple.

In the proposed transistor structure, the nanowires are arranged in a hexagonal pattern with a spacing between the nanowires, \(s_{\text{nw}}\), equal to the nanowire diameter, \(d_{\text{nw}}\). At the edges of the array there is an additional space equal to the nanowire diameter, in order to add some fabrication margin to the electrode alignment. The gate length, \(L_{g}\), is scaled to avoid short-channel effects, as described in Chapter 4, with a gate dielectric thickness, \(t_{\text{ox}}\), of one
tenth of the nanowire diameter. The electrode thickness at the source, \( t_S \), and drain, \( t_D \), is set to three times the nanowire diameter, which is much larger than the transfer length of the nanowire-electrode interface. A schematic illustration of the structural parameters is presented in Figure 6.2, together with a summary of the scaling rules in Table 6.3.

![Figure 6.2: Schematic cross-sectional view of the transistor structure, illustrating the structural parameters.](image)

### 6.1.4 OPTIMIZATION

Some of the structural parameters are identified as having no clear optimum value and require further optimization. One of them is the spacer layer thickness, \( t_{gs,gd} \), which sets the compromise between the parasitic capacitances and the series resistance. Interestingly, the series resistance and the overlap capacitance are both dependent on the spacer layer thickness, although they have a completely opposite scaling. A thick spacer layer would reduce the overlap capacitance, whereas the series resistance is increased, and vice versa. This means that the spacer layer thickness is the most important structural parameter to optimize as it sets the balance between these two parasitic elements.

Another parameter, for which the optimum is not obvious is the number of parallel nanowires in the array, as well as the aspect ratio of the array. Increasing the number of nanowires decreases the series resistance, although the overlap capacitances are increased. However, there are also fringe capaci-
Table 6.3: Scaling rules for the structural parameters in the nanowire transistor architecture.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Roadmap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nanowire diameter</td>
<td>$d_{nw} = {50, 35, 22, 16, 12, 8}$ nm</td>
</tr>
<tr>
<td>Gate dielectric thickness</td>
<td>$t_{ox} = 0.1 \cdot d_{nw}$</td>
</tr>
<tr>
<td>Nanowire spacing (e-e)</td>
<td>$s_{nw} = d_{nw} + 2 \cdot t_{ox}$</td>
</tr>
<tr>
<td>Nanowire array size</td>
<td>$n_{x,y}$ (To be optimized)</td>
</tr>
<tr>
<td>Electrode length</td>
<td>$L = d_{nw} n_{y} + s_{nw} (n_{y} + 1)$</td>
</tr>
<tr>
<td>Electrode width</td>
<td>$W = d_{nw} n_{x} + s_{nw} ((n_{x} - 1) \cos(\pi/6) + 2)$</td>
</tr>
<tr>
<td>Electrode extension</td>
<td>$L_e = 100 \cdot d_{nw}$</td>
</tr>
<tr>
<td>Gate length/thickness</td>
<td>$L_g$ (See Table 6.1)</td>
</tr>
<tr>
<td>Source/drain thickness</td>
<td>$t_{s,D} = 3 \cdot d_{nw}$</td>
</tr>
<tr>
<td>Spacer layer thickness</td>
<td>$t_{gs,gd}$ (To be optimized)</td>
</tr>
</tbody>
</table>

tances from the edges of the array. As the fringe capacitances are scaling with the circumference instead of the area, they play a relatively large role for small arrays, where the circumference to area ratio is high. Furthermore, there are resistances within the electrodes, which also scale with the array size. This complexity means that the optimum array size and electrode configuration must be simulated.

For further optimization, the series resistance from the electrodes may partly be moved to the gate resistance and vice versa, depending on the array aspect ratio and electrode configuration. The same is true for the parasitic fringe capacitances, which get larger for long and thin arrays, due to the longer circumference for the same area.

6.2 RESULTS

In Figure 6.3, the result of an optimization of both the spacer layer thickness and the number of nanowires in the array is presented. This optimization is repeated for each technology node as the device is scaled down and the optimized roadmap device dimensions are presented in Figure 6.4(a). It is clearly observed that the spacer layer thickness should be decreased linearly, similarly to the gate length and the spacing between the nanowires. It is evident that a spacer layer thickness thinner than the nanowire diameter is recommended in order to reduce the resistance from the ungated segments of
InAs Nanowire Devices and Circuits

The nanowires. This is much thinner than the thickness of the spacer layer in most fabricated vertical nanowire MOSFETs [20,21,75,76].

Figure 6.3: Optimization of the spacer layer thickness and nanowire array size for the 22 nm node.

The resulting optimized $f_T$ and $f_{\text{max}}$ for each technology node is presented in Figure 6.4(b), with simultaneous scaling of the nanowire diameter, gate length and nanowire spacing [15]. It is clearly observed that the performance is improved as the device dimensions are scaled down. This is primarily due to the increased normalized transconductance for thin nanowires, while the parasitic capacitances are reduced.

![Image of Figure 6.3 and Figure 6.4](image.png)

Figure 6.4: (a) Roadmap for the scaling of the transistor dimensions. (b) Predicted $f_T$ and $f_{\text{max}}$ performance for each technology node as the transistor is scaled down.

The predicted parasitic capacitances are presented together with the intrinsic gate capacitance in Figure 6.5(b). It is evident that the parasitic capacitances are decreased as the structure is scaled down and that they are about the same size as the intrinsic gate capacitance. The series resistance, on the other hand, is increasing exponentially as the device is scaled down, as
illustrated in Figure 6.5(a). This exponential increase is a serious issue that gets increasingly problematic for thinner nanowires, due to the geometrical decrease in cross-sectional area combined with the mobility degradation from surface scattering.

The most critical part is the resistance at the nanowire-electrode interface, which consists of a distributed network of contact and series resistances. This resistance is not reduced by a thinner spacer layer. Furthermore, it is not significantly reduced by increasing the thickness of the electrode, as it is already much thicker than the transfer length. In order to decrease this resistance, the contact resistivity or the resistivity within the nanowire must be reduced. The development of high-quality nanowire-electrode interfaces is imperative together with an improved doping control of the nanowires. An alternative would be to increase the nanowire diameter outside the channel by regrowth or by digitally etching the gated segment to a thinner diameter.

6.2.1 ITRS ROADMAP

The predicted parasitic elements are compared to the ITRS roadmap in Figure 6.6. It is predicted that the series resistance for the nanowire transistor architecture is lower than in the ITRS roadmap, although the exponential increase as the structure is scaled down is a cause of concern. The total capacitance in the transistor structure is comparable to the ITRS roadmap. However, the intrinsic gate capacitance is lower in the nanowire transistor in comparison to the parasitic capacitances, as seen in in Figure 6.7(a). The predicted development of $f_T$ and $f_{max}$ for high-performance multi-gate MOSFETs is presented in 6.7(b). The data is compiled from the ITRS roadmap and is compared to the performance for the 8 nm nanowire technology node. According to this prediction, the performance of the nanowire transistor
architecture remains competitive at least until the year 2026.

![Figure 6.6: Series resistance (a) and total gate capacitance (b), compared to the ITRS roadmap.](image)

![Figure 6.7: (a) Capacitances in the transistor structure compared to the ITRS roadmap. (b) Predicted $f_T$ and $f_{\text{max}}$ performance according to the ITRS roadmap in comparison to the optimized 8 nm nanowire transistor design.](image)

### 6.2.2 COMPETING TECHNOLOGIES

A summary of the performance for different transistor technologies is presented in Table 6.4. The best demonstrated RF performance for a vertical nanowire transistor is an $f_T$ of 141 GHz and an $f_{\text{max}}$ of 155 GHz [27]. As shown in this chapter, it is expected that an optimized transistor architecture may achieve a performance about 5–10 times higher than this, exceeding one terahertz. Currently, the best III-V MOSFETs demonstrate a performance of a few hundred gigahertz, using InAs or InGaAs as channel material [77,78]. As demonstrated by [79], this may be combined with a FinFET layout, with a demonstrated $f_T$ of 281 GHz and an $f_{\text{max}}$ of 365 GHz. A conventional 45 nm
**Table 6.4**: Benchmarking with competing transistor technologies.

<table>
<thead>
<tr>
<th>Technology</th>
<th>$f_T$ (GHz)</th>
<th>$f_{\text{max}}$ (GHz)</th>
<th>$g_m$ (S/mm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>This Work (12 nm)</td>
<td>1730</td>
<td>1230</td>
<td>6.12</td>
</tr>
<tr>
<td>This Work (22 nm)</td>
<td>1380</td>
<td>1040</td>
<td>4.86</td>
</tr>
<tr>
<td>This Work (35 nm)</td>
<td>800</td>
<td>640</td>
<td>2.88</td>
</tr>
<tr>
<td>InAs Nanowire MOSFET [27]</td>
<td>103</td>
<td>155</td>
<td>0.73</td>
</tr>
<tr>
<td>Planar InAs MOSFET [77]</td>
<td>245</td>
<td>355</td>
<td>1.73</td>
</tr>
<tr>
<td>Planar InGaAs MOSFET [78]</td>
<td>370</td>
<td>280</td>
<td>2.00</td>
</tr>
<tr>
<td>InGaAs FinFET [79]</td>
<td>281</td>
<td>365</td>
<td>2.85</td>
</tr>
<tr>
<td>45 nm Si MOSFET [80]</td>
<td>485</td>
<td>-</td>
<td>1.60</td>
</tr>
<tr>
<td>65 nm Si MOSFET [81]</td>
<td>360</td>
<td>420</td>
<td>1.55</td>
</tr>
<tr>
<td>InAs Pseudomorphic HEMT [82]</td>
<td>644</td>
<td>681</td>
<td>1.90</td>
</tr>
<tr>
<td>InGaAs Metamorphic HEMT [83]</td>
<td>688</td>
<td>800</td>
<td>2.70</td>
</tr>
<tr>
<td>InP HEMT [84]</td>
<td>385</td>
<td>1200</td>
<td>2.30</td>
</tr>
<tr>
<td>Graphene Transistor [85]</td>
<td>427</td>
<td>10</td>
<td>1.33</td>
</tr>
</tbody>
</table>

silicon-on-insulator CMOS technology has demonstrated an $f_T$ of 485 GHz [80]. An $f_{\text{max}}$ of 420 GHz has been achieved for a 65 nm Si MOSFET [81].

The highest RF performance is still achieved by HEMT designs, although the limited scalability of these devices means that the III-V MOSFETs are rapidly closing the gap. The best devices have a very high $f_{\text{max}}$, exceeding one terahertz, as demonstrated for an InP HEMT [84]. A high $f_T$ of 688 GHz has been demonstrated for an InGaAs HEMT [83]. Another emerging technology is transistors based on graphene, where the best demonstrated $f_T$ is 427 GHz, but with a very low $f_{\text{max}}$ [85].

It is clear that III-V nanowire transistors offers a very high potential performance, which in the future may surpass that of even the best HEMT devices. If most design and fabrication issues are solved, the performance of these devices may approach, or even surpass, the terahertz limit. Furthermore, the performance may be further increased by scaling the device dimensions, at least down to a nanowire diameter of 8 nm. It is probable that nanowire MOSFETs operating at a frequency of a few hundred gigahertz will be fabricated within a few years.
In this chapter, the previously performed optimization is used as a basis for further investigation of alternative transistor structures, with a focus on the 22 nm technology node. Material changes are introduced for the spacer layers and electrodes and the influence on the performance is quantified. The studied materials are chosen to represent what may be plausible alternatives for fabrication in the near future. Additionally, the impact from asymmetric spacer layers is investigated, as well as the possible gains from epitaxial regrowth on the drain side. A schematic illustrating some of these modifications is presented in Figure 7.1.

7.1 ASYMMETRIC SPACER LAYERS

As demonstrated in the previous chapter, the performance of the nanowire transistor architecture is primarily limited by the large series resistance in the nanowire. This resistance may be reduced by a number of different approaches, of which the most obvious solution is to increase the doping level in the ungated segment. However, the proposed doping level is already very high at $10^{20}$ cm$^{-3}$. Furthermore, it is very difficult to achieve a sharp doping interface between the ungated segment and the channel. An alternative is to further decrease the spacer layer thickness at the expense of higher parasitic capacitances. However, in the previous chapter the optimum spacer layer thickness was calculated and a further reduced thickness would only reduce the performance. Furthermore, the highest resistance is located at the nanowire-electrode interface, which is not reduced by a thinner spacer layer.

One design consideration, which may be advantageous in some circuit layouts is to move some of the resistance from the source to the drain side. This may be done introducing asymmetric spacer layers, by changing the ratio between the source and drain spacer layers thickness. This results in a device with an improved performance in one direction at the cost of a reduced performance in the other direction, as illustrated in Fig. 7.2. It is clear that the gain is relatively small compared to the degradation of the performance in the other direction. Asymmetric structures may be useful in circuit applications, where reducing either the parasitic capacitances or the series resistance on
Figure 7.1: (a) Schematic illustration showing different design variations. In this case, the bottom and top spacer layers differ in both material and thickness. Furthermore, the nanowire has been regrown to a larger diameter at the top and an etched mesa from the semiconductor substrate is used as the bottom electrode. In (b), the transistor in the reverse direction is shown.

one side may be useful. An example would be to maximize the performance of e.g. an amplifier circuit. However, as fabricating a structure with a varying spacer layer thickness over the sample is very difficult, all devices are most likely limited to the same asymmetry. In a large scale circuit, great care must thus be given to the alignment of the individual transistors, in order to avoid an overall performance degradation.

7.2 THICKER UNGATED SEGMENTS

Another method to reduce the series resistance is to use a larger nanowire diameter in the ungated segment. This would decrease the series resistance as the cross-sectional area is increased. A further improvement is expected as the mobility in thick nanowires is higher due to lower surface scattering. A larger diameter of the ungated segment may be achieved by introducing an epitaxial regrowth step into the fabrication process. An alternative is to start with a thick nanowire and instead thin down the channel segment using digital etching.

The effect of increasing the nanowire diameter on the top side of the structure is presented in Figure 7.3. It is evident that a higher performance may be achieved in the reverse direction, with the source positioned at the top of the structure. This is due to the transistor performance being more sensitive to the series resistance at the source side, which degrades the internal gate-
source voltage. A lower parasitic capacitance, on the other hand, is more important on the drain side, due to its degradation of $f_{\text{max}}$. It is seen that a regrowth step may help to regain some performance for doping levels lower than about $10^{18} \text{ cm}^{-3}$. However, for higher doping levels than about $10^{19} \text{ cm}^{-3}$, the increased parasitic capacitance is degrading the performance to a larger degree than the gain from a further decreased series resistance. For a doping concentration of $10^{17} \text{ cm}^{-3}$, the performance may be almost doubled by the regrowth.

### 7.3 MATERIAL CHANGES

It is also interesting to investigate alternative materials in the structure. Changing the spacer layer material is interesting not only from a performance perspective, but also due to fabrication considerations. Important properties include a low permittivity and a high mechanical stability, as well as the control of the targeted thickness and high etching selectivity. An organic spacer has a relatively low permittivity, although it is very difficult to control the thickness, which is very uneven on the substrate. For spacers thinner than about 50 nm, a deposited material is most likely required, such as Si$_3$N$_4$. The best would be to remove the spacer material altogether and instead use air, although mechanical stability may be a problem.

The impact on the performance for a few plausible material changes is presented in column (a)-(c) in Table 7.1. A change to Si$_3$N$_4$ as source spacer degrades the performance, whereas replacing the drain spacer layer with air improves the performance. This is due to the higher permittivity for Si$_3$N$_4$ at 7.5, in comparison with the generic spacer layer with a permittivity of 3, used as a reference.
Figure 7.3: Change in RF performance by increasing the nanowire diameter for the ungated segment at the top of the structure. This is presented for the 12 nm node, with the source at either the bottom or the top of the structure.

Another modification is to replace the Au source electrode with an InAs layer. This would remove the contact resistance at the cost of a higher electrode resistance. As most of the series resistance for narrow nanowires is due to the contact resistance, there will be a point in the scaling below which an InAs contact would be better than an Au electrode. This crossover point is highly dependent on the InAs resistivity, as illustrated in Fig. 7.4, where a comparison to the conventional Au electrode is presented.

Figure 7.4: Comparison of the series resistance from using a source electrode composed of either Au or an InAs layer with different resistivities.
Table 7.1: Performance for alternative 22 nm transistor structures.

<table>
<thead>
<tr>
<th>Design</th>
<th>Ref.</th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
<th>(e)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source spacer</td>
<td>Resist</td>
<td>Si₃N₄</td>
<td>Resist</td>
<td>Resist</td>
<td>Si₃N₄</td>
<td>Si₃N₄</td>
</tr>
<tr>
<td>Drain spacer</td>
<td>Resist</td>
<td>Resist</td>
<td>Air</td>
<td>Resist</td>
<td>Resist</td>
<td>Air</td>
</tr>
<tr>
<td>Source contact</td>
<td>Au</td>
<td>Au</td>
<td>Au</td>
<td>InAs</td>
<td>InAs</td>
<td>InAs</td>
</tr>
<tr>
<td>NW spacing</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>100</td>
<td>22</td>
</tr>
<tr>
<td>$f_T$ (GHz)</td>
<td>1380</td>
<td>960</td>
<td>1750</td>
<td>540</td>
<td>250</td>
<td>550</td>
</tr>
<tr>
<td>$f_{\text{max}}$ (GHz)</td>
<td>1040</td>
<td>810</td>
<td>1670</td>
<td>300</td>
<td>180</td>
<td>340</td>
</tr>
<tr>
<td>$f_{T,i}$ (GHz)</td>
<td>1420</td>
<td>950</td>
<td>1740</td>
<td>760</td>
<td>390</td>
<td>500</td>
</tr>
<tr>
<td>$f_{\text{max},i}$ (GHz)</td>
<td>1020</td>
<td>560</td>
<td>1160</td>
<td>720</td>
<td>320</td>
<td>400</td>
</tr>
</tbody>
</table>

Below the 12 nm node, contacting the nanowire with an InAs layer would be better than an Au electrode if the InAs resistivity is lower than about 1 $\Omega$ µm. For a resistivity of 0.1 $\Omega$ µm it would be advantageous regardless of the device scale. This is due to the elimination of contact resistance between the Au electrode and the nanowire. However, if the resistivity of the InAs layer is high, the performance is instead degraded due to the increased electrode resistance. This is the case in Table 7.1(c), where the performance is severely degraded by the introduction of an InAs source contact with a relatively high resistivity of 10 $\Omega$ µm, as reported by [26]. This is more than a factor of ten too high to be advantageous in comparison to an Au electrode.

7.4 CASE STUDIES

In Table 7.1(d), a structure that is plausible for fabrication in the near future is presented, featuring a Si₃N₄ source spacer and a conventional organic layer as drain spacer. In this case an InAs source contact, with a resistivity of 10 $\Omega$ µm is used and the spacing between the nanowires is set to 100 nm. This structure is predicted to achieve an $f_T$ and $f_{\text{max}}$ about 200 GHz in the forward direction. It is also evident that it would be more advantageous to use this device in the reverse direction. This is due to the improvement of the intrinsic gate-source voltage, resulting from moving the highly resistive InAs contact from the source to the drain side.

The more refined structure in Table 7.1(e) completely eliminates the drain spacer material in favour for air, as well as using a nanowire spacing of
InAs Nanowire Devices and Circuits

22 nm. This would increase the performance almost a factor of two in the forward direction. The gain in the reverse direction is much less, as keeping the low permittivity air spacer on the drain side is more beneficial than the performance gained by reducing the resistance on the source side.

In Fig. 7.5, an optimization is presented of the drain spacer layer thickness and array size for the transistor structure in Table 7.1(e). The only difference from the table is that regrowth has been introduced on the drain side, increasing the nanowire diameter by 50%. The optimum number of nanowires is relatively low, which is mostly due to the high resistivity of the InAs source contact. The structure is relatively insensitive to the drain spacer layer thickness, due to the combination of low parasitic capacitances from the air spacer and a low series resistance thanks to the regrowth. The absolute maximum performance for both \( f_T \) and \( f_{\text{max}} \) is achieved for a spacer layer thickness of 8 nm.

![Figure 7.5](image)

**Figure 7.5:** Optimization of (a) \( f_T \) and (b) \( f_{\text{max}} \) for the 22 nm node with an InAs source contact and a drain regrowth factor of 50%. A source spacer of Si\(_3\)N\(_4\) is used, with a fixed thickness of 15.8 nm. Optimization is performed of the thickness of the drain spacer, which consists of air. The width of the nanowire array is fixed at two nanowires.

The expected performance is about 1 THz, which is almost twice as high as the unoptimized performance of the transistor in Table 7.1(e). This demonstrates the need to repeat the optimization for each modification of the transistor structure in order to maximize the performance. This is due to changes in both the resistivities and permittivities of the materials, which influence the compromise of the spacer layer thickness, as well as the optimal nanowire array size.
MOS capacitors are essentially MOSFETs but without the drain terminal. These devices may be used to evaluate the device physics and properties such as the quality of the gate dielectric interface. Furthermore, it is interesting to evaluate the performance of these devices as they have applications in circuit design.

In this chapter, a vertical nanowire capacitor structure is studied both by simulations and measurements on fabricated devices. The focus is on the intrinsic capacitance-voltage characteristics, as well as the frequency response from DC up to 67 GHz. It is proposed that the nanowire may be modelled as a distributed RC-network consisting of both the intrinsic capacitance and the series resistance within the nanowire. This allows the high-frequency properties to be derived and the quality factor of the device is determined. It is demonstrated that the distributed model may be used to extract the resistivity of the nanowires from RF measurements. Furthermore, a complete model of the parasitics inside a fabricated capacitor structure is devised, which is used to separate the parasitic elements from the intrinsic capacitor characteristics.

Figure 8.1: Schematic illustration of a vertical nanowire capacitor structure.
8.1 DESIGN

The nanowire capacitor design is similar to the transistor structure, although with the exclusion of the top contact. Instead, the length of the gate contact is significantly extended in order to maximize the capacitance. A schematic illustration of the nanowire capacitor structure is presented in Figure 8.1. As the capacitor is oriented in the vertical dimension, this design offers a high capacitance with a small footprint.

8.2 FABRICATION

As the capacitance for a single nanowire is in the range of femtofarads, reliable measurements using a conventional impedance analyser is not possible. In order to circumvent this limitation, the capacitors are typically based on arrays exceeding a thousand nanowires. This brings the total capacitance up to the picofarad range. However, the use of large arrays introduces problems if there are large variations in the length of the nanowires. In order to ensure that all nanowires have the same gate length, the top contact is etched back a few hundred nanometers. This also eliminates any potential leakage through the gold seed particle, which may not be fully covered by the gate dielectric.

![Figure 8.2: Device fabrication yield as a function of the growth standard deviation and the number of nanowires in the array.](image)

The problem gets worse as the number of nanowires in the array is increased, and it is therefore important to reduce the standard deviation of the nanowire lengths from the growth. This is illustrated in Figure 8.2, for a targeted nanowire length of 800 nm, where 200 nm of the top contact is removed. If the standard deviation in the growth is high, the device yield...
quickly deteriorates as the number of nanowires increases. In this case, the yield is defined as the proportion of devices, for which all included nanowires are contacted with the same gate length. Another contributing factor is the uniformity of the etching of the top contact. If one in a thousand nanowires is too short, with a targeted yield of more than 50%, this requires a growth standard deviation of less than about 60 nm. If more of the top contact is removed, a higher standard deviation from the growth is possible, as illustrated in Figure 8.3. Today, the quality of the best grown arrays achieves a standard deviation of less than 30 nm, although in many cases it is much worse due to unreliable growth conditions.

Figure 8.4: Influence of the pad capacitance on a low-resistivity substrate.
In order to reduce the parasitic capacitance from the measurement pad, it is positioned on a very thick spacer layer. This is essential for low frequency measurements if the substrate has a low resistivity. Otherwise the measurement will be completely dominated by the pad capacitance, as illustrated in Figure 8.4. The parasitic capacitances are further reduced by EBL patterning of the gate electrodes, typically arranged in a finger configuration. In order to de-embed the parasitic capacitances and resistances from the electrode structure, reference samples are fabricated without any nanowires. The capacitor structure is designed to be RF-compatible in order to allow measurements up to 67 GHz.

### 8.3 RF CHARACTERISTICS

The resistance in very long and thin nanowires degrades the effective capacitance of the device at high frequencies. This effect is modelled by a distributed RC-network along the direction of the nanowire, as illustrated in Figure 8.5. The effective impedance of the RC-network may be calculated using different approaches. One method is to do a nodal analysis of the network and solving it numerically using a MATLAB interface to LTSpice IV [86]. This method is very flexible for changes in the design but is relatively slow. This method performs sufficiently well to simulate a single nanowire, although the nodal analysis is too slow to be suitable for arrays of nanowires. For the simple case of a single nanowire, it is also possible to do derive an analytical solution.

In a nanowire array, each nanowire is connected by an electrode, which in itself is modelled as a distributed RC-network. This results in tens of thousands of nodes for an array, if a discretization of 100 nodes is used for each nanowire. Instead, the network is modelled as a transmission line using

![Small signal model of the intrinsic part of the nanowire capacitor, consisting of a single nanowire.](image)
matrix operations on transmission line parameters. This method is very fast, although it requires more manual work to setup the transformation between the different transmission line parameters and arrange them into the correct network. This also places restrictions on the type of network that may be modelled, but the method has proven sufficiently flexible for this structure.

### 8.4 SIMULATIONS

The intrinsic properties are derived from the band structure simulations described in Chapter 4. From the derived carrier concentration, the intrinsic capacitance is calculated, as well as the resistivity of the nanowire. The resulting capacitance-voltage characteristics are presented in Figure 8.6 for a nanowire capacitor with a diameter of 12 nm and 22 nm, respectively.

![Figure 8.6: Simulated CV for a nanowire capacitor with a diameter of (a) 12 nm and (b) 22 nm. The gate length is 500 nm.](image)

It is clearly observed that the CV exhibits steps for the thin nanowire due to quantization of the fewer active subbands. In the thicker nanowire more subbands are active, which causes a CV characteristic more reminiscent of bulk behaviour. Due to the resistance within the nanowire, the effective capacitance is degraded at high frequencies. For the 22 nm, this effect is evident at frequencies higher than about 100 GHz. As the resistance is higher in the 12 nm nanowire, this degradation is shifted to lower frequencies.

### 8.5 MEASUREMENT DATA

Capacitors with different nanowire diameters have been fabricated. In order to extract the data for a single nanowire, the parasitic capacitances and resistances from the surrounding structure must be removed. A schematic
Figure 8.7: Schematic illustration of the fabricated nanowire capacitor. The parasitic elements in the structure are marked in red colour.

Illustration is presented in Figure 8.7, where the parasitic elements are marked in red. The parasitics include a large pad capacitance, as well as a series resistance in the InAs source layer. Furthermore, the finger gates are modelled as a distributed network of the capacitances and the resistances in the electrode. These elements are combined and the total impedance is calculated using transmission line parameters. By using this model, it is possible to sort out the contribution of all parasitic elements in the structure and separate the intrinsic capacitor characteristics.

By using reference samples without nanowires and assuming that the parasitic elements are constant, it is possible to extract the contribution from a single nanowire. The model assumes that the intrinsic capacitance and resistivity are independent of the frequency, which is a reasonable assumption in the accumulation region. However, this analysis is less reliable for negative voltages, due to the significant modulation of the intrinsic capacitance by the voltage.

The measurement data for nanowire capacitors with 33 and 48 nm is presented in Figure 8.8(a-b), together with the fitted device model, including parasitics. It is evident that the model corresponds well to the measurement data for all frequencies and voltages. The extracted negative capacitance at high frequencies is due to a parasitic inductance in the measurement setup, as confirmed by measurements on reference samples. This systematic error is subtracted from the data as a constant series inductance. By fitting the model to the measurement, it is possible to subtract all parasitic elements and extract the characteristics for a single nanowire, as presented in Figure 8.8(c-d). As the intrinsic capacitance may be accurately extracted from the low frequencies, the resistivity remains the only free parameter. Based on the distributed
Figure 8.8: Measurement data for capacitors based on nanowires with a diameter of (a) 33 nm and (b) 48 nm. The simulation data, including all parasitic elements, is presented as the dashed lines. The extracted data after removal of the parasitic elements is presented in (c) and (d), corresponding to a single nanowire, normalized with the gate length.

RC model for a single nanowire, it is thus possible to extract the nanowire resistivity. The resulting resistivities for different nanowire diameters are presented in Figure 8.9. The resistivity is relatively independent on the nanowire diameter, with a value of about 0.01 Ω cm. This result corresponds well to previously reported data in [87].

The modelling also means that the intrinsic capacitance-voltage characteristics of the nanowire may be extracted, as presented in Figure 8.10. This represents the ideal capacitance that would be achieved for a nanowire without any series resistance. The frequency dispersion in the accumulation region is very small is and not significant with the accuracy of the extraction method. The deviation for lower frequencies is due to uncertainty in the measurement setup, caused by measuring small capacitances outside the calibrated range of the equipment. For the 33 nm device, the intrinsic capacitance is about 0.7 aF/nm, which may be compared to the expected geometric capacitance of
**Figure 8.9:** (a) Extracted resistivity for nanowires with varying diameters for different biasing conditions. (b) Summary of the nanowire resistivity at −2.5 V, 0 V, and 2.5 V plotted against the nanowire diameter.

**Figure 8.10:** Extracted intrinsic CV of a single nanowire capacitor with a nanowire diameter of (a) 33 nm and (b) 48 nm.

about 2.7 aF/nm. This indicates a significant contribution from the quantum capacitance, consistent with the prediction from the simulations.

### 8.6 QUALITY FACTOR

In circuit design it is advantageous to have a high quality factor for the reactive components, i.e. capacitors and inductors. The quality factor essentially describes how large the reactance for the component is in comparison to the parasitic resistance. In the case of an ideal nanowire capacitor structure, this resistance is located inside the nanowire. As the reactive part for a capacitor decreases with the frequency and the resistance is constant, the quality factor
is degraded at high frequencies. Furthermore, it is reduced at lower gate voltages due to the depletion of carriers, which increases the resistance of the nanowire, as illustrated in Figure 8.11(a).

For longer nanowires, the resistance along the nanowire is higher. This means that the quality factor is degraded for nanowire capacitors with a long gate length. This effect is illustrated in Figure 8.11(b), where it is evident that a nanowire capacitor may achieve a quality factor exceeding $10^3$ at a frequency of 1 GHz for nanowires shorter than about 200 nm. The maximum quality factor in the short nanowire limit is close to $10^4$. The quality factor is higher...
in thick nanowires due to the lower series resistance.

The quality factor for the fabricated nanowire capacitors, after subtraction of the parasitic elements, is presented in Figure 8.12. The devices have a gate length of about 700 nm and demonstrate a quality factor of about 100 at a frequency of 1 GHz. This corresponds to a predicted maximum quality factor of about $10^4$ in the short nanowire limit, consistent with the simulation results. It is reasonable to assume that an optimized nanowire capacitor structure with short nanowires may achieve a quality factor of 100 at a frequency of 90 GHz. This may be compared to a conventional thin-gate oxide MOS varactor, which typically has a quality factor of about 20 at a frequency of 20 GHz [88], whereas the ITRS roadmap targets a quality factor exceeding 50 at a frequency of 5 GHz [4].

Another advantage of a vertical nanowire capacitor structure is the small footprint of the device structure, which is much smaller than for a conventional planar capacitor. According to the ITRS roadmap, the capacitance per area for a conventional MOS capacitor is about 5–10 fF/µm² [4]. The fabricated nanowire structure demonstrates a capacitance of about 5 fF/µm². However, an optimized structure with an electrode configuration according to the nanowire transistor roadmap would achieve a capacitance density exceeding 1 pF/µm². It is concluded that there is potential for a vast improvement in capacitance density in comparison to a planar MOS capacitor architecture.
In this chapter, the performance of a number of benchmarking circuits is investigated. This is done by combining the intrinsic device models with the calculated parasitic elements in the previous chapters. The circuit simulations are performed in Cadence, which is a software package commonly used by both industry and academia.

9.1 DEVICE MODELS

In order to use the device models in Cadence, they need to be converted into Verilog-A code, which is a programming language used to describe the operation of electronic devices. After compilation of the Verilog-A code, an associated circuit symbol is generated, which may be graphically connected with other components. This allows for rapid testing of different circuit layouts without doing a manual nodal analysis, which quickly gets unwieldy for large circuits with more than a few electrical nodes.

9.1.1 NANOWIRE TRANSISTOR

For the intrinsic performance, the transistor characteristics derived from the band structure simulations are used. As these simulations are very complex and computationally expensive, it is not possible to use them directly in the Cadence environment. Instead, a look-up table is generated from the simulation data. This means that not all physical parameters of the transistor may be changed in the circuit model. However, it is possible to shift the threshold voltage and change the gate width by the number of nanowires in the array, without rendering the lookup table invalid. In most cases, this is sufficient to do a proper optimization of the circuit design. Generating a new lookup table is necessary if the nanowire diameter, gate oxide thickness, or any material properties are changed.

The intrinsic transistor model is combined with the parasitic capacitances derived from the field-solver simulations in FastCap. From these simulations, analytical expressions for the parasitic capacitances are compiled, with the number of nanowires in the array as a free parameter. The series and gate resistances are simple analytical expressions and are easily implemented as
well. This means that the parasitic elements are able to scale with the number of nanowires in the array, and thus the gate width may be controlled. However, if the spacing between the nanowires or the electrode configuration is changed, the FastCap simulation needs to be rerun in order to generate a new set of parameters.

The circuit performance is studied for a transistor architecture based on nanowires with diameters of 12 nm and 22 nm. The array is composed of 4x8 nanowires, as illustrated in Figure 9.1. The characteristics, including parasitic resistances, for a 22 nm nanowire transistor, are presented in Figure 9.2. It is seen that a maximum transconductance of about 3 S/mm is achieved at a gate-source voltage close to 0.9 V. The transconductance is degraded compared to the intrinsic value calculated in Chapter 4, which is due to the influence of the series resistances. The extrinsic transconductance is actually higher in the 22 nm nanowire than in a thinner nanowire of 12 nm due to the lower series resistance, in contrast to the conclusions about the intrinsic transistor performance. However, as the current level is lower in a thinner nanowire, this would result in a reduced power consumption. The lowest output conductance is found in the saturation region, where a voltage gain of about 25 dB may be achieved for a large voltage range.

**Figure 9.1:** Side and top view of the nanowire transistor structure used for the circuit simulations. The dimensions are mostly to scale.
9.1.2 NANOWIRE CAPACITOR

The nanowire capacitor structure is implemented similarly to the nanowire transistor, by the use of a lookup table generated by the band structure simulations. It is possible to change the number of nanowires in the array and thus scale the capacitance. As with the transistor, a separate lookup table must be generated for each nanowire diameter.

9.2 COMMON-SOURCE AMPLIFIER

In order to investigate the performance of nanowire transistors in amplifier circuits, a simple common-source architecture is studied. The circuit is based on a single common-source amplifier stage, as illustrated in Figure 9.3(a). The focus is on maximizing the power gain at a frequency of 90 GHz, while maintaining a low power dissipation and a high linearity.
Figure 9.3: (a) Circuit schematic for a simple common-source amplifier. The maximum power gain is presented for the amplifier based on nanowires with a diameter of (b) 12 nm and (c) 22 nm, together with influence of using different supply voltages and sizes of the nanowire array.

Figure 9.4: Power gain at 90GHz for a common-source amplifier using a nanowire transistor with a diameter of (a) 12 nm and (b) 22 nm.

The simulated maximum power gain for different frequencies is presented in Figure 9.3. It is seen that a power gain of about 20 dB is expected at a frequency of 90 GHz, with a slightly lower performance in the thinner nanowire. More nanowires in the array increases the performance, although it saturates for arrays larger than about 32. This is due to the lower series resistance, as well the improved ratio between the current level and the parasitic capacitances, due to the reduced influence from fringe capacitances. The maximum power gain for different bias conditions is presented in Figure 9.4. It is clearly observed that a higher supply voltage increases the performance, due to a higher current level and transconductance. A suitable DC bias at the input is about 40 % of the supply voltage.
9.3 CASCODE AMPLIFIER

In order to further increase the power gain, a cascode amplifier is investigated. The circuit is consists of a common-source stage used as input to a common-gate stage, as illustrated in Figure 9.5(a). As the cascode amplifier consists of two separate transistors, it is interesting to investigate how these should be aligned in an optimized circuit layout. There are four possible combinations of the transistor alignment. However, only two remain if a connection between the top of one transistor to the bottom of the other is to be avoided. By simulations, it is confirmed that the best layout is to have the source of the common-source transistor at the bottom of the structure and the reverse for the common-gate transistor, as illustrated in Figure 9.5(b). The use of a cascode topology increases the power gain with about 7 dB at a frequency of 100 GHz, as shown in Figure 9.5(c). However, at higher frequencies than about 600 GHz, the common-source topology is better.

![Circuit schematic](image)

**Figure 9.5:** Circuit schematic (a) and layout (b) for a cascode amplifier using nanowire transistors. (c) Comparison of the maximum power gain for a common-source and cascode amplifier.

9.4 RATIOED LOGIC INVERTER

As no PMOS transistors have been simulated, the performance of a ratioed logic inverter is investigated. In this architecture, the pull-up network consists of a resistive load in the form of an NMOS, with the gate connected directly to the supply voltage. This architecture has the advantage of using only n-type transistors, but have significant drawbacks with a high static power consumption, as well as non-ideal output voltage levels. However, the switching speed is comparable to that of CMOS. Although not very suitable for large-scale digital circuits due to the power consumption, it is useful for
smaller complementary digital circuitry using a technology where no PMOS transistors are available.

The inverter circuit is simulated in Cadence, following the optimization in [89]. A linear chain of 15 inverters is simulated in order to obtain the self-driven characteristics and ensure regeneracy of the inverter. From this, the propagation delay and power consumption may be calculated, as well as output voltage levels. Consistent with the results in [89], a good ratio between the number of nanowires is 3 in the pull-down network (PDN) for each nanowire in the pull-up network (PUN). More nanowires is required in the PDN in order to pull down the voltage level for a low output. This is due to the PUN transistor always being on, which causes the output to be a voltage division between the PUN and the PDN. During this phase, a direct path between the supply voltage and ground causes a high power consumption. Furthermore, it is important to maintain regeneracy, by ensuring that the threshold voltage is located between the low and high output levels. An

![Figure 9.6: Input and output voltage of a ratioed logic inverter using transistors with a nanowire diameter of 12 nm, with a supply voltage of (a) 0.5 V and (b) 1.0 V. The same is presented for a 22 nm design in (c) and (d).](image-url)
optimization of the threshold voltage is performed for each inverter design and for different supply voltage.

The inverter characteristics are presented in Figure 9.6, for an inverter with a nanowire diameter of 12 nm and 22 nm. This is shown for a supply voltage of 0.5 V and 1.0 V. A summary of the inverter performance is presented in Table 9.1. It is evident that the propagation delay is a few picoseconds, whereas the power consumption is a few tens of microwatts. This may be compared to the ITRS roadmap, which predicts a propagation delay of about 1 ps after the introduction of III-V/Ge into a CMOS process [4]. Both the power-delay product and the energy-delay product indicate that an optimal performance is achieved for thin nanowires with a low supply voltage of 0.5 V. The performance is significantly better for the 12 nm nanowires than for 22 nm, which indicates that a further scaling of the transistor dimensions is advantageous.

| Table 9.1: Performance of a ratioed logic inverter based on nanowire transistors. |
|---------------------------------|---|---|---|---|
| Nanowire Diameter (nm)          | 12 | 12 | 22 | 22 |
| Number of PDN Nanowires         | 3x4 | 3x4 | 3x4 | 3x4 |
| Number of PUN Nanowires         | 1x4 | 1x4 | 1x4 | 1x4 |
| Supply Voltage (V)              | 0.5 | 1.0 | 0.5 | 1.0 |
| Threshold Voltage (mV)          | 183 | 363 | 195 | 375 |
| Propagation Delay (ps)          | 1.49 | 1.27 | 2.27 | 2.06 |
| Power Consumption (µW)          | 19.5 | 81.9 | 28.5 | 137 |
| PDP (aJ)                        | 28.9 | 104 | 64.7 | 282 |
| EDP (10⁻³⁰ Js)                   | 43.0 | 132 | 147 | 579 |
| Output High (mV)                | 314 | 727 | 343 | 846 |
| Output Low (mV)                 | 107 | 197 | 89 | 151 |
| Noise Margin (mV)               | 76 | 166 | 103 | 224 |
Conclusion

The physical properties of nanowire-based transistors offer a compelling argument for their introduction into future nanoelectronic circuits. The architecture solves some of the most challenging problems faced by the semiconductor industry by introducing a structure that is both inherently three-dimensional and offers the possibility to integrate III-V channel materials on Si substrates. The vertical alignment offers a natural way to fabricate a gate-all-around structure, which provides a near optimal gate control. Furthermore, a vertical orientation enables a potentially higher integration density, as the footprint of the device is independent on the gate length. This is especially prominent in the case of nanowire capacitors, where an extremely high capacitance per die area is possible. A vertical alignment even offers the prospect of true 3D circuits by stacking devices on top of each other.

10.1 CHALLENGES

A vertical nanowire transistor architecture introduces a number of design challenges. The most immediate problem to address is the large series resistances in the ungated segments of the nanowire, as well as the contact resistance at the interface between the nanowire and the metal electrodes. It is probable that the diameter of the ungated segment should be made thicker than the channel region, either by regrowing the ungated segment or by digital etching of the channel. Furthermore, a very high doping level is required in the ungated segment in order to reduce the resistivity.

The series resistance of the ungated segment may also be reduced by using very thin spacer layers. However, this increases the parasitic capacitances from the direct overlap of the metal electrodes. This means that a compromise is necessary and this work has presented an optimization of these structural parameters in the structure. It is demonstrated that an optimized structure may reach parasitics comparable to those predicted in the ITRS roadmap, by packing the nanowires in large and compact arrays, with very thin spacer layers. However, much process development is needed before this structure may be fabricated. An ideal solution for reducing the parasitic capacitances would probably be to use metal plugs to instead contact the nanowires.
vertically, thereby avoiding much of the overlap capacitance. It would also be interesting to investigate the influence of spacer layers with a non-uniform thickness and more intricate electrode structures, such as T-gates.

The main challenge faced today in the fabrication of these devices is to accurately control the gate length, as well as the thickness of the spacer layers. An accurate control of these structural parameters is imperative in order to fabricate optimized high-performance devices. Much improvement of the nanowire growth is also necessary, where very dense nanowire arrays with uniform lengths are yet to be demonstrated. Furthermore, very sharp doping interfaces are necessary in order to decrease the series resistance, while at the same time avoid unintentional doping of the channel. It is probable that these problems may be solved, given sufficient time for process development and access to proper and well-calibrated equipment.

It would also be desirable to develop a p-type nanowire MOSFET, which is a requirement to do CMOS logic. From a simulation point of view, it would be interesting to apply the models developed in this work to e.g. a GaSb material system and evaluate the potential CMOS performance. Other interesting concepts to investigate include simulations of core-shell structures or nanowires with different doping profiles. The models developed in this work are very flexible and may be applied to different material systems with relatively minor modifications to the simulations.

10.2 OUTLOOK

It is most likely that future transistor technologies venture beyond a mostly planar architecture in favour of more intricate 3D structures. The interest in multi-gate technologies from the industry is clear and an important step towards this goal was the introduction of the tri-gate transistors into commercial CMOS production. The gradual replacement of Si with higher mobility III-V semiconductor materials is another important step and the ITRS roadmap predicts an introduction of III-V materials into CMOS within the next few years. The proposed nanowire transistor architecture incorporates both of these key technologies and promises a performance that surpasses the ITRS prediction until the year 2026.

In this work, it has been demonstrated that an optimized nanowire transistor, with a realistic footprint, may reach an RF performance in the terahertz region. In comparison to a traditional Si-based CMOS technology, this architecture offers both higher switching speeds and a lower power consumption. It is probable that HEMTs will continue to hold the RF performance record for FET devices in the near future. However, these devices suffer from limited scalability and the performance has saturated in recent years.
The development of III-V MOSFETs, on the other hand, is progressing at a remarkable pace, with a performance rapidly approaching that of even the best HEMT devices.

Large-scale adoption of nanowires in commercial production is ultimately up to the semiconductor companies. It is difficult to predict what companies such as Intel have planned for the future and it is possible that they have a completely different solution in store. However, even if the nanowire technology is found to be prohibitively expensive for general use, it may find a niche in e.g. space or military applications, where an extremely high performance or low power consumption is critical.
Bibliography


[83] D.-H. Kim, B. Brar, and J. del Alamo, “fT = 688 GHz and fmax = 800 GHz in Lg = 40 nm In0.7Ga0.3As MHEMTs with $gm_{max} = 2.7$ mS/µm,” in *Electron Devices Meeting (IEDM), 2011 IEEE International*, Dec. 2011, pp. 13.6.1–13.6.4.


Performance Evaluation of III–V Nanowire Transistors

Kristofer Jansson, Erik Lind, and Lars-Erik Wernersson

Abstract—III–V nanowire (NW) transistors are an emerging technology with the prospect of high performance and low power dissipation. Performance evaluations of these devices, however, have focused mostly on the intrinsic properties of the NW, excluding any parasitic elements. In this paper, a III–V NW transistor architecture is investigated, based on a NW array with a realistic footprint. Based on scaling rules for the structural parameters, 3-D representations of the transistor are generated, and the parasitic capacitances are calculated. A complete optimization of the structure is performed based on the RF performance metrics $f_T$ and $f_{\text{max}}$, employing intrinsic transistor data combined with calculated parasitic capacitances and resistances. The result is a roadmap of optimized transistor structures for a set of technology nodes, with gate lengths down to the 10-nm-length scale. For each technology node, the performance is predicted, promising operation in the terahertz regime. The resulting roadmap has implications as a reference both for benchmarking and for device fabrication.

Index Terms—Field-effect transistor (FET), InAs, modeling, nanowires (NWs), roadmap.

I. INTRODUCTION

Gate-all-around III–V nanowire (NW) transistors are a potential candidate for use in future electronic devices, showing great promise for high performance and scalability. Both vertical and lateral implementations have been demonstrated with large transconductances and high drive currents, combined with good subthreshold characteristics [1]–[3]. This is mainly due to excellent electrostatic control in the cylindrical geometry combined with the advantageous transport properties of III–V materials [4]. However, one fundamental drawback for these vertical devices is the relatively large parasitic capacitances, caused by a large overlap between electrodes. Large series resistances from the ungated parts of the NWs add to the problem, making the use of thick source and drain spacers troublesome.

Crossbar arrangements of the electrodes have previously been considered for vertically arranged devices, concluding that large NW arrays are effective to reduce the effect of the fringing capacitances [5]. It has also been shown that closely packed NWs may screen the electric fields within the array, effectively reducing the electrode capacitances [6]. It is also important to evaluate the performance of NW-based transistors in detail and benchmark it against alternative technologies [7], [8]. Previous simulation work on the subject has focused primarily on the intrinsic performance of the transistors, promising high performance as the device is scaled down [9], [10]. There are also indications that the NW geometry may offer a higher margin for process variations [11]. However, a full optimization of the transistor design has not yet been performed.

In this paper, a complete optimization of all structural parameters is presented for a proposed transistor architecture based on vertically arranged InAs NWs, with diameters from 50 down to 8 nm. The focus is on evaluating the RF performance of the optimized devices, including the impact from parasitic elements, using a small-signal intrinsic transistor model. An evaluation based on digital metrics, such as $I_{\text{on}}/I_{\text{off}}$ ratio, subthreshold slope, and drain-induced barrier lowering, is beyond the scope of this paper.

In order to study the scaling behavior of the transistor architecture, a set of scaling rules for the structural parameters is proposed. Optimization for a set of technology nodes is performed with respect to the main RF metrics $f_T$ and $f_{\text{max}}$, employing simulated and reported values for intrinsic parameters. For simplicity, a technology node is defined here as equal to the NW diameter. Based on these optimizations, the optimal device structure and performance are predicted, and the magnitudes of the parasitic elements are quantified.

In this paper, the focus is on a vertical transistor arrangement, although it is realized that similar conclusions will be found for NWs arranged in a lateral geometry. The scaling rules should also be valid for transistors of different III–V materials such as GaN, with the appropriate changes to the intrinsic transistor performance and material parameters.

II. DESIGN

The evaluated transistor architecture is based on vertically grown InAs NWs, positioned in a hexagonally arranged array. A 3-D schematic illustration of the architecture is shown in Fig. 1(a), with a detailed cross section in Fig. 1(b) and top view in Fig. 1(c). This layout is consistent with a recent implementation of InAs NW transistors on a Si substrate [12]. In this fabrication process, the NWs are placed by electron beam lithography patterning of Au particles on III–V substrates, such as InP and InAs, or more recently on Si [13]. In this paper, a generic insulating substrate with the permittivity of Si $\varepsilon_{\text{sub}} = 11.7$ is considered. The NWs are covered by a high-$\kappa$ gate...
dielectric, such as Al₂O₃ or HfO₂. In this paper, a relative permittivity of ε₀ = 15 is assumed.

Vertically arranged metal electrodes are used to contact the NWs. Alternatively, an InAs epitaxial layer, if introduced below the NWs, may be utilized as a source contact. In the following analysis, source and drain contacts of Au and a gate electrode of W are considered. Spacer layers are introduced between the electrodes, ideally consisting of a mechanically stable material with a low relative permittivity. Common spacer layer materials include BCB and photoreists or inorganic films, such as SiO₂ and Si₃N₄ [14], [15]. The focus here is on a generic spacer layer material with a relative permittivity of εₑₑ = 3.

III. Method

In this paper, the transistor architecture is evaluated based on the main RF performance metrics f_T and f_max. To calculate f_T and f_max, the small-signal transistor model in Fig. 2 is used, corresponding to the structure in Fig. 1. The y-parameters are derived symbolically by employing modified nodal analysis using the MatLab tool SCAM to process a netlist of the model [16], [17]. As it is only possible to find an implicit analytical expression for the voltage over C_{gs,i} which controls the current source, an iterative numerical model is used. For each iteration, this voltage is changed semirandomly until it converges. From the y-parameters, the short-circuit current gain h_{21} and the unilateral power gain U are calculated to get f_T and f_max.

A. Parasitic Capacitances

Due to the complexity of the 3-D structure, an explicit analytical solution is not available to calculate the parasitic capacitances. Instead, it is necessary to use a numerical fast field solver, utilizing the boundary-element method [18]. In this paper, the FastCap algorithm is used, which has a high accuracy and is commonly utilized for capacitance calculations [19]. In order to simulate a wide range of designs rapidly, a FastCap interface is implemented within a MatLab application, designed to generate 3-D models from structural parameters. By automatically generating the 3-D structures from a script, instead of modeling them manually, it is possible to perform extensive sweeps of each structural parameter. The generated 3-D model contains the surfaces of the complete electrode structure, as well as the NW array, defining interfaces to regions with a specified dielectric constant. The NWs are modeled as open cylinders with a dielectric interface of zero permittivity inside. Although real NWs may be closer to a hexagonal shape, the difference in capacitance is negligible. The results should be applicable even on other shapes, such as square-shaped or triangular NWs.

The parasitic capacitances consist of two main parts, capacitances between the gate and the source and drain electrodes C_{gs,e} and C_{gd,e} and capacitances between the gate and the ungated parts of the NWs. To represent the distributed nature of the capacitances and resistances from the NWs, these parts are modeled by pi networks. This means that the gate–source and gate–drain capacitances from the NWs are split in one internal part C_{gs,wi} and C_{gd,wi} and one external part C_{gs,we} and C_{gd,we}. These parts, as well as the electrodes, are tagged as separate surfaces in the generated 3-D structure. As the fringing capacitance between the source and drain C_{ds} is relatively large in these devices, it is also included in the model. In the following analysis, the transistor is assumed to operate in saturation, i.e., the intrinsic gate–drain capacitance C_{gd,i} is zero. The intrinsic gate–source capacitance C_{gs,i}, in saturation, is two-thirds of the total intrinsic gate capacitance C_{gs,i}.

However, using a direct method to simulate the capacitances is problematic, as the simulation time increases exponentially with the size of the structure. To remove the dependence of...
the simulation time on the size of the NW array, a method based on linear extrapolation is used. In this case, the arrays are considered to be extended in only the y dimension (increasing \( n_y \)), which means that only a few increasingly larger arrays need to be simulated. It is found that each of the parasitic capacitances in Fig. 2 may be represented by one constant part, representing the ends of the array, and one part linearly dependent on the number of NWs, representing the core of the array. An almost perfect linearity of the parasitic capacitance scaling has been verified by comparison with the complete brute-force simulations. The linearization error is found to be smaller than the random error introduced by FastCap, which is a few percent at the used accuracy.

### B. Parasitic Resistances

The parasitic series resistances are calculated analytically and consist of three components: the resistances in the ungated parts of the NWs \( R_{S,w} \) and \( R_{D,w} \), the contact resistances in the metal–semiconductor interface \( R_{S,c} \) and \( R_{D,c} \), and the resistances in the electrodes \( R_{S,e} \) and \( R_{D,e} \). The gate resistance has only the electrode part and is denoted by \( R_G \).

The electrode resistances are calculated conventionally by the dimensions of the electrodes, employing the electrode resistivity. The gate electrode resistivity \( \rho_G \) is set to the bulk value \( \sqrt{2\rho_{\text{Au}} \mu_{\text{Au}}} \). It is realized that these resistivities may differ somewhat from those of thin films.

The NW resistivity \( \rho_w \) is assumed to be dependent on the doping concentration, as well as the mobility of the InAs NWs. This resistivity is manifested as series resistances \( R_{S,w} \) and \( R_{D,w} \), which scale with the NW diameter and array size, as well as the spacer layer thicknesses. This gives the following expression for the NW resistance on each side:

\[
R_w = \frac{4t_{gs,g} \rho_w}{d_{gw} \pi n_x n_y} \left[ \rho_w = \frac{1}{N_{D,w} \mu_t} \right].
\]

(1)

Here, \( \rho_w \) is the resistivity of the NWs, \( n_x \) and \( n_y \) are the number of NW columns and rows in the array, and \( t_{gs,g} \) denotes the thickness of the source or drain spacer. The intrinsic mobility of the NWs \( \mu_t \) differs from the bulk value and is assumed to be dependent on the diameter, due to the confined cylindrical geometry. Values for the intrinsic mobility of InAs NWs for different diameters, including surface scattering, are employed from [20] and is shown in Table I. In order to counteract the increasing resistance of thinner NWs, the doping of the NWs \( N_{D,w} \) is scaled geometrically with the cross-sectional area, from \( 10^{19} \, \text{cm}^{-3} \) for the 50-nm node to \( 10^{20} \, \text{cm}^{-3} \) for the 8-nm node. This is represented by the following expression, where \( \alpha = 0.0016 \, \text{nm}^{-2} \) and \( \beta = 0.0069 \, \text{nm}^{-2} \):

\[
N_{D,w} = \left( 1 + 9 \frac{4d_{nw}^2 - \alpha}{\beta} \right) \cdot 10^{19} \, \text{cm}^{-3}.
\]

(2)

It is assumed that the surface scattering dominates the introduced impurity scattering and that the consequential mobility degradation is negligible.

For the contact resistivity \( \rho_c \) Table II, data from [21] are used, which gives \( \rho_c = 1.0 \, \Omega \cdot \text{cm}^2 \). The contact resistivity is manifested as additional series resistances, dependent on the contact area between the NWs and the source and drain electrodes. The interface between the NW and the electrode is modeled as a distributed resistance network as described by [22]. This allows the effective contact resistance on each side \( R_{S,e} \) and \( R_{D,e} \) to be expressed as a function of the source or drain electrode thickness \( t_{S,D,e} \) and the transfer length \( L_T \):

\[
R_c = \frac{4\rho_w L_T}{\pi d_{gw} n_x n_y} \coth \left( \frac{t_{S,D}}{L_T} \right), \quad [L_T = \left( \frac{d_{nw} \rho_c}{4 \rho_w} \right).]
\]

(3)

### IV. Scaling Rules for III–V NW Transistors

In order to reduce the number of independent parameters, a set of scaling rules is proposed in Table III. In the scaling rules, three different types of parameters are identified: parameters derived from intrinsic properties, parameters limited by fabrication challenges, and parameters that are unknown and need to be optimized.
A. Parameters Derived From Intrinsic Properties

Some of the parameters are derived directly from intrinsic properties. One such parameter is the gate length $L_g$, which ideally should be as short as possible to reduce the gate capacitance. However, the gate length should also be long enough to avoid short-channel effects such as degradation of the subthreshold slope and drain-induced barrier lowering, as described by [23]. Short-channel effects degrade the intrinsic transistor performance, which would change the optimal transistor structure slightly. However, the optimization process would remain unchanged, regardless of short-channel effects.

It follows from [23] that $L_g > 4.6\lambda$, where $\lambda$ is the natural length for the InAs NW transistor. The natural length for a cylindrical geometry has been derived by [24] and is given by

$$\lambda = d_{nw} \sqrt{\frac{2\varepsilon_{InAs} \ln(1 + 2t_{ox}d_{nw})}{16\varepsilon_{ox}}}.$$  (4)

Here, $d_{nw}$ is the diameter of the NW, and $t_{ox}$ and $\varepsilon_{ox}$ denote the thickness and relative permittivity of the gate dielectric, whereas $\varepsilon_{InAs}$ is the relative permittivity of InAs. The resulting gate lengths for different NW diameters are presented in Table I. The gate dielectric thickness is set to 10% of the NW diameter but no thinner than 1 nm.

Values of the intrinsic gate capacitance $C_{gg,i}$ are derived from band structure calculations in [23]. This capacitance consists of a bulk geometric oxide capacitance in series with a semiconductor capacitance. Also, presented in [23] are intrinsic data for the transconductance $g_{mn}$ based on simulations of the ballistic transport. As there are no data available for the output conductance $g_{DL}$, they are simply assumed to be 10% of the transconductance. This corresponds to a voltage gain of ten, which is a typical value for highly scaled HEMTs [25]. Deviations from this value of the output conductance are shown later in Fig. 6(c). All data presented in Table I are given for a bias of $V_{gs} = 0.5$ V.

B. Parameters Limited by Fabrication Challenges

Many of the structural parameters are limited primarily by fabrication challenges and should ideally be as small as possible. In the scaling rules, these parameters are related to the NW diameter by a fixed scaling factor. One such parameter is the spacing between the NWs $s_{nwa}$, which is set to a single NW diameter, including the gate dielectric thickness. Note that the NW spacing is given as the edge-to-edge spacing. One additional spacing is included at the edges of the array, defining the internal electrode length $L$ and width $W$. This extra space provides margin for alignment in the fabrication process. The combined area defines the size of the internal source, drain, and gate electrodes. Additionally, the electrodes are extended 100 NW diameters $L_e$ outwards from the array, representing the external connection to the device. The gate and drain extend from the wide sides of the array, while the source extends in two directions from the thin sides.

The thicknesses of the source and drain electrodes $t_s$ and $t_d$ are set high enough for the distributed contact resistance to saturate, i.e., $t_{S,D} \gg L_T$, as seen in (3). This is the case above a few NW diameters for all technology nodes. In the scaling rules, the source and drain thicknesses are fixed at three times the NW diameter, whereas the gate electrode has a uniform thickness equal to the gate length.

V. Optimization of the Transistor Design

As presented in the last section, most of the structural parameters are either fixed by intrinsic properties or should be as small as possible. This leaves two parameters to be optimized: the spacer layer thickness and the configuration of the NW array. In this section, an optimization of these parameters is performed, and the performance is evaluated for each technology node, using $f_T$ and $f_{max}$ as figures of merit.

A. Spacer Layer Thickness and Array Size

The compromise between the parasitic capacitances and resistances is mainly set by the configuration of the NW array and the thicknesses of the spacer layers. The parameter values for each technology node are used, and $f_T$ and $f_{max}$ are calculated for a wide range of spacer layer thicknesses and array sizes. At this stage, the optimization is simplified by assuming the same spacer layer thickness at source and drain, and the width of the array is fixed at two NWs.

In Fig. 3, the result from this analysis for the 22-nm-diameter node is shown. For both designs, the optimal spacer layer thickness is around the size of the NW diameter, which is much thinner than what is usually demonstrated in fabricated devices [27], [28]. It is evident that an optimized $f_{max}$ design requires much thinner spacer layers than an $f_T$-oriented design. This is due to a higher sensitivity to the parasitic resistances compared to the parasitic capacitances.

Also, shown in these figures are the tolerance windows for 99%, 95%, and 75% performances. The tolerance has implications for the fabrication, specifying how good control of the fabrication process is required. It is obvious that above 95% of maximum $f_T$ performance is maintained for any spacer layer thicker than about 10 nm. Another observation is that $f_{max}$ is more sensitive to variations in the spacer layer thickness with 95% performance maintained up to about 30 nm. However, it is still possible to maintain over 75% of the maximum performance for almost any spacer layer thickness.

Concerning the optimal array size, the performance increases with a larger array size until around 30 NWs. This is due to a reduced influence of the parasitic capacitances from the edges of the array. However, as the array size is further increased, the influence from the edge capacitances diminishes. From this point, an increased parasitic capacitance is more degrading than the gain from a reduced series resistance. Another degradation for large array sizes is due to the constant source electrode resistance, as the width of the array is fixed at two NWs. It is also evident that a high $f_T$ is maintained for a slightly wider range of array sizes than $f_{max}$, allowing more room for customization of, e.g., drive current and power dissipation.

In Fig. 3, it is seen that there is a large overlap between the $f_T$ and $f_{max}$ optimizations. This means that the optimization...
JANSSON et al.: PERFORMANCE EVALUATION OF III–V NANOWIRE TRANSISTORS 2379

Fig. 3. Optimization of (a) $f_T$ and (b) $f_{\text{max}}$ with respect to the spacer layer thicknesses and the number of NWs in the array. The arrays are two NWs wide, and the source and drain spacer layer thicknesses are equal.

TABLE IV
TRANSISTOR PERFORMANCE FOR DIFFERENT ASPECT RATIOS OF THE NANOWIRE ARRAY AT THE 22-nm NODE

<table>
<thead>
<tr>
<th>NWs</th>
<th>$C_{gs}$ (aF)</th>
<th>$C_{gd}$ (aF)</th>
<th>$C_{ds}$ (aF)</th>
<th>$R_G$ (Ω)</th>
<th>$R_{S,D}$ (Ω)</th>
<th>$f_T$ (GHz)</th>
<th>$f_{\text{max}}$ (GHz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1x64</td>
<td>1340</td>
<td>1050</td>
<td>16</td>
<td>6.61</td>
<td>20.6</td>
<td>950</td>
<td>750</td>
</tr>
<tr>
<td>2x32</td>
<td>970</td>
<td>790</td>
<td>20</td>
<td>12.1</td>
<td>13.1</td>
<td>1180</td>
<td>920</td>
</tr>
<tr>
<td>4x16</td>
<td>810</td>
<td>680</td>
<td>29</td>
<td>2.37</td>
<td>9.10</td>
<td>1330</td>
<td>1020</td>
</tr>
<tr>
<td>8x8</td>
<td>750</td>
<td>650</td>
<td>47</td>
<td>4.54</td>
<td>7.00</td>
<td>1400</td>
<td>1040</td>
</tr>
</tbody>
</table>

The increase in $f_{\text{max}}$ for a square-shaped array may seem counterintuitive, as $f_{\text{max}}$ should be highly dependent on the gate resistance. This is also the result when using an approximation of $f_{\text{max}}$, such as that in [29]. However, this model does not fully take into account the effect of the large parasitic series resistances in these devices, thus overestimating $f_{\text{max}}$ and the dependence on the gate resistance. The result is that a reduction of the fringing capacitances, as the perimeter-to-area ratio is the smallest for a square-shaped array, is more important for the performance than a reduced gate resistance.

As the optimization suggests the use of square-shaped arrays, it is realized that it would be advantageous to revise the electrode structure. By letting the gate, instead of the source, extend in two directions from the array, the gate resistance is cut in half. This modification increases $f_{\text{max}}$ by about 6.4% while reducing $f_T$ with only about 1.5%. The data presented in Table IV are based on the double-gate simulations.

C. Roadmap

The full optimization done for the 22-nm node is repeated for all other technology nodes in the roadmap. The result is presented in Table V with detailed information about the parasitic elements provided in Fig. 4. It is evident that the conclusion about the optimal array configuration is maintained as the structure is scaled. Additionally, the overlap between the $f_T$ and $f_{\text{max}}$ optimizations increases as the optimal spacer layer thicknesses converge. Regarding the performance, an improvement is predicted for all nodes down to 8 nm, with a maximum $f_T$ of 2.7 THz and an $f_{\text{max}}$ of 1.8 THz.

As the device is scaled, the series resistance from the thin NWs increases exponentially. It is imperative to counteract this resistance by increasing the doping. Otherwise, the performance may saturate below the 22-nm node, as the degradation from the exponentially increasing series resistance is much higher than the gains from the linearly decreasing parasitic capacitances and a slightly higher intrinsic transconductance.

VI. Benchmarking

To evaluate the magnitude of the parasitic elements for the vertical InAs NW transistors, they are compared to the ITRS for conventional Si transistors, as presented in Table VI [26]. The comparison is made for devices of the same gate length with interpolation of the ITRS data when necessary. The intrinsic and total gate capacitances, as well as the total parasitic series resistances, are normalized by the total gate width. It is evident

Paper I 105
that the intrinsic gate capacitances in the NW transistor are about half of those in the ITRS, which is due to the small density of states for InAs. However, the parasitic capacitances are higher, and this contribution increases as the device is scaled down. Nevertheless, the total gate capacitance is comparable to or even lower than that of the ITRS. This fact, combined with a higher transconductance and better gate control, ensures that vertical III–V NW transistors are highly competitive to traditional planar silicon-based devices.

Another observation is that the basic scaling behavior for an optimized vertical NW FET is different from the planar MOSFETs in the ITRS. In the ITRS, the series resistance is kept almost constant, whereas the total gate capacitance is decreased for each generation. Conversely, in the NW transistor scaling, the total gate capacitance decreases slightly, whereas the series resistance increases due to the higher NW resistance. Even so, in this aggressive scaling, the series resistance is lower than that in the ITRS. However, it is evident that one of the main challenges with these devices is to further reduce the series resistance in the NWs.

VII. EVALUATION OF DESIGN PARAMETERS

To study deviations from the roadmap, a number of independent parameters are varied systematically, and the results are compared with the optimization for the 22-nm node. A material parameter that severely influences the performance is the resistivity of the ungated parts of the NWs. By varying the resistivity, it is possible to study the effects caused by the NW mobility, doping, or the contact resistance in a single parameter. This is important as there is a relatively large uncertainty about the values of these parameters, which may vary considerably between different fabrication processes.

In Fig. 5, the NW resistivity is varied between 0.55 and 55 \( \mu \Omega \cdot \text{m} \) for the 22-nm node. It is evident that the performance is degraded and that the fabrication tolerance is very much stricter when the resistivity is increased, with demands of spacer layer thicknesses down to less than half the NW diameter. This gets particularly problematic for \( \rho_{\text{nw}} = 55 \mu \Omega \cdot \text{m} \), where most of the performance is lost for spacer layers thicker than a few tens of nanometers.

One key parameter for decreasing the parasitic capacitances in the device is the distance between the NWs. Note that changing the NW distance also changes the size of the electrodes, as described by the scaling rules. In Fig. 6(a), the change in \( f_T \) and \( f_{\text{max}} \) by varying this parameter is presented. It is apparent that densely packed NWs are essential to achieve the highest possible performance. On the other hand, frequencies near the terahertz regime may still be achieved even for a NW spacing a few times larger than the NW diameter.

Finally, the influence of the intrinsic transistor performance is studied by varying the transconductance and output conductance, as shown in Fig. 6(b) and (c), respectively. This makes it possible to predict the impact from a degradation of the intrinsic performance, due to, e.g., short-channel effects. It is observed that the performance increases almost linearly with \( g_m \), as expected, with a slight saturation for large values due to the parasitics. Furthermore, it is seen that an increase in \( g_d \) may severely degrade the performance, particularly \( f_{\text{max}} \). The effect is much smaller on \( f_T \), which is expected, as a high output conductance primarily limits \( f_{\text{max}} \).

A method to reduce the series resistance is to increase the NW diameter by epitaxial regrowth. This decreases the resistance at the cost of increased parasitic capacitances. The effect of regrowth on the drain side is shown in Fig. 7 for the 22-nm node. It is evident that regrowth is unnecessary and even degrading for the performance at doping levels above \( 10^{19} \text{cm}^{-3} \). However, if the doping level is \( 10^{18} \text{cm}^{-3} \) or lower, it is possible to regain some performance by increasing the diameter. Another observation is that the source and drain should be inverted, in order to have the low resistance on the source side, which helps to increase the voltage at the internal \( V_{\text{gs}} \).
Fig. 5. Effects on the geometrical mean of $f_T$ and $f_{\text{max}}$ from changes in the NW resistivity, at the 22-nm node.

Fig. 6. Influence on $f_T$ and $f_{\text{max}}$ at the 22-nm node from changing (a) NW spacing and derived parameters, (b) $g_m$ with $g_d = 0.1 \cdot g_m$, and (c) $g_d$ for $g_m = 4.86 \, S/mm$.

Fig. 7. Effect on the geometrical mean of $f_T$ and $f_{\text{max}}$, when the diameter of the NWs on the drain side is increased by epitaxial regrowth from 22 up to 44 nm. This is presented for different NW doping levels in per cubic centimeter, with the source at either the top or bottom contact.

VIII. Conclusion

In this paper, an optimization has been performed of all structural parameters in a vertical III–V NW transistor architecture, for NWs with diameters from 50 down to 8 nm. This optimization provides a roadmap of desired device dimensions to achieve the optimum balance between parasitic capacitances and resistances at each technology node. Additionally, the performance has been predicted for each nod, and it is concluded that an $f_T$ and $f_{\text{max}}$ performance beyond 1 THz may be achieved and that it increases as the devices are scaled.

It has been shown that the thickness of spacer layers should be similar to the NW diameter, on the order of a few tens of nanometers. Furthermore, it is clear that it is more important to have good control of the spacer layer thickness than to increase the density of the NW array. An increase in the resistivity of the NWs narrows the tolerance window for the spacer layers, demanding layers thinner than 10 nm. The arrays should consist of about 50 NWs, in order to reduce the series resistance, and be arranged in a square-shaped array, in order to reduce the fringing capacitances.

By benchmarking against the ITRS, it has been demonstrated that the parasitic capacitances may be reduced to levels below the ITRS by a dense packing of the NWs. Even though the parasitic capacitances are higher, this is compensated by the lower intrinsic gate capacitance. However, the performance degradation, due to the series resistance, increases as the device dimensions are scaled. It is concluded that a good control of the doping, contact resistance, and the spacer layer thickness is imperative.

ACKNOWLEDGMENT

The authors would like to thank M. Berg for his collaboration to produce the illustration of the 3-D structure in Fig. 1.
REFERENCES


Kristofer Jansson received the M.S. degree in engineering nanoscience from Lund University, Lund, Sweden, in 2010, where he is currently working toward the Ph.D. degree in the Department of Electrical and Information Technology.

Erik Lind received the Ph.D. degree in solid-state physics from Lund University, Lund, Sweden, in 2004. He is currently an Associate Professor with the Department of Electrical and Information Technology, Lund University.

Lars-Erik Wennberg received the Ph.D. degree in solid-state physics from Lund University, Lund, Sweden, in 1998. Since 2005, he has been a Professor in nanoelectronics at Lund University.
Intrinsic Performance of InAs Nanowire Capacitors

Kristofer Jansson, Erik Lind, and Lars-Erik Wernersson

Abstract—The intrinsic properties of vertical InAs nanowire (NW) capacitors are investigated. The band structure is simulated using a Schrödinger–Poisson solver, taking the conduction band nonparabolicity into account. This is combined with a distributed RC model to simulate the current–voltage characteristics. It is found that the influence from the nonparabolicity is substantial for devices with a small nanowire diameter, resulting in an increased carrier concentration, a shift in the threshold voltage, and a higher intrinsic capacitance. These NW capacitors may be a suitable alternative in high frequency applications approaching 100 GHz, while maintaining a quality factor above 100.

Index Terms—Capacitor, InAs, modeling, nanowires (NWs).

I. INTRODUCTION

NANOWIRE (NW) devices are gaining increasing attention based on the improved electrostatic control demonstrated for Si as well as III–V NW devices. Besides transistor implementations, NW capacitors have also been realized [1]–[6]. The electrical properties of the NW capacitors provide information about the materials utilized and the data can be used to determine the charge density in the channel. For narrow diameter NWs, the electrical properties are also influenced by quantum confinement and band structure effects [7], [8]. It is hence essential to quantify the contribution from such effects and to calculate the intrinsic capacitance in order to interpret the experimental data correctly.

In this paper, we simulate the intrinsic properties of thin InAs NW capacitors. In order to account for the quantum confinement, due to the small dimensions of the NWs, a customized self-consistent Schrödinger–Poisson solver has been developed. Additional complexity is introduced by the cylindrical geometry and from considering conduction band nonparabolicity of high-energy states. The influences from a cylindrical geometry, as well as from the nonparabolicity, have been quantified in previous work [9]–[14]. However, a complete, numerical band-structure simulation, incorporating both of these effects simultaneously, has not been presented.

Furthermore, a distributed RC model is introduced in order to consider the effect of the series resistance inside the NW. The band structure simulations are subsequently combined with the distributed RC circuit model, in order to simulate the frequency-dependent current–voltage (C–V) characteristics of the complete NW capacitor structure.

Experimental verification of the model is very difficult at this stage, due to challenges both in the fabrication and characterization of these devices. The effects from the nonparabolicity are likely to be completely overshadowed by the uncertainty in various process parameters, such as the doping concentration and the quality of the dielectric interface. It is known that traps at the dielectric interface may degrade the performance of the device [15]–[17]. However, the effects of traps are not considered in this paper as the focus is on the intrinsic physics and the ultimate performance of the device architecture. For experimental data, please refer to other published works in the field [1] and [18].

II. STRUCTURE AND MATERIALS

In this paper, a capacitor structure consisting of a single InAs NW is investigated, focusing on three NW diameters 12, 22, and 35 nm. The NWs are aligned in a vertical configuration, as shown in the cross-sectional schematic in Fig. 1. The lengths of these NWs are typically a few hundred nanometers and they are grown epitaxially from seed particles on a semiconductor substrate. This structure is similar to previously fabricated capacitor and transistor structures [1], [2].

The NWs are covered with a thin gate dielectric, in this paper fixed at 3.5 nm. Common gate dielectrics with a high relative permittivity include Al2O3 or HfO2. Here, the gate dielectric is assumed to have a relative permittivity of 15. The NWs are contacted by metal electrodes, arranged vertically around the NW.

Isolating spacer layers are introduced between the electrodes, typically consisting of an organic layer or an inorganic material, such as SiO2 or Si3N4. A good spacer layer material should have high mechanical stability, while maintaining a low relative permittivity to minimize the parasitic capacitances. In this paper, the focus is on the intrinsic performance of the capacitor, although it would be easy to add these parasitic elements to the simulation, as described in [19].

III. SCHRODINGER–POISSON SOLVER

In order to calculate the band structure inside the NW, a self-consistent Schrödinger–Poisson solver has been developed. The foundation for this method is the traditional Schrödinger–Poisson solver, which has been extensively utilized to study the band structure of a wide range of devices [20]. In order to be applicable on NWs, the equations are formulated in the cylindrical geometry [11]. Also, the strong nonparabolicity within the InAs conduction band needs to be modeled. For this, a modified approach of the solution method is used in [12].

Self-consistency is ensured by subsequently solving the Poisson equation to calculate the electrostatic potential, and...
the Schrödinger equation to calculate the band structure and thus the carrier concentration. The result from the Poisson solver is used as input to the Schrödinger equation and vice versa until convergence is achieved. A schematic overview of the complete solution method is shown in Fig. 2.

A. Poisson Solver

The NW geometry is modeled by a radially varying electrostatic potential, \( V(r) \), which is symmetrical around the center of the NW. The potential is assumed to also vary along the NW and this influence will be modeled by a distributed RC network. In order to calculate the electrostatic potential along the radial dimension, \( r \), the Poisson equation is formulated in polar coordinates

\[
\frac{d^2}{dr^2} + \frac{1}{r} \frac{d}{dr} V(r) = -\frac{g(N_d - n(r))}{\varepsilon_r \varepsilon_0}
\]

where \( N_d \) is the NW doping concentration, \( n \) is the carrier concentration, and \( \varepsilon_r \) is the relative permittivity of the semiconductor material. The correct carrier concentration is unknown at this stage and will later be found by iterating the Schrödinger–Poisson solver. In order to seed the first iteration, an estimation is performed using Fermi–Dirac statistics

\[
n(r) = N_c F_{1/2} \left( -\frac{q V(r)}{kT} \right)
\]

where \( N_c \) is the effective conduction band density of states and \( F_{1/2}(E) \) is the half-order Fermi integral. The Fermi level is used as a reference for the energy at 0 eV. This equation is solved using the shooting method, utilizing a fourth-order Runge–Kutta solver with repeated interval bisection. This numerical solution is very fast compared with the solution of the Schrödinger equation and is not a bottleneck in the simulations.

B. Schrödinger Solver

Using the electrostatic potential calculated from the Poisson equation, the corresponding eigenvalues and eigenfunctions are calculated using the Schrödinger equation. Each of these correspond to a subband in the NW band structure.

The eigenvalues of each subband, \( E_{n,\nu} \), consist of two quantum numbers: 1) one radially dependent, \( n \) and 2) one azimuthal, \( \nu \). The azimuthal quantum numbers, corresponding to the angular component of the geometry, are derived from symmetry and must be a positive integer [11]. For each azimuthal quantum number, a set of radial quantum numbers, along with the associated eigenfunctions, \( \psi_{n,\nu}(r) \), are calculated using the Schrödinger equation

\[
E_{n,\nu} \psi_{n,\nu}(r) = \left[ -\frac{\hbar^2}{2m^*} \frac{d^2}{dr^2} + \frac{1}{r} \frac{d}{dr} - \frac{\nu^2}{r^2} \right] + V(r) \psi_{n,\nu}(r).
\]

The eigenvalue problem of the Schrödinger equation is solved by utilizing the shooting method, where different eigenvalues are tested until matching boundary conditions are found. An advantage with the shooting method is that the simulation complexity increases linearly with the number of bands, as well as with the resolution. Furthermore, only subbands that are actually influencing the carrier concentration are included. This is done by excluding states with an energy above a threshold of 0.5 eV above the Fermi level.

Normalization of the wave function is performed, after the final boundary condition is found, which also allows for proper normalization of the initial boundary condition. This step is essential to correctly incorporate the nonparabolicity.

C. Nonparabolicity

In order to account for conduction band nonparabolicity, a factor \( \alpha \) is introduced in the Schrödinger solver, as described in [12]. For InAs, \( \alpha = 1.4 \text{ eV}^{-1} \), which describes the magnitude of deviation from a parabolic band shape [21]. It is known that this parameter may be dependent on the device structure, although this effect is not considered in this paper. The validity of this approximation has been verified by comparison to the tight-binding simulations in [7], where a good correspondence is demonstrated.

The resulting equation, compensated for nonparabolicity is presented below, where the wave vector, \( K \), is introduced as
Fig. 3. (a) Energy dispersion relation and (b) wave functions for each subband up to 0.5 eV above the Fermi level. In (c), the resulting density of states is shown. These data are presented for a NW with a diameter of 12 nm at a surface potential of 0.5 V. Similarly, the results for a NW with a diameter of 22 nm are shown in (d)–(f) and for 35 nm in (g)–(i).

an additional parameter

\[(1 + \bar{a}) E_{n,\nu}(K)|\psi_{n,\nu}(r)\rangle = \left[ -\frac{\hbar^2}{2m^*} \left( \frac{d^2}{dr^2} + \frac{1}{r} \frac{d}{dr} \frac{v^2}{r} - K^2 \right) + (1 + 2\bar{a})V(r) \right. \langle \psi_{n,\nu}(r)(r)|a V(r)|\psi_{n,\nu}(r)\rangle \bigg|_{\nu=\nu}\right. \langle \nu=\nu\rangle.\]  

\[(4)\]

\[\bar{a} = \langle \psi_{n,\nu}(r)|a V(r)|\psi_{n,\nu}(r)\rangle E_{n,\nu}(K).\]  

\[(5)\]

In order to calculate the carrier concentration, the density of states must be calculated and this requires the complete dispersion relation for each subband. By solving the modified Schrödinger equation in (4) and (5), for different values of \(K\), it is possible to calculate the complete nonparabolic dispersion relation, \(E_{n,\nu}(K)\). For each specific \(K\), a similar solution method as for the simpler model in (3) is utilized, by iterating (4) and (5) until convergence. As an initial guess, the result from the parabolic solution of (3) is used.

D. Calculation of the Carrier Concentration

From the dispersion relation, the corresponding density of states, \(g(E)\), for each subband, \(E_{n,\nu}\), is calculated. This is combined with the Fermi–Dirac distribution, \(f(E)\), which allows the number of occupied states to be calculated by integrating over the subband energies. This should theoretically be integrated from the bottom of each subband up to infinite energy, but in practice it is sufficient to limit the numerical calculations to 0.5 eV above the Fermi level.

Combined with the probability distribution, i.e., the square of the absolute eigenfunction, \(\left|\psi_{n,\nu}(r)\right|^2\), the radial distribution of carriers in the NW, \(n(r)\), is calculated for each subband. Finally, the contributions from all subbands are added together to get the total distribution of the carrier concentration

\[n(r) = \sum_{n,\nu} \left|\psi_{n,\nu}(r)\right|^2 \int_{E_{n,\nu}(0)}^{\infty} f(E) g(E) dE\]  

\[(6)\]

\[g(E) = \frac{2}{E} \left( \frac{dE}{dK} \right)^{-1} f(E) = \frac{1}{1 + e^{E/kT}}.\]  

\[(7)\]
Fig. 4. (a) Resulting electrostatic potential and (b) carrier concentration in a NW with a diameter of 12 nm, for a surface potential of 0.5 V. In (c), the potential at the surface, as well as in the center, of the NW is shown as a function of the gate voltage. In (d), the carrier concentration in the middle of the wire is contrasted to the mean carrier concentration for different gate voltages. Similarly, the results for a NW with a diameter of 22 nm are shown in (e)–(h) and for 35 nm in (i)–(l).

E. Self-Consistency

As the carrier concentration is dependent on the electrostatic potential and vice versa, a method to ensure self-consistency is required. This is done by restarting the solution scheme using the carrier distribution calculated by (6) instead of (2) as input to the Poisson solver, which in turn will give a new electrostatic potential. The complete solution method in Fig. 2 is iterated until convergence is achieved, i.e., the new electrostatic potential is within a given tolerance level of the potential from the last iteration.

IV. Band Structure

The dispersion relations for all subbands, up to 0.5 eV above the Fermi level, are shown in Fig. 3(a), (d), and (g) for a NW with a diameter of 12, 22, and 35 nm, respectively. It is clearly observed that the shape gets less parabolic and more linear as the energy is increased, when nonparabolic effects are considered.

An overview of all the subband eigenfunctions are shown in Fig. 3(b), (e), and (h). It is evident that most of the contribution from the eigenfunctions is located in the middle of the NW, especially in the 12-nm case. In the 22- and 35-nm case, a lot more subbands with higher harmonics are populated, which allows more carriers near the surface of the NW.

The calculated density of states is shown in Fig. 3(c), (f), and (i). The nonparabolicity increases the number of occupied subbands drastically, as well as allowing more states per subband. This results in a much higher density of states for the nonparabolic case. The effect gets stronger as the energy levels are increased further up in the band structure.

It is also evident that the nonparabolicity shifts the subbands to lower energies and this effect gets increasingly stronger for higher energy subbands. The lower confinement energy is due to a change in the apparent effective mass from $m^*$ in the parabolic approximation, to $m^*(1 + \alpha)$, when conduction band nonparabolicity is considered [12]. This effect is most prominent in the device with the relatively thin diameter of 12 nm and is clearly visible for the second subband in Fig. 3(c).

However, this effect is counteracted by a stronger bending of the electrostatic potential, caused by a higher carrier...
Fig. 5. (a) Resulting electrostatic potential and (b) corresponding carrier concentration distribution for both the parabolic and nonparabolic models, calculated for varying surface potentials. The resulting gate-source voltage is illustrated with the green line, for 0, 0.5, and 1 V. Similarly, the results are shown for a 22-nm NW in (c) and (d) and for 35 nm in (e) and (f).

In Fig. 4(a), (e), and (i), the strongest band bending is found close to the surface where most of the high-energy carriers will be located. This causes the quantum well close to the surface to be thinner and deeper, which pushes the subbands to higher energies. This effect is more prominent in the thicker 35-nm NW, as most of the carriers are located near the surface due to the bad gate control of the potential in the center of the NW.

In Fig. 4(c), (g), and (k), the effect of an applied gate voltage on the surface and center potential is illustrated. It is easier to control the potential in the center of a thin NW in comparison to a thicker one, as expected, whereas there is not much difference in the control of the surface potential. When accounting for nonparabolicity, it is predicted that the potential will be more difficult to control, which is due to the increased number of carriers in the NW.

Similarly, in Fig. 4(d), (h), and (l), it is seen that most of the carriers are concentrated in the center for the 12-nm case. This is contrasted in the 35-nm case, where most of the carriers are concentrated near the NW boundary. The NW with a 22-nm diameter shows a transition between these two extremes, where the carriers are more evenly distributed. This reduces the influence of the nonparabolicity as less high-energy states are occupied. It is also evident that nonparabolic effects become more prominent as the gate voltage is increased.

V. CARRIER DISTRIBUTION

The result of the full band structure simulation for a NW with a 12-nm diameter is shown in Fig. 4(a) and (b). The data show that most of the carriers in the NW are concentrated in the middle of the NW with a volume distribution. For comparison, the classical solutions are also included, predicting a substantial carrier density at the NW boundary, violating the quantum mechanical solution. Furthermore, it is evident that the nonparabolicity increases the carrier density with ∼30%. This change is clearly observed in the shape of the electrostatic potential as well.

The process is repeated for the 22- and 35-nm NWs and the results are shown in Fig. 4(e)–(h) and (i) and (j), respectively. In these cases, and particularly in the 35-nm NW, the majority of the carriers is instead close to the boundary. The more complicated shape of the distribution is due to the influence of many more subbands in the wide diameter NWs. It is also evident that there is much less gate control of the electrostatic potential in the center of the NW, in comparison to the 12-nm case.

In Fig. 4(c), (g), and (k), the effect of an applied gate voltage on the surface and center potential is illustrated. It is easier to control the potential in the center of a thin NW in comparison to a thicker one, as expected, whereas there is not much difference in the control of the surface potential. When accounting for nonparabolicity, it is predicted that the potential will be more difficult to control, which is due to the increased number of carriers in the NW.

Similarly, in Fig. 4(d), (h), and (l), it is seen that most of the carriers are concentrated in the center for the 12-nm case. This is contrasted in the 35-nm case, where most of the carriers are concentrated near the NW boundary. The NW with a 22-nm diameter shows a transition between these two extremes, where the carriers are more evenly distributed. This reduces the influence of the nonparabolicity as less high-energy states are occupied. It is also evident that nonparabolic effects become more prominent as the gate voltage is increased.

A. Surface Potential Sweep

By calculating the electrostatic potential and the carrier concentration for different surface potentials, the results shown in Fig. 5(a)–(f) are achieved. It is noted that the influence from the nonparabolicity is strongest for high surface potentials. This is expected due to the presence of more occupied high-energy states, where the nonparabolic effects are most influential. A deeper potential well also results in a larger difference in the amount of subbands, as the previously
described increase of the effective mass shifts the subbands to lower energies.

VI. C–V CHARACTERISTICS

From the surface potential sweep, all information required to calculate the C–V characteristics is available. In order to perform this calculation, a circuit model has been developed consisting of a distributed RC network. All capacitive and resistive elements in this model may be calculated from the distribution of the carrier concentration in the surface potential sweep. For the resistivity calculations, electron mobilities of 1350, 3460, and 6200 cm²/Vs are employed for InAs NWs with diameters of 12, 22, and 35 nm, respectively [22], [19].

The complete capacitor circuit model is shown in Fig. 6. For the following analysis, a distribution of 100 nodes along the vertical dimension is used. In order to calculate the resulting impedance of the circuit model, the netlist solver LTSPICE IV is utilized [23]. By using a customized interface, developed in MATLAB, netlists with arbitrary circuit elements are constructed, which subsequently are simulated in LTSPICE IV and the results fetched back into MATLAB. This enables automation of all simulation steps, which is essential to be able to perform an extended parameter analysis.

The resulting C–V characteristics are shown in Fig. 7(a) for a NW with a diameter of 12 nm and a gate length of 500 nm. The capacitance value remains essentially the same for frequencies up to ~100 GHz. For higher frequencies, the capacitance is reduced due to the series resistance.

It is evident that the effects originating from the nonparabolicity will influence the C–V characteristic, as shown in Fig. 7(b) at a low frequency. One effect is a flattening of the C–V curve, due to the presence of more states in each band. This effect gets more prominent for higher subband energies, where the capacitance degrades less than in the parabolic case. Also evident is the increased amount of subbands in the voltage range, due to the shift of the subbands to lower energies. The combination of these effects results in a higher predicted capacitance than in the parabolic approximation. The shift of the lowest subband to a lower energy also causes a slight downward shift of the threshold voltage.

For NWs with a diameter of 22 and 35 nm, both with a 500-nm gate length, the results are similar, as shown in Fig. 7(d) and (g), respectively. However, due to much larger amount of subbands, the shape of the C–V characteristic resembles the planar behavior, where the contributions from individual subbands are difficult to distinguish. It is also noted that the 22- and 35-nm NWs maintain the capacitance up to higher frequencies, thanks to the much lower series resistance in the wider NWs.

Noteworthy is also that the effect of the nonparabolicity is weaker in the 35-nm case, as shown in Fig. 7(b). This is due to less influence of high-energy states as a higher proportion of the carriers occupy low-energy states in any of the numerous subbands. Interestingly, the nonparabolicity is even less influential in the 22 nm-case, as shown in Fig. 7(e), due to a more even distribution of carriers within the NW. This may be contrasted to the 12-nm NW, where most of the carriers are located in the center of the NW, and the 35-nm NW, where most of the carriers are located close to the NW surface.

VII. QUALITY FACTOR

A conventional figure of merit to evaluate the performance of a capacitor is the quality factor, $Q$, defined as the ratio between the reactive, $X_C$, and resistive, $R_S$, parts of the effective impedance, $Z$

$$Q = \frac{X_C}{R_S} = \frac{1}{\omega R_S C_S} \left[ Z = R_S + j X_C \right].$$  

In Fig. 7(c), (f), and (i), the variation of the quality factor with the gate voltage and frequency is presented. It is clear that the quality factor varies greatly with the frequency, ranging from ~1 at 100 GHz to ~10³ at 1 MHz, for a NW with a 12-nm diameter. For the 22- and 35-nm NWs, the quality factor is about a factor three and six higher, respectively. This is mainly due to the much lower series resistance. Furthermore, it is obvious that the quality factor is greatly degraded for low gate voltages due to the increased series resistance.

A detailed overview of the quality factor, and how it is affected by the gate length, is shown in Fig. 8(a) for a 12-nm diameter NW. The quality factor is determined at a surface potential of 0.5 V and the effects of a nonparabolic conduction band is considered. It is evident that the increased NW resistance causes a degradation of the quality factor at long gate lengths. In order to maximize the performance, the gate length should be no longer than a few tens of nanometers. When the gate length is increased above a few hundred nanometers, the performance seems to slowly saturate. The quality factors of the 22- and 35-nm NWs are shown in Fig. 8(b) and (c). The conclusions are similar, but the quality factor is increased for wider diameters due to the lower series resistance.

It is evident that the quality factor may exceed 100 at 60 GHz for short gate lengths. This may be compared with a conventional thin-gate oxide MOS varactor, with a quality factor of ~20 at 20 GHz [24]. Other implementations have shown quality factors close to ten at 60 GHz [25]. The quality factor may be substantially improved by the use of two contact electrodes at either end of the NW or by placing multiple NWs in an array.
Fig. 7. Resulting $C-V$ characteristic at different frequencies for a NW with a diameter of (a) 12, (d) 22, and (g) 35 nm, when nonparabolicity is considered. The difference between the parabolic and nonparabolic $C-V$ at a low frequency is shown in (b), (e), and (h) for NW with a diameter of 12, 22, and 35 nm, respectively. In (c), (f), and (g), the variation of the quality factor with the gate voltage and frequency is presented.

Fig. 8. Quality factor for capacitors with NW diameters of (a) 12, (b) 22, and (c) 35 nm, for varying frequencies and gate lengths. The quality factor is extracted at a surface potential of 0.5 V.

VIII. CONCLUSION

The intrinsic capacitance of InAs NW capacitors has been studied in detail taking the conduction band nonparabolicity into account. It is concluded that the influence from nonparabolicity is substantial, especially for high gate voltages, as the increased density of states increases the carrier concentration. It is also evident that the nonparabolicity introduces a small shift in the threshold voltage, due to the shift of the subbands to lower energies.

It has been demonstrated that optimized InAs NW capacitors will show little intrinsic frequency dispersion up to $\sim 100$ GHz, while maintaining a quality factor above 100. These values will be reduced as the diameter is scaled from
35 to 12 nm due to increased series resistance. The high quality factor at high frequencies, combined with a high capacitance modulation, presents these devices as a viable alternative for high-performance circuit implementations.

REFERENCES


Kristofer Jansson received the M.S. degree in engineering nanoscience from Lund University, Lund, Sweden, in 2010, where he is currently pursuing the Ph.D. degree with the Department of Electrical and Information Technology.

Erik Lind received the Ph.D. degree in solid-state physics from Lund University, Lund, Sweden, in 2004. He is currently an Associate Professor with the Department of Electrical and Information Technology, Lund University.

Lars-Erik Wernersson received the Ph.D. degree in solid-state physics from Lund University, Lund, Sweden, in 1998. He has been a Professor of nanoelectronics from Lund University since 2005.
Ballistic Modeling of InAs Nanowire Transistors
Kristofer Jansson, Erik Lind, Lars-Erik Wernersson

Department of Electrical and Information Technology, Lund University, Sweden

Abstract

In this work, the intrinsic performance of InAs nanowire transistors is evaluated in the ballistic limit. A self-consistent Schrödinger-Poisson solver is utilized in the cylindrical geometry, while accounting for conduction band non-parabolicity. The transistor characteristics are derived from simulations of ballistic transport within the nanowire. Using this approach, the performance is calculated for a continuous range of nanowire diameters and the transport properties are mapped. A transconductance exceeding 4 S/mm is predicted at a gate overdrive of 0.5 V and it is shown that the performance is improved with scaling.

Furthermore, the influence from including self-consistency and non-parabolicity in the band structure simulations is quantified. It is demonstrated that the effective mass approximation underestimates the transistor performance due to the highly non-parabolic conduction band in InAs. Neglecting self-consistency severely overestimates the device performance, especially for thick nanowires. The error introduced by both of these approximations gets increasingly worse under high bias conditions.

Keywords:
Nanowire, MOSFET, InAs, Ballistic, Simulation, Modeling

1. Introduction

Nanowire transistors are developed to extend the performance scaling of conventional transistors [1], [2], [3], [4]. Combining the high mobility of III-V materials with the excellent gate control provided by a gate-all-around architecture, the performance of these devices has been predicted to reach the THz regime [5], [6], [7].

The band structure of highly scaled nanowires has been extensively studied using various approaches [8], [9], [10], [11], [12], [13]. A common approach is to use a tight-binding model to simulate the band structure [14], [15], [16], [17], [18]. However, the use of a tight-binding model is very computationally expensive, especially for thick nanowires with many subbands. This limits a reasonable analysis to a small number of data points with thin nanowires. In order to reduce simulation time, self-consistency between the band structure and the electrostatic potential is occasionally disregarded [15]. Others have utilized kp perturbation theory or resorted to the basic effective mass approximation, which is a relatively rough approximation, especially under high bias conditions [19], [20], [21]. Although this approximation may be reasonable for Si, the highly non-parabolic conduction band in InAs introduces a significant error [22].

In this work, the intrinsic performance of an InAs nanowire transistor structure is studied, following the approach previously used in [22], where the intrinsic properties of an InAs nanowire capacitor were determined by the development of a Schrödinger-Poisson solver adapted to InAs nanowire structures [23], [24], [25]. This solution method is customized for the cylindrical geometry and is fully self-consistent, while accounting for the non-parabolic conduction band in InAs. Furthermore, the simulations are repeated using the effective mass approximation, as well as for a constant electrostatic potential, in order to quantify the effect of these approximations.

The calculated band structure is used to derive the transistor characteristics, using a conventional ballistic transistor model [26], [27]. The ballistic transport is modeled self-consistently with regard to the indirect control of the gate barrier height, including the influence of the quantum capacitance [26]. The model is sufficiently fast to simulate even relatively thick nanowires on a standard laptop. This allows an analysis to be performed for a continuous range of nanowire diameters between 5 nm and 35 nm. Furthermore, this work provides a complete insight into the transport properties of differently scaled nanowires, by mapping the location of the carrier concentration and velocities within the nanowire.

2. Structure and materials

The transistor structure considered in this study consists of a single InAs nanowire, covered by a high-κ dielectric with an assumed permittivity of 15, roughly corresponding to HfO₂. The thickness of the gate dielectric is scaled as one tenth of the nanowire diameter, although no thinner than one nanometer. The gate is aligned in a gate-all-around configuration, in order to maximize the control of the electrostatic potential. A schematic illustration of a generic nanowire transistor structure is presented in Fig. 1. This transistor structure is similar to the nanowire transistors fabricated in [28], [29], [30], [31].
results in this work are relevant for both vertical and lateral transistor architectures.

Nanowire diameters from 5 nm to 35 nm are simulated, with emphasis on 12 nm and 22 nm. In this work, no parasitic elements are considered, as the focus is on the ultimate intrinsic performance of the transistor structure. An overview of the influence of the parasitic capacitances and resistances is presented in [5].

3. Band structure simulations

In order to calculate the electrostatic potential within the nanowire, \( V(r) \), Poisson’s equation is solved in polar coordinates along the radial dimension, \( r \), for a given surface potential:

\[
\frac{d^2}{dr^2} + \frac{1}{r} \frac{d}{dr} V(r) = \frac{q(N_d - n(r))}{\epsilon_r \epsilon_0},
\]

(1)

where \( N_d \) is the doping concentration, \( n(r) \) the carrier concentration, and \( \epsilon_r \) the permittivity of InAs.

The resulting electrostatic potential is used as input to the Schrödinger-like solver, which calculates the corresponding eigenvalues, \( E_{\alpha \nu} \), and eigenfunctions, \( \psi_{\alpha \nu}(r) \), for each subband in the nanowire [24], [25], [22]. No penetration of the wavefunction inside the gate dielectric is allowed, which means that the carrier concentration is zero at the nanowire surface. Due to the cylindrical geometry, there are two quantum numbers associated to each eigenvalue: one radial, \( r \), and one azimuthal, \( \nu \). In order to account for non-parabolicity, the parameter \( \alpha \) is introduced and the solver is repeated for each wave vector, \( K \):

\[
(1 + \alpha E_{\alpha \nu}(K)) E_{\alpha \nu}(K) \psi_{\alpha \nu}(r) = \left[-\frac{\hbar^2}{2m^*} \left( \frac{d^2}{dr^2} + \frac{1}{r} \frac{d}{dr} \right) + \frac{\hbar^2 \nu^2}{2r^2} - K^2 \right] + (1 + 2\alpha E_{\alpha \nu}(K)) V(r) - \alpha \langle \psi_{\alpha \nu}(r) \mid V(r) \rangle \mid \psi_{\alpha \nu}(r) \rangle.
\]

(2)

By adding the contribution of all subbands, the carrier distribution within the nanowire may be determined from the density of states, \( g(E) \), and the Fermi-Dirac distribution, \( f(E) \):

\[
n(r) = \sum_{\alpha, \nu} \left| \psi_{\alpha \nu}(r) \right|^2 \int_{E_{\alpha \nu}(0)}^{\infty} f(E)g(E) \, dE.
\]

(3)

\[
g(E) = \frac{2}{\pi} \left( \frac{dE}{dK} \right)^{-1}, \quad f(E) = \frac{1}{1 + e^{E/kT}}
\]

(4)

Consistency is ensured between the electrostatic potential and the band structure by solving the Schrödinger-like equation together with Poisson’s equation self-consistently in an iterative loop until convergence.

In this work, the model has been extended to calculate the carrier velocities required for the ballistic transistor simulations. The velocity, \( v(E) \), may be calculated by considering the kinetic energy of the carriers, which is dependent on the occupied states in the band structure. The velocity is proportional to
the slope of the energy dispersion relation:

\[ v(E) = \frac{1}{\hbar} \left( \frac{dE}{dk} \right). \] (5)

In the effective mass approximation, the bands are fully parabolic, which means that the effective mass of the carriers is constant. Thus higher energy states will have increasingly higher velocities. When considering the non-parabolicity of the conduction band, the shape of the dispersion relation is no longer completely parabolic and is approaching a linear slope of 1.4 eV\(^{-1}\) for InAs [32], [33]. This results in an increasingly higher effective mass for high energy carriers and the velocity saturates at a constant value. In order to quantify the influence of the non-parabolicity, simulations have also been performed using the simple effective mass approximation.

Simulations have also been performed without self-consistency, by solving the Schrödinger-like equation in a constant potential well. If self-consistency is ensured, the charge from the carriers within the nanowire deforms the electrostatic potential, thereby pushing the subbands to higher energies. This effect is illustrated in Fig. 2, where it is evident that this effect is more prominent for high surface potentials, due to the increased number of carriers. As more carriers fit inside non-parabolic bands, this effect is larger than in the effective mass approximation. This is true for all nanowire diameters between 5 and 35 nm.

4. Transport properties

In Fig. 3(a), the mean carrier concentration within a 12 nm nanowire is presented. It is clearly observed that the effective mass approximation underestimates the number of carriers. This is due to the effect of more active subbands in the non-parabolic model, as well as a higher density of states in the non-parabolic bands. On the other hand, the effect from the self-consistency reduces the carrier concentration due to shifting of the subbands to higher energies. This effect is stronger in the non-parabolic model, due to the higher charge from the more numerous carriers.

However, the mean velocity of the carriers is overestimated in the effective mass approximation, as illustrated in Fig. 3(b). As the effective mass is constant in the parabolic bands, the velocity gets increasingly higher for high energy states, whereas it is approaching a constant value in non-parabolic bands. The effect from the self-consistency is relatively small.

In Fig. 3(c), the product of the carrier concentration and the velocity is presented. This figure-of-merit is closely related to the expected current, which is derived from this product. It is expected that the current in a nanowire transistor will be higher than predicted by the effective mass approximation, primarily due to more active subbands. It is evident that the underestimation of the carrier concentration in the effective mass approximation is more severe than the overestimation of the carrier velocity. However, neglecting self-consistency overestimates the expected current due to the increased number of carriers in the static band structure.

The quantum capacitance for different gate-source voltages...
is presented in Fig. 3(d). This is compared to the geometric gate capacitance, $C_{ox}$, derived from the coaxial geometry, where the gate dielectric thickness, $t_{ox}$, is fixed at one tenth of the nanowire radius, $R$, with a relative permittivity, $\varepsilon_{ox}$, of 15:

$$C_{ox} = \frac{2\pi \varepsilon_{ox} \varepsilon_0}{R} \ln \left( \frac{R + t_{ox}}{R} \right) \approx 4.57 \text{ aF/nm} \quad (6)$$

It is evident that the quantum capacitance is much lower than the geometric gate capacitance, which means that the transistor operates close to the quantum capacitance limit. Furthermore, the sequential population of the first subbands in the nanowire is clearly visible. Each plateau corresponds to the transition to the next subband in the band structure. It is clear that three subbands are active in a 12 nm nanowire, at a gate-source voltage of 1.0 V.

Similar results are obtained for a thicker nanowire, with a diameter of 22 nm, as shown in Fig. 3(e-h). In this case the underestimation of the carrier concentration is slightly lower. This effect is primarily due to the higher number of subbands in the thicker nanowire, which causes the carriers to occupy lower energy states, where the effect of the non-parabolicity is smaller. However, the error introduced by not ensuring self-consistency is much larger, due to the severe overestimation of the control of the electrostatic potential in the center of the nanowire.

The velocity is greatly overestimated by the effective mass approximation also in the 22 nm nanowire. Furthermore, neglecting self-consistency introduces a larger error in comparison to the thinner nanowire. When the two metrics are combined, the overall effect of the non-parabolicity is smaller than in the thinner nanowire. However, the error introduced by not maintaining self-consistency is larger. There are no prominent plateaus in the quantum capacitance, due to the many occupied subbands. This produces a more continuous, as opposed to quantified, characteristic, closer the expected bulk behaviour.

5. Transistor modeling

From the carrier concentration and their corresponding velocities, it is possible to calculate the transistor characteristics. This is performed using a ballistic transistor model, based on the work in [34], [26]. In this model, the ballistic current is limited by a potential barrier between the source and drain, $U_B$, which is imposed by the band structure. This barrier is regulated indirectly by the gate-source voltage, $V_{GS}$, which influences the surface potential under the gate. However, the barrier height is also influenced by the influx of carriers from the source, $n^+$, and drain, $n^-$. These carriers reduce the gate control of the barrier height as described by [26]:

$$U_B = -qV_{GS} + \frac{q^2}{C_{ox}}(n^+ + n^-). \quad (7)$$

The effect on the barrier height from the mobile charge is dependent on the relative size between the geometric gate capacitance and the quantum capacitance. In this work, this effect is implemented by self-consistently calculating the barrier height using an iterative method described in [26]. Earlier works have established that highly scaled InAs nanowire transistors operate close to the quantum capacitance limit, consistent with the results in the previous section [35], [36], [37].

By applying a drain-source voltage, $V_{DS}$, the states in the band structure are populated asymmetrically by carriers traveling in the positive and negative direction. The positive states are filled up to the source Fermi level, $E_{F,s}$, whereas the negative states are filled up to the drain Fermi level, $E_{F,d}$. The current density in the positive, $J^+$, and negative direction, $J^-$, is calculated by multiplying the carrier concentration and the corresponding velocity for all filled states in each subband:

$$J^+/-(r) = \frac{1}{2} \sum_{n^+} \left| \psi_{n^+,r}(r) \right|^2 \int_{E_{n^+,r}(0)}^{\infty} f(E - E_{F,s/d}) \cdot g(E - U_B) \cdot v(E - U_B) \, dE.$$ \quad (8)

The current, $I_D$, is calculated from the difference in the current density in the positive and negative direction.

6. Transistor characteristics

The resulting transistor characteristics are presented in Fig. 4, for a nanowire with a diameter of 12 nm and 22 nm, respectively. For the 12 nm nanowire, the current is about 1.4 A/mm, with a transconductance of 4.1 S/mm, at a gate
overdrive of 0.5 V and a drain-source voltage of 1.0 V. For the 22 nm nanowire, the corresponding current is 1.1 A/mm, with a transconductance of 3.4 S/mm. The values are normalized with the gate width, which in this structure is equal to the nanowire circumference.

As shown in Fig. 4(c), the current is quantified in primarily three subbands in the thin nanowire, whereas the conduction in the thick nanowire is more continuous, due to the presence of a lot of subbands. Note that the threshold voltage is about 0.15 V higher for the thinner nanowire, as the first subband is pushed to higher energies in the stricter confinement.

Neglecting either the non-parabolicity of the conduction band or the self-consistency results in less accurate values of the current and transconductance, as well as a shift in the threshold voltage. This deviation is increasing for higher gate-source voltages, in agreement with the conclusions from the band structure simulations.

In order to investigate the influence of the quantum capacitance, the gate dielectric thickness is varied in Fig. 4(d). It is evident that the transistor operates close to the quantum capacitance limit, as the drive current is only slightly improved by increasing the gate capacitance.

7. Transistor scaling

In order to study the scaling properties of the transistor structure, a parameter analysis of the nanowire diameter is performed in the interval between 5 nm and 35 nm, at a constant surface potential of 0.5 V. For each diameter, the mean carrier concentrations and velocities is calculated, as well as the product of these metrics. Furthermore, the equivalent current is derived, which is defined as the maximum current, assuming perfect gate control and no carrier injection from the drain side.

The results are presented in Fig. 5(a-d). It is evident that there are almost no carriers in nanowires with a diameter thinner than about 6 nm. This is due to the confinement in the very thin nanowires, which pushes the lowest subband above the Fermi level. At a diameter of about 10 nm, the transition from transmission in a single subband to multiple subbands is visible as oscillations in the carrier concentration. It is evident that neglecting the self-consistency introduces large errors, especially for thick nanowires. Furthermore, the non-parabolicity allows more carriers in each band, consistent with the earlier analysis. The mean velocity of the carriers is relatively constant regardless of the nanowire diameter. However, the effective mass approximation, as well as neglecting self-consistency, systematically overestimates the carrier velocity.

The combined metric is presented in Fig. 5(c), where it is evident that the most advantageous transport properties are present for a nanowire with a diameter of about 17 nm. For thicker nanowires there is an almost linear decrease in performance. The effective mass approximation underestimates the performance for thin nanowires, whereas the error from neglecting self-consistency is very large and increases for thick nanowires.

In Fig. 5(d), the equivalent current is presented. It is clearly observed that the normalized current is about 1.8 A/mm for thick nanowires. However, the threshold voltage gets higher as the structure is scaled down. At the same gate overdrive, the normalized current will be significantly higher in the thinner nanowires.

8. Mapping of transport properties

Finally, a complete mapping of the location of the carriers within the nanowire and their velocities is performed for nanowire diameters between 10 nm and 35 nm. This provides valuable information about where in the nanowire most of the conduction is located.

The control of the electrostatic potential is presented in Fig. 6(a), for a constant surface potential of 0.5 V. The center potential is almost the same as the surface potential in thin nanowires, whereas the control of the center potential is degraded for thicker nanowires. The center potential is about 0.45 V for a nanowire with a diameter of 10 nm and 0.1 V at a diameter of 35 nm.

In Fig. 6(b), a mapping of the carrier concentration within the nanowire is presented. Most of the carriers are located close to the center in thin nanowires. As the diameter is increased, the gate control of the center of the nanowire is reduced and most of the carriers accumulate close to the nanowire surface.

Figure 5: The variation of the mean carrier concentration with the nanowire diameter is shown in (a) and the mean carrier velocity in (b). The product of these two metrics is presented in (c). The resulting normalized current, assuming no drain injection and perfect gate control, is presented in (d). The surface potential is fixed at 0.5 V.
Considering the carrier velocities in Fig. 6(c), it is evident that the fastest carriers are located within the center of the nanowire. However, the variation in carrier velocity is much smaller than the variation in the carrier concentration. The fastest carriers are found in a nanowire with a diameter of 16 nm.

The combined metric in Fig. 6(d), shows that the center conduction is dominating for nanowire diameters below 13 nm. For thicker nanowires, most of the conduction is located away from the center. However, the relative contribution to the current is dependent on the cross-sectional segment area at that location. The area for the center segment approaches zero, whereas it is largest at the surface of the nanowire.

The equivalent current is presented in Fig. 6(e), where it is clear that most of the contribution to the transistor current is actually close to the surface of the nanowire. For a more fair comparison, the current is normalized with the nanowire circumference in Fig. 6(f). In thick nanowires most of the current pass in the segment about 80% out from the center. As the diameter is scaled down, the maximum approaches the center of the nanowire and in a 10 nm nanowire, most of the conduction is located at about 35% out from the center.

9. Conclusion

In this work, simulations of the intrinsic performance and scaling behaviour of InAs nanowire MOSFETs are presented. A transconductance of 4.1 S/mm and 3.4 S/mm is predicted for a nanowire diameter of 12 nm and 22 nm, respectively, at a gate overdrive of 0.5 V. It is evident that the performance is improved as the diameter of the nanowire is scaled down.

It is concluded that both non-parabolicity and self-consistency must be considered in order to get an accurate prediction of the device performance. This is true for all nanowire diameters between 10 nm and 35 nm and the deviation gets larger for higher surface potentials. Ensuring self-consistency is very important, especially in thick nanowires, whereas the effect from the non-parabolicity is most prominent in thin nanowires.

The highest carrier concentration is located near the center in thick nanowires, whereas it is located closer to the surface in thicker nanowires, with a transition at a diameter of about 13 nm. The carrier velocity within the nanowire is more uniform than the carrier concentration, with a maximum velocity in the center of a nanowire with a diameter of 16 nm. However, the actual contribution to the current from the center of the nanowire is relatively small, due to the small cross-sectional area.

The conclusions presented in this work have implications regarding the estimation of the ultimate performance of III-V nanowire MOSFETs. Furthermore, the results are useful for understanding the device physics and may be used as guidelines for optimizing the device architecture for fabrication. Finally, the calculated transistor characteristics may be utilized in circuit simulations and benchmarking.

[1] R. Chau, S. Datta, M. Doczy, B. Doyle, B. Jin, J. Kavalieros, A. Majumdar, M. Metz, M. Radosavljevic, Benchmarking nanotechnology for high-performance and low-power logic transistor applica-
Amplifier Design Using InAs Nanowire Transistors
Kristofer Jansson, Erik Lind, and Lars-Erik Wernersson

Abstract—In this work, amplifier design using ballistic vertical InAs nanowire transistors is investigated, focused on a simple common-source amplifier. The maximum power gain at 90 GHz is evaluated for different nanowire transistor architectures together with the power dissipation. The linearity of the amplifier is evaluated by estimating the IIP3 and 1 dB compression point. Furthermore, the impact of the parasitic capacitances and resistances are quantified and it is demonstrated that the gain may be increased by a cascode design. It is concluded that a power gain exceeding 20 dB at 90 GHz may be achieved by a common-source amplifier based on an InAs nanowire transistor architecture. A power consumption below 1 mW is possible, while still maintaining a high power gain. Furthermore, an IIP3 of about 10 dBm is predicted. The combination of these qualities makes the nanowire transistor architecture an attractive prospect for low-power amplifiers at millimeter wave frequencies.

Index Terms—Nanowires (NWs), InAs, Transistor, Amplifier.

I. INTRODUCTION

NANOWIRE transistors in III-V materials are emerging with the potential to continue the improvement of high-performance integrated circuits [1]–[4]. These devices combine the high mobility of InAs with the excellent electrostatic control provided by a gate-all-around structure [5], [6]. The vertical orientation provides a natural way to fabricate gate-all-around devices, with the prospect of a high integration density [7]. The result is a transistor with a high transconductance and low power dissipation, which shows great potential for use in amplifier circuits. Due to the vertical orientation, however, the parasitic overlap capacitances between the electrodes can be large. Furthermore, there are series resistances in the structure, mainly due to the ungated segments of the nanowires, combined with the contact resistance between the nanowires and the electrodes. It has been demonstrated that large arrays of densely packed nanowires decreases the parasitic elements down to the level of the ITRS roadmap and that these devices show impressive performance at high frequencies, with a projected $f_T$ and $f_{max}$ exceeding 1 THz [1], [8]. Excellent linearity has also been predicted for these devices, although it is highly dependent on the bias conditions due to the quantized conduction in only a few subbands [9]–[11].

In this work, a transistor architecture based on vertically aligned InAs nanowires is investigated for use in RF amplifier circuits. For simplicity in the benchmarking and optimization, a standard common-source amplifier is studied at a frequency of 90 GHz. The maximum gain is predicted together with the power consumption and the influence on the performance from the parasitic capacitances and series resistances is quantified. The linearity of the amplifier is evaluated by the IIP3 and 1 dB compression points. Furthermore, an optimization of the transistor structure and bias conditions is performed. The analysis is further extended to a cascode amplifier, where the potential improvement of the gain and power consumption is investigated. The optimal alignment of the cascode transistors and the impact from asymmetric spacer layers are investigated in order to minimize the impact of the parasitic elements.

II. TRANSISTOR STRUCTURE

The transistor structure consists of vertically aligned InAs nanowires, arranged in an array of 4x8 nanowires. The nanowires are covered by a high-$\kappa$ dielectric with a permittivity of 15, with the thickness scaled as one tenth of the nanowire diameter. Vertically aligned electrodes are used to contact the nanowires, separated by spacer layers consisting of a dielectric material with a low permittivity, in this case assumed to be 3. This results in a gate that completely surrounds the nanowire in a gate-all-around configuration, resulting in excellent electrostatic control. A schematic of the architecture is shown in Fig. 1. This transistor structure is similar to fabricated devices [12]–[15].

In this work, the focus is on nanowires with a diameter of 12 nm and 22 nm. The intrinsic transistor performance is based on simulations of the ballistic transport in [16], [17], which is combined with the parasitic elements predicted in [8]. The dimensions of the structure follows the scaling roadmap and optimization described in [8]. An array spacing of one nanowire diameter is assumed and very thin spacer layers are used. The source and drain electrodes are considered to be made of Au, whereas the gate is made of W. The gate length is set long enough to minimize short-channel effects and the source and drain electrodes are significantly longer than the transfer length at three times the nanowire diameter [18]. A summary of the structural parameters is presented in Table I.

III. TRANSISTOR MODEL

The transistor model is based on the band structure simulations in [19], [20], [17]. This model considers self-consistency between the electrostatic potential and the band structure, as well as the nonparabolicity of the InAs conduction band [21]. The transport in the nanowire is modeled by a conventional ballistic model, based on a top of the barrier approach [22], [23]. In this work, the transistor model has been extended to model the influence of the source and drain potential on the barrier height. This introduces output conductance into the transistor characteristics from the effect of drain induced barrier lowering. These improvements have been derived by
The intrinsic transistor characteristics are combined with the parasitic capacitances and resistances calculated in [8]. The parasitic elements include parasitic overlap and fringe capacitances between the vertically aligned electrodes and the nanowires. Series resistances are primarily located within the ungated segment of the nanowire, together with the contact resistance at the interface between the nanowire and the electrode. Further resistance is added from the electrode structure.

The resulting transistor characteristics for a 4x8 nanowire array, including parasitic resistances, are presented in Fig. 2 for nanowires with diameters of 12 nm and 22 nm, respectively. A mapping of the transconductance, output conductance and self gain is performed in order to investigate the biasing conditions for the transistor. This is presented in Fig. 3 for all gate-source and drain-source voltages between 0 and 1 V. The maximum transconductance is found at the maximum drain-source voltage of 1 V, at a gate-source voltage of 0.885 V and 0.915 V for the 12 nm and 22 nm nanowires, respectively. The voltage gain is relatively constant at around 20–25 decibel, regardless of the biasing conditions.

A summary of the performance metrics for the nanowire transistor is presented in Table II. The data is presented for the biasing with the maximum transconductance. The extrinsic transconductance is severely degraded due to the relatively large series resistance, especially for the thinner nanowire.

### Table I

**Structural parameters in the nanowire transistor architecture.**

<table>
<thead>
<tr>
<th>Diamete</th>
<th>12 NM</th>
<th>22 NM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Nanowires</td>
<td>4x8</td>
<td>4x8</td>
</tr>
<tr>
<td>Gate Dielectric Thickness (nm)</td>
<td>1.2</td>
<td>2.2</td>
</tr>
<tr>
<td>Gate Length (nm)</td>
<td>16.1</td>
<td>29.6</td>
</tr>
<tr>
<td>Nanowire Spacing (nm)</td>
<td>14.4</td>
<td>26.4</td>
</tr>
<tr>
<td>Spacer Layer Thickness (nm)</td>
<td>7.9</td>
<td>15.8</td>
</tr>
<tr>
<td>Source/Drain Electrode Thickness (nm)</td>
<td>36</td>
<td>66</td>
</tr>
</tbody>
</table>

### Table II

**Summary of the nanowire transistor performance.**

<table>
<thead>
<tr>
<th>Diameter</th>
<th>12 NM</th>
<th>22 NM</th>
</tr>
</thead>
<tbody>
<tr>
<td>VGS at max gmax (V)</td>
<td>0.885</td>
<td>0.915</td>
</tr>
<tr>
<td>Current (A/mm)</td>
<td>1.34</td>
<td>1.96</td>
</tr>
<tr>
<td>Transconductance (S/mm)</td>
<td>2.35</td>
<td>2.90</td>
</tr>
<tr>
<td>Intrinsic Transconductance (S/mm)</td>
<td>5.24</td>
<td>3.96</td>
</tr>
<tr>
<td>Output Conductance (S/mm)</td>
<td>0.17</td>
<td>0.29</td>
</tr>
<tr>
<td>Self Gain (dB)</td>
<td>23.0</td>
<td>19.9</td>
</tr>
<tr>
<td>Intrinsic Self Gain (dB)</td>
<td>24.5</td>
<td>23.3</td>
</tr>
</tbody>
</table>

In order to investigate the use of the nanowire transistor architecture in amplifier applications, a simple common-source amplifier is studied. A schematic of the amplifier design is presented in Fig. 1(c). As the focus is on the transistor performance, a perfect inductor is used for the simulations, acting as an RF choke. Furthermore, a perfect and lossless matching is assumed at both the input and output for maximum power gain.

The amplifier performance is benchmarked primarily based on the maximum power gain at a frequency of 90 GHz, which is a commonly targeted frequency for millimeter wave amplifiers. Furthermore, the power consumption is evaluated.
The linearity of the device is benchmarked using IIP3 and the 1 dB compression point.

Simulations are performed in Cadence, using customized Verilog-A models for the transistors, including both the intrinsic transistor characteristics, as well as the distributed parasitic elements. The simulation data is complemented by numerical calculations in MATLAB. The maximum power gain is calculated using S-parameter analysis in Cadence, and the power consumption is easily extracted at the specified bias point.

In order to benchmark the circuit technology itself, a perfect impedance matching at the input and output is assumed. This makes it difficult to simulate the linearity in Cadence, as it is dependent on the designed matching network. The IIP3 is instead estimated directly from the transistor characteristics, including the series resistance, as described by [9], [10]. If only third order distortion is considered, the 1 dB compression point is approximately 9.6 dB lower than IIP3 as described by [24].

A. Results

In Fig. 4, the power gain and power consumption at 90 GHz are mapped for supply voltages between 0 and 1 V. This is performed for nanowire transistors with a diameter of either 12 nm and 22 nm. Furthermore, the input DC voltage, corresponding to the transistor gate-source voltage is varied. The array size is fixed at 4x8, with a total of 32 nanowires. This array size was chosen as the smallest array size, which still exhibits a maximum gain close to the saturated value for infinitely large arrays.

It is seen that the maximum power gain is not achieved at the point of maximum transconductance, but instead about 0.4 V. At this bias point and for a supply voltage of 1 V, a power gain at 90 GHz of 19.2 dB and 20.4 dB is achieved for the 12 nm and 22 nm designs respectively. The power consumption at this point is relatively low at 1.7 mW and 3.6 mW. The predicted linearity is excellent, with an IIP3 of about 10 dBm and an 1 dB compression point close to 0 dBm.

In order to reduce the power consumption, operation at a supply voltage of 0.5 V is investigated. This means that the optimum input DC level is reduced as well, in this case to 0.2 V. This results in a much lower power consumption of 0.23 mW and 0.64 mW, respectively. However, the gain is slightly reduced, especially for the thinner nanowire, with a power gain of 17.9 dB. The gain in the thicker nanowire is better with 18.8 dB. Note that the power consumption will be increased by the inclusion of a matching network. The input and output impedance for the different amplifier designs are presented in Fig. 4(e).

The maximum power gain for different frequencies is presented in Fig. 5. It is evident that operation in the THz regime is possible while maintaining a positive power gain and that the gain may be increased by increasing the supply voltage, using larger nanowire arrays, or by using the thicker nanowires. In Fig. 5(c) and 5(d), the performance impact from the parasitic capacitances and resistances is investigated. Fig. 5(c) shows the 12 nm design with a 4x8 nanowire array at a supply voltage of 0.5 V. This design is clearly very limited by the parasitics. For the thicker nanowire in Fig. 5(d), in an array of 4x8 nanowires at a supply voltage of 1 V, the parasitic elements...
degrade the performance to a lesser degree.

V. CASCODE AMPLIFIER

In order to further increase the power gain, a cascode amplifier structure is investigated. The cascode structure enables a reduction of the gate-drain parasitics in the common-source stage due to the Miller effect, thus resulting in a higher performance. The cascode implementation is based on a common-source input stage, followed by a common-gate stage. This results in a lower current. The supply voltage is set to 1 V, which in the cascode configuration is divided between the two transistors in the circuit. The orientation of the common-gate transistor does not influence the gain significantly. In order to allow a connection between the stages in a single metal layer, it is concluded that the common-gate transistor should have the source located at the bottom of the nanowire, as illustrated in Fig. 7.

A cascode structure introduces interesting design considerations regarding the vertical alignment of the two transistors. Each transistor may be oriented either with the source or the drain at the bottom, resulting in four possible layouts. The power gain for these four layouts are compared in Table IV. It is clear that the source of the common-source stage should be located at the bottom of the nanowire for optimum performance. This avoids placing the extra parasitic capacitances from the substrate at the more performance sensitive drain side. The orientation of the common-gate transistor does not influence the gain significantly. In order to allow a connection between the stages in a single metal layer, it is concluded that the common-gate transistor should have the source located at the top of the nanowire, as illustrated in Fig. 7.

The resulting power gain at 90 GHz is about 26 dB for the 22 nm design, which is about 7 dB higher than in the common-source amplifier. The power consumption of 1.5 mW is lower than for the common-source amplifier at the same supply voltage. This is due to the division of the supply voltage between the two transistors, resulting in a lower current. The maximum power gain for different frequencies is presented in Table IV.

Further improvements of the performance may be achieved by introducing asymmetric spacer layers in the cascode structure, as presented in Fig. 7(c-d). It is clear that the bottom spacer...
layer should be thin in order to reduce the source series resistance for the common-source stage. Even though the cascode design reduces the gate-drain capacitance for the common-source stage thanks to the Miller effect, it is still advantageous to use a thicker spacer layer at the top side. This also reduces the capacitive coupling between the output and the intermediate node, while maintaining a low series resistance at the source side.

Finally, the influence of higher parasitic capacitances and resistances in the structure is studied. When the resistance within the nanowire is increased up to a factor of ten, as illustrated in Fig. 6(c), the gain is greatly reduced. Furthermore, a thinner top spacer gets more advantageous. This is due to the high resistance at the intermediate node between the two transistors, which shifts the transition to unconditional stability to lower frequencies. The performance degradation from a factor ten higher parasitic capacitances degrades the performance to a smaller degree than the resistance, as illustrated in Fig. 6(d). In this case, the optimal spacer layer thickness at the bottom is only slightly thicker than in the standard case.

VI. CONCLUSION

An optimization of an InAs nanowire common-source amplifier architecture has been performed and the performance evaluated for a full range of biasing conditions between 0 and 1 V. This analysis has been extended to a cascode structure,
which has been shown to improve both the power gain and the power consumption in comparison to the common-source amplifier. The optimal alignment of both transistors in the cascode has been predicted and the influence from asymmetric spacer layers has been quantified.

It is concluded that a common-source amplifier using InAs nanowire transistors with a realistic footprint may achieve a power gain of about 20 dB at an operating frequency of 90 GHz. A very low power dissipation of about 1 mA is predicted. The linearity has been estimated, with an IIP3 of about 10 dBm. These properties make InAs nanowire MOSFETs highly suitable for low-power amplifiers at millimeter wave frequencies.

REFERENCES

RF-Characterization of Vertical Wrap-gated InAs/high-κ Nanowire Capacitors

Jun Wu, Kristofer Jansson, Aein S. Bahadi, Martin Berg, Erik Lind, and Lars-Erik Wernersson

Abstract—This paper presents radio-frequency (RF) as well as low-frequency capacitance-voltage (C-V) characterization of vertical wrap-gated InAs/high-κ nanowire metal-oxide-semiconductor (MOS) capacitors. A full equivalent circuit model for traps is used to fit the low-frequency C-V characteristics from which the interface trap density \(D_{it}\) and border trap density \(N_{bt}\) are evaluated separately. The results show comparable \(N_{bt}\) but far lower \(D_{it}\) (< 10^{12} 	ext{eV}^{-1}	ext{cm}^{-2} \text{near the conduction band edge}) for a nanowire MOS gate stack compared to planar references. In the RF domain, the influence of nanowire series resistances become significant, and by introducing a distributed RC-model, the nanowire resistivity is evaluated from the capacitance data as a function of the gate bias. An on/off resistivity ratio of 10^{-2} is obtained for the best device. Using the measured data, the quality factor is finally evaluated both for fabricated and ideal cases. The results agree well with simulated data.

Index Terms—Nanowire, InAs, High-κ, CV, RF, Trap density, Resistivity, Quality factor

I. INTRODUCTION

TARGETING the next generation of high-speed, low-power electronics with prospective ultra-scaled devices, vertical wrap-gated InAs/high-κ nanowire metal-oxide-semiconductor field-effect transistors (MOSFETs) are considered [1]-[2]. The wrap-gate geometry provides superior electrostatic control [3]-[4]. The use of InAs offers excellent transport properties, such as high electron mobility and injection velocity [5]-[6]. However, there are critical challenges for vertical devices. Large series resistances in thin nanowires can severely deteriorate the extrinsic transconductance [7]. Large interface trap and border trap densities easily restricts the Fermi level movement, degrading the sub-threshold slope and the on-current [8]. Characterization and reduction of trap densities is hence of importance. These methods both require symmetric and ohmic contacts, which can be difficult to realize, and provide inferior electrostatic control, as compared to the vertical wrap-gate geometry. For characterization of \(D_{it}\) and \(N_{bt}\) on planar devices, the most commonly used method is the capacitance-voltage (C-V) technique [12]-[13]. For vertical wrap-gate structure, however, large parasitic capacitances are usually present, which have to be minimized for a reliable analysis [14]-[15]. Besides, a full trap model [16]-[17] is necessary to separate \(D_{it}\) and \(N_{bt}\), since the influence of both types of traps on measured C-V curves is usually combined.

In this work, we provide, to the author’s best knowledge, the first RF C-V measurements using 1-port Vector Network Analyzer (VNA) [18]-[19], in combination with low-frequency C-V measurements using an Impedance Analyzer on InAs/high-κ vertical wrap-gated nanowire MOS capacitors. By adopting a finger-gate device layout, the parasitic capacitances in the transistor structure are minimized enabling accurate determination of the intrinsic properties. From the low frequency C-V characteristics, \(D_{it}\) and \(N_{bt}\) profiles are determined using a full trap model adopted from planar MOS capacitors. In the RF domain, where the trap response is reduced, the influence of the nanowire series resistances has a stronger effect. A complete distributed RC-network model is developed to fit the measured capacitance-frequency (C-f) characteristics, from which the nanowire resistivities at various gate voltages are determined. This work hence presents a method to determine \(D_{it}\), \(N_{bt}\) and \(\rho_{nm}\) for vertical wrap-gated nanowire MOSFET simultaneously using an extended C-V technique. Finally, the measured data is used to determine the intrinsic C-V characteristics and quality factors as a function of frequency of nanowire MOS capacitors. The measured quality factors are consistent with simulation results in the literature.

II. DEVICE FABRICATION

A finger-gate technique was adopted [20] in order to fabricate vertical nanowire MOS capacitors with minimized parasitic capacitances. The schematics in Figure 1 (a)-(f) illustrate the device fabrication steps. Devices were fabricated on a highly resistive Si(111) substrate covered by a 300-nm-thick InAs buffer layer serving as source contact, as shown in Figure 1(a). Au seed particles were deposited using an electron beam lithography (EBL) lift-off process. Subsequently, InAs nanowires were grown at 420°C in an Aixtron low-pressure metalorganic vapor phase epitaxy (MOVPE) chamber with arsenic (AsH₃) and trimethylindium (TMIn) as precursors. The molar fraction of AsH₃ and TMIn was 1.92 x 10^{-4} and 2.79 x 10^{-6}, respectively. A roughly 100-nm-long, highly doped segment was first grown with tetraethyltin (TESn) as dopant precursor. High doping ensures a minimized source series resistance \(R_{ss}\), which is crucial, since a large \(R_{ss}\) would conceal the high-frequency behavior of the nanowires. For the second growth step, the TESn precursor supply was closed in order to grow the intrinsic nanowire segment. Some doping is, however, still expected due to a memory effect.
originating from residual Sn present in the Au seed particle. After growth, a high-κ bilayer dielectric was deposited in an atomic layer deposition (ALD) chamber with 10 cycles of Al₂O₃ at 260°C followed by 50 cycles of HfO₂ at 100°C. The high-κ was then patterned using UV-lithography to serve as an etch mask for etching of the buffer layer in a later step, Figure 1(b). The wafer was spin-coated by S1818, followed by reactive ion etching (RIE) to define a 100-nm-thick gate-to-source spacer, Figure 1(c). A 200-nm-layer of W was sputtered and subsequently patterned together with the spacer beneath to form the finger-shaped gate contact, Figure 1(d). To achieve this geometry, PMMA A8 resist was exposed by Deep UV lithography and EBL. The former removed the excessive area across the sample, whereas the latter removed the fine areas near the nanowires. After development, the W film together with the spacer beneath was etched by RIE. This resulted in a gate contact consisting of 13 fingers in parallel, each 1.3 μm wide and about 10 μm long. Finally, Ti/W/Au pads were deposited using lift-off to define ground/signal/ground (GSG) structures for RF measurements, which was followed by InAs buffer-layer mesa etching, Figure 1(e)-(f). Ground and signal pads were isolated after the mesa etching with a part of the gate fingers suspended in the air due to under-etching of the InAs. Each MOS capacitor had 1300 nanowires placed in zigzag rows with a pitch of 200 nm. The nanowire diameter varied from 30 nm to 50 nm across the sample, which gave a length variation from 700 nm to 1000 nm. Reference devices without nanowires for de-embedding were also fabricated next to each MOS.
capacitor. Figure 1(g) shows a scanning electron micrograph of one of the fabricated nanowire MOS capacitors. Due to the reduced overlap area between gate and source, parasitic capacitances were minimized.

The nanowire MOS capacitors were measured using a conventional 4294A Impedance Analyzer for low-frequency measurements and an Agilent E8261A VNA for RF-measurements. The frequency ranges were 1 kHz-100 MHz and 100 MHz-67 GHz, respectively.

III. LOW FREQUENCY C-V CHARACTERISTICS

Figure 2(a) shows the low-frequency nanowire MOS capacitor C-V characteristics measured by the Impedance Analyzer. The plotted nanowire capacitance (C_{nw}) was extracted via a lumped RC parallel model as shown by the inset. Before the extraction, the admittance of the reference device was subtracted. The parasitic capacitance e was about 30-80 fF, which is negligible compared with the 1.3 pF of C_{nw}. Furthermore, the parasitic C-V does not show voltage dependence, which is crucial for the accurate C-V analyses. A clear C-V modulation is obtained for the nanowire MOS capacitor, indicating an un-pinned Fermi level. Besides, effects of trap responses on the frequency modulated C-V characteristics are also present. The frequency dispersion in the accumulation region originates mainly from the majority carrier interaction with border traps [17]. In depletion, interface traps, as well as border traps, are both influential. Towards weak inversion, the hole interaction with both types of traps as well as the hole generation/recombination through bulk diffusion further contributes to the C-V frequency dependence. To model these effects, a full trap equivalent circuit model, developed for planar MOS capacitors [17], is adopted as shown in Figure 2(b). By adjusting the interface and border trap profiles, good fits between measured and simulated C-V curves are obtained, as shown in Figure 2(c). The surface electron/hole capture cross-sections (σ_{0e}/σ_{0h}) used to fit the measured capacitance and conductance data are 1 x 10^{-14} cm^{-2} and 1 x 10^{-15} cm^{-2}, respectively. The oxide thickness and effective x value are 6 nm and 14, respectively, which gives C_{ox} = 2 μF/cm². The carrier concentration is about 10^{13} cm^{-3} for undoped InAs nanowires [21]. The fitting is less accurate towards the valence band edge. By fitting the curves, both D_{1} and N_{bt} are evaluated as shown in Figure 2(d). It can be seen that the total trap density mainly consists of N_{bt} (5-10 times larger than D_{1}). This also highlights the significance of an accurate N_{bt} evaluation. Planar InAs capacitors, with comparable high-κ deposition conditions [17], show similar N_{bt} projected to the surface (1.38 x 10^{13} eV^{-1} cm^{-2}). Nanowire MOS gate stacks, however, exhibit a far lower D_{1}, reaching below 10^{12} eV^{-1} cm^{-2} close to the conduction band edge [17].

IV. RADIO FREQUENCY C-V CHARACTERISTICS

In the RF domain, the S11 parameters were obtained as a function of gate biases using the 1-port VNA measurement, as shown by the inset in Figure 3(a). The device-under-test (DUT) admittance can be obtained from measured S11 parameter via

\[ y = (1 - S11)/(1 + S11). \]

After de-embedding the parasitic admittance, C_{nw} was extracted using the parallel model as shown in Figure 2. Figure 3(a) shows the nanowire MOS capacitor C-V characteristic from 144 MHz to 13 GHz. As the measurement frequency increases, the trap response to ac signal will diminish. As can be seen from the figure, beyond 1.3 GHz, a significant capacitance drop is observed. This is due to the unique electrical and geometrical properties of vertical wrapped nanowire devices, where the effect of series resistances along the nanowire increases with the input signal frequency.
The frequency dependence of $C_{nw}$ is more clearly shown in figure 3(b) for three biases corresponding to accumulation, depletion and weak inversion. The drops at high frequencies reflect the influence of series resistance along the nanowire. It shows the C-V curves of a unit segment of the nanowire.

The RC distributed network behavior provides an method to evaluate $\rho_{nw}$. For this purpose, a complete RC model is built as shown in Figure 4(a) [22]. The measured capacitor is modeled using a circuit including the contribution from both the nanowires and the surrounding structure. A single nanowire is modeled by a distributed RC-network including the intrinsic nanowire capacitances ($C_{t}$) and the series resistances within the nanowire ($R_{s}$). The surrounding parasitics include the finger gates, which are also modeled as a distributed RC-network, consisting of parasitic overlap capacitances ($C_p$) and series resistances in the gate electrodes ($R_f$). Further parasitics include the outer series resistances of the gate finger ($R_{gw}$) and the InAs buffer layer ($R_{InAs}$). All of these elements are combined in a circuit model and the equivalent impedance is calculated using transmission line parameters. By using reference samples, the parasitic elements in the surrounding structure are extracted and shown in Table I. These values are subtracted from the measurements of the capacitors, which leaves the C-f characteristics resulting from the distributed RC-network inside the nanowire, as shown in figure 4(c). As the maximum capacitance are determined at low frequencies, the resistivity can be obtained from the effective capacitance degradation in the frequency sweep, for each bias point. This method assumes that both the capacitances and the resistances are frequency independent, which is a reasonable assumption in the accumulation region. However, for negative voltages, the effect from the RC-network is difficult to distinguish from the intrinsic capacitance modulation through $G_{ch}$ in Figure 4(b). This degrades the validity of the model in which $C_{t}$ is assumed to be constant at fixed bias voltages. A good agreement to the measurement data is achieved, as shown in Figure 4(b).

The simulations are repeated for different nanowire diameters and the resulting resistivities are presented in Fig. 4(d) as a function of the gate bias. Resistance values extracted in this way correspond well to previously reported data, but provide far better gate bias modulation compared with the $\rho_{nw}$ extracted using conventional back-gated lateral I-V measurements [11]. The best device shows a $\rho_{nw}$ on/off-state ratio of about $10^{-2}$. Besides, no clear trend with the diameter is observed, suggesting that there is little variation in material quality within the diameter interval among the nanowires studied [11].

<table>
<thead>
<tr>
<th>TABLE I</th>
<th>PARAMETERS FOR THE DEVICE AND MODEL PRESENTED IN FIG.4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{f}$/ finger</td>
<td>Total $C_{t}$/ finger</td>
</tr>
<tr>
<td>3 $\Omega$</td>
<td>70 $\Omega$</td>
</tr>
</tbody>
</table>

Finally, the intrinsic high-frequency C-V characteristics of
Fig. 5. Quality factors as a function of measurement frequency extracted at gate voltage equal to 2.5V. The real cases are calculated from the equivalent impedance of the distributed RC network within a single nanowire, whereas the ideal ones are calculated from the intrinsic R and C of a nanowire, removing the effect of the series resistance, which are equivalent to an infinitely thin slice of the nanowire. The real cases are shown in Figure 4(f). The measurements were extended from about 1 kHz to 67 MHz to about 10 GHz. The capacitances extracted in this way correspond to an infinitely thin slice of the nanowire without any frequency dependence originating from the distributed RC-network. By extracting both the intrinsic capacitance and resistance of the nanowire, the ideal quality factor corresponding to the shortest gate length can be easily calculated and is presented as a function of frequency in Figure 5. Also presented are the quality factors for the fabricated capacitors with gate lengths of 600~800nm. It is clearly shown that the values vary from about 10^5 at 1 MHz to about 1 at 100 GHz, with a strong frequency dependence. Scaling down the gate length improves the performance significantly due to the reduced series resistance. It is evident that the quality factors can exceed 100 at 60 GHz for very short gate lengths, which may be compared with conventional thin-gate oxide MOS transistors with a quality factor of about 20 at 20 GHz [23]. The quality factor evaluated here fits very well to the simulation data published in [22].

V. CONCLUSION

In this work, we demonstrate low-frequency and RF C-V characterization of InAs/high-κ vertical wrap-gated nanowire MOS capacitors. The measurement frequency was extended from 1 kHz to 67 GHz by using a VNA in combination with a conventional Impedance Analyzer. From low-frequency C-V characteristics, \( D_{it} \) and \( N_{tr} \) were evaluated based on a full trap model adopted from planar MOS capacitors. The results indicate comparable \( N_{tr} \) with planar references dominates the total trap density, and that a far lower \( D_{it} \) can be obtained for InAs nanowires compared with their planar counterpart. In the RF domain, the influence of nanowire series resistances on the C-V characteristics becomes strong. By introducing a complete distributed RC-network model, nanowire resistivities was evaluated under different gate biases for vertical wrap-gated nanowire devices. The evaluated resistivities are consistent with literature, but with a far better on/off ratio compared with those extracted using a back-gated four-point I-V method. The work hence provides an approach to simultaneously evaluate \( D_{it} \), \( N_{tr} \) and \( \rho_{ox} \) of vertical wrap-gated nanowire MOSFET, based on an extended C-V technique. Finally, the quality factor was evaluated, which exceeds 100 at 60 GHz for short gate length.

REFERENCES


Extrinsic and Intrinsic Performance of Vertical InAs Nanowire MOSFETs on Si Substrates

Karl-Magnus Persson, Member, IEEE, Martin Berg, Mattias B. Borg, Member, IEEE, Jun Wu, Sofia Johansson, Johannes Svensson, Kristofer Jansson, Erik Lind, and Lars-Erik Wernersson

Abstract—This paper presents dc and RF characterization as well as modeling of vertical InAs nanowire (NW) MOSFETs with $L_G = 200$ nm and $\text{Al}_2\text{O}_3/\text{HfO}_2$ high-$\kappa$ dielectric. Measurements at $V_{DS} = 0.5$ V show that high transconductance ($g_m = 1.37 \text{ mS}/\mu\text{m}$), high drive current ($I_{DS} = 1.34 \text{ mA}/\mu\text{m}$), and low on-resistance ($R_{ON} = 287 \Omega/\mu\text{m}$) can be realized using vertical InAs NWs on Si substrates. By measuring the $1/f$-noise, the gate area normalized gate voltage noise spectral density, $S_{V/g}/g^2$, is determined to be lowered by one order of magnitude compared with similar devices with a high-$\kappa$ film consisting of $\text{HfO}_2$ only. In addition, with a virtual source model we are able to determine the intrinsic transport properties. These devices ($L_G = 200$ nm) show a high injection velocity ($v_{inj} = 1.7 \times 10^7 \text{ cm/s}$) with a performance degradation for array FETs predominantly due to an increase in series resistance.

Index Terms—InAs, MOSFET, nanowire (NW), RF.

I. INTRODUCTION

STRIVING toward ultrascaled devices, the transistor architecture and the material properties need consideration. The nanowire (NW) geometry offers advantageous electrostatic scaling [1] and the use of a high-$\kappa$ gate dielectric allows reduced equivalent oxide thickness (EOT) without large gate leakage currents. InAs has a high injection velocity, $v_{inj}$, and allows simple fabrication of low resistance contacts [2], [3]. High current densities have been demonstrated for thin InAs transistor channels [4].

In this paper, we demonstrate dc characterization, modeling, and RF characterization of FETs consisting of arrays of vertical NWs (a-FETs) [5]–[7]. Our data for a-FETs show a threefold improvement on earlier data and we achieve transconductance values above $1 \text{ mS}/\mu\text{m}$ for individual NWs. This is attributed to an improved bi-layer gate dielectric using $\text{Al}_2\text{O}_3/\text{HfO}_2$, as well as more reliable processing with a thin $\text{Si}_3\text{N}_4$ film, reducing the source series resistance [8]. From dc modeling, we identify the main limiting factors in the transistor layout.

II. DEVICE FABRICATION

Devices are fabricated on Si substrates cut from a 4-in Si wafer covered by a 300-nm-thick InAs contact layer (ICL). The Si is highly resistive and the InAs layer is grown without intentional doping ($n = 3 \times 10^{18} \text{ cm}^{-3}$) demonstrating a sheet resistance of 36 $\Omega/\square$. Definition of size and position of gold seed particles is made with electron-beam lithography and growth of InAs NWs is performed in an metal-organic-vapor-phase-epitaxy growth chamber at 420 °C and with a Sn dopant flow corresponding to a doping level of $\sim 1 \times 10^{18} \text{ cm}^{-3}$ [9]. For a-FETs, NWs are placed in zigzag rows to reduce the parasitic capacitance within the array [10], with 300-nm spacing between each NW [Fig. 1(a) and (b)]. Devices are fabricated with 1, 52, 96, and 192 NWs and with two NW diameters; $D_{NW} = 45$ nm and $D_{NW} = 28$ nm [8]. In addition to the difference in size of catalytic gold particles, the rest of the growth conditions are identical. The dielectric is deposited in a Cambridge Savannah ALD process with 10 cycles of $\text{Al}_2\text{O}_3$ at 250 °C and 60 cycles of $\text{HfO}_2$ at 100 °C, with a total thickness of 7 nm translating to an estimated EOT of 1.8 nm. The $D_{NW} = 28$ nm have a thinner HfO$_2$ layer, yielding an estimated EOT = 1.3 nm. Subsequent to the ALD, the $2 \times 2 \text{cm}^2$ pieces are cleaved into four samples and each sample is thereafter processed individually. Devices are isolated by etching out source-mesas from the ICL in a wet-etch procedure. A 60-nm-thick source-to-gate spacer layer is fabricated in a process where a
plasma-enhanced-chemical-vapor-deposited (PECVD) Si$_3$N$_4$-film [Fig. 1(a)] is conformally formed on the NWs. The film is then selectively removed on the sides of the wires, while keeping the lateral film intact by protecting it with a polymer etch mask in a SF$_6$ dry-etch process [Fig. 1(b)]. The sputtered W-gate is defined to 200-nm length by the thickness of an etched down polymer etch mask, using a SF$_6$ dry-etch process similar as with the nitride. The gate-drain top spacer is fabricated with a spin-coated polymer resist, which is etched-back to $\sim 300$ nm. The top contact metal consist of 250-nm sputtered Ti/W/Au. A schematic cross section of a device is shown in Fig. 1(c). The ICL acts as source, and the top metal as the drain contact.

### III. Measurements and Small-Signal Model

DC characterization is performed with a Keithley 4200-SCS. For the $D_{NW} = 45$-nm devices, transfer characteristics of a single NW FET (s-FET) and two a-FETs (52 and 192 NWs, respectively) are shown in Fig. 2(a)–(c), respectively. The peak transconductance is reduced for a-FETs as compared with the s-FETs, and the data are presented in Table I. Plots of output characteristics are shown in Fig. 3(a) and (b) for an s-FET and an a-FET (52 NWs), respectively. In Fig. 3(c), $g_{m}$ versus $V_{GS}$ for three s-FETs is shown. At $V_{DS} = 0.5$ V, the best device shows $g_{m} = 1.37$ mS/μm. The s-FET device referred to as NW1 is the same in Fig. 3(a) and (c). In Fig. 3(d), transconductance for several devices have been plotted against the $ON$-resistance ($R_{ON}$) to correlate the transconductance with the access resistance. All our performance measures are normalized to circumference [$I_{DS,norm} = I_{DS}/(n \cdot \pi \cdot D_{NW})$, where $n$ is the number of NWs. For the $D_{NW} = 45$-nm NWs, the uniform doping coupled with the large diameter causes a substantial band bending inside the NWs, which limits the off performance. This is strongly dependent on the NW doping and radius, as the Fermi level movement is restricted toward the valence band due to the increasing $D_{NW}$ toward the valence band of InAs [11]. In Table I, we have as a reference included data for a 28-nm-diameter a-FET that shows a substantially improved subthreshold swing (SS). The fact that the thinner NWs show better off-state performance and lower DIBL is attributed to increased electrostatic control due to its smaller diameter and the thinner EOT. In contrast, a difference in conduction related to confinement effects is only expected for sub-15-nm NWs, as concluded from a simulation study of the band structure as a function of NW diameter scaling [12].

#### Table I

**Peak Values Listed for 45-nm s-FETs and 45- and 28-nm a-FETs With 52, 96, and 192 NWs, Respectively. All Our Measures Are Normalized to NW Circumference. Transconductance Characteristics of the Tabled S-FETs Are Shown in Fig. 3(c). As Reference, Data for a XOIFET* [14], a Radial NW FET* [15], a Lateral NW FET* [4], and a Top Down GAA FET* [16] Are Added. All Values Given for $V_{GS} - V_T \leq V_{DS} = 0.5$ V, Which Is Also Here Chosen as the Definition for $I_{ON}$. Some Values Are Estimates Made by the Authors**

<table>
<thead>
<tr>
<th>Diameter / # of NWs</th>
<th>$I_{DS}$ (mA)</th>
<th>$L_{EO}$ (μm)</th>
<th>$R_{ON}$ (Ω·μm)</th>
<th>$g_{m}$ (mS/μm)</th>
<th>$I_{ON}$ (mA)</th>
<th>SS (mV/dec)</th>
<th>DIBL (mV)</th>
</tr>
</thead>
<tbody>
<tr>
<td>28 nm / 96</td>
<td>~300</td>
<td>1.3</td>
<td>~1000</td>
<td>0.15</td>
<td>0.06</td>
<td>140</td>
<td>110</td>
</tr>
<tr>
<td>45 nm / 192</td>
<td>~300</td>
<td>1.8</td>
<td>1630</td>
<td>0.23</td>
<td>0.13</td>
<td>540</td>
<td>240</td>
</tr>
<tr>
<td>45 nm / 52</td>
<td>~300</td>
<td>1.8</td>
<td>1240</td>
<td>0.32</td>
<td>0.17</td>
<td>480</td>
<td>320</td>
</tr>
<tr>
<td>45 nm / 1 (NW3)*</td>
<td>~200</td>
<td>1.8</td>
<td>310</td>
<td>1.09</td>
<td>0.59</td>
<td>420</td>
<td>260</td>
</tr>
<tr>
<td>45 nm / 1 (NW1)*</td>
<td>~200</td>
<td>1.8</td>
<td>210</td>
<td>1.19</td>
<td>0.67</td>
<td>500</td>
<td>320</td>
</tr>
<tr>
<td>45 nm / 1 (NW2)*</td>
<td>~200</td>
<td>1.8</td>
<td>190</td>
<td>1.37</td>
<td>0.75</td>
<td>800</td>
<td>600</td>
</tr>
<tr>
<td>13 nm / N/A*</td>
<td>2.4</td>
<td>1.70</td>
<td>1.72</td>
<td>0.90</td>
<td>180</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>25 nm / 1*</td>
<td>~1.70</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>55 nm / 1*</td>
<td>100</td>
<td>1.35</td>
<td>1.23</td>
<td>0.6</td>
<td>140</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>20 / 30 nm / 1*</td>
<td>20 / 1.2</td>
<td>~500</td>
<td>1.65</td>
<td>0.54</td>
<td>88</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>
Comparing $V_T$ for our $DNW = 45$-nm s-FETs and a-FETs, devices shows similar values (typically between $-0.4$ and $-0.2$ V) and there is no relation between $V_T$ and peak transconductance, $g_m$-peak. Quantifying $g_m$-peak widths, the values for s-FETs and a-FETs do not differ, suggesting there is no significant spread in $V_T$ within arrays that would cause a degraded maximum $g_m$. For reference, we have benchmarked our data with other high performance InAs and InGaAs NW FETs: a 13-nm planar XOI FET, $LG = 230$ nm [14], a 25-nm diameter InAs/InP radial NW FET [2-gate, $I_{DS,norm} = I_{DS}/(n \cdot 0.75 \pi \cdot D_{NW})$], $LG = 170$ nm [15], a 15 nm in diameter lateral NW, $L_G = 100$ [4], and a top-down rectangular GGA FET (20 nm $\times$ 30 nm), $L_G = 20$ nm [16].

Low frequency noise (LFN) characterization is made with a Stanford SR830 lock-in amplifier in series with a Stanford SR830 lock-in amplifier. The current noise spectral density, $S_{IV}$, is measured for $V_{DS} = 50$ mV and $f = 10$ Hz. A common measure of comparison is the deduced gate voltage noise spectral density, $S_{VG} = S_{ID}/\mu_m^2$, normalized to the gate area ($S_{VG} \cdot L_G \cdot W_G$). Data for two s-FETs and two a-FETs are plotted in Fig. 4. The lowest levels for all the measured devices are determined to range in the interval $S_{VG} \cdot L_G \cdot W_G = 310-410 \mu m^2/\mu V^2/Hz$. Another common measure of comparison is the $S_{ID}$ normalized to the drain current, $S_{ID/ID}$. The lowest measured levels are $S_{ID}/I_{DS} = 0.9 \times 10^{-10}-1.5 \times 10^{-10}$ Hz$^{-1}$ and $S_{ID}/I_{DS} = 0.9 \times 10^{-8}$, $2.0 \times 10^{-8}$ Hz$^{-1}$, for two 52-NW a-FETs and two s-FETs, respectively. The measurements indicate that when accounting for the difference in total gate width, the extracted $S_{VG}$ for s-FETs and a-FETs show undeviating values. Further, comparing the $1/f$-noise to similar devices but fabricated with a high-$\kappa$ of HfO$_2$ only [15], the new devices improved the noise performance about one order of magnitude in terms of $S_{VG} \cdot L_G \cdot W_G$ of $5700 \mu m^2/\mu V^2/Hz$ and even greater for $S_{ID}/I_{DS} = 7.3 \times 10^{-5}$ Hz$^{-1}$. The decrease in LFN can be related to the introduction of an Al$_2$O$_3$ film, which more efficiently can reduce the amount of suboxides [11]. For comparison, the ITRS roadmap states that for the year 2015, multigate MOSFET devices for mm-wave applications should comply with $S_{VG} \cdot L_G \cdot W_G = 10.3 \mu m^2/\mu V^2/Hz$ [17]. Hence, further work is required to reach the target goals within the roadmap.

The RF characterization is performed using an Agilent E8361A network analyzer from 60 MHz to 40 GHz at an RF power of $-27$ dBm [5]. For measurements, a-FETs with 192 NWs was chosen on the premises of having a higher absolute $g_m$ and a scanning electron microscopy (SEM) image of the InAs mesa structure of one such device is shown in Fig. 5(a).
current gain, $h_{21}$, and unilateral power gain, $U$, for the best device is shown in Fig. 5(b), for which $f_t = 14.6$ GHz and $f_{max} = 30.0$ GHz at $V_{DS} = 1.0$ V, and $V_{GS} = -0.45$ V and $-0.60$ V, respectively. The RF data are fitted to a standard MOSFET small-signal model [5] where the parasitic elements are included in Fig. 5(c). For simplicity, impact ionization was not accounted for as it turned out to have negligible effect in fitting the measured data. Statistics over four devices show $f_t$ in an interval of 9.7–17.4 GHz and $f_{max}$ of 20.3–30.0 GHz, at $V_{DS} = 1.0$ V.

IV. DC DEVICE MODELING

To describe the transistor characteristics of the $D_{NW} =$ 45-nm devices, a virtual source (VS) model is fitted to the measured dc data [18]. The model is based on a channel charge equation and also provides a coherent intrinsic gate capacitance [19]. Modeled devices are chosen on the premises of low $R_{ON}$ in combination with less pronounced impact ionization, thus being able to neglect the later. Since a direct measurement of the NW contact and channel resistivity is difficult, we do a simultaneous fit of the VS-model parameters with estimates of the source/drain resistance. For a s-FET (referred as NW1), assuming a NW resistivity of $1.0 \times 10^{-5}$ Qcm [20] and a first spacer layer thickness of 67 nm (60-nm Si$_3$N$_4$ + 7 nm high-α), we obtain $R_S = 65 \Omega$μm. The metal-ICL contact resistance is measured to be in the order of 2 Ω and for simplicity it is counted as a constituent of the ICL. For the s-FET, the 18 Ω series resistance contribution from the ICL, corresponding to 2.5 $\Omega$μm, can be ignored. Fitting the data in Fig. 3(a) gives $R_D = 141$ Ωμm, equivalent to 300 nm resist top spacer thickness [translating to 160-nm nongated resistive NW segment length, see Fig. 1(c)] and in agreement with the experimental implementation. At a gate length of $L_G = 200$ nm, we obtain values of mobility, $\mu_{LOW-FIELD} = 1300 \text{cm}^2/\text{V s}$ and $v_{inj} = 1.7 \times 10^7 \text{cm/s}$. While the rather low mobility can be expected for a highly doped surface channel, the injection velocity is in the same range as reported for an InAs HEMT of similar gate length [2] ($v_{inj} \sim 2 \times 10^7 \text{cm/s}$).

To analyze a-FETs, the fitted model for the s-FET is expanded and fitted to an a-FET with 52 NWs. The significant increase in $R_{ON}$ for a-FETs compared with s-FETs can in part be explained by the undoped ICL that adds 18 Ω source series resistance, see Fig. 6. To obtain a good fit, however, it is needed to reduce $v_{inj}$ by ~30% and also increase the resistivity of the NWs by a factor of around four, and this fitting is shown in Fig. 3(b). The decrease in $v_{inj}$ can be explained by longer channels for a-FETs as compared with the s-FET (as well as thicker spacer layers) due to the defining polymer resist, used for the etch-back of the gate that tends to become thicker around arrays. The steep increase in series resistance can, however, not be explained simply by a thicker gate-drain separation. Instead, it must be considered that the doping incorporation of the surfactant Sn is higher for single NWs is thought to be related to diameter variation (NWs in arrays typically vary between 40–45 nm) in combination with variations of separation layer thicknesses over the sample. To compare the effect of the reduced $v_{inj}$ with the effect of external resistances, we plot calculated values for the extrinsic transconductance, $g_{m,ext}$, as a function of $R_{ON}$, in Fig. 3(d) using the VS model. In the plot there
Fig. 6. Schematic illustration of the distribution of the parasitic series resistance extracted from dc measurements and VS modeling where the subscripts \( R_{S,ICL} \) and \( R_{S,NW/R_{D,NW}} \) are the series resistances associated with the ICL and the NW, respectively. Values are shown for (a) 52-NW a-FET and (b) 192-NW a-FET.

are two lines representing two values of \( V_{\text{inq}} \) (the fitted s-FET and a-FET) corresponding to \( L_G = 200 \text{ nm} \) \((g_{m,\text{fit,}\text{fit}})\) and \( L_G = 300 \text{ nm} \) \((g_{m,\text{fit,}\text{fit}})\), respectively. Comparing the simulated data with the experimentally measured \( g_{m,\text{ext}} \), a good fit can be obtained. This indicates that the external resistances are the major source for device performance degradation of a-FETs as compared with s-FETs. This can essentially be described using the conventional formula relating the extrinsic and intrinsic transconductance \( (1 + g_{m,\text{int}}) R_s + g_{d,\text{int}} (R_s + R_D) \) \([21]\). Here, \( g_d \) is the output conductance. From the average experimental data for the s-FETs shown in Fig. 5(c), we deduce a value for the intrinsic transconductance, \( g_{mi} = 2.4 \text{ mS/\mu m} \). For the a-FETs, this value range between 0.5-0.7 \text{ mS/\mu m}.

The total parallel resistance contributions from the un-gated segments of the NWs are shown in Fig. 6(a) and (b) for a 52-NW a-FET and a 192-NW a-FET, respectively. Comparing the normalized series resistance contributions attributed to the un-gated NW segments (omitting the ICL), \( R_S = 265 \Omega/\mu m \) and \( R_D = 845 \Omega/\mu m \) for the 52-NW a-FET, and \( R_S = 270 \Omega/\mu m \) and \( R_D = 900 \Omega/\mu m \) for the 192-NW a-FET. These values are quite close as expected (2% and 6% in difference for \( R_S \) and \( R_D \), respectively). For a 52-NW a-FET, the ICL \((18 \Omega)\) adds another 12% to the total series resistance while for a 192-NW a-FET, the series resistance contribution from the ICL adds another 42%. This explains why a 52-NW a-FET has better normalized dc performance than a 192-NW a-FET.

V. RF DEVICE MODELING

Performing electrostatic simulations of the capacitances of the de-embedded structure of \( 12 \times 12 \mu m^2 \) \([5]\), that is the overlapping source–gate and gate–drain areas, as shown in Fig. 5(a) with the overlapping rectangles, the parasitic capacitances \( C_{ph,p} \) and \( C_{ph,p} \) are determined to 131 fF and 16 fF, respectively. The substantially larger \( C_{ph,p} \) originates from the thinner source spacer layer and higher permittivity of the Si3N4 as compared with the drain spacer. Considering the low density of states for InAs, the high frequency intrinsic gate capacitance, \( C_{g,i} \), for 192 NWs is simulated to be in the interval of 6–23 fF for \( V_{DS} = V_T \) varying between 0.5 and 0.5 V, at \( V_{DS} = 1 \text{ V} \). Considering the low measurement gate bias and assuming that the \( C_{g,ext} \) contribution lies in the middle of the given range, the simulated number \( C_{g,ext} = 5.95 \text{ fF/\mu m} \) is 13% larger than the number extracted from the measurement \( C_{g,ext} = 5.27 \text{ fF/\mu m} \). Actual pad size difference compared with the patterned area as well as uncertainty in the spacer layer thicknesses and dielectric constants are likely the reasons for the deviation between measured and calculated capacitances. The extracted \( g_{m,\text{int}} \) of 0.53 mS/\mu m corresponds to a \( \delta_{\text{ext}} \) of 0.36 mS/\mu m, which can be related to the maximum \( g_{m,\text{int}} \) measured at dc for the same device \((g_{m,\text{int}} = 0.30 \text{ mS/\mu m}, 20% \text{ lower})\). The low \( V_{DS} \) in the measurement for maximum \( f_t \) and \( f_{\text{max}} \) suggest a shift in \( V_T \) as compared with dc \((V_T = -0.45 \text{ V} \text{ for } V_{DS} = 1 \text{ V}) \). Both the shift in \( V_T \) and the increase in \( g_m \) can be related to slow traps that are not responding at higher frequencies, thus explaining the lowering of the \( V_T \) and a frequency dependent \( g_m \) \([22]\). Comparing the series resistances from the RF model extraction with those from the VS model fitting of the dc measurements, they are found to be fairly similar. The RF extraction gives \( R_S = 34 \Omega \) and \( R_D = 33 \Omega \) while the dc measurements translates into \( R_S = 28 \Omega \) and \( R_D = 33 \Omega \).

VI. DISCUSSION

Based on the data presented in this paper we draw some general conclusions regarding III–V NWs and their implementation; although high intrinsic performance, that is high transconductance and drive currents, has been demonstrated that the implementation of RF-compatible devices still is a challenge. Analysis of planar III–V MOSFETs \([23]\) show that the control of the access resistance is one of the key parameters to increase the transconductance and to obtain attractive RF-properties. The method used in the implementation presented here, that is uniform doping of the transistor channel, needs further refinement beyond diameter scaling in order to avoid increase in the access resistance. Increased doping control may be one alternative to reduce the source/drain resistance \([9]\). Alternatively, the access regions may be regrown, although at the cost of parasitic capacitance \([10]\). A heterobarrier may also be included into the channel to increase the carrier control \([24]\). It appears that a combination of these approaches may be required to fully exploit the potential of the high III–V injection velocity for RF-compatible MOSFETs in either vertical or lateral geometry.
VII. Conclusion

The performance of s-FET has been evaluated ($\tau_{\text{on,ext}} = 1.37 \text{ mS/\mu m}$). Based on a VS model we are able to accurately model and determine intrinsic transport properties including $\tau_{\text{on}} = 1.7 \times 10^7 \text{ cm/s}$. The performance is degraded by series resistance for devices consisting of arrays of NWs. The introduction of a thin Al$_2$O$_3$ film combined with an outer HfO$_2$ film may be responsible for the measured reduction in 1/f noise compared with similar devices fabricated with HfO$_2$ only.

Acknowledgment

The authors would like to thank M. Egard, V. Drakinsky, and T. Bryllert for assistance and H. Sjöland for valuable discussions.

References


[18] A. Khakifirooz, O. M. Nayfeh, and D. Antoniadis, “A simple semime- 

consistent current and charge FET models: From ballistic to drift- 

erties of n-type InAs nanowires grown from metal-organic precursors,” Nanotechnology, vol. 21, no. 20, pp. 1–9, Apr. 2010.


Karl-Magnus Persson (M’12) is currently pursuing the Ph.D. degree in nanoelectronics with Lund University, Lund, Sweden.

His current research interests include fabrication and modeling of III-V nanowire MOSFET devices and circuits.

Martin Berg is currently pursuing the Ph.D. degree with Lund University, Lund, Sweden.

His current research interests include nanowire-based MOSFETs in analog circuits.

Mattias Borg (M’12) received the Ph.D. degree in physics from Lund University, Lund, Sweden, in 2012. He is currently a Post-Doctoral Researcher with IBM Research, Zurich, Switzerland.

Jun Wu received the M.S. degree in engineering nanoscience from Lund University, Lund, Sweden, in 2011, where he is currently pursuing the Ph.D. degree in nanoelectronics.

His current research interests include nanowire-based MOSFETs.

Sofia Johansson received the M.S. degree in engineering nanoscience from Lund University, Lund, Sweden, in 2008, where she is currently pursuing the Ph.D. degree. Her current research interests include fabrication and characterization of III-V nanoelectronic devices.
Johannes Svensson received the M.S. degree in physics and the Ph.D. degree from the University of Gothenburg, Göteborg, Sweden, in 2004 and 2010, respectively. He is a Researcher with Lund University, Lund, Sweden.

Kristofer Jansson received the M.S. degree in engineering nanoscience from Lund University, Lund, Sweden, in 2010, where he is currently pursuing the Ph.D. degree with the Department of Electrical and Information Technology.

Erik Lind received the Ph.D. degree in solid state physics from Lund University, Lund, Sweden, in 2004. He is an Associate Professor of electrical and information technology with Lund University.

Lars-Erik Wernersson received the M.S and Ph.D. degrees from Lund University, Lund, Sweden, in 1993 and 1998, respectively. He has been a Professor of nanoelectronics with Lund University since 2005.