Modeling, Fabrication and Characterization of Silicon Tunnel Field-Effect Transistors

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Diplom-Physiker

Christian Philipp Sandow
aus Göttingen

Berichter: Universitätsprofessor Dr. Siegfried Mantl
Universitätsprofessor Dr. Lutz Feld

Tag der mündlichen Prüfung: 8. Juni 2010

Abstract

Over the last decades, the continuous down-scaling of metal-oxide-semiconductor field-effect transistors (MOSFETs) enabled faster and more complex chips while at the same time the space and power-consumption was kept under control. However, in the future, the further reduction of the power consumption per unit area will be restricted by a fundamental limit of the inverse subthreshold swing of MOSFETs, which relates its on/off-current-ratio to the operation voltage. Since logic devices operate at a given on/off-current-ratio, the limited subthreshold swing will prevent further reduction of the operation voltage, which is the main parameter to reduce the power consumption.

In this thesis, the Tunnel-FET (TFET) is studied as an alternative switching device which could overcome the physical limit of the subthreshold slope in MOSFETs. After introducing the working principle of the TFET, device parameters are studied extensively in quantum simulations based on the non-equilibrium-Green's-function method. It is found that the performance of a nanowire device geometry is superior to that of planar structures and that the gate dielectric should be as thin as possible. Moreover, the impact of doping concentration on the switching behavior is investigated. For very large doping concentrations, the subthreshold swing is expected to deteriorate while smaller doping concentrations lead to reduced on-currents. Therefore, the doping concentrations need to be tailored to a specific application. Finally, TFETs with different substrate materials are simulated and the influence of bandgap and effective masses is illustrated. A small bandgap improves band-to-band tunneling currents, therefore, the on-currents of the TFET increase. However, due to the ambipolar behavior of the TFET, the off-currents increase as well. Therefore, an optimal TFET is proposed, which is a heterostructure nanowire that utilizes a small bandgap material at the source/channel-junction and a large bandgap material at the drain/channel-junction.

The extensive simulations are complemented by a study on different experimental realizations of the TFET: As a first step, planar silicon TFETs were fabricated on ultra-thin-body silicon-on-insulator substrates. The resulting TFETs exhibit minimal inverse subthreshold slopes of 325 mV/dec and on-currents of the order of 10^{-2} μA/μm. Since these results are inferior to MOSFET performance, optimizations of the doping concentration and gate dielectric thickness are investigated and both parameters are found to impact the performance as predicted by the simulations. Furthermore, the lateral steepness of the source doping profile is identified as an important parameter, which limits the switching slope.

To benefit from the improved electrostatics of nanowires, in a second step, silicon nanowire array TFETs with widths of < 20 nm were fabricated using a top-down approach. In order to optimize the slope of the doping profile, for the first time laser annealing was employed for dopant activation in TFETs. To find the optimum annealing conditions, the impact of different laser energies in combination with a thermal post-anneal treatment on the TFET performance is studied.

The electrical characterization of the nanowire TFETs shows an improvement of the subthreshold swing by about 10% and of the on-currents by one order of magnitude when compared to the planar TFETs. To deepen the understanding of TFET operation, low temperature measurements have been performed and band-to-band tunneling is found to be the dominant conduction method. Moreover, for the first time possible parasitic recombination mechanisms are identified in a TFET which might limit the switching slope in silicon.

Since small-band gap heterostructure nanowires might offer largely improved tunneling probabilities, in this thesis, a first experimental realization of InSb nanowire MOSFETs is presented. As the bandgap is the most important property for TFET applications, it is carefully extracted from the electrical characteristics and it is found to match the value known from bulk InSb very well.

In summary, this thesis presents quantum simulations and two experimental realizations of TFETs in silicon are studied in detail. Variations of device parameters show a path for further optimizations of silicon TFETs.

As a first step beyond silicon, InSb nanowire MOSFETs are fabricated successfully for the first time and the potential of InSb for TFET operation is discussed.

Kurzfassung

In den vergangenen Jahrzehnten erlaubte die kontinuierliche Verkleinerung der Metal-Oxid-Halbleiter-Feldeffektransistoren (MOSFETs) die Herstellung schnellerer und zunehmend komplexer Mikrochips. Mit der Verkleinerung der MOSFETs wurde gleichzeitig die Versorgungsspannung reduziert, was dazu führte, dass die Leistungsaufnahme pro Fläche konstant blieb. In Zukunft wird diese Reduktion der Versorgungsspannung jedoch durch die physikalisch beschränkte minimale Unterschwellensteigung des MOSFETs begrenzt werden.

Im Rahmen dieser Promotionsarbeit wird der Tunnel-Feldeffektransistor (TFET) als Alternative zum MOSFET untersucht, da der TFET keine Beschränkung der minimalen Unterschwellensteigung aufweist und somit eine weitere Reduktion der Leistungsaufnahme erlauben könnte. Zu Beginn der Arbeit wird zunächst das Konzept des TFETs eingeführt und wichtige Parameter werden anhand von umfangreichen Bauelementesimulationen auf Basis der Nicht-Gleichgewichts-Greens-Funktions-Methode untersucht.

Zunächst wird der Einfluss der Bauelementgeometrie studiert. Zwei Realisierungen des TFETs, zum einen als planarer TFET und zum anderen als Nanodraht-TFET werden verglichen. Aufgrund seiner überlegenen Gateelektrostatik ist der Nanodraht-TFET der planaren Realisierung überlegen und erlaubt größere Anströme sowie verbesserte inverse Unterschwellensteigungen. Die Dicke des Gatedielektrikums hat ebenfalls Einfluss auf die Leistungsfähigkeit des TFETs. Je dünner das Gatedielektrikum, desto besser wird die Gateelektrostatik und damit die Tunneleigenschaften.

Ein weiterer wichtiger Parameter ist die Dotierstoffkonzentration in Source und Drain. Für die Dotierstoffkonzentration können zwei Grenzfälle unterschieden werden. Für sehr große Konzentrationen wird die Unterschwellensteigung schlechter, bei zu kleinen Konzentrationen hingegen ist der Anstrom reduziert. Somit sollte die Dotierstoffkonzentration der jeweiligen Anwendung angepasst werden.

Da die relativ große Bandlücke von Silizium hohe Tunnelströme verhindert, werden abschließend verschiedene TFETs mit unterschiedlichen Bandlücken simuliert und verglichen. Es zeigt sich, dass Materialien mit geringer Bandlücke zwar größere Anströme aufweisen, jedoch aufgrund der Ambipolarität des TFETs die Ausströme des TFETs stärker als die Anströme zunehmen. Deshalb wird als optimale Lösung ein Nanodraht-TFET basierend auf einer Heterostruktur vorgeschlagen, bei dem am Source/Kanal-Übergang ein Halbleiter mit kleiner Bandlücke und am Drain/Kanal-Übergang ein Halbleiter mit großer Bandlücke zum Einsatz kommt.

Im zweiten Teil der Arbeit werden die Simulationen durch experimentell realisierte TFETs auf Silizium ergänzt:

Zunächst wurden TFETs auf ultra-dünnen Silizium-auf-Isolator-Substraten hergestellt und charakterisiert. Die gemessenen Kennlinien zeigen eine minimale inverse Unterschwellensteigung von 325 mV/dec und Anströme in der Größenordnung von 10^{-2} µA/µm. Da somit die Leistungsfähigkeit dieser TFETs geringer ist als diejenige vergleichbarer MOSFETs, wird zur weiteren Optimierung der TFETs der Einfluss der Dotierstoffkonzentration in Source und Drain sowie der Dicke des Gatedielektrikums studiert. Beide Parameter beeinflussen die Kennlinien der TFETs entsprechend den durchgeführten Simulationen. Darüber hinaus zeigen die experimentellen Daten, dass die Steilheit des Dotierprofils am Source/Kanal-Übergang die Tunnelwahrscheinlichkeit begrenzt.

Um eine Verbesserung der Tunneleigenschaften zu erreichen, wurden im nächsten Schritt TFETs in Form von parallelen Nanodrähten mit einem Drahtdurchmesser von < 20 nm unter Nutzung eines „Top-Down" Prozesses hergestellt. Zudem wurde zur Vergrößerung der Steilheit der Dotierstoffprofile erstmalig Laser Annealing zur Dotierstoffaktivierung in TFETs eingesetzt. Die elektrische Charakterisierung zeigt eine Verbesserung der inversen Unterschwellensteigung um ca. 10% und des Anstroms um eine Größenordnung im Vergleich zu den planaren TFETs.

Um das Verständnis der physikalischen Grundlagen der Funktionsweise des TFETs zu verbessern wurden Tieftemperaturmessungen durchgeführt, welche bestätigen, dass Band-zu-Band-Tunneln der dominante Leitungsprozess ist. Darüber hinaus zeigen die Tieftemperaturmessungen erstmalig, dass parasitäre Rekombinationsmechanismen in Silizium TFETs existieren, die die Leistungsfähigkeit der Tunneltransistoren beeinträchtigen könnten.
Da Halbleiter mit kleiner Bandlücke im Vergleich zu Silizium deutlich größere Tunnelwahrscheinlichkeiten ermöglichen, wurden im Rahmen dieser Arbeit erstmalig InSb-Nanodraht-MOSFETs hergestellt. Bei der elektrischen Charakterisierung dieser MOSFETs wurde der Schwerpunkt auf die Extraktion der Bandlücke gelegt, da diese der wichtigste Parameter für zukünftige TFETs aus InSb ist.
Zusammenfassend wird in dieser Arbeit die Leistungsfähigkeit von Silizium-TFETs durch Quanten-Simulationen und zwei verschiedene experimentelle Realisierungen detailliert untersucht. Anhand von Parametervariationen werden Wege zur weiteren Optimierung von Silizium-TFETs aufgezeigt. Als ein erster Schritt in Richtung zukünftiger Tunneltransistoren auf Basis von III-V-Halbleitern mit geringer Bandlücke werden erstmalig MOSFETs aus InSb präsentiert.

Contents

List of Abbreviations

AZ5214	Optical photo resist with a nominal thickness of 1.4 µm
AZ5206	Optical photo resist with a nominal thickness of 0.6 µm
BOX	Buried Oxide (isolation layer of an SOI structure)
BTBT	Band-to-Band-Tunneling
CMOS	Complementary MOS technology
DOS	Density of States
E-Beam	Electron Beam
ELA	Excimer Laser Annealing
FET	Field-Effect-Transistor
FIB	Focused Ion Beam
HSQ	Hydrogen Silsesquioxane, negative resist for electron beam lithography
ICP	Inductive Coupled Plasma etching machine
IT	Information Technology
ITRS	International Technology Roadmap for Semiconductors
LDOS	Local Density of States
LPCVD	Low-Pressure Chemical Vapor Deposition
MFP	Mean free path length
MOS	Metal-Oxide-Semiconductor
MOSCAP	Metal-Oxide-Semiconductor Capacitor
MOSFET	Metal-Oxide-Semiconductor Field-Effect-Transistor
NEGF	Non-Equilibrium Green's Function formalism
NW	Nanowire
NWFET	Nanowire Field-Effect-Transistor
PECVD	Plasma-Enhanced Chemical Vapor Deposition
PMMA	Polymethylmethacrylat, positive resist for electron beam lithography
QCL	Quantum Capacitance Limit
RCA	Radio Corporation of America, cleaning technique for Si wafers
RIE	Reactive Ion Etching maching
RTA	Rapid Thermal Annealing
RTP	Rapid Thermal Processor
SEM	Scanning Electron Microscope
SOI	Silicon on Insulator
SPER	Solid Phase Expitaxial Regrowth
SRH	Shockley-Read-Hall recombination model
sSOI	Strained Silicon on Insulator
TCAD	Technology Computer Aided Design
TEM	Transmission Electron Microscope
TFET	Tunnel Field Effect Transistor

TRIM	The Stopping and Range of Ions in Solids, monte-carlo algorithm for implantation simulations
UTB	Ultra-Thin-Body
UV6	Deep-UV optical photo resist with a nominal thickness of 0.6 µm
VLS	Vapor-Liquid-Solid method
WKB	Wenzel-Kramer-Brillouin approximation

List of Symbols

Γ	Broadening function, inverse life-time
ε	On-site energy in finite difference calculations
ε_0	Dielectric constant of the vacuum: $8.854 \cdot 10^{-12}$ As/(Vm)
ε_r	Relative dielectric constant
λ	Wave length
λ_{dop}	Screening length for potential variations due to the presence of ionized dopants
λ_{mfp}	Mean free path length (momentum relaxation length)
Λ	Screening length for potential variations
μ	Chemical potential
μ_D	Chemical potential of the drain
μ_s	Chemical potential of a semiconductor
μ_S	Chemical potential of the source
$\rho(x)$	Charge density
Σ	Self-energy
τ_G	Gate delay time
ϕ_S	Surface potential
Φ_f^0	Distance between source conduction band and channel valence band
ψ_E	Solution to the time-independent SCHRÖDINGER equation for energy E
ψ	Wave-function
a	Spatial distance of simulation grid nodes
$A(E)$	Spectral function
C_G	Gate capacitance
C_q	Quantum (semiconductor) capacitance
$D(E)$	Density of states
D_{As}	Implantation dose for arsenic
D_B	Implantation dose for boron
E	Energy
E_c	Energy of the conduction band edge
E_g	Bandgap energy
E_t	Transverse energy
E_v	Energy of the valence band edge
f	FERMI-DIRAC distribution
f_S	FERMI-DIRAC distribution in the source
f_D	FERMI-DIRAC distribution in the drain
G	GREEN's function
G^n	Correlation function
h	PLANCK constant: $4.135 \cdot 10^{-15}$ eV/s
H	Hamilton operator
k	One-dimensional crystal momentum,

	also in kT: BOLTZMANN constant: $8.617 \cdot 10^{5}$ eV/K
k_t	transverse crystal momentum
$\vec{k}$	Vector of crystal momentum
I_D	Drain current
I_G	Gate leakage current
I_{off}	Off-state current
I_{on}	On-state current
I_S	Source current
L	Channel length
m^*	Effective carrier mass
m^*_c	Effective carrier mass in the conduction band
m^*_v	Effective carrier mass in the valence band
N_D	Number of ionized donor states
N_A	Number of ionized acceptor states
q	Elementary charge: $1.6022 \cdot 10^{-19}$ C
Q_{ox}	Oxide charge
Q_q	Semiconductor charge
R_S	Source resistance
R_{shunt}	Parallel shunt resistance
R_D	Drain resistance
S	Inverse subthreshold slope
t_0	Nearest neighbor coupling parameter
t_{Ox}	Oxide thickness
t_{Si}	Silicon body thickness
T	Transmission, in kT also: Temperature
T_{BTBT}	Band-to-band tunneling transmission
T_{scat}	Transmission due to scattering
$U(x)$	Self-consistent potential
V	Voltage
V_{bi}	Built-in voltage of a junction
V_{DD}	Operating voltage
V_{DS}	Source to drain voltage
V_{GD}	Gate to drain voltage
V_{GS}	Gate to source voltage
V_T	Threshold voltage

Chapter 1
Introduction

During the last three centuries, technology developed at unprecedented speed towards a point at which nobody in developed countries could imagine to live without its products. Among the most important inventions, the computer stands out due to its versatility. In conjunction with the invention of the transistor, the computer dramatically changed the way how information is handled. This revolution of the information technology (IT) is unique and nowadays, nearly every profession makes use of computers, every person in developed countries uses mobile phones and more than one billion people have access to the worldwide internet [1], which simplifies their day-to-day lives.

Unfortunately, this tremendous success of technology does not come without a cost. The yearly consumption of fossil fuels is at its height and burning those energy sources leads to global warming which will become one of the biggest challenges for humankind in the 21st century.

Apart from the search for new sources of energy, saving energy is an important task for everybody. Nowadays, consumers start to take power consumption into consideration when buying new technology products. Therefore, saving energy becomes a more and more important challenge when designing new products.

In contrast to most industries, reducing the power consumption of microchips has been a main driver of research and development in the microchip industry for at least two decades. Until the late 1980's, the microchip industry successfully reduced the transistor dimensions to increase the transistor packing density and hence the computing performance of microprocessors. At the same time, the clock frequency of the chips was increased and the power consumption per unit area grew. Starting with Intel's 80486 microprocessor in the late 1980's, central processing units need active cooling to remove the dissipated heat from the circuitry and keep it operational. At this point, it became clear that further down-scaling has to include the operation voltages to keep the power consumption per unit area constant. Consequently, for the first time since the introduction of the 8086 in 1978, the core voltage was reduced from 5 V to 3.3 V in 1992. During the next chip generations, scaling was continued successfully until the SiO_2 gate oxide thickness became too thin ($\approx$ 1 nm) to efficiently prevent leakage currents. But, SiO_2 has been replaced by HfO_2-based gate oxides and down-scaling has continued until today. In 2010, the minimum core voltage of the Intel Core i5 processor is 0.65 V and the minimum feature size shrank from 3 µm in 1978 to 45 nm in 2010. Despite this enormous success of scaling, the classical down-scaling strategy

(constant-field scaling [2]) cannot be continued indefinitely. There are a number of demanding challenges which may turn out to be impregnable. On transistor level, the most fundamental are:

1. ***Variability.*** For shrinking device dimensions, the statistical nature of various processes like doping and layer depositions will become more important, since the number of atoms involved reduces drastically.
2. ***Limited switching slope.*** In modern microprocessors, standard metal-oxide-semiconductor field-effect-transistors (MOSFETs) are predominantly used for switching. While an ideal switch provides an abrupt transition from the off- to the on-state, at room temperature there is a physical limit for the switching slope in FETs. Together with a fixed on-/off-current ratio of 10^5, this limitation implies that the minimum operation voltage of a MOSFET is 0.3 V. Therefore, the current operation voltage can only be scaled by another factor of two before the physical limit is encountered.
3. ***Tunneling currents.*** The third fundamental challenge arising from transistor scaling is the increase of tunneling currents. For ultra-thin gate-oxides, this problem has already been faced and replacing SiO_2 gate oxides by high-κ (e.g. HfO_2) dielectrics is a promising solution. However, for ultimately scaled MOSFETs, the potential barrier which prevents current flow in the off-state can become transparent due to tunneling. In this case, switching may become impossible.

While theses elemental limitations demand a change of the basic MOSFET structure, there are several technological challenges related to scaling, which may as well be hard to overcome, but still could be solved using MOSFET devices.

In the framework of this thesis, one of the three most fundamental questions of MOSFET scaling is addressed, the limited switching slope. This thesis is structured as follows:

As the starting point of the present work, a brief overview of MOSFET operation is given and the physical limit of the switching slope is discussed. Since the limited switching slope is inherent to the MOSFET, an alternative device concept, namely the Tunnel-FET (TFET) is introduced and the strategy how the TFET could overcome the limited switching slope of the MOSFET is presented. To substantiate the qualitative reasoning, a simple model for TFET switching is presented and the prospected performance of a one-dimensional silicon nanowire TFET is discussed.

Since there are many factors which can influence the performance of the TFET, a more complex device simulation, based on first principles, is presented in Chapter 3. The results of these simulations demonstrate that the ultimate choice for realizing a TFET in a single material is a parallel nanowire array with < 10 nm wire diameter in the case of silicon.

Chapter 4 presents experimental results on planar TFETs which are realized on ultra-thin-body silicon-on-insulator-substrates. The impacts of gate-oxide thickness and doping concentration on the TFET performance are studied and the predictions from the simulations are confirmed qualitatively.

These results indicate that the steepness of the doping profile is a key factor for improving the TFET performance. Therefore, in Chapter 5, laser annealing is introduced as a method to obtain steeper doping profiles. In order to further improve the TFET performance, laser annealing is studied on silicon nanowire TFET arrays (nanowire width about 20 nm) which are fabricated using a top-down process that includes electron beam lithography and dry etching. To date, these nanowire TFETs are the smallest nanowire TFETs available in literature, the only ones manufactured using a top-down process and the only TFETs which use laser annealing for dopant activation.

Since the TFET relies on band-to-band tunneling, the size of the bandgap of the semiconductor material in use is very important. The semiconductor with the smallest known bandgap is InSb.

In order to eventually fabricate a nanowire InSb TFET, in Chapter 6 first InSb nanowire MOSFETs are presented and the bandgap of the material is extracted carefully. These transistors have been fabricated in the group of Prof. J. Appenzeller at Purdue University, IN, USA. In conjunction with the simulation results of Chapter 3, the usability of InSb for TFETs is assessed.

Chapter 2
Theory of the Tunnel-FET

2.1. The MOSFET

2.1.1. Overview

The following chapter focuses on the theory of Metal-Oxide-Semiconductor Field-Effect Transistors (MOSFETs), the basic element of all modern logic circuits. After covering MOSFET basics, the Tunnel-FET will be introduced as an alternative and its operation will be discussed in depth.
In digital applications, complementary MOSFETs act as a voltage driven current switch that ideally do not consume any power in equilibrium. The three terminals of a MOSFET are source and drain, the two current terminals, and the gate which allows for the modulation of a potential barrier in the channel that either blocks or permits current flow. The two digital states '0' and '1' are represented by two voltages, the common ground (0 V) and the supply voltage (V_{DD}) of the MOSFET. The simplest logic circuit is the inverter that consists of two complementary MOSFETs which permit current flow either in the '0'-state or the '1'-state. In this CMOS (Complementary MOS)-configuration, current (I_D) only flows during the transition from one state to the other during which the current is needed to charge the gate-capacitors (C_G) of the subsequent logic stage. Hence, a fundamental measure of MOSFET performance is the intrinsic switching time, the so-called gate delay of the MOSFET:

$$\tau_G = \frac{C_G V_{DD}}{I_D} \tag{2.1}$$

This gate delay can be visualized as the time constant associated with the charging of a transistor gate with a charge $Q_G = C_G V_{DD}$ by the output current of a preceding transistor I_D.

2.1.2. The MOS-Capacitor

To understand MOSFET-operation, it is instructive to study the MOS-Capacitor (MOSCAP, see Figure 2.1) first. A MOSCAP consists of a stack of a semiconductor (e.g. Si), an insulating layer (e.g. SiO_2) and a metallic gate (e.g. Poly-Si). Electrically, the MOSCAP features three distinct regions of operation: accumulation, depletion and inversion. Depending on the wafer background dop-

ing of the semiconductor, the voltages necessary to achieve these states reverse. For a p-type Si-substrate as shown, a negative bias voltage leads to an accumulation of majority charge carriers (holes) at the Si/insulator interface.

For a moderate positive bias voltage, the holes are repelled from the interface and no free charge carriers are left, the semiconductor is depleted. On the other hand, for large positive bias, minority charge carriers (electrons) are attracted to the interface thus forming a conductive inversion layer. This behavior is a consequence of the field-effect in semiconductors, which allows for the modulation of the charge carrier density by an electric field.

The reason for the field effect becomes apparent in the band-diagram. In moderately doped p-type silicon, the electrochemical potential μ_s is located near the valence band edge E_v which yields a moderate number of intrinsically free charge carriers (e.g. $\approx 10^{17}\,cm^{-3}$). In the presence of an electric field, the bands (conduction band edge E_c, valence band edge E_v) are bend, which leads to a varying $E_{c/v}$- μ_s. If one of the band-edges lies within a few kT of μ_s, charge carriers accumulate and a charge builds up that terminates the electric field lines and reduces the electric field.

This also illustrates the main difference of semiconductors and metals. In metals, a huge number of charge carriers screens any electric field within a distance of a few Ångstrom, but due to the lack of free charge carriers, the effective screening length in weakly doped semiconductors depends on the field strength and can range over micrometers. In insulators, however, there is no screening and an electric field is constant throughout the material.

In the MOSCAP, these three material characteristics are exploited. The insulating dielectric is used to convey the electric field originating from the metal to the semiconductor surface while blocking any current flow. In a specific MOSCAP, the external voltage needed to create a certain electric field depends on the work-function-difference between the metal and the semiconductor. The voltage that yields zero electric field is referred to as 'flatband'-voltage and also depends on the number of trap states in the insulator and at the semiconductor/insulator interface. In case

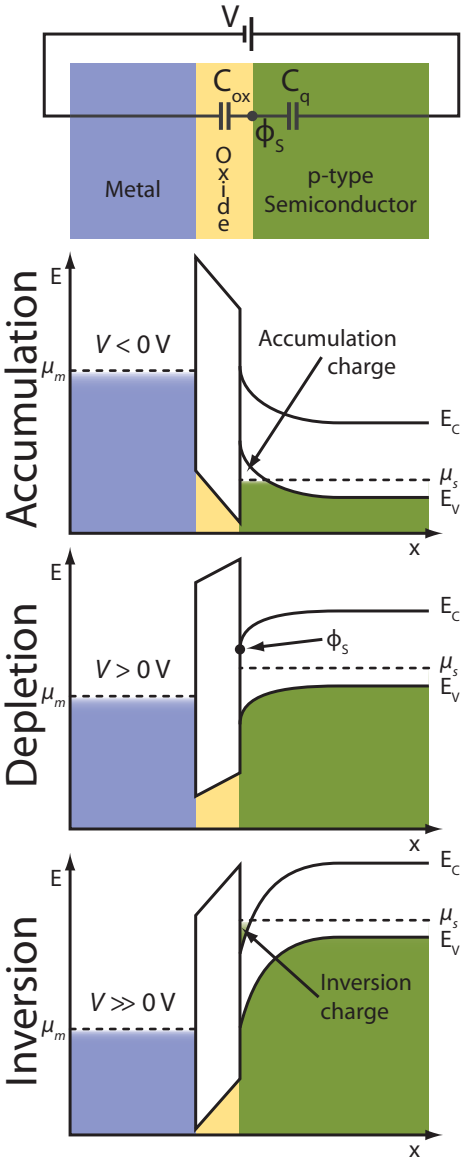


Figure 2.1. MOS-capacitor structure and band profiles for a p-type semiconductor. C_{ox} is the oxide capacitance, C_q, the semiconductor capacitance, ϕ_s is the surface potential of the semiconductor. $E_{C/V}$ are the conduction and valence band edges, while $\mu_{m/s}$ are the chemical potential of metal and semiconductor, which differ by V, the applied external voltage.

of poor interface quality, a large number of charged interface trap states can terminate electric field lines and thus reduce the effective electric field.

Apart from the field effect, the central property of the MOSCAP is the relation between external voltage V and surface potential ϕ_S (see Figure 2.2), which in this work is identified with the conduction band edge E_c of the semiconductor at the insulator interface. This relation is governed by a capacitive voltage divider that is a result of the series combination of the constant insulator capacitance C_{ox} and the capacitance of the semiconductor (also known as quantum capacitance):

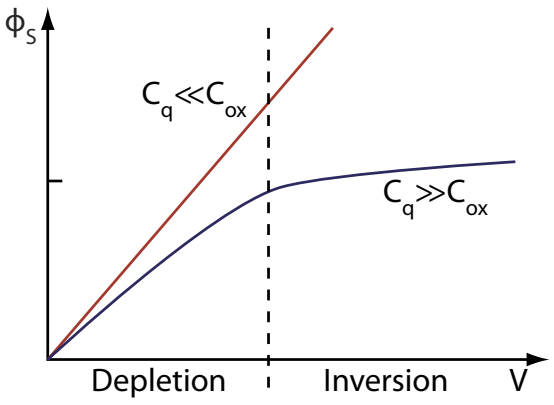


Figure 2.2. Dependence of the surface potential ϕ_S on the applied external voltage V. Two cases can be distinguished. If C_q is larger than C_{ox}, the band movement stops in inversion. In the quantum capacitance limit, C_q is much smaller than C_{ox} and the surface potential follows the voltage 1:1.

$$C_q = \frac{dQ_q}{d\phi_S} \tag{2.2}$$

which is defined as the derivative of the semiconductor charge Q_q with respect to the surface potential ϕ_S. This quantum capacitance is voltage dependent and it differs for the three operating regimes. In accumulation, the quantum capacitance is large because the large number of free majority carriers can easily follow any change of the applied voltage. In depletion, however, there are no free carriers and the capacitance is negligible. For inversion, two cases have to be distinguished. For small signal frequencies, typically smaller than 100 Hz, the number of minority carriers can follow the applied voltage, but for larger frequencies the number of carriers is limited by the small thermal generation rate. Therefore, for typical operation frequencies, the capacitance in inversion is very small.

In a MOSFET, minority carriers are usually supplied by the source and drain regions, therefore the inversion quantum capacitance is of the same order as the accumulation capacitance. Depending on the size of the quantum capacitance, the applied gate voltage drops either on the oxide capacitance or on the quantum capacitance. For large C_q, like in accumulation and inversion, the main voltage drop occurs at the oxide capacitor which means that the surface potential can hardly be moved by the external voltage. In depletion, however, the quantum capacitance is small and in case of excellent device electrostatics, the surface potential follows the applied external voltage one-to-one.

Quantum capacitance limit

As MOSFET switching is based on the relation between surface potential and gate voltage, it is an interesting question if the one-to-one correspondence could be extended to the accumulation and inversion regimes. To achieve this, the ratio between oxide capacitance and quantum capacitance in accumulation and inversion needs to be increased to a point where the quantum capacitance is much smaller than the oxide capacitance. This condition is called 'quantum capacitance limit (QCL). One way to achieve QCL is to increase the oxide capacitance by using thinner oxides and materials that provide a larger dielectric constant than the commonly used SiO_2. The other way is

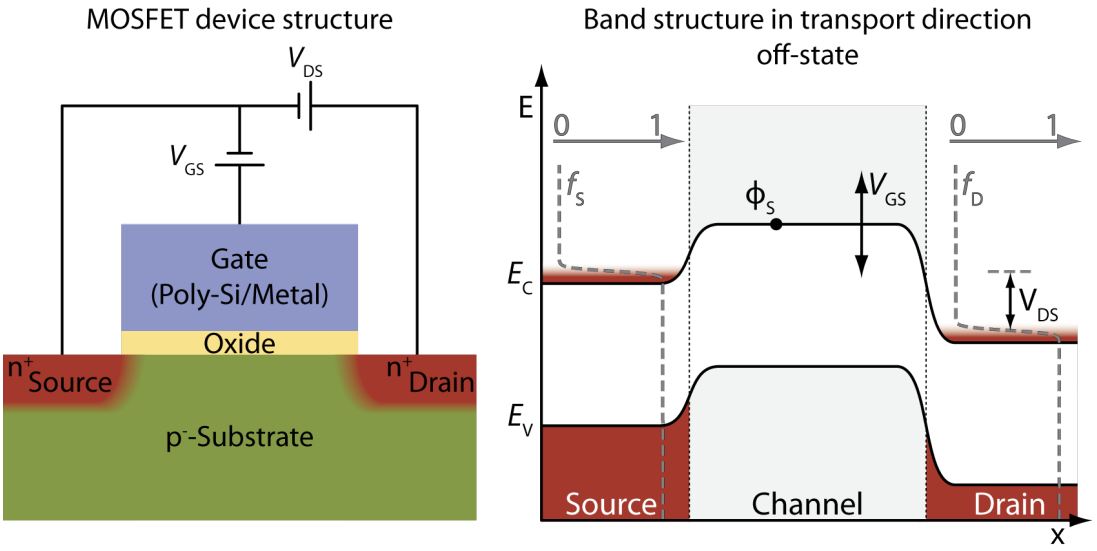


Figure 2.3. Left: Device structure of a MOSFET, *Right*: Band-structure of a MOSFET in the off-state with FERMI-Distributions shown for source and drain.

to reduce the quantum capacitance, which can be accomplished by reducing the density of states of the semiconductor, for example by reducing its dimensionality or by replacing the semiconductor material. The quantum capacitance is directly related to the channel density of states $D(E)$:

$$C_q = q^2 \underbrace{\int_{-\infty}^{\infty} dE D(E) \frac{df}{dE}}_{D_0} \tag{2.3}$$

D_0 is the channel density of states averaged over a few kT around μ in the channel.
Once, the quantum capacitance limit is reached, it can lead to superior scaling behavior of MOSFETs (see reference [3]).

2.1.3. MOSFET operation

Similar to the bipolar transistor, the MOSFET is a three-terminal barrier-controlled device (see Figure 2.3). However, in a MOSFET the barrier is modulated by capacitive coupling to the gate terminal in a MOSCAP configuration. The MOSCAP is complemented by the two heavily doped current terminals source and drain. The identical doping of source and drain is of the opposite type with respect to the low background doping.
Due to the built-in potentials at the source/channel and drain/channel interfaces, current flow is prevented even without any gate bias (normally off-transistor). Under gate bias, the MOSFET is dominated by the three operation modes of the MOSCAP. In accumulation, an increased potential barrier at the source/channel interface blocks all current flow except for a very small off-state leakage.
When V_{GS} drives the channel into depletion, the bands in the channel follow the external potential, the barrier at the source/channel interface is reduced and the source to drain current rises exponentially with V_{GS}. Eventually in inversion, the band movement nearly stops due to the large

number of free carriers in the channel and for $V_{GS}>V_{DS}$, the maximum surface potential ϕ_s in the channel is effectively pinned by μ_S, the chemical potential of the source. In this on-state, the current depends to a power between one and two on V_{GS}. The transition between the exponential and this linear gate voltage dependence occurs at the threshold voltage V_T.

Subthreshold regime

Figure 2.4 depicts the band-profile of an n-MOSFET along the channel below and above threshold. The bands in source and drain are filled according to the respective FERMI-distributions:

$$f_{S/D}(E) = \frac{1}{1 + e^{\frac{E-\mu_{S/D}}{kT}}} \tag{2.4}$$

which are characterized by the electrochemical potentials in source and drain. Only at energies with a non-zero difference of the FERMI occupation probabilities between source and drain, current can flow if it is not blocked by the potential barrier in the channel. This relation was first expressed mathematically by LANDAUER (for a detailed treatment see [4]):

$$I_D = \frac{q}{h} \int_{-\infty}^{\infty} \mathrm{d}E\overline{T}(E)(f_S - f_D) \tag{2.5}$$

where the effect of the potential barrier is included in a general energy dependent transmission

$\overline{T}(E)$. Here, the transmission is evaluated at the top of the potential barrier (ϕ_s) and in case of single-mode transport, it is approximated as zero at all energies below the barrier and one at all energies above. The constant of proportionality is q, the elementary charge, divided by h, the PLANCK constant. This ratio is called the "quantum of conductance" which represents the conductance of an ideal single-mode conductor. With this simple model, that also ignores scattering, the limits of the switching performance of the MOSFET can be investigated. Ideally, in an n-type device in saturation, f_D becomes very small in the range where the transmission is non-zero. The transmission is modeled as a step-function which, assuming the quantum capacitance limit, moves one-to-one with the applied gate voltage V_{GS}. In this case, the inverse slope of the I_D-V_{GS} characteristic is:

$$\begin{aligned} S &= \left(\frac{d\log_{10} I_D}{dV_{GS}}\right)^{-1} = \left(\frac{d\log_{10}}{dV}\int_0^{\infty} dE \frac{1}{1+\exp\left(\frac{E+qV}{kT}\right)}\right)^{-1} \\ &= \left(\frac{d\log_{10}}{dV}\frac{kT}{q}\ln\left[1+\exp\left(\frac{-qV}{kT}\right)\right]\right)^{-1} \\ &= \frac{\ln(10)\cdot\frac{kT}{q}\cdot\ln\left[1+\exp\left(\frac{-qV}{kT}\right)\right]\left[1+\exp\left(\frac{-qV}{kT}\right)\right]}{\exp\left(\frac{-qV}{kT}\right)} \end{aligned} \tag{2.6}$$

While this equation yields the inverse slope for all regions of operation of the model structure, for a real device it is only applicable to the subthreshold regime where the assumption that the channel bands moves one-to-one with the applied gate potential is justified. For the subthreshold region, the minimal inverse slope becomes a constant:

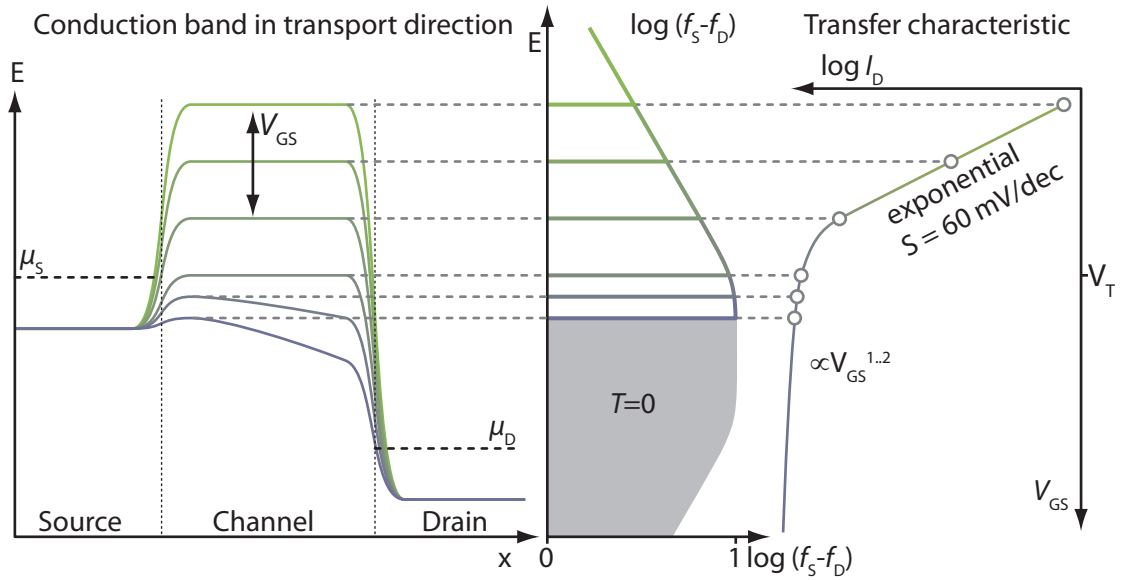


Figure 2.4. Left: Conduction band profile of a MOSFET along the channel. The potential barrier moves according to V_{GS}. Center: Logarithmic plot of the FERMI difference showing that the fermi occupation probability drops exponentially. The grey shading visualizes the varying potential barrier below which no current can flow. Right: Transfer characteristic of a MOSFET that features an exponential increase and a linear to parabolic saturation region.

$$\lim_{V\to\infty}\left[\frac{\ln(10)\cdot kT/q\cdot\ln\left[1+\exp\left(\frac{-qV}{kT}\right)\right]\left[1+\exp\left(\frac{-qV}{kT}\right)\right]}{\exp\left(\frac{-qV}{kT}\right)}\right] \tag{2.7}$$
$$= \ln(10)\cdot kT/q \overset{T=300\mathrm{K}}{\approx} 60\,\mathrm{mV/dec}$$

This result reveals, that the thermal broadening and especially the high-energy tail of the source-FERMI-distribution is responsible for the limited switching slope which is inherent to any barrier-controlled device independent of its scale or material.
A direct consequence of this limitation is that switching between two given current levels I_{on} and I_{off} that represent the two digital states requires a certain minimal voltage range. For example, according to the ITRS (International Technology Roadmap for Semiconductors) 2008 [5], CMOS should provide a minimal I_{on}/I_{off} ratio of ~10^5, for which even in the ideal case at least 300 mV of gate voltage swing is necessary.

On-state
When the gate-voltage has lowered the potential in the channel to the point where the conduction band edge in the channel E_c is lower than μ_S, the chemical potential in the source, the transistor turns on, and a large current emits thermally over the remaining barrier (see Figure 2.5). Depending on V_{DS}, two cases have to be distinguished. For $V_{DS} \ll V_{GS}$, the device is in the "linear regime", that is I_D depends linearly on V_{DS} because in the significant region above E_c, the FERMI-difference f_S-f_D changes with V_{DS}. Any change of the difference of the FERMI probabilities below E_c

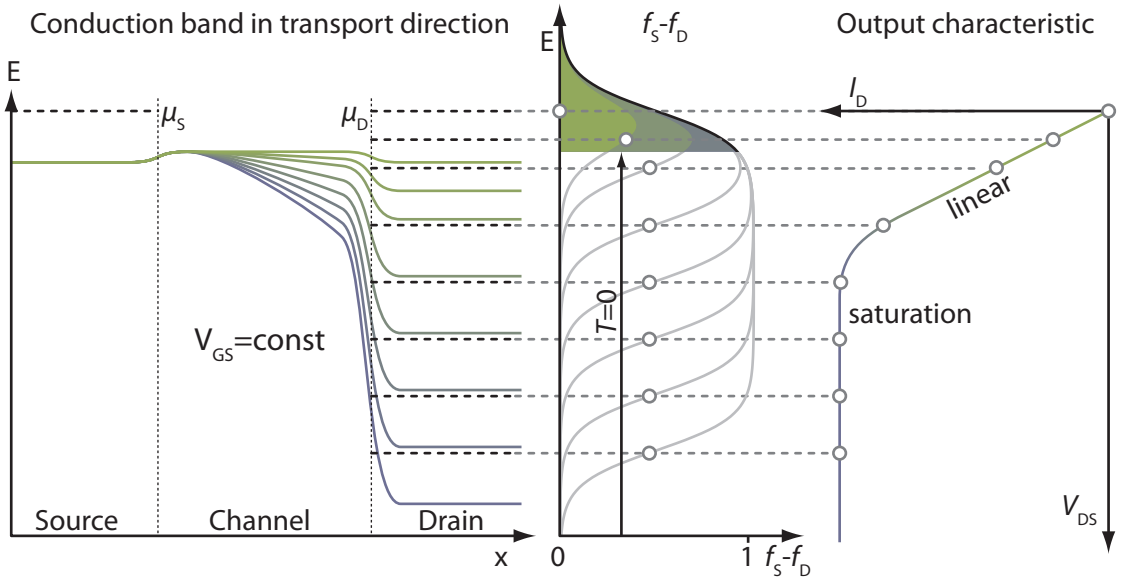


Figure 2.5. Left: Conduction band profile along the channel. The chemical potential of the drain moves with the applied V_{DS}. *Center*: According to the LANDAUER equation, the current is proportional to the convolution of the Transmission T and the difference of the FERMI probabilities f_S-f_D. *Right*: Resulting output characteristic with linear and saturation regions.

does not affect the device current, because the transmission T is zero for this regime. For large V_{DS}, the FERMI-difference still changes, but only in regions with T=0, therefore the current saturates.

Limitations of the simple model

While the model already covers the qualitative behavior of the MOSFET very well, several aspects are not taken into account:

- Dimensionality: The assumption of single mode transport only holds true in very small devices like nanowires. For all other devices, several transverse modes contribute to the current and they can either be treated independently and summed up or if scattering between the modes should be included, a full quantum simulation has to be performed. The number of transverse modes contributing to conduction does not only depend on the size of the channel but also on its material because the effective mass of the charge carriers directly translates into the energy spacing of the modes ($E \propto 1/m^*$).
- Electrostatics: While the quantum capacitance limit and therefore a one-to-one correspondence of ϕ_s to V_{GS} is a reasonable assumption for the subthreshold regime, in the on-state the relation between ϕ_s and V_{GS} depends on the involved capacitances. To date, there are no experimental realizations of devices operating in the QCL. Another aspect that has been neglected is the effect of the source and drain potentials on ϕ_s. Because the band-profile is continuous at both ends of the channel, there are transition regions, where the coupling of the channel potential to the gate is exceeded by the coupling to the source and drain potentials. For small channel length, these transitions regions can overlap which then leads to short channel effects.

- Scattering: For devices with channel lengths L larger than the mean free path length λ_{mfp}, scattering has to be incorporated in the model. A simple way to include scattering is proposed by [6], where a single scattering event in the channel with a transmission

$$T_{scat} = \frac{\lambda_{mfp}}{L + \lambda_{mfp}} \tag{2.8}$$

represents all scattering in the channel. The mean free path of a material is related to the low-field mobility μ^0 of the channel material.

2.2. The Band-to-Band-Tunneling Transistor

2.2.1. Introduction

Although the principle of band-to-band tunneling was already discovered in 1957 by L. Esaki [7] and the first gated *p-i-n* structure was proposed in 1978 [8], the interest in the first results on TFETs [9]-[15] was limited. This changed rapidly after W. Hansch and I. Eisele et al. [16] started to investigate the TFET in 2000 and J. Appenzeller et al. found in 2004 [17] that the TFET might provide a means to overcome the 60 mV/dec switching limit of the classical MOSFET. Following these initial results, several groups started to study the theoretical aspects of TFET operation. Among those were the impact of the channel dimensionality [18], power consumption [19], phonon scattering [20], temperature dependence [21], gate overlap [22], threshold voltage [23], performance comparison to CMOS [24], heterostructure TFETs [25]-[28], strain [29] and general modeling [30]. While Appenzeller's results were obtained with carbon-nanotube FETs, the adoption of the operating principle to silicon FETs seems to be more attractive because the mature silicon technol-

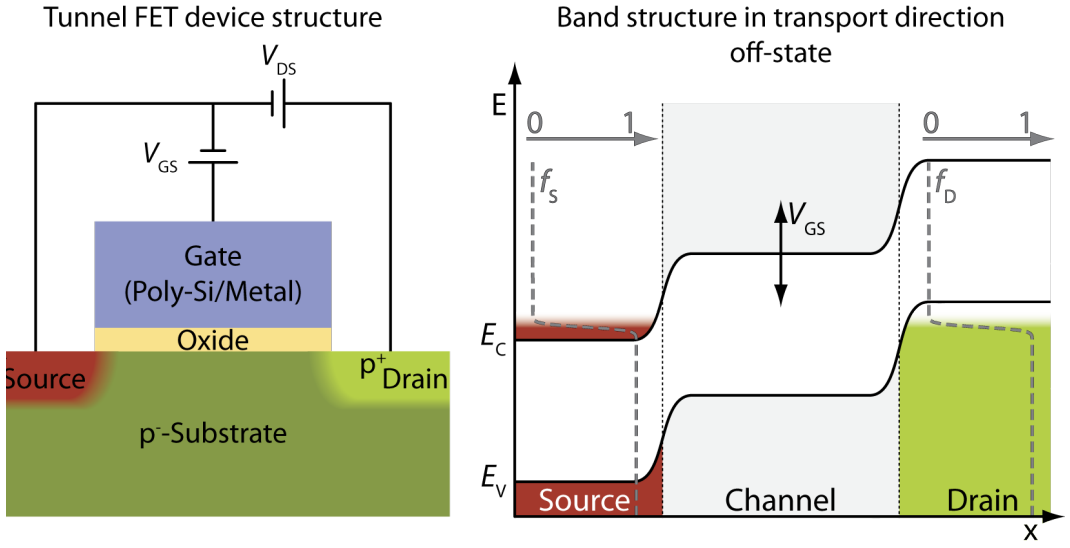


Figure 2.6. Left: Structure of a *n-i-p* Tunnel FET with external voltage sources. *Right:* Band profile of the Tunnel FET in the off-state without applied V_{DS}. The bandgap blocks any current flow between source and drain. The FERMI-distributions in source and drain are plotted in gray.

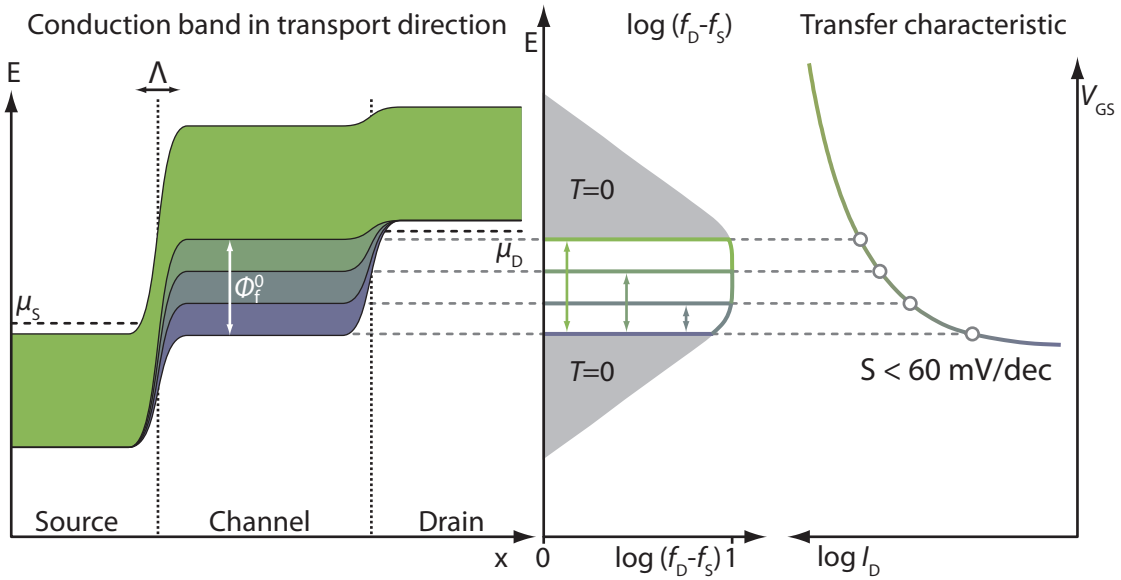


Figure 2.7. Left: Band profile for a TFET in the on-state. In p-type operation, tunneling occurs at the source/channel junction in an energy window determined by the gate voltage. *Center*: Difference of Fermi occupation probabilities of source and drain. The tunneling junction acts like a band-pass filter that only transmits current in the energy range given by the distance of the band-edges. *Right*: Subthreshold characteristics of the TFET. Because of the band-pass filtering, subthreshold swings steeper than 60 mV/dec are possible.

ogy could offer a straight route to industrial application. Therefore, similar to the early work, most recent experiments, including this work, focus on silicon and germanium [31]-[36]. Furthermore, C. Hu proposed a slightly modified TFET structure, which exploits the vertical electric field to create a tunnel junction [37]. To answer the question how the TFET could overcome the 60 mV/dec switching-limit of the MOSFET, the following section introduces its principle of operation.

2.2.2. Principle of operation

The device structure of the TFET resembles that of the MOSFET with one exception. While in the MOSFET, source and drain are doped with the same type of dopant, in the TFET, source and drain are of opposite doping types. The device structure and band profile in equilibrium are drawn in Figure 2.6. In equilibrium, the built-in potentials of the *p-i* and *n-i* junctions result in a staircase-like band-profile. If a small V_{DS} is applied to the equilibrium state, electron and hole currents are blocked by the built-in potential barriers. This is the off-state of the TFET. Now, if a negative gate-bias is applied, the bands in the channel move up. For negative V_{DS}, charge carriers can tunnel through the bandgap at the source/channel junction as soon as the valence band in the channel is lifted above the conduction band in the source. This operating mode is the p-channel on-state of the TFET because holes accumulate in the channel. The n-channel on-state can be created by applying a positive V_{GS} and negative V_{DS}. In this case, the bands in the channel move down and tunneling occurs at the channel-drain junction as soon as the conduction band in the channel is pushed below the valence band in the drain.

For larger negative V_{DS} values, saturation occurs similar to the MOSFET case, but for large V_{DS} values in combination with only moderate V_{GS}, it is even possible to create a potential profile, where tunneling occurs at both junctions simultaneously, resulting in an exponential increase of I_D with V_{DS}. Devices like the TFET, that feature n- and p-channel operation in a single device, are called ambipolar.
To complete this overview of different operation regimes of the TFET, positive V_{DS} bias has to be considered. For sufficiently large positive V_{DS}, the *p-i-n*-structure is forward biased, which means that without applied gate voltage both electron and hole currents can flow and result in exponential diode characteristics. With gate bias applied, either the electron or the hole current can be blocked by a potential barrier but not both at the same time. Therefore this mode of operation cannot be used for switching, though it might be possible to create a negative differential resistance as observed in Esaki-diodes (see [7]).

Subthreshold regime
Since the TFET is based on tunneling rather than thermal emission, the subthreshold behavior differs substantially from that of the MOSFET. Again, equation (2.5) can be used to analyze the subthreshold conduction. This time, in the off-state, the transmission T is approximated to be zero for all energies, but as soon as the transistor is turned on, the transmission becomes significant for the energy range in which tunneling is possible (see Figure 2.7). Thus, the high- and low-energy parts of the Fermi-distribution do not contribute to the current flow, which can be described as a band-pass filter behavior. The prior analysis for the MOSFET illustrates that these high- and low energy regions of the Fermi-distribution are responsible for the limited slope of the MOSFET. In principle, the band-pass filtering leads to much steeper subthreshold swings and therefore enables further down-scaling of the operation voltage. To quantify the subthreshold swing, the simple model for the transmission has to be refined because with the assumption that T jumps discontinuously from zero to one, an infinitely steep subthreshold swing is predicted. In reality, the transmission depends on the band-bending at the source/channel junction. One way to describe T more accurately is to use the Wenzel-Kramer-Brillouin (WKB) (see for example reference [38]) approximation for the band-to-band tunneling (BTBT) process (see eq. (2.9)). For this purpose, the tunnel barrier is described by a triangularly shaped potential barrier whose height depends on the applied gate voltage. The width of the barrier depends on the parameter Λ, the natural length scale for potential variations.

$$T_{BTBT} = \exp\left(\frac{-4\Lambda\sqrt{2m^* E_g^{3/2}}}{2\hbar(E_g + \Phi_f^0(V_{GS}))}\right) \tag{2.9}$$

Other important parameters are the effective mass of the carriers m^* and the size of the bandgap E_g. The gate voltage enters the tunneling transmission indirectly by changing Φ_f^0, the distance between the conduction band in the source and the valence band in the channel. Similar to the MOSFET case, in the subthreshold regime, a one-to-one movement of the channel bands with the applied gate voltage is assumed.
To evaluate the inverse subthreshold slope S, the inverse of the logarithmic derivative of equation (2.5) has to be evaluated. The result is plotted in Figure 2.8 for several silicon TFETs with different Λ. For simplicity, the source potential is set to zero, the chemical potentials μ_S and μ_D are assumed to coincide with the respective band edges and ϕ_S is identified with $-qV_{GS}$. The gate voltage V_{GS} is swept from 0 V to -3 V and a V_{DS}= -3 V is chosen to restrict this discussion to the non-saturated

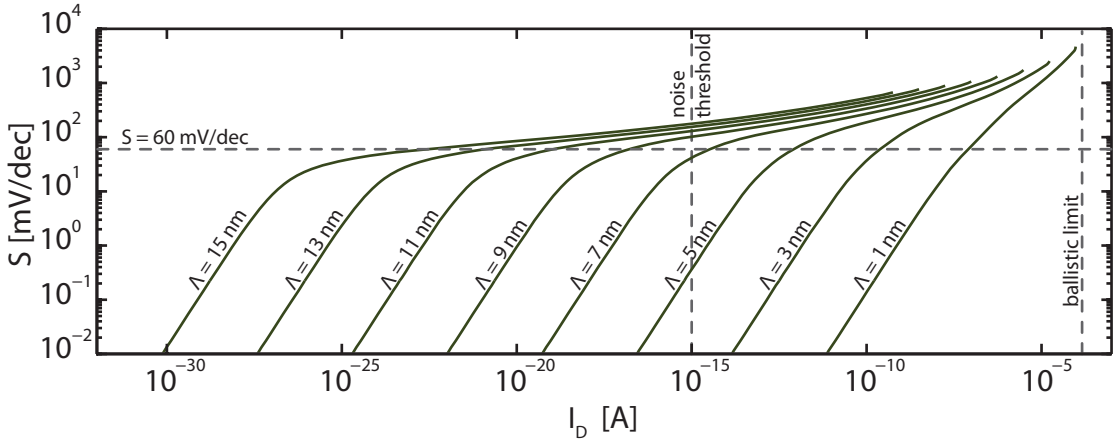


Figure 2.8. Visualization of the implicit dependence of the inverse subthreshold slope S on the drain current and the natural scaling length Λ. Slope and current depend directly on V_{GS}, but this plot makes it possible to evaluate the coincidence of steep slopes and high drain currents.

regime. Instead of plotting S versus V_{GS}, the abscissa shows the drain current because a steep subthreshold swing might be useless if it does not occur in a usable drain current range.
The first thing to note is that in contrast to the constant subthreshold slope of the MOSFET, S for the TFET shows a strong dependence on the gate voltage, which manifests in Figure 2.8 as a drain current dependence. For small currents, S depends linearly on the drain current, but as soon as the slope approaches 60 mV/dec, the gate voltage moves the valence band edge in the channel above the thermally broadened transition region of the drain Fermi-distribution and the slope deteriorates quickly.
Obviously the natural length scale for potential variations Λ is of utmost importance for the device performance and it is the only parameter in equation (2.9) that depends on the device geometry. While the definition and optimization of Λ is discussed in more detail in Chapter 3, at this point, two examples may illustrate the concept. A silicon-on-insulator transistor with a body thickness of 20 nm and an SiO_2 thickness of 4 nm features a Λ of 15 nm. At the other end, a 2 nm thick silicon-nanowire-FET with a wrap-around-gate and an SiO_2 thickness of 1 nm provides a Λ of 1 nm. While with current technology, the first example can be manufactured relatively easily, the realization of the second example might not be realistic.
Since the research on TFETs focuses on the optimization of subthreshold slope, the WKB model is used to evaluate the conditions under which the TFET can act as a steep slope switch (see Figure 2.8). As already argued in equation (2.1), for fast switching, high current levels are necessary. For a device with single-mode transport, the maximum possible current in the absence of scattering is given by the ballistic resistance of 12.9 kΩ. On the other hand, the lowest measurable currents of about 1 fA are limited by the noise of the measurement setup. Therefore, to benefit from several decades of sub-60 mV/dec switching, the natural scaling length Λ has to be smaller than 7 nm in the case of silicon, whose realization requires the use of advanced process technology. To improve the current levels and therefore the performance for a given Λ, it is beneficial to increase the number of conducting modes. While the simplest approach, the use of a two-dimensional channel does increase the number of modes, it also enlarges Λ which compensates the performance

boost. A better way to increase the number of modes while this time maintaining a small Λ is the use of a multi-nanowire array.
The example of the parameter Λ illustrates the general impact of a change in the transmission T_{BTBT}. A decrease of the effective mass m^* or the bandgap E_g result in similar effects. Table 2.1 summarizes effective masses and bandgap energies for common semiconductor materials. For most of the materials, an improvement compared to silicon is expected. Nevertheless, due to its well controllable manufacturing processes, silicon is still the first choice for semiconductor mass products.

On-state
The main difference in the output characteristic of the TFET and the MOSFET is the exponential onset for the TFET (see Figure 2.9) that replaces the linear onset for the MOSFET. This exponential onset is a result of the device structure of the *p-i-n* TFET in combination with the MOSCAP characteristic and it only occurs if C_{ox} is smaller than C_q. For small V_{DS} in combination with large V_{GS}, the channel potential Φ_f^0, does not follow the applied gate potential, but as soon as the valence band in the channel would be lifted above μ_D, holes accumulate in the channel that prevent the channel bands from further movement. Under these bias conditions the channel potential depends heavily on V_{DS}, which therefore also controls the band bending at the source/channel junction. By this mechanism, the drain potential modifies the tunnel probability which manifests as an exponential onset.

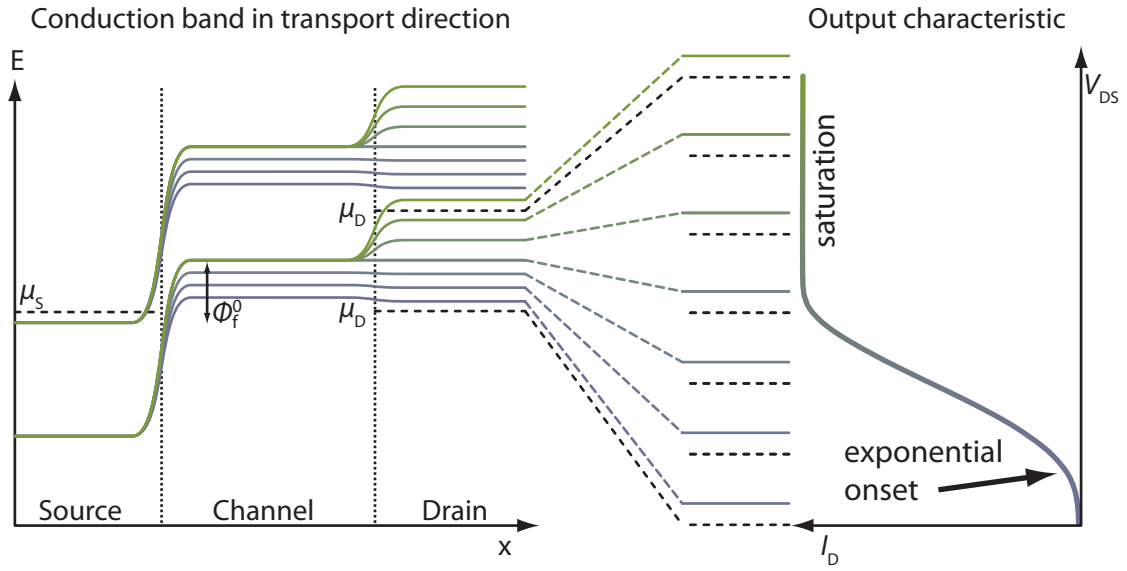


Figure 2.9. On-state of the TFET. For small V_{DS}, the channel potential is pinned by the drain chemical potential. As a result, the tunneling probability at the source/channel junction becomes V_{DS} dependent. For larger V_{DS}, saturation occurs because the energy interval for current flow is controlled only by channel potential.

Material	Bulk Mobility		Bandgap	Effective mass	
	electron	hole		electron	l/h-hole
	[cm²/Vs]	[cm²/Vs]	[eV]	m^*/m_0	m^*/m_0
Si	1400	450	1.12	0.19	0.16 / 0.49
Ge	3900	1900	0.66	0.082	0.043 / 0.33
GaAs	8500	400	1.42	0.063	0.082 / 0.51
InGaAs	12000	300	0.74	0.041	0.026 / 0.41
InAs	40000	450	0.35	0.023	0.026 / 0.41
InSb	70000	850	0.17	0.015	0.015 / 0.43

Table 2.1. Summary of bandgaps, mobilities and effective masses for common semiconductors. For the effective hole masses both, light hole and heavy hole masses are given.

For larger V_{DS}, the chemical potential μ_D is lifted above the valence band edge in the channel and the drain current starts to saturate because the transmission T_{BTBT} becomes independent of V_{DS} and the energy interval in which tunneling is possible is fixed by the source and channel potentials. Furthermore, the difference f_S-f_D becomes independent of V_{DS} because it only changes at energies where the transmission is zero.
Concerning the expected on-state performance of the TFET, again Figure 2.8 is helpful. The saturation-currents at V_{GS}=V_{DS}= -3 V can be identified with the maximum current that is plotted for different Λ. Depending on Λ, the maximum on-currents range from 10^{-9} A to 10^{-4} A which is always smaller than the ballistic limit of $2{\cdot}10^{-4}$ A which could be achieved by a MOSFET. This means that in terms of on-state current, because of the non-ideal transmission, even the most aggressively scaled TFET is inferior to the MOSFET. However, the on-current is not the only and not the most important performance metric. The gate delay, see eq. (2.1), also includes the amount of charge that the on-current has to supply for switching. It has been shown in reference [18] that in case of the quantum capacitance limit, the gate capacitance of the TFET can become very small which also relaxes the necessity for large on-currents. In fact the main result is that in terms of the gate delay the TFET can perform as good as the MOSFET, but in terms of slope and therefore power-consumption it can outperform the MOSFET.

Chapter 3
Modeling of the Tunnel-FET

3.1. A brief primer on NEGF simulations

3.1.1. Choice of simulation approach

In order to go beyond the WKB model of Chapter 2 there are two relevant competing modeling techniques, one is the Monte-Carlo technique (see for example [39] and [40])and the other one is the direct solution of the SCHRÖDINGER and POISSON equations by using the Non-Equilibrium Green's Function (NEGF) (see for example [41] and [42]) approach. Monte-Carlo techniques are most suited for the simulation of long devices where transport is dominated by scattering. The starting point for Monte-Carlo simulations is usually the BOLTZMANN-transport equation which as a classical transport equation does not include quantum effects. Therefore, tunneling models are introduced phenomenologically and are not necessarily based on first principles. However, Monte-Carlo simulations provide excellent descriptions of all particle interactions, i.e. COULOMB-, phonon-, and photon-scattering.
The NEGF approach on the other hand is fully based on first principles and naturally includes quantum effects like quantization and tunneling. However, due to computing limitations, NEGF is usually limited to devices on the nanometer scale and the consideration of scattering and two and three dimensional device structures is even more challenging in terms of computation time. As band-to-band tunneling is a quantum effect which occurs on the nanoscale, NEGF is more appropriate for the description of the TFET, which is the reason why in the following, NEGF simulations for the TFET are presented.

3.1.2. Basic one-dimensional equations

The most basic physical equation to describe the electronic structure of matter is the SCHRÖDINGER equation. Though it neglects spin and relativistic effects its time-dependent form can describe the propagation of non-relativistic electrons in vacuum very well. In the solid-state, however, the sheer number of atoms and electrons renders any exact numerical solution impossible. To enable numerical simulations, some simplifications and approximations have to be made:

- If energy is conserved, i.e. all scattering processes are elastic, the time evolution can be separated and instead of the time-dependent SCHRÖDINGER equation, the time-independent SCHRÖDINGER equation is solved. The remaining problem is to find the energy eigen-values and -states of the HAMILTON operator *H*.
- The periodicity of the crystal lattice is used to reduce the spatial dimension of the lattice to one primitive cell. The corresponding solution of the SCHRÖDINGER equation is called 'BLOCH-wave'. This assumption implies the conservation of the wave-vector $\vec{k}$, which also implies that scattering conserves the crystal momentum. The only quantity left that is not explicitly conserved is phase.
- Electron-electron interactions are neglected, but a self-consistent field approximation is used instead. The self-consistent potential describes the average potential that one electron experiences due to the presence of all other electrons (HARTREE-potential without exchange term). This simplification for example precludes the simulation of COULOMB-blockade.
- When using single atom wave functions, the first three assumptions allow for the calculation of the band-structure of crystals. Though band structure calculations are a very active and important field of research, including the full band structure in transport simulations is computationally very intensive and goes beyond the scope of this work. To simplify the band structure, only one conduction and one valence band are included in the simulation. The shape of the bands in $\vec{k}$-space is approximated by the effective mass approximation, which is a second order polynomial fit to the real band structure. By using the effective mass approximation, the SCHRÖDINGER equation adopts the form of a free particle equation, the only difference being that instead of the free electron mass, the effective mass is used. As another consequence of the effective mass approximation, independent of the real band structure, all semiconductors are modeled with a direct bandgap. That means that the phonon-assisted nature of band-to-band tunneling in indirect semiconductors (Si, Ge, ..) is neglected. However for room temperature simulations, we can assume that the phonon density is very large and does not limit the tunneling rates. According to [20], even the opposite is the case because phonons can exchange energy with electrons and therefore allow for tunneling that is forbidden in ballistic transport, thereby increasing the device currents.

The resulting one-dimensional SCHRÖDINGER equation is:

$$\underbrace{-\frac{\hbar}{2m^*}\frac{d^2}{dx^2}\psi_E + U(x)\psi_E}_{H\psi_E} = E\psi_E \tag{3.1}$$

The first term, which comprises the kinetic energy of the wave-function ψ_E with eigen-energy *E*, only depends on the PLANCK constant and the effective mass. The second term represents the potential energy, here in the form of the self-consistent potential *U(x)*. To determine the self-consistent potential, the second basic equation, the POISSON-equation needs to be solved:

$$\frac{d}{dx}\left(\varepsilon_r(x)\frac{d}{dx}U(x)\right) = -\frac{\rho(x)}{\varepsilon_0} \tag{3.2}$$

This equation states the proportionality of the second derivative of the self-consistent potential to the electric charge density *ρ(x)* divided by the dielectric constant of vacuum ε_0. For later use, the relative dielectric constant $\varepsilon_r(x)$ is assumed to vary spatially.

Length scale of potential variations
In order to describe a two-dimensional transistor channel correctly, in principle a two-dimensional Poisson equation is solved. While this method has to be employed for bulk-transistors, for devices with a limited vertical extent like thin body silicon-on-insulator or nanowire structures an even simpler approach can be used. Under the assumption that the vertical potential varies much slower than the lateral potential, the vertical potential can be approximated by a second order polynomial, with the only parameter being Λ, the effective screening length for lateral potential variations (see [43]). The result is an effective one-dimensional Poisson equation, that incorporates the effect of the gate:

$$\frac{d^2U}{dx^2} - \frac{U - V_{GS} + V_{bi}}{\Lambda^2} = -\frac{\rho}{\varepsilon_r \varepsilon_0} \tag{3.3}$$

Here, V_{GS} is again the gate voltage and V_{bi} is the built-in potential which is due to the work-function difference in the MOSCAP structure. From this differential equation, the importance of Λ for the TFET becomes apparent, since Λ determines the steepness of the potential profile. Λ is calculated differently for different device geometries. For a thin body SOI device, Λ is:

$$\Lambda_{SOI} = \sqrt{d_{Semi}\, d_{Insu}\, \varepsilon_{Semi} / \varepsilon_{Insu}} \tag{3.4}$$

For a coaxial gate-all-around nanowire, however, Λ is smaller:

$$\Lambda_{NW} = \sqrt{\varepsilon_{Semi} / 8\varepsilon_{Insu}\, d_{Semi}^2 \ln\left(1 + 2\, d_{Insu} / d_{Semi}\right)} \tag{3.5}$$

In both expressions, ε_{Semi} and ε_{Insu} represent the relative dielectric constants of the semiconductor and the gate insulator, respectively. d_{Semi} and d_{Insu} are the thicknesses of the SOI or nanowire and the gate insulator.

3.1.3. Finite difference method

Schrödinger equation
Now, that the two basic differential equations needed for transport simulations are specified, in the following the numerical method for solving those equations will be presented.
The first step for a numerical solution of the Schrödinger and Poisson equation is to define a spatial lattice on which the wave-function and the self-consistent potential are defined.

$$\{\psi\} = \begin{pmatrix} \psi_1 \\ \vdots \\ \psi_N \end{pmatrix} = \begin{pmatrix} \psi(x_1) \\ \vdots \\ \psi(x_N) \end{pmatrix} \tag{3.6}$$

Here, $\{\psi\}$ is defined as a vector that represents the wave-function ψ evaluated at the sites of a one dimensional lattice $x_1 \ldots x_N$ with spacing a. The self-consistent potential U is expanded in the same way but for simplicity it is set to a constant value U_0 in the following discussion. In this expansion, the Hamilton operator is represented by a matrix $[H]$ which in its simplest form only regards nearest neighbor couplings:

$$[H]\{\psi\} = E\{\psi\} \tag{3.7}$$

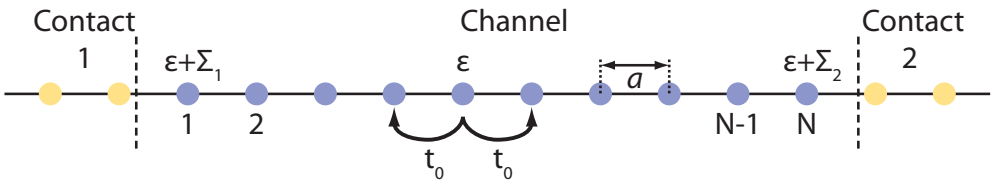


Figure 3.1. One dimensional NEGF simulation domain. All lattice sites with spacing a are connected by the nearest neighbor coupling t_0. The semi-infinite contacts are considered by a self-energy term Σ that adds to the first and last site of the simulation lattice.

with

$$[H] = \begin{pmatrix} \varepsilon & -t_0 & & & \\ -t_0 & \varepsilon & \ddots & & \\ & \ddots & \ddots & -t_0 & \\ & & -t_0 & \varepsilon & -t_0 \\ & & & -t_0 & \varepsilon \end{pmatrix} \tag{3.8}$$

By comparison with the differential form of the stationary SCHRÖDINGER equation, the on-site energy ε and the nearest neighbor coupling strength t_0 can be evaluated to:

$$t_0 = \frac{\hbar^2}{2m^* a^2} \text{ and } \varepsilon = 2t_0 + U_0 \tag{3.9}$$

For a single lattice site, the matrix equation corresponds to:

$$E\psi_n = -t_0\psi_{n-1} + \varepsilon\psi_n - t_0\psi_{n+1} \tag{3.10}$$

In this form, it becomes obvious that only nearest neighbors with the indices n-1 and n+1 are considered in the numerical solution, which is therefore also called a 'tight-binding' approach. To derive an *E-k* relation for the matrix representation of the SCHRÖDINGER equation, a plane-wave ansatz for ψ_n is used.

$$\psi_n = e^{ikx_n} = e^{ikna} \tag{3.11}$$

The resulting *E-k* relation is:

$$E = \varepsilon + 2t_0\cos(ka) \tag{3.12}$$

This solution represents a cosine-shaped band, which for small *ka* resembles the parabolic E-k relation of the free-electron SCHRÖDINGER equation:

$$E = \varepsilon + 2t_0 - t_0^2 a^2 k^2 = \frac{\hbar^2 k^2}{2m^*} + U_0 \tag{3.13}$$

While in vacuum the parabolic *E-k* relation is exact, in the solid-state the cosine-shaped band of equation (3.12) is a more accurate description, because it takes into account the finite width of the bands as well as the different curvatures at the top and the bottom of the band.

Poisson equation
Following this example of the Schrödinger equation, the Poisson equation, which is also a second order differential equation, can be turned into a matrix equation in a similar fashion, here for constant ε_r:

$$\frac{1}{a^2}[U(x_{i-1}) - 2U(x_i) + U(x_{i+1})] = -\frac{\rho(x_i)}{\varepsilon_r \varepsilon_0} \tag{3.14}$$

This matrix representation makes the numerical solution of those equations very simple and automatic. For any arbitrarily shaped potential, the only challenge is to write down the corresponding matrix. The actual solution is reduced to an eigen-value problem which can be computed easily.

3.1.4. Boundary conditions – Schrödinger equation

So far, the finite-difference scheme that has been described suffers from one limitation. It is restricted to a lattice of finite dimension, which means that the simulation domain is encompassed by impenetrable regions that the wave-function cannot enter. While this behavior is appropriate for most problems involving a potential well, for transport in semiconductors it is unacceptable. Instead the concept of semi-infinite leads, which assumes that the simulation domain ('channel') continues infinitely on one or both ends, is very useful. To include semi-infinite domains, without at the same time increasing the dimension of the corresponding matrices to infinity, the 'self-energy' method has been developed (see Figure 3.1).
To illustrate the concept of 'self-energy' Σ, we assume a channel with only one lattice point $n = 0$ and a semi-infinite contact that is attached on the left site: $n = -1 \ldots -\infty$:

$$E\psi_0 = \varepsilon\psi_0 - t_0\psi_{-1} \tag{3.15}$$

The semi-infinite contact itself fulfills equation (3.10) and with an ansatz that is similar to equation (3.11), ψ_{-1} can be expressed as:

$$\psi_{-1} = e^{ika}\psi_0 + B(e^{-ika} - e^{+ika})$$

While the origin of the first term can be understood immediately from eq. (3.11), the second term is a result of a possible reflection at the contact/channel interface. Only a fraction B of the incident wave is transmitted into the channel. Substituting this expression into (3.15) shows that the effect of the contact can be included in the on-site energy ('self-energy') of the channel lattice site and a source term S that describes the excitation of the channel by an incident wave from the contact:

$$E\psi_0 = (H + \Sigma)\psi_0 + S \text{ with } \Sigma = -t_0 e^{ika} \text{ and } S = it_0 2B\sin(ka)$$

The main consequence of adding the 'self-energy'-term to the Hamilton operator is an open boundary condition. Therefore, any wave-package acquires a finite lifetime, i.e. it can escape the simulation domain and if it escapes, it never returns. In the energy domain, the finite lifetime causes a broadening of formerly discrete energy-levels. The broadening of the levels, or the inverse life-time of a state, is given by the broadening function:

$$\Gamma(E) = i[\Sigma(E) - \Sigma^{\dagger}(E)] \tag{3.16}$$

3.1.5. Boundary conditions – Poisson equation

While for the Schrödinger-equation open boundary conditions are favorable, the solution of the Poisson-equation should fulfill a zero field condition, i.e. the first derivative of the self-consistent potential should vanish (Neumann condition). This condition is necessary to maintain a smooth continuation of the potential into the contacts, which at the same time ensures that no charge builds up at the boundary. Mathematically, the zero field condition is expressed as follows:

$$U(x_{-1}) - U(x_0) = U(x_{N+1}) - U(x_N) = 0$$

3.1.6. Local Density of States

One of the central properties that the solution of the Schrödinger equation should yield is the charge density. In analogy to the classical calculation of the charge density, we first calculate the Local Density of States (LDOS) and in a second step 'fill' these states according to an equilibrium Fermi-distribution. The LDOS weights each eigen-state by the square of its wave-function at a location *r*:

$$D(r,E) = \sum_{\alpha} \left| \psi_{\alpha}(r) \right|^2 \delta(E - \varepsilon_{\alpha}) \tag{3.17}$$

This quantity can be identified with the diagonal elements of the spectral function *A(E)*.

$$A(r,r',E) = 2\pi \sum_{\alpha} \psi_{\alpha}(r)\, \delta(E - \varepsilon_{\alpha})\, \psi_{\alpha}^{*}(r') \tag{3.18}$$

The off-diagonal elements of *A(E)* can be interpreted as transition probabilities between to positions *r* and *r'* at a given energy *E*. This interpretation is the starting point to calculate device current from transmission, which is essentially given by the off-diagonal elements of *A(E)*. To evaluate the spectral function, it is convenient to make use of the (retarded) Green's function:

$$G(E) = \left[(E + i\eta)I - H\right]^{-1} \tag{3.19}$$

In this definition, *I* is the identity matrix and η is an infinitesimally small quantity that ensures causality. The Green's function acts as an 'inverse' of the Hamilton operator. For example, inhomogeneous solutions of the Schrödinger equation can be calculated from the Green's function and the source term. Using the Green's function, the spectral function becomes:

$$A(E) = i\left[G(E) - G^{\dagger}(E)\right] \tag{3.20}$$

3.1.7. Charge density

While the local density of states is a very valuable quantity to understand device operation, it is the charge density that is the necessary input to solve the Poisson equation self-consistently. Starting with the spectral function *A(E)*, the calculation of the charge density for a single reservoir in equilibrium follows the lines of the classical calculation. The charge density ρ equals the integral of the product of the density of states and the Fermi-distribution of the reservoir.

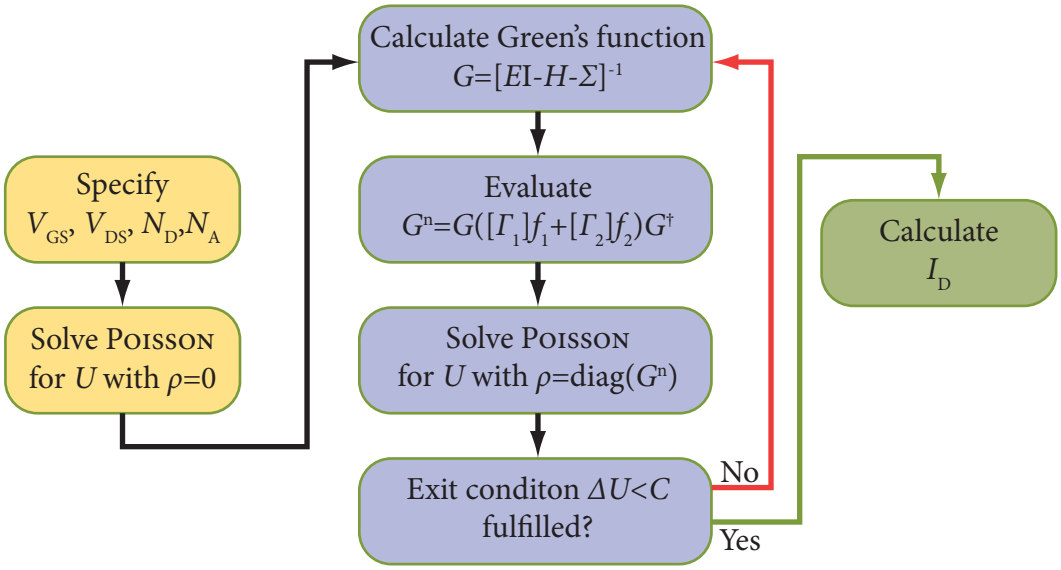


Figure 3.2. Flow chart for the self-consistent solution of the Schrödinger and Poisson-equations

$$[\rho] = \int_{-\infty}^{\infty} \frac{dE}{2\pi} f_0(E-\mu)[A(E)] \tag{3.21}$$

In a transistor device, however, the channel is in a non-equilibrium state, therefore equation (3.21) cannot be used. Instead, the spectral function $A(E)$ is split into several spectral functions $A_1(E)$... $A_N(E)$ that represent the subset of states which is accessible by the contacts 1...N.

$$A_n(E) = G\Gamma_n G^\dagger \text{ and } A(E) = \sum_i A_i(E) \tag{3.22}$$

Therefore, in the non-equilibrium case, the charge density is given by:

$$[\rho] = \int_{-\infty}^{\infty} \frac{dE}{2\pi} \sum_i f_0(E-\mu_i)\,\mathrm{diag}[A_i(E)] \tag{3.23}$$

At this point, it is important to be aware of the fact that this charge density [ρ]is one-dimensional. The charge density that enters the Poisson equation, however, is three-dimensional. For example, to calculate the three-dimensional charge density for a nanowire, the one-dimensional density has to be divided by width and thickness.

Another aspect that has to be taken into account is spin. As without a magnetic field, electronic states are two-fold spin-degenerate, even single mode transport includes two modes, one for each spin state. This is accounted for by including another factor of two in the charge density. A more generalized version of the charge density is the density matrix or correlation function G^n.

$$[G^n] = 2\pi\psi_E\psi_E^\dagger = \sum_i [A_i]f_i \tag{3.24}$$

From a quantum-field theory perspective, the correlation function describes the correlation of subsequent annihilation and creation of electrons at two positions.

3.1.8. Current

The most important observable quantity in a semiconductor device is the current. Though there are several ways to calculate the current within the NEGF formalism, we continue to use the Landauer equation (2.5). The only parameter that needs to be redefined is $\overline{T}(E)$, the transmission. It already has been mentioned in Section 3.1.6 that the off-diagonal elements of the spectral function include information on the transition probability. The complete definition is:

$$\overline{T}_{i\to j}(E) = \mathrm{Trace}\left[\Gamma_i G \Gamma_j G^\dagger\right] = \mathrm{Trace}\left[\Gamma_i A_j\right] = \mathrm{Trace}\left[\Gamma_j A_i\right] \tag{3.25}$$

At the element level, because each Γ has only one non-zero element, only the off-diagonal elements of A contribute to the transmission. The current at a single contact therefore is:

$$I_j = {}^{q}\!/_{h} \sum_{i\neq j} \int_{-\infty}^{\infty} dE\, \overline{T}_{i\to j}\left[f_i(E) - f_j(E)\right] \tag{3.26}$$

To account for spin degeneracy, again a factor of two has to be added.

3.1.9. Self-consistent calculation

At this point, all essential building blocks of NEGF calculations have been described, therefore this section focuses on how to combine these blocks into a self-consistent calculation. The overall procedure is shown in Figure 3.2.

The only step which has not been discussed yet is how the Poisson-equation is used to calculate a correction factor for the self-consistent potential. Because the straightforward approach that just solves Schrödinger and Poisson-equation alternatingly leads to poor convergence, a Newton-Raphson iteration (see Appendix C of [44])scheme is used instead. The Newton-Raphson algorithm usually converges sufficiently in less than ten iterations. One requirement to employ the Newton-Raphson algorithm is that the derivative $d\rho/dU$ can be calculated. This work follows the approach of Roger Lake et al., who propose in [44] to use a classical model for the carrier density based on Fermi-statistics to evaluate the derivative analytically. This derivative is calculated separately for electrons and holes, where the following expression applies to electrons:

$$\frac{d\rho}{dU} = \frac{2q}{kT}\underbrace{\left(\frac{m_e^* kT}{2\pi\hbar^2}\right)^{3/2}}_{N_c} \mathcal{F}_{-1/2}\left(\frac{\mu - U}{kT}\right) \tag{3.27}$$

It depends on the effective density of states of the conduction band N_C and the Fermi-Dirac-integral $\mathcal{F}_{-1/2}$ of the order -1/2, which is the derivative of $\mathcal{F}_{1/2}$, that describes one-dimensional carrier statistics.

The exit condition for the self-consistent iterations is that the change of the self-consistent potential ΔU is smaller than a given value, which is typically of the order of 50 meV.

3.2. Extension of the basic NEGF formalism

3.2.1. Two-band model.

The way the NEGF formalism was presented so far only allows for the calculation of transport in a single band. TFET operation, however, relies on band-to-band tunneling which means that at least two bands have to be present. The textbook approach to include a second band is to use a pair of wave functions at each lattice site, similar to the way spin is included in the PAULI-equation.While this approach has the benefit of being extensible to a full-band model, where all relevant bands are included with their appropriate effective masses, the computational load increases tremendously. Therefore, this work follows another approach, which makes use of a one-dimensional, one-band basis. To nevertheless include the effect of the bandgap and the second band, the dispersion relation becomes energy dependent. This concept of an energy dependent dispersion relation was introduced by FLIETNER [45]:

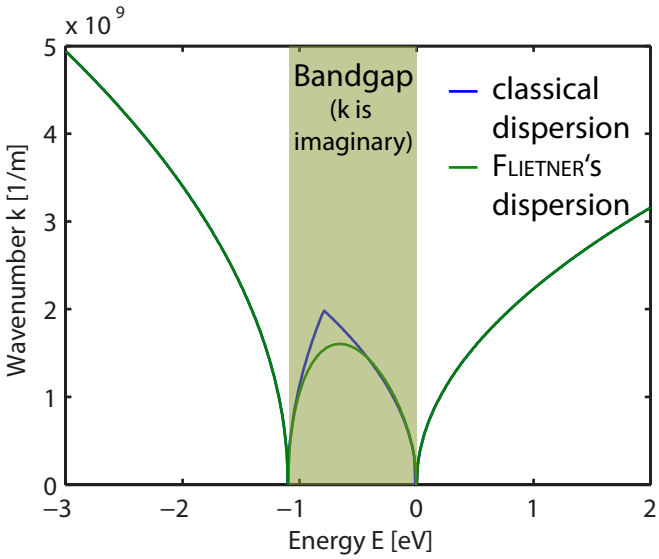


Figure 3.3. Comparison of the classical dispersion relation with FLIETNER's version that provides a smooth transition from conduction to valence band.

$$\frac{\hbar^2 k^2}{2m^*_{c/v}} = E \qquad \begin{cases} E > E_c \\ E < E_v \end{cases}$$
$$\frac{\hbar^2 k^2}{2m^*_{c/v}} = E\left(1 - E/E_g\right)\left(1 - \left(1 - \sqrt{m^*_{c/v}/m^*_{v/c}}\right)E/E_g\right) \qquad \begin{cases} E_{br} < E < E_c \\ E_{br} > E > E_v \end{cases} \tag{3.28}$$

While k is real in the conduction and valence bands, in the bandgap k acquires an imaginary value that acts as a damping factor for the wave function. The transition from conduction to valence band behavior occurs at the charge neutrality point E_{br}. Though it would be possible to describe the bandgap by using the same dispersion relation as in the bands, a smooth transition between both bands would not be possible in this case because the first derivative at the charge neutrality point is discontinuous. The given dispersion relation (see Figure 3.3) avoids this undesired behavior and is known to describe band-to-band tunneling accurately [46].

The energy-dependent dispersion relation enters the NEGF calculation by the modification of the coupling parameter t, which therefore becomes energy dependent, too. Though the HAMILTON-operator is defined for a single energy only, different t can contribute to the same *[H]* because the potential energy U might vary along the lattice. Special care has to be taken in order to ensure

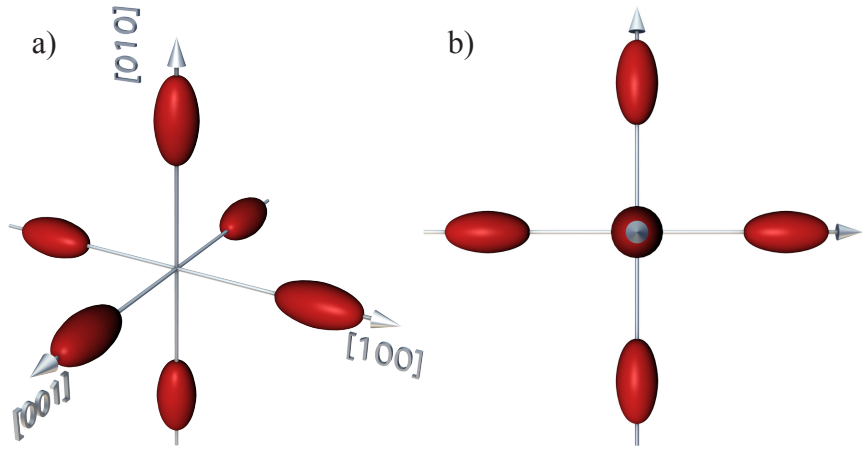


Figure 3.4. a) Effective mass ellipsoids for the 6-fold degenerate Si conduction band. b) The same dispersion relation viewed from the [001] direction.

the hermiticity of *[H]* at points where *t* changes. The correct, hermitian version of [*H*] for two different coupling parameters t_1 and t_2 is:

$$[H] = \begin{bmatrix} \ddots & \ddots & 0 & 0 & 0 \\ \ddots & U+2t_1 & -t_1 & 0 & 0 \\ 0 & -t_1 & U+t_1+t_2 & -t_2 & 0 \\ 0 & 0 & -t_2 & U+2t_2 & \ddots \\ 0 & 0 & 0 & \ddots & \ddots \end{bmatrix} \tag{3.29}$$

3.2.2. Silicon band structure

The first task at hand is the simulation of silicon nanowire devices. Consequently, the specifics of the silicon band structure have to be included in the model. Figure 3.4 (a) shows six constant energy surfaces near the conduction band minimum that lead to a six-fold degeneracy of the conduction band in bulk material. Because of the parabolic dispersion relation, the width of the ellipsoids corresponds to the effective mass in the respective direction. In nanowires, the six fold degeneracy is lifted because two dimensions are confined. To understand the effect of confinement, it is important to distinguish between those effective masses that are responsible for the confinement energy and those responsible for transport. For transport in the [001] direction, the constant energy surfaces are sketched in Figure 3.4 (b). If the [100] and [010] directions are confined, the confinement energy for ellipsoids on the [001]-axis is given by the transverse effective mass for both confinement directions. The other four ellipsoids have smaller confinement energies, because for those, the larger longitudinal effective mass contributes, too.
Consequently, the six-fold degeneracy turns into a four-fold degeneracy, and the transport mass is given by the transverse effective mass, because the ellipsoids that contribute with the longitudinal effective mass are not occupied since their quantization energy is larger. For density of states and charge density calculations, the degeneracy is four-fold as well, but the effective mass used in this case is the density of states effective mass.

The case of the valence band is simpler, because the contributing bands are isotropic and their confinement mass equals their transport mass. Therefore, the degeneracy is lifted and only the heavy-hole band contributes to the current flow and density of states.

3.2.3. Two and three-dimensional structures

What has been neglected so far is the fact that most semiconductor devices do not provide one-dimensional transport. Therefore the questions arise under what conditions does one-dimensional transport occur and how can multi-dimensional transport be included in the NEGF formalism. Qualitatively the first question can be answered by using a 'toy'-model of quantization. For three-dimensional structures, there is no quantization, which results in continuous bands in all three directions. As soon as one dimension is confined, size quantization occurs which leads to a subband structure in that dimension. The subband splitting can be calculated by a simple 'particle-in-a-box' model for a box with length L, where the states are numbered with the quantum number n.

$$E_n = \frac{\hbar^2 \pi^2}{2m^* L^2} n^2 \tag{3.30}$$

A simple criteria for one-dimensional transport is E_2-$E_1 > 5\,kT$. For silicon, this condition is met at a width $L \approx 4\,\mathrm{nm}$, which means that for larger confinement width, more than one subband contributes to the current flow. For two-dimensional transport, the second transverse dimension, which is not confined leads to continuous bands. Only if both transverse dimensions are confined, for example in a nanowire, single-mode transport occurs which can be described accurately with the model developed so far.

To extend the model to two dimensional transport, the continuous band in the transverse direction has to be included. Usually this is done by introducing a two-dimensional lattice, which increases the dimension of all involved matrices by a power of two. As a consequence, the computation time also increases by a power of two, therefore even relatively small lattices (e.g. 100 points) become unreasonable to solve on desktop computer systems. One way to reduce the computation time is to neglect interaction between states with different transverse crystal momentum. In this case, the transverse states form independent parallel channels for each k_T-value that can be summed up. If only the conduction or valence band are included in the simulation, there is another trick to reduce the computation time. Instead of calculating the LDOS for each k_T-value and then filling those according to a one-dimensional Fermi-distribution, only one LDOS for k_T=0 is calculated and filled according to a two-dimensional supply function (Fermi-Dirac-integral $\mathcal{F}_1$). However, this approach fails in combination with the two-band model that was introduced in the previous section, because in the calculation of charge transport it does not respect the conservation of energy and crystal momentum at the same time.

3.3. The 1D Si TFET at a single bias point

Using the NEGF model that has been summarized in the previous two sections, the fundamental simulation results for a one-dimensional Si nanowire TFET are presented in this section. The basic parameters used for the simulation are given in Table 3.1.

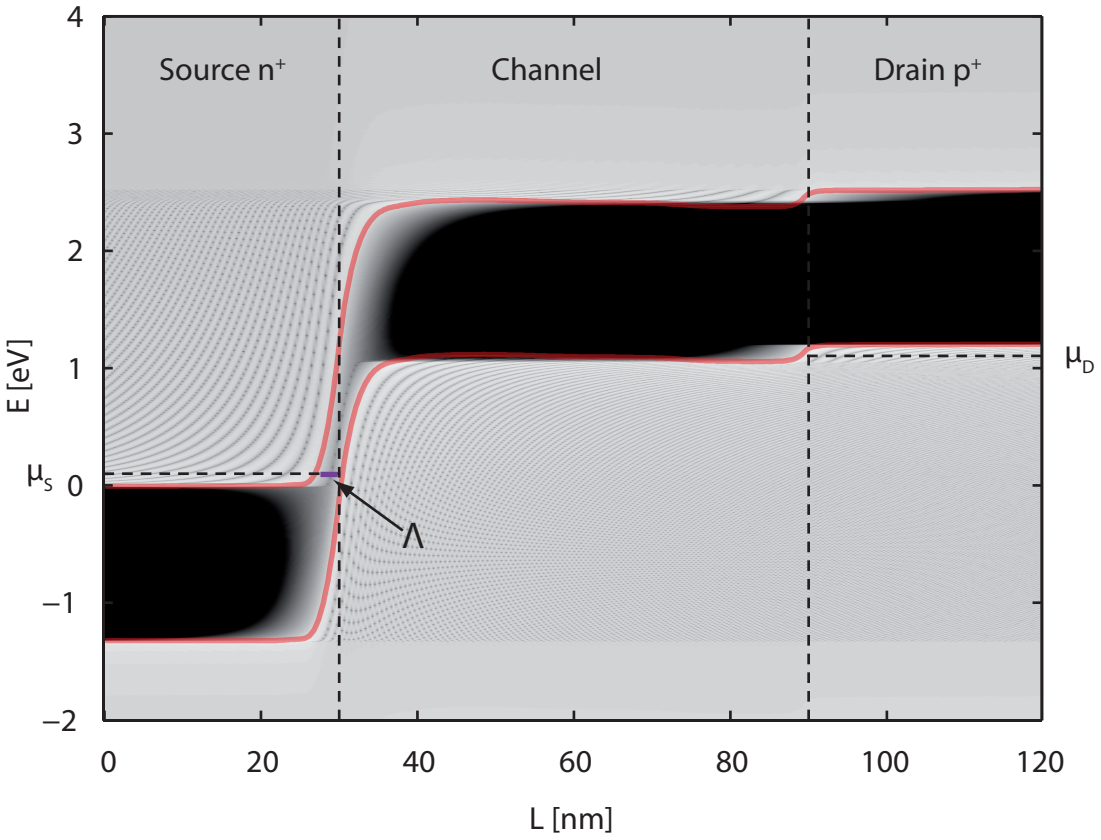


Figure 3.5. Gray-scale plot of the local density of states (white represents large local density of states, black small in arbitrary units) in transport direction of a rectangular Si nanowire TFET at V_{DS}=-1 V and V_{GS}=-2.5 V. The red lines depict the underlying potential profile.

Dimensions		**Physical parameters**	
Wire length	120 nm	Effective electron mass	$0.19 \cdot m_0$
Channel length	60 nm	Effective hole mass	$0.49 \cdot m_0$
Source/Drain length	30 nm	Bandgap energy	1.1 eV
Wire width	5 nm		
Wire height	5 nm	**Numerical parameters**	
Physical parameters		Total number of nodes	400
Temperature	300 K	Nodes in the channel	200
Source FERMI-level	0.1 eV	Nodes in Source/Drain	100
Drain FERMI-level	0.1 eV	Lattice spacing	3 Å
Λ	2.2 nm	# of energy grid points	4096

Table 3.1. Simulation parameters for the one dimensional rectangular nanowire described in this section.

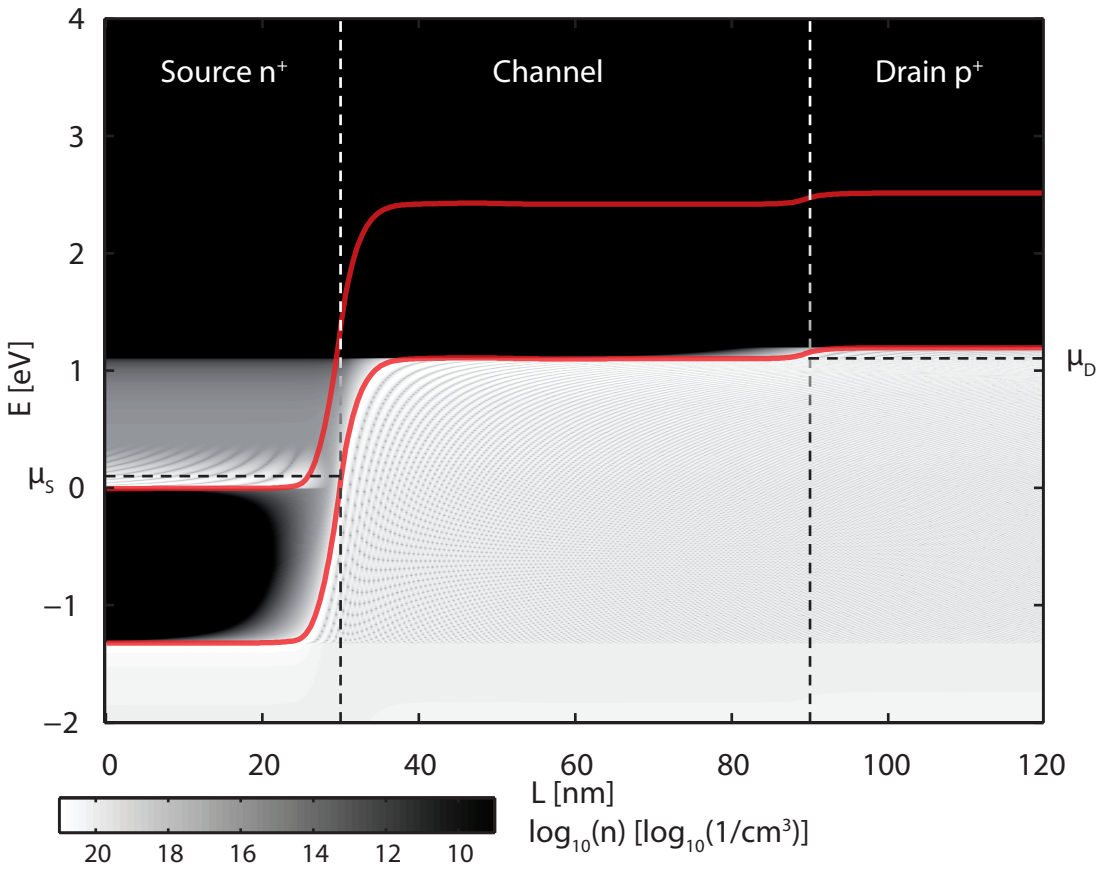


Figure 3.6. Gray-scale plot of the energy resolved local electron density.

The width and height of the nanowire are chosen such that the subband spacing is larger than 5 kT. Therefore, at least for small V_{DS} single-mode transport is guaranteed. The Λ-value corresponds to an SiO_2 thickness of about 1 nm and together with the small wire dimensions and ballistic transport guarantees that the nanowire TFET operates in the quantum capacitance limit (see discussion in Section 3.5). The doping level is specified by the FERMI-level, which equals the distance of the chemical potential to the respective band-edge. A positive value means that the chemical potential is located within the band and therefore degenerate doping is assumed.

3.3.1. Local density of states

The local density of states for the given structure at V_{DS}=-1 V and V_{GS}=-2.5 V is calculated using eq. (3.20) and is shown in Figure 3.5 . In this figure, the conduction band edge in the source is set to ground, while the electrochemical potential μ_S is fixed at 0.1 eV. The dark band which represents a small LDOS displays the bandgap of the material. The regions below the source valence band edge and above the drain conduction band edge exhibit the undisturbed density of states of a one-dimensional conductor. However, the conduction band in the source and the valence band in the drain feature interference patterns that are a result of the incoming 'particle-waves' interfering with their reflections off the potential barriers.

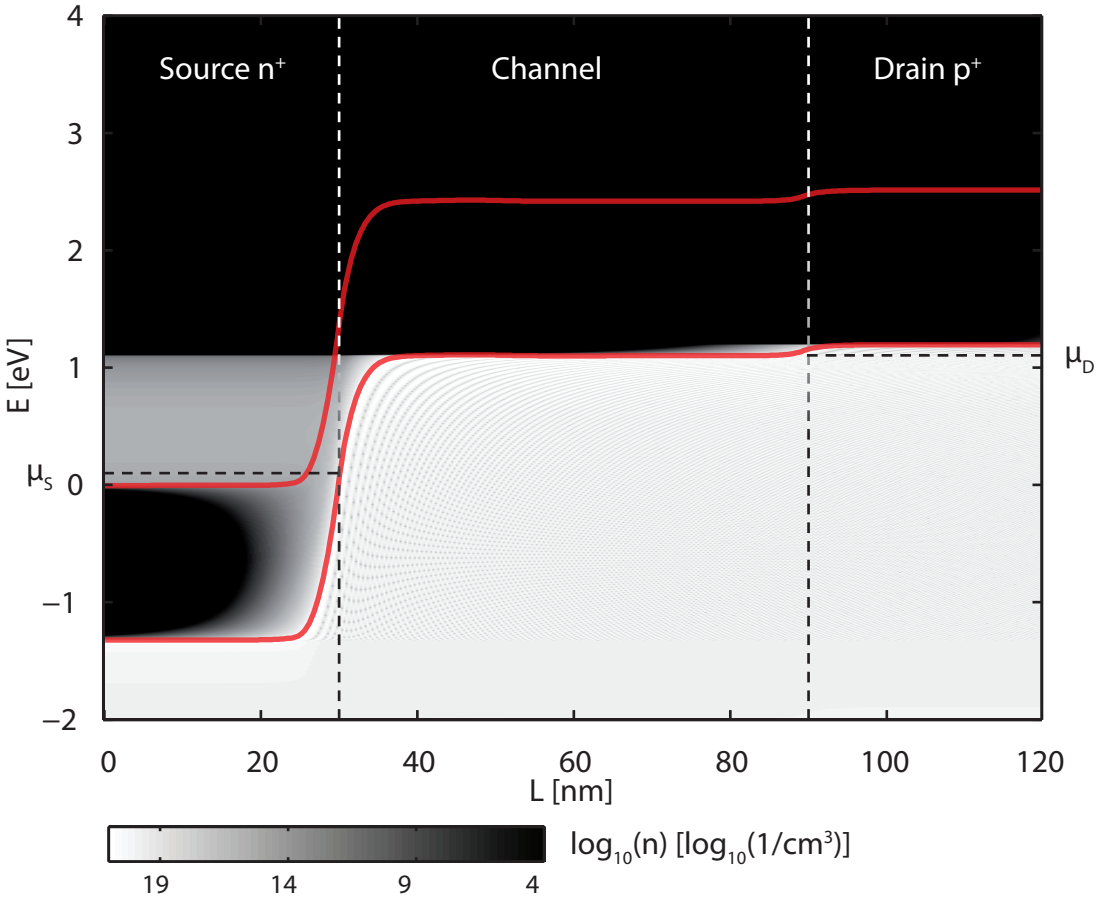


Figure 3.7. Electron density originating from the drain contact. The electron density in the conduction band of the source is a result of band-to-band tunneling.

For the given bias conditions, the TFET is in the on-state. This fact shows up in the LDOS at the source/channel junction, where the LDOS in the bandgap is significantly enhanced in comparison to the other parts of the device. In the energy range between 0 eV and 1 eV, the LDOS of the conduction and valence bands come very close because the wave-functions penetrate into the bandgap. It is in this energy range, where tunneling occurs and Λ (see Figure) is a good measure for the tunneling distance close to the source chemical potential. Yet, it is also obvious that the tunneling distance is not constant over the energy range where tunneling is possible. For example, close to the conduction band edge in source, the tunneling distance is larger than at an energy of 0.5 eV. This is a consequence of the smooth transition between the horizontal bands in the source and the nearly vertical bands at the junction. Furthermore, this observation demonstrates a first inaccuracy of the WKB approximation that was presented in Chapter 2.

3.3.2. Charge carrier densities

Energy resolved local charge density
Although studying the local density of states is very instructive to verify the correctness of the simulation, the density of charge carriers is more important because it can prove that the simu-

lation indeed captures tunneling. Furthermore, the charge carrier density is needed as input to solve the Poisson-equation and achieve self-consistency. Figure 3.6 shows the electron density as a function of position and energy. As expected, the valence band is fully populated while the conduction band is only filled in the n-doped source. The total free charge which is needed as input to the Poisson equation can be calculated from Figure 3.6 by integration over the energy. However, it is important to only include the electron density in the conduction band, while in the valence band only holes contribute to the free charge density. According to eqn. (3.24), the total electron density is the sum of those states populated by the source and those states populated by the drain. Now, one way to check for tunneling is to study if the states that are filled by the drain extend into the conduction band in the source. Figure 3.7 presents the partial electron density that is originating from the drain reservoir. Clearly, the source conduction band is partly filled by the drain and the carrier density is modulated by the tunneling probability.

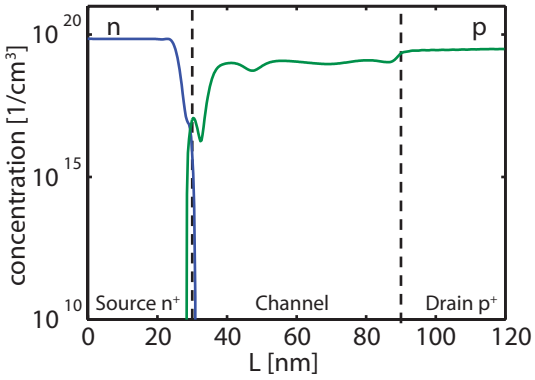


Figure 3.8. Total electron and hole concentration as a function of position. The local maximum of the hole concentration at the source/channel interface is due to holes injected from the source.

Total charge density

Figure 3.8 shows the integrated electron and hole densities for the device at hand. The first point to note is that the carrier densities in source and drain are constant and reflect the doping of these regions. Still, the values of the electron and hole densities differ, which is a result of the way doping is introduced in the device. The parameter that determines the doping level is the distance between

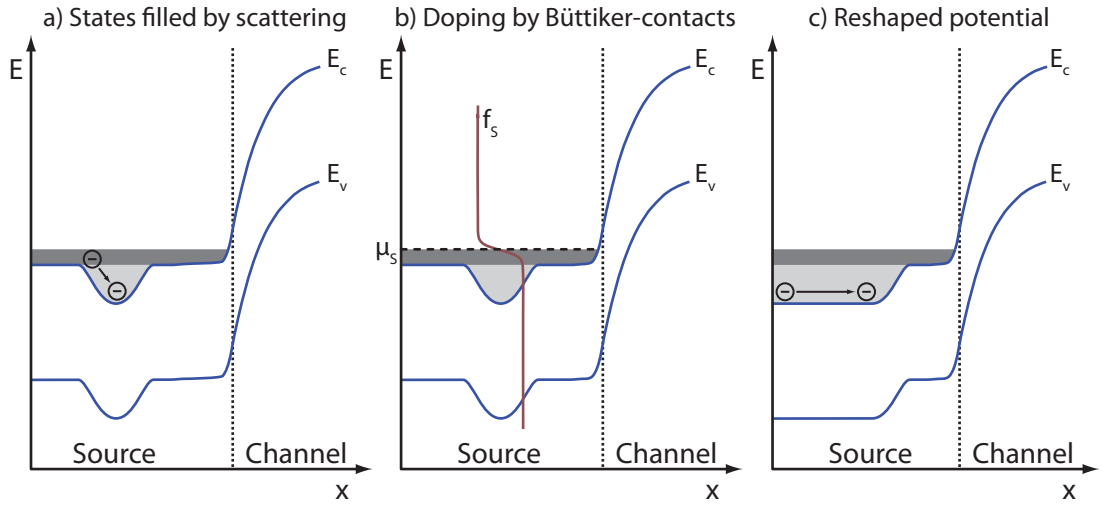


Figure 3.9. Comparison of different approaches that fix the 'potential-pocket' problem which can occur in self-consistent simulations.

the band-edge and the chemical potential, which is set to 0.1 eV for both contacts. However due to the different density of states which reflect the different effective masses of electrons and holes and the different degeneracy of the conduction and valence band, the doping concentrations corresponding to this FERMI level are different. The hole density in the channel is smaller than in the drain because of the built-in potential drop at the drain/channel junction. This potential drop pushes the band below μ_D, which reduces the number of states in the channel that are available around μ_D. Near the source/channel junction, we note an unexpected local maximum of the hole concentration which is caused by band-to-band tunneling, similar to the behavior observed for electrons, which spill through the bandgap (see Figure 3.7). This means, some conduction band states from the source, filled according to μ_S, extend into the valence band in the channel and therefore contribute to the total hole density.

How to achieve self-consistency

All results that are presented in this chapter are calculated in a self-consistent simulation. Although the general procedure as described in Section 3.1.9 is correct, some properties of the model can lead to divergence. The main reason for convergence problems is that the semi-classical model that is used to calculate $d\rho/dU$ does not reflect the way charge density is created in a ballistic device. In a ballistic simulation, the only sources to populate the density of states are the equilibrium reservoirs that the device is connected to. Scattering or thermal generation and recombination are not included. Therefore, a potential well that might form during the self-consistent iteration is not filled by charge carriers and does not lead to an increase of charge density. Yet, this behavior contradicts the semi-classical model for $d\rho/dU$ and leads to divergence at the potential well.
There are several ways to deal with this problem:

- ***Include scattering in the simulation (see Figure 3.9 a)***. When scattering is included in the simulation, a potential pocket will be populated by charge carriers that lost energy due to scattering. While this approach even improves the accuracy of the model, it increases the computation time tremendously for two reasons. First, scattering introduces another layer of self-consistency to the calculation, which increases the number of iterations by a factor of about ten for each bias point. Second, without scattering not all elements of the GREEN's function have to be computed and the number of needed elements scales linearly with the number of nodes. However, if scattering is included all elements of the GREEN's function are necessary, thus their number scales quadratically with the number of nodes.
- ***Model doping by using BÜTTIKER-contacts (see Figure 3.9 b)***. The issue of potential pockets only leads to divergence in the doped source and drain regions. Therefore, it is possible to populate the unfilled density of states by using BÜTTIKER-contacts, which inject or extract charge carriers according to a fixed FERMI-distribution. Yet, although this method does not introduce another layer of self-consistency, the computation time increases because all source and drain elements of the GREEN's function are needed.
- ***Reshape potential pockets to open them to the reservoirs (see Figure 3.9 c)***. The only place where potential wells are physically expected in TFETs under certain conditions is the channel. Consequently, one way to solve the issue of potential pockets is to only scan the source and drain potentials for them and remove the pockets artificially. This is done by setting the potential between the boundary and the local minimum to the value of the local minimum. Because this approach does not increase the computing time, it is used throughout this work.

3.3.3. Current

The next and most important property to discuss is the current flow in the device which is calculated according to eqn. (3.26) and presented in Figure 3.10. The energy dependent plot shows that the only significant current is indeed flowing in a small energetic window which is limited by the conduction band edge in the source at 0 eV and the valence band edge at 1.1 eV. The smoothness of the curve, i.e. the absence of spikes, confirms that the transition between conduction and valence band states has been modeled correctly. After integration over the energy, the device delivers a total current of I_D=6.83 nA.

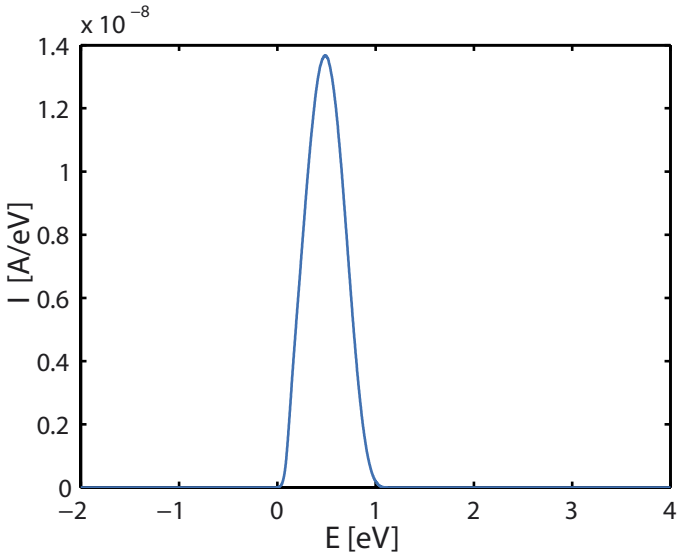


Figure 3.10. This current spectrum demonstrates that the only significant current flow occurs in the region where tunneling is possible.

3.4. 1D Si TFET characteristics

3.4.1. Transfer/output characteristics

After presenting the properties that can be extracted for a single bias condition, Figure 3.11 presents full transfer characteristics of the modeled device.

Subthreshold regime
The NEGF treatment covers the full ambipolar behavior of the TFET. This simulation allows for the following observations:

- For small negative V_{DS}, the p-branch is suppressed very well resulting in a wide gap between the n- and the

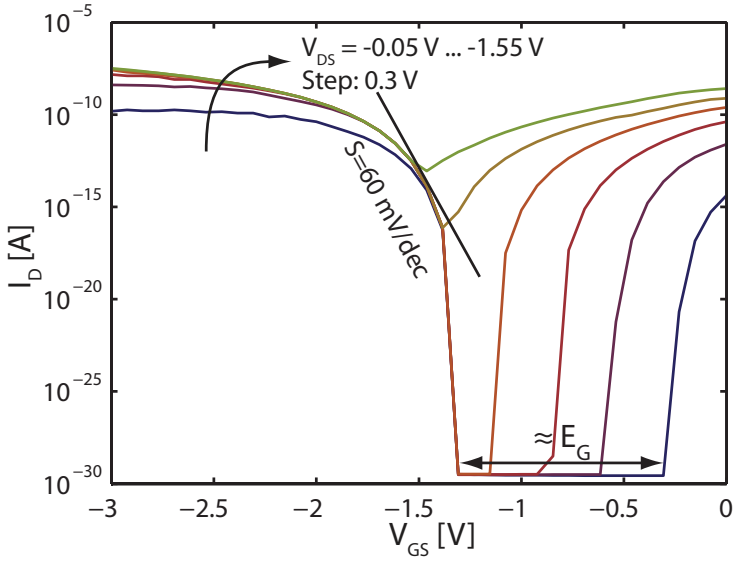


Figure 3.11. Transfer characteristics of a one-dimensional nanowire Si-TFET with a Λ of 2.2 nm and a width/thickness of 5 nm.

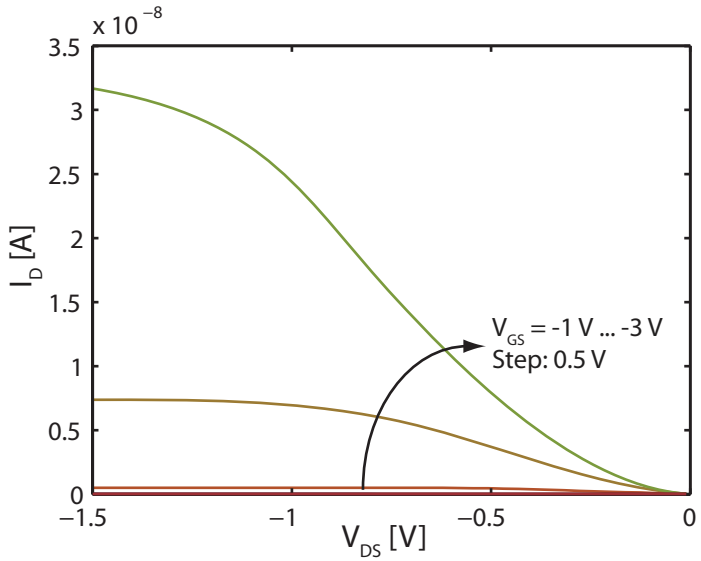


Figure 3.12. Output characteristics of a one-dimensional nanowire Si-TFET. Despite operating in the quantum-capacitance-limit, a non-linear onset is apparent.

p-branch, that is as large as the bandgap of E_G (see arrow).

- For larger negative V_{DS}, the gap between n- and p-branch is reduced by the value of V_{DS} and the off-state current starts to increase for $V_{DS} \approx -E_G$.
- The subthreshold slope exhibits the expected V_{GS}-dependence and deteriorates quickly for large negative V_{GS} voltages. The current level at which the slope becomes worse than 60 mV/dec is around 10^{-14} A, which is just slightly above the common measurement limits of 10^{-15} A.

These observations confirm the diagnosis from Chapter 2 that in single nanowires the region of the steepest slope only occurs at current levels that cannot be used for device operation. Nevertheless, even in the case of silicon, a parallel wire array architecture which employs thousands of wires will raise the current levels correspondingly and therefore might make the steep slopes exploitable.

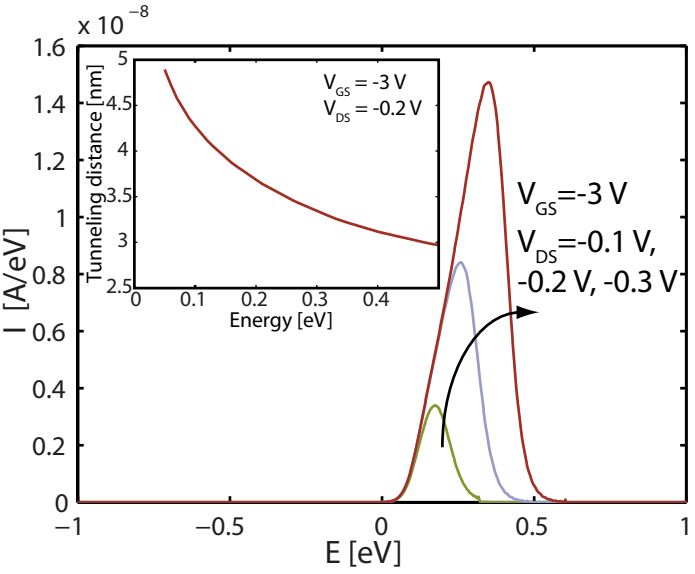


Figure 3.13. Current spectra for three different V_{DS} voltages. In contrast to the WKB the peak current is not constant. The reason for this is the energy dependent tunneling distance (*inset*).

On-state

To evaluate the on-state of the TFET, the output characteristics are simulated and presented in Figure 3.12. We note that saturation occurs at different V_{DS}-values depending on the applied gate voltage. The on-current is about two orders of magnitude below the ballistic limit, which reflects the fact that the tunneling probability is in the 10^{-2} range. In contrast to our expectations for a one-dimensional device in the quantum capacitance limit, we observe a non-linear onset for small V_{DS}. This observation raises the question if there are other reasons for a non-linear onset than those discussed in Section 2.2.2. The energy spec-

tra of the current for three different V_{DS} values (see Figure 3.13) show that the peak current levels rise with larger V_{DS}, which in this case is a result of an energy dependent tunneling probability. The reason for the energy-dependence is the non-constant tunneling distance for the different energies (see Section 3.3.1 and inset of Figure 3.13). Close to the conduction band edge in the source, the curvature of the potential is small, which results in a large tunneling distance. For larger energies the curvature increases and the tunneling distance shrinks. This means that in the quantum-capacitance limit, the output characteristics do not necessarily become linear. However, we expect that the extent of the non-linear region depends on the source doping level, because the doping concentration determines the curvature of the potential at the source/channel-junction. To verify this assumption, the impact of the doping concentration is studied in the following section.

3.4.2. Impact of doping concentration

The doping concentration, which in this simulation framework is represented by the distance of the chemical potential to the respective band edge, has a strong influence on the TFET performance.

- The doping concentration determines the built-in potential V_{bi} of the *p-i* and *i-n*-junctions, which acts in conjunction with the applied V_{DS} and V_{GS} to create the band-offsets necessary for tunneling. The larger V_{bi} is, the smaller V_{DS} needs to be to create the same band-bending.
- The charge carriers in source and drain screen any electric field that penetrates these regions. For larger doping concentrations, the related screening length λ_{dop}, that adds to the electrostatic Λ, reduces.
- For device structures that rely on doped leads, a larger doping concentration also reduces the amount of parasitic series resistances, because the resistance of a doped semiconductor reduces with increasing doping concentration. Although this effect could be observed in experimental data, the present NEGF study does not include parasitic resistances.

The first result becomes apparent in the simulated subthreshold characteristics in Figure 3.14. Here, the distance between chemical potential and respective band-edge is varied from -50 meV to 350 meV at a V_{DS} of -0.1 V. The changing built-in potential shows up as a variation of the distance between the n- and p-branches. The larger the doping level is, the stronger the ambipolar effect becomes. For the smallest doping concentration ($\mu_{S/D}$-$E_{c/v}$=-0.05 V) , the relatively small V_{DS} cannot compensate for the built-in potential, therefore the current flow is blocked. With increasing doping concentration the current level rises, but there is another undesirable effect that balances the benefits of larger doping. Around V_{GS}=-1.5 V, the subthreshold slope deteriorates strongly for larger doping concentrations. This behavior is a result of the fact that for larger doping levels, the broadened Fermi-distribution moves more and more into the source conduction band. Therefore, the high-energy tails of the Fermi-distribution contribute more and more to the current flow, which ultimately leads to the 60 mV/dec limit which

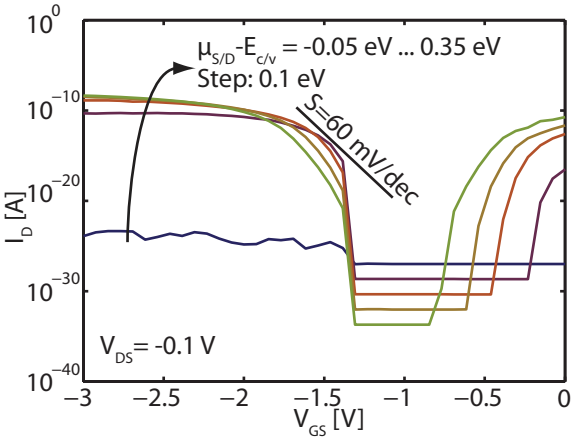


Figure 3.14. Impact of doping level on the subthreshold performance.

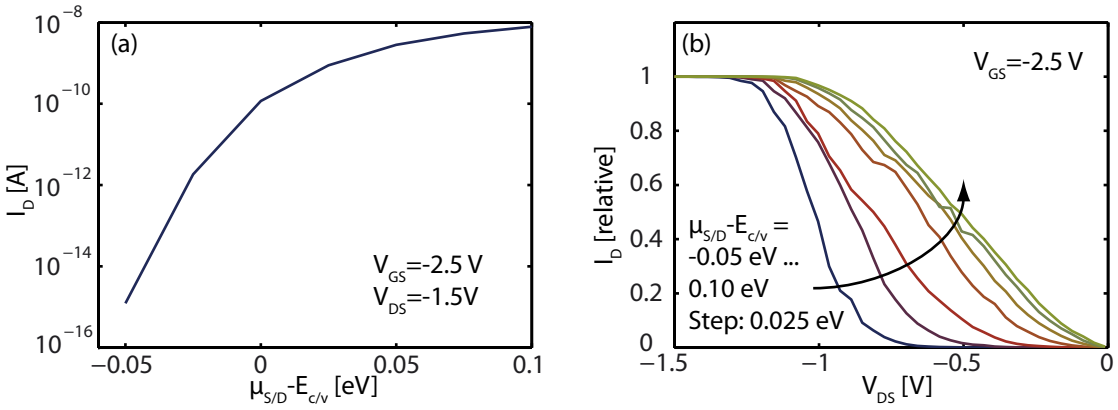


Figure 3.15. a) Saturation current as a function of doping level. The doping concentration affects the built-in potential of the p-i-n structure as well as the band-bending at the junctions. b) Normalized output current as a function of V_{DS}. The nonlinearity of the onset is suppressed for increasing doping level because the curvature of the band is increased.

is prevalent in MOSFETs. This means that a trade-off needs to be made for the doping concentration. On one hand, large current levels require large doping levels, but on the other hand, a steep slope limits the range of reasonable doping levels.
The difference in on-current is not only a function of the built-in potential but also of the screening length. Their combined effect on the saturation current is shown in Figure 3.15 (a). The curve is dominated by the different screening lengths λ_{dop}, which are extracted from the simulation to vary from about 50 nm for the smallest doping concentration to 2 nm for the largest. Figure 3.15 (b) shows the normalized output characteristics of the same devices. In agreement with the findings of the previous section, we observe that the non-linear part of the output characteristics becomes suppressed for larger V_{DS}, which is another direct consequence of the smaller screening length for higher doping levels.
However, for a device which does not operate in the quantum capacitance limit, there is also another explanation, why larger doping levels can reduce the non-linearity. If a device has a large quantum capacitance, carriers injected from the drain pin the channel potential close to μ_D. As a consequence, the channel potential depends on V_{DS} instead of V_{GS}. Therefore, the width of the tunneling junction is determined by V_{DS}, which leads to an exponential onset. Apart from suppressing the nonlinearity, optimizing the doping concentration can be used to reduce the ambipolar behavior of the TFET by choosing different doping concentrations for source and drain. If a large doping concentration is used in the source, the tunneling probability is large, while a small doping concentration in the drain suppresses the ambipolar branch.
In summary, the doping concentration is a very important parameter that impacts many aspects of TFET performance. The optimal range is very limited, but from the presented analysis, a Fermi-level of 50 meV to 100 meV in the band seems to be most promising in the case of silicon.

3.4.3. Λ-scaling

As we have already assessed in Chapter 2, the device electrostatics that is represented by the parameter Λ, is of utmost importance for TFET device performance. To verify this observation, Figure 3.17 presents transfer characteristics in the Si nanowire for different Λ-values from 1 nm to

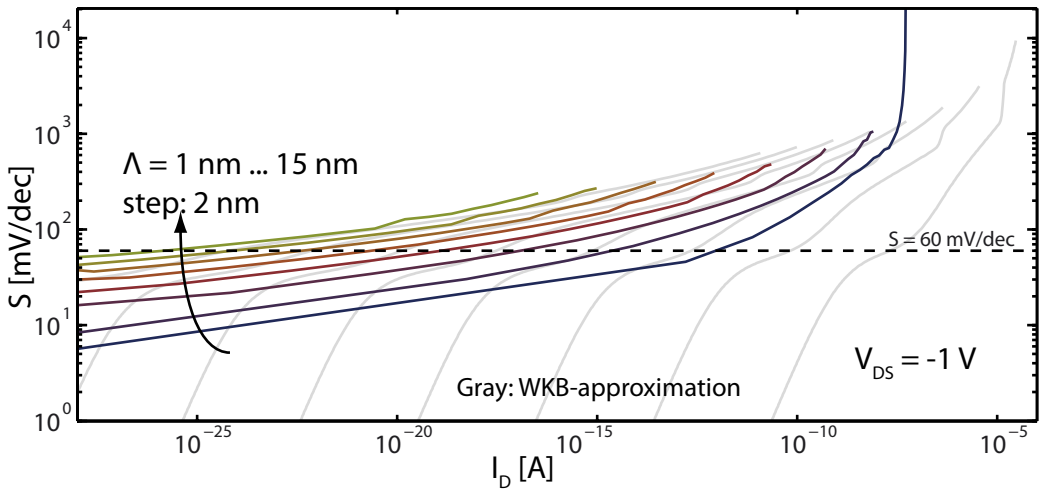


Figure 3.16. Inverse subthreshold slope versus drain current for different values of Λ at V_{DS}=-1 V. The colored lines are the result of NEGF calculations, while the gray lines are calculated by the WKB approximation in the fashion of Chapter 2.

15 nm and V_{DS}=-0.1 V. Clearly, the inverse subthreshold slope increases with increasing Λ and the point at which S equals 60 mV/dec shifts to smaller current values for larger Λ. Moreover, as Λ also affects the overall tunneling probability, the current level depends exponentially on Λ.

Figure 3.16 shows a current versus subthreshold slope plot similar to Figure 2.8. This time, the colored lines are the result of a NEGF simulation, while the gray lines in the background show the result of the corresponding WKB calculation. The general trends are similar to what has been described in Chapter 2: The current levels at which a steep subthreshold slope is observable are too small to be exploited for device operation. Nevertheless, an optimized device structure like a parallel wire array could raise the currents to a reasonable level in the range of 1 µA/µm to 10 µA/µm in the on-state. We observe that above the 60 mV/dec line, the WKB calculation reproduces the slopes qualitatively, but the current levels are about three orders of magnitude higher than the results of the NEGF simulation. The reason for this behavior is two-fold: on the one hand, the value of Λ that is used in the WKB approximation neglects the screening length in the doped source which has to be added to the electrostatic Λ. On the other hand, the WKB approximation assumes an energy independent tunneling probability, which is unrealistic. Therefore, the tunneling probability and the resulting currents are overestimated.

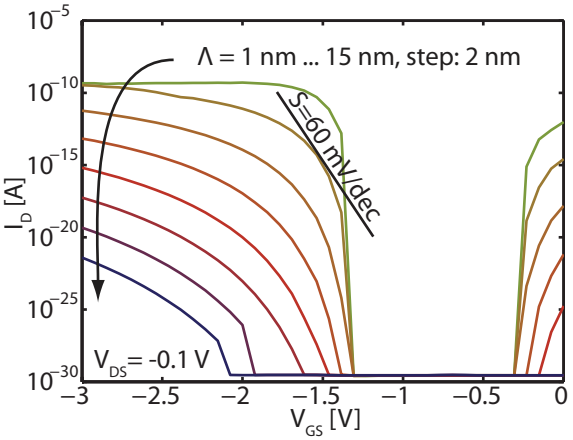


Figure 3.17. Transfer characteristics for Λ-values from 1 nm to 15 nm. A decrease of the inverse subthreshold slope with decreasing Λ is apparent.

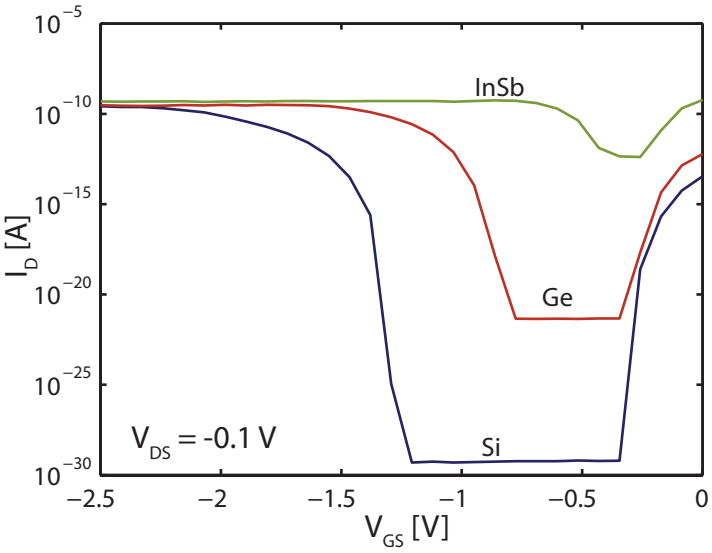
3.4.4. Choice of material

The choice of the transistor material changes the effective masses, the dielectric constant and the size of the bandgap. All three parameters influence the tunneling probability, either directly or via a change of the tunneling distance. Apart from these direct consequences, there are also implications for the manufacturability of devices. It is very helpful to reach the quantum capacitance limit in order to obtain the steepest subthreshold slopes [18]. One way to achieve the quantum capacitance limit is to reduce the density of states in the channel and thus also the quantum capacitance. The proper choice of material makes the fabrication easier, because in a material with small effective masses, the subband splitting induced by confinement is larger, which relaxes the requirements on the nanowire dimensions. Furthermore, a smaller effective mass reduces the density of states directly, too. In this section, we will compare the transfer characteristics for three different materials, namely Si, Ge and InSb, because these three materials cover the range of suitable parameters well. Figure 3.18 shows single transfer characteristics for the three materials at V_{DS}=-0.1 V and different nanowire diameters, that all correspond to subband splitting of 5 kT for the respective conduction band.

As expected, the different bandgaps are reflected in the distance of n- and p-branch as well as in the off-state current. Also, for large V_{GS}, the current levels line up inversely proportional to the effective masses. In contrast to the expectations, a comparison of the slopes reveals that the Si-nanowires feature the steepest slopes.

This observation is a result of the methodology used for the comparison. The two meaningful methods are:

- ***Compare characteristics for a given diameter and calculate the Λ according to eq. (3.5).*** This method has the advantage that the electrostatics for all wires is nearly identical, but due to the vastly different confinement energies, the lineup of the effective bandgaps of the materials might not reflect the nominal bandgaps in the bulk case at all. For example at a diameter of 5 nm, the bandgap of an InSb-wire is larger than that of a Si-wire. As a result, the InSb-wire would feature smaller tunneling currents than the Si-wire. Furthermore, the advantage of larger diameters for smaller effective mass materials is not exploited with this approach.



	Si	Ge	InSb
d_{nw} [nm]	5.5	8.4	19.6
Λ [nm]	2.3	3.4	6.5
E_q	5 *kT*	5 ***kT***	5 *kT*

Figure 3.18. Transfer characteristics for Si, Ge and InSb nanowires of diameters that correspond to a confinement energy of 5 ***kT***. The table lists the nanowire diameters and the Λ for an oxide thickness of 1 nm.

- ***Compare characteristics for constant confinement energy, i.e. different diameters, and calculate Λ according to eq. (3.5).*** With this method, the relaxed diameters for Ge and InSb can be exploited and the advantage of smaller bandgaps applies. However at larger diameters, the electrostatics, that means Λ, deteriorates. This case is more realistic and illustrates that the benefits from a small effective mass channel material can only be beneficial if at the same time the gate electrostatics are improved, for example by the use of a higher-κ dielectric ($\kappa > 30$).

The comparison in Figure 3.18 is done according to the second method and it predicts worse slopes for Ge and InSb. Another factor that leads to a seemingly worse slope is the ambipolar behavior. At the same time as the n- and p-branches move closer with smaller bandgap, also the minimum current level rises, which does not only deteriorate the I_{on}/I_{off} ratio, but also the range in which small slopes would occur.

This means that in an optimal device structure the ambipolar behavior needs to be suppressed. We already have seen that the doping level affects the ambipolar behavior. Another more promising way is the adoption of a heterostructure that suppresses tunneling at the channel/drain-junction. Because heterostructures are beyond the experimental scope of this thesis, no simulations are presented on heterostructures.

3.5. The 2D Si TFET at a single bias point

All results of the Sections 3.3 and 3.4 are obtained using a one-dimensional model that assumes single-mode transport. Although this type of model is well suited for parametric studies, because of the reasonable computation time, most experimental results are obtained in higher-dimensional systems. Even the majority of today's Si nanowires has a diameter of 10 nm to 20 nm, which only allow for quasi-one-dimensional transport. Therefore, in this section the SCHRÖDINGER-part of the model is solved for a two-dimensional system. Following the description in Section 3.2.3, the SCHRÖDINGER-equation is solved individually for each transverse wave-number k_t and then integrated over k_t to evaluate observables like charge density and current (see eq.).

$$I_{2D} = \int_0^{k_t(E_{\max})} \frac{dk_t}{2\pi} I_{1D}(k_t) \tag{3.31}$$

This means that k_t is conserved in the simulations, which is not true for the phonon-assisted tunneling in indirect bandgap semiconductors. Therefore, the simulation results present a lower limit for the current, because any non-conserving currents are excluded from the simulation. Furthermore, the energy transferred by phonons (maximally of the order of a few kT) allows for largercurrent levels especially in the off-state, which can deteriorate the subthreshold swing (see [20]).

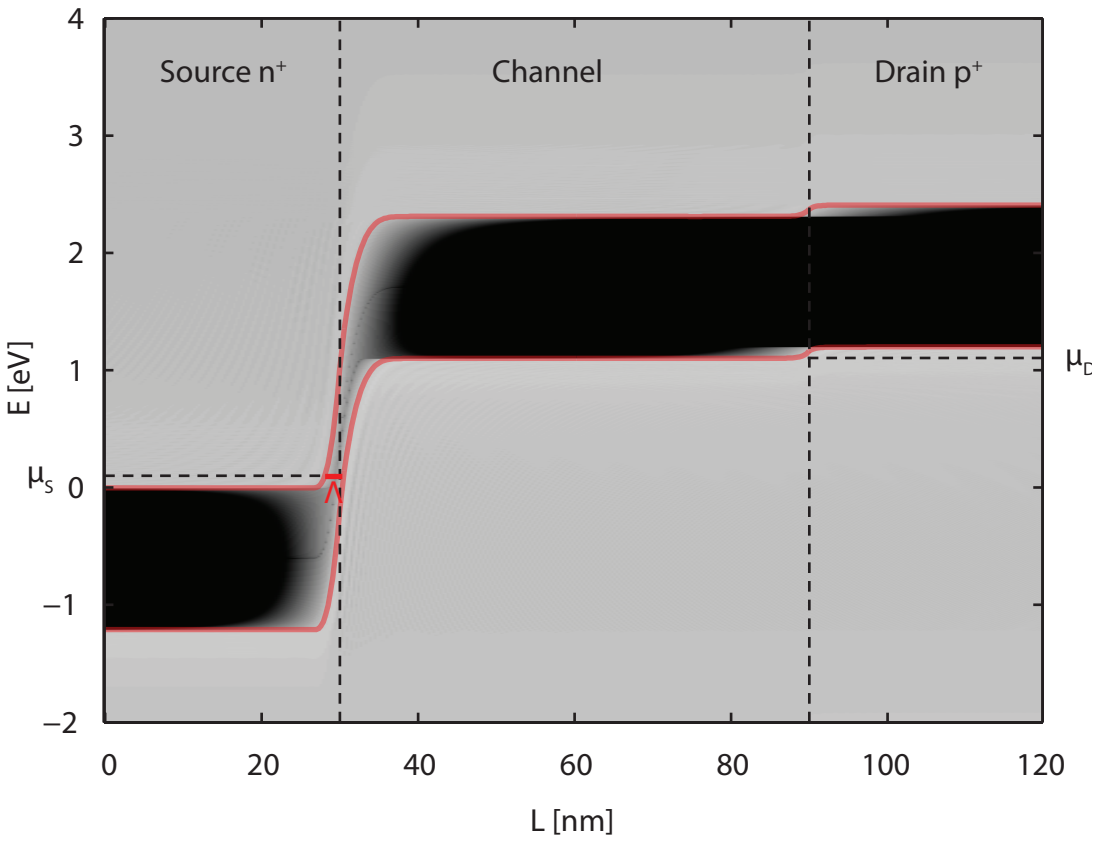


Figure 3.19. Local density of states for a 2D ultra-thin body TFET with a body thickness of 5 nm.

The parameters used for the simulation are identical to those of Table 3.1., the only differences being the infinite device width and the additional integration over k_t, which is done numerically for n_t=30 discrete values of k_t that are equally spaced up to a maximum transverse energy of E_t=0.5 eV. The number of integration nodes and the maximum transverse energy have been minimized to keep the computation time small while at the same time the maximum deviation from a reference simulation running at n_t=500 and E_t=1 eV is maintained below 5%. To disentangle the effects of Λ and LDOS, the Λ-value is chosen to be the same as in the one-dimensional simulations. However in a planar structure, the effective gate dielectric thickness has to be scaled down drastically to produce a small Λ of 2.2 nm (compare eq. (3.4) and (3.5)). A reasonable solution is to substitute the 1 nm thick SiO_2 by a 1.7 nm thick HfO_2.

3.5.1. Local density of states

The local density of states for a 2D TFET (see Figure 3.19) at V_{GS}=-2.5 V and V_{DS}=-1 V can be viewed as the superposition of several 1D-local density of states for different effective bandgaps. Compared to the pure 1D-LDOS, there are several major differences:

- The 2D-LDOS is in principle energy independent, however, in this simulation there is a limit for the integration over k_t, which makes the LDOS decrease above that limit. Nevertheless,

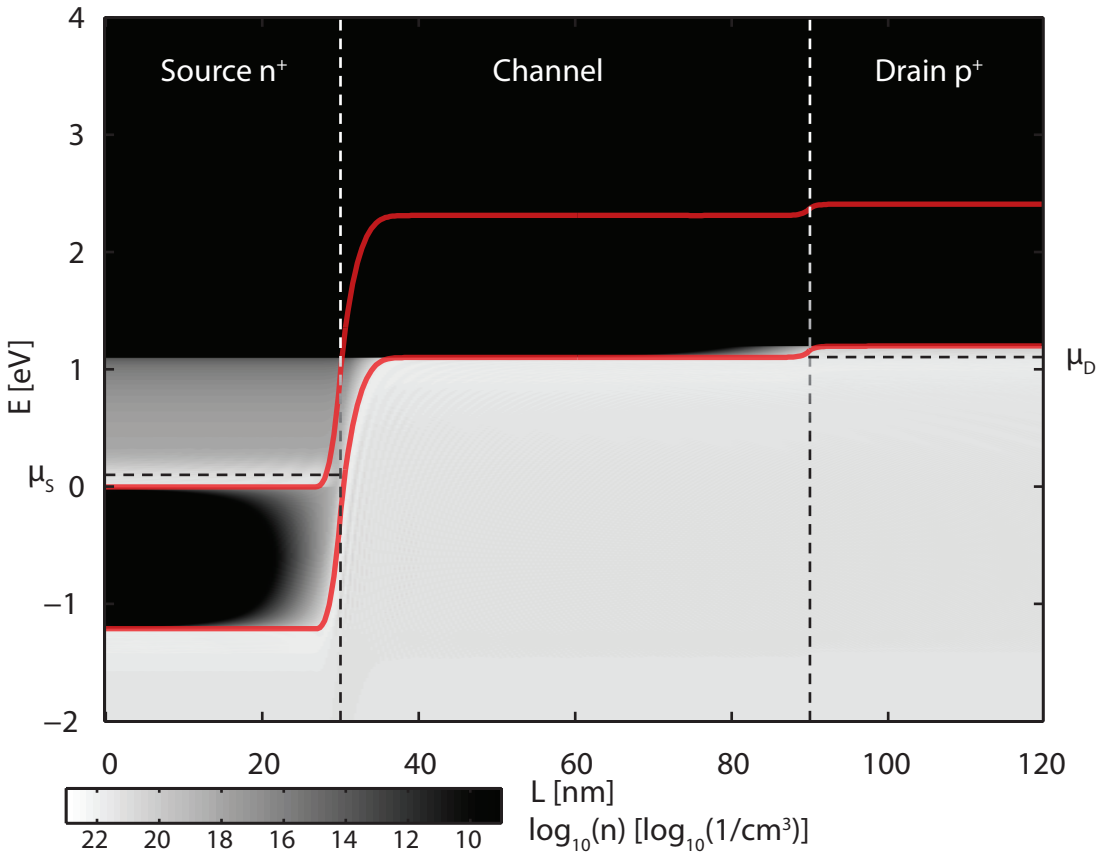


Figure 3.20. Electron density for a 2D ultra-thin body TFET with a body thickness of 5 nm.

the number of states available close to μ_S and μ_D is larger compared to the one-dimensional case. The effects on carrier density will be studied in more detail in the next section.

- The interference patterns which are a result of incident waves scattering off the potential barriers are less distinct. This is a another consequence of the superposition of different 1D-LDOS, because their interference patterns do not add coherently.
- The size of the bandgap is smaller, because quantization occurs only in one dimension.

Furthermore, the potential profile is steeper at the junctions, but we will see that this is a result of the different carrier concentration during the self-consistent calculation. Another effect that will be discussed in a later section is the impact of the density of states on the quantum capacitance.

3.5.2. Charge carrier densities

Figure 3.20 displays the corresponding energy resolved local charge density. Compared to the one-dimensional case, again the less distinct interference patterns and the larger overall charge density stand out. We also observe a strong tunneling component of the charge density in the source conduction band.

However, the most interesting quantity is the total electron and hole concentration as a function of position (see Figure 3.21 (a)). Here, the difference between the one- and two-dimensional densities of states becomes very obvious. For the same μ_S-E_c or E_v-μ_D = 0.1 eV, the carrier concentration differs by a factor of 3 for the n-type and a factor of 4.5 for the p-type. This difference explains why the potential profile is steeper in 2D, in fact λ_{dop} is reduced. Similar to the source and drain regions, the charge concentration in the channel increases by about a factor of 3. This is a first indication that the quantum capacitance of the 2D device is indeed larger than that of the 1D nanowire.

Figure 3.21 (b) shows an example of the numerical integration over the different 1D charge densities. As mentioned before, the integral runs from a transverse energy E_t=0 eV to E_t=0.5 eV in 30 steps of constant k_t. This step number is plotted on the abscissa while the left ordinate gives the ratio of the charge density for the respective step and the resulting total charge density after integration in percent. The shape of the curve is smooth and well-behaved which confirms that the number of integration steps is sufficient. The right ordinate shows the logarithm of the ratio

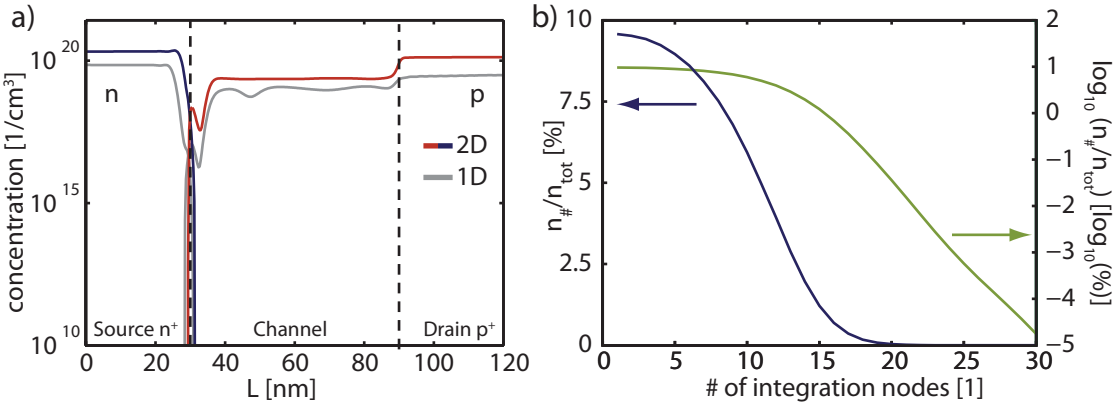


Figure 3.21. a) Total electron and hole concentration as a function of position for a 2D ultra-thin body TFET. For reference, the 1D results are plotted in grey. b) Relative contribution of each transverse wave-vector integration node to the total charge density.

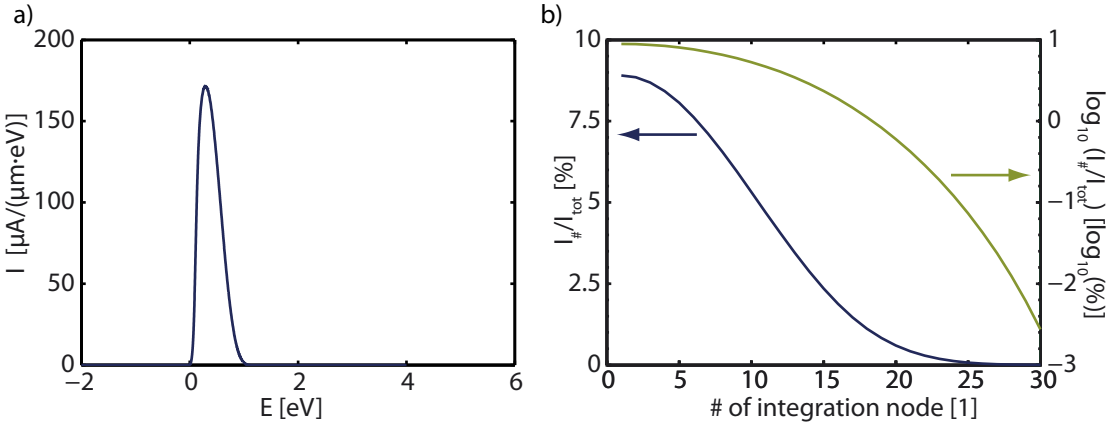


Figure 3.22. a) Current spectrum for a 2D ultra-thin body TFET. b) Relative contribution of each transverse wave-vector integration node.

and we note that within the integration range, the value drops below 10^{-4} %, which verifies that the range is large enough to include all significant contributions.

3.5.3. Current

Figure 3.22 a) displays the current spectrum for the 2D device at V_{GS}=-2.5 V and V_{DS}=-1 V. Compared to the 1D current spectrum which features a peak current of about 2.8 µA/(µm·eV), the current level of 167 µA/(µm·eV) is much higher in the 2D spectrum. This again is a result of the reduced λ_{dop} in 2D and the increased tunneling probability that comes with it. Additionally, the 2D device possesses a smaller bandgap, which also improves the transmission. But, from a manufacturing point of view, this comparison is unfair, because a 2D with the same gate dielectric thickness as the 1D device would have Λ=3.9 nm and thus deliver much less current than the optimized 2D device. However, the normalization of the 1D single-mode wire current is questionable because it is not easy to define the 'width' of a single mode. Since for all comparisons with 2D results, this problem is prevalent, the comparisons should be interpreted qualitatively rather than in a quantitative way. The integrated total current is 83.6 µA/µm.
The purpose of Figure 3.22 b) is to confirm the proper integration over the transverse wave-vector. Similar to the case of charge density, the relative current contribution of each integration node with respect to the total current is given. On the left ordinate, the percentage value is given and the corresponding curve is again smooth and well-behaved. The curve belonging to the right ordinate, which shows the logarithm of the percentage, drops to 10^{-2} %, confirming the contributing transverse energy range is well-covered.

3.6. 2D Si TFET characteristics

3.6.1. Quantum capacitance limit

Before subthreshold and output characteristics are discussed in the following sections, the issue of quantum capacitance should be discussed briefly. Since the local density of states in the

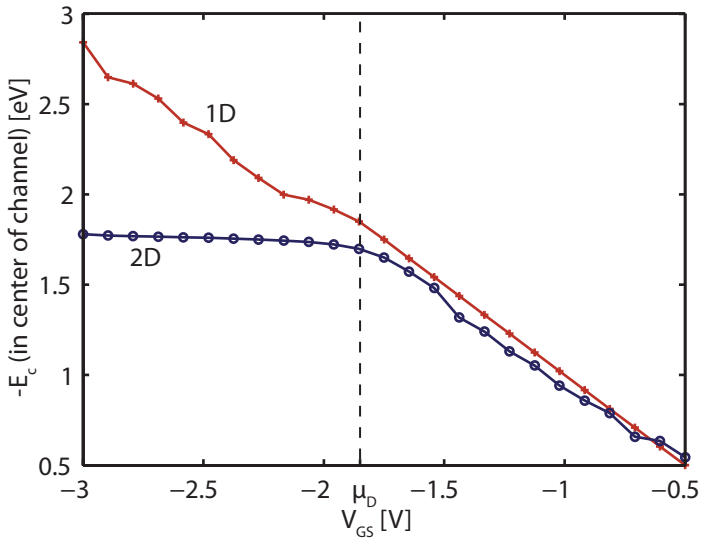


Figure 3.23. Comparison of the movement of the center of the channel conduction band edge with applied V_{GS} for the presented 1D and 2D devices. In the 2D case, the band movement stops in inversion, while in the 1D case, it continues.

channel is much larger in the 2D case than in the 1D case, the 1D structure is more likely to operate in the quantum capacitance limit. Figure 3.23 shows a comparison of the dependence of the conduction band energy in the center of the channel on V_{GS}. We observe that in the 2D case, the channel band movement essentially stops at about 1.8 eV, the energy of the drain chemical potential μ_D, but in the 1D case, it continues to follow V_{GS}. This observation confirms that the 1D device appears to operate in the quantum capacitance limit while the 2D device does not.

At this point it has to be mentioned, that the ballistic nature of the simulation plays an important role when reaching the quantum capacitance limit. As there is no scattering in the channel, additional states which form in the channel "potential well" are not filled, similar to the potential wells in source and drain discussed in Section 3.3.2. Therefore, in a ballistic simulation, the carrier concentration in the channel is smaller than if scattering was included and the pinning of the channel bands due to the quantum capacitance is less likely. These considerations lead to the conclusion that the simulated 1D device might only behave like operating in the quantum capacitance limit, but a real device might not due to scattering. A second conclusion is that a device made from a material with a large mean free path, which operates in the ballistic regime might be a good candidate for reaching the quantum capacitance limit. Nevertheless, the presented results on the 1D device are still representative for operation in the quantum capacitance limit.

3.6.2. Transfer/output characteristics

Subthreshold regime

Figure 3.24 presents the subthreshold characteristics of the same 2D TFET which has been discussed in Section 3.5. The current is, as before, normalized to the width of the device and for reference the normalized 1D results are plotted. The comparison of both characteristics reveals that the 2D TFET provides much larger currents, but again, the issue of the normalization of the 1D current applies. Independent of any normalization problems, we observe a shift of the "threshold voltage" to smaller V_{GS} for the 2D case, which is a consequence of the smaller bandgap in 2D. The analysis of the subthreshold swing reveals very similar behavior for both cases, when compensated for the threshold voltage shift. The only difference becomes apparent at V_{DS}=-50 mV, where the minimum slope in 2D is S_{2D}=15 mV/dec and in 1D it is S_{1D}=6 mV/dec. The reason for this difference is possibly that the 1D structure is effectively operating in the quantum capacitance limit, which leads to improved channel electrostatics.

The 2D-TFET transfer characteristics allows for another important observation, the current levels at which sub-60 mV/dec slope occurs (< 10^{-2} µA/µm) coincides with the lower part of the operating regime of today's MOSFETs (10^{-4} µA/µm to 10^{3} µA/µm).

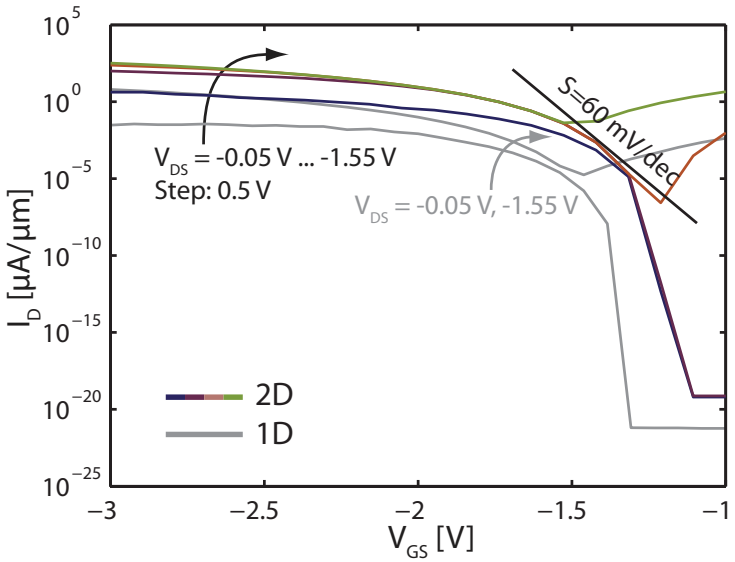


Figure 3.24. Transfer characteristics of a 2D ultra-thin-body TFET Si-TFET with a Λ of 2.2 nm and a thickness of 5 nm.

On-state

Figure 3.25 shows the output characteristics of the 2D-device under study. For the chosen gate voltage range (V_{GS}=-1 to -3 V), the output currents are again comparable to classical MOSFETs, but compared to the 1D-device (see Figure 3.12), despite the increased carrier concentration, the saturation occurs at larger V_{DS} values in 2D. This is a consequence of the smaller bandgap in the 2D device. For the same V_{GS}, the energy of the channel valence band edge is larger for a smaller bandgap, which means a larger V_{DS} is needed to achieve saturation in 2D. The degree of non-linearity in the onset is smaller for the 2D device, which can be explained by the larger carrier concentration in the 2D source, that reduces the energy dependence of the transmission (compare Figure 3.15). The effect of the larger quantum capacitance in 2D, which increases the influence of μ_D on the channel potential and leads to an increased non-linearity, seems to be overruled.

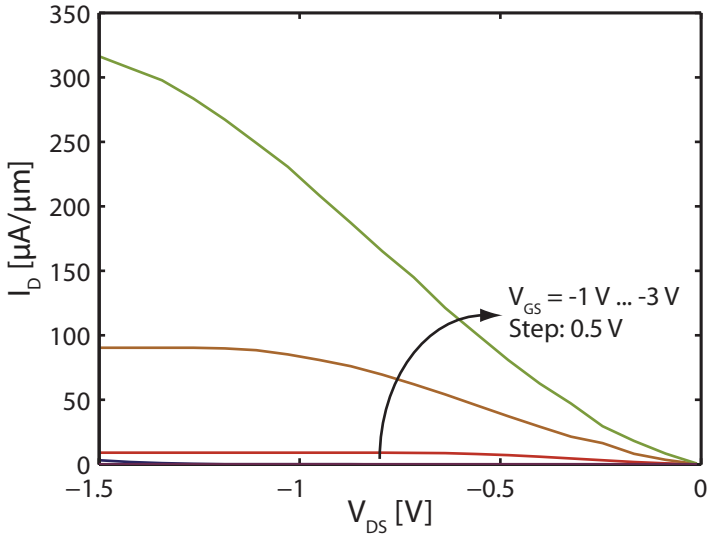


Figure 3.25. Output characteristics of a 2D ultra-thin-body TFET.

3.7. Conclusion

In this chapter, first the basic non-equilibrium Green's function method is introduced, followed by some extensions that account for multiple bands and higher device dimensions. The main part of the chapter then focuses on the simulation of a one-dimensional silicon nanowire TFET and the impact of several parameters like doping concentration and gate electrostatics is discussed. The simulation is extended to a two-dimensional ultra-thin body TFET. Both devices are compared to disentangle the effects of dimensionality from other external factors like electrostatics and doping concentration.

To summarize, the conclusions for each optimization parameter are listed below:

- The starting configuration, a one-dimensional nanowire of 5 nm diameter, provides steep subthreshold slopes, but unfortunately the sub-60 mV/dec slope only occurs at current levels that are not usable for digital logic applications. As a remedy for this problem, parallel multi-wire arrays are proposed. Furthermore, despite the one-dimensional nature of the nanowire, we find a non-linearity in the output characteristics which, depending on the amount of non-linearity, might limit the speed for switching applications.
- To reduce the non-linearity of the output characteristics, the impact of doping concentration is studied and indeed, the non-linearity can be suppressed by increasing the doping concentration. As a welcome side-effect of a larger doping concentration, we state an increase in the on-state current. However, the larger on-currents come at a cost, which is a deteriorated subthreshold slope. Therefore, the doping level needs to be optimized and an optimal range of 50 meV < μ_S-Ec = μ_D-Ev < 150 meV is found.
- The next subject to optimization is the gate-electrostatics in form of the screening length Λ. At first glance, a smaller Λ-value is always beneficial, but a second glance reveals that in order to achieve small Λ-values, the diameter of the wire has to be reduced. For very small dimensions, this leads to an increase of the bandgap, which degrades the on-state performance of the TFET. Because of this trade-off, the optimal Λ depends on the choice of material.
- The issue of confinement induced bandgap widening becomes especially important when studying the influence of the base material. The comparison of Si, Ge and InSb reveals that low-dimensional transport can be achieved much easier in materials with a small effective mass. Although the wire diameter can be relaxed for small effective mass materials, the thickness of the gate-dielectric has to be reduced by a similar factor to keep the screening length Λ constant. In addition to the larger Λ, also the off-state degrades because of the stronger ambipolar behavior that comes with the small bandgap. For example, when to avoid this problem an InSb wire is scaled aggressively to a diameter of 5 nm, the bandgap is larger than that of a Si wire of the same diameter, but because of the different dielectric constant, Λ for the InSb wire is worse. These considerations demonstrate that there is a trade-off between small effective mass and bandgap on the one hand and a small Λ on the other hand.
- Finally, the impact of dimensionality is studied by comparing the one-dimensional silicon nanowire with a two-dimensional ultra-thin-body TFET with a body thickness of 5 nm. The main advantage of a 1D wrap-around gate is that much smaller Λ-values can be realized, which makes the performance of the 1D structure inherently superior. When comparing 1D and 2D devices of the same Λ, the current level of the 1D devices is about two orders of magnitude smaller than that of the 2D TFET. However, the subthreshold slope for the nanowire is about a factor of 3 steeper, which is interpreted as an improvement due to improved gate control in the quantum capacitance limit. In reality, the improved current level of the 2D TFET cannot

be realized because to achieve the same Λ as in a nanowire, the gate dielectric needs to be scaled down much beyond the state-of-the-art.

Chapter 4
TunnelFETs on SOI

4.1. Introduction

This Chapter presents the experimental realization of the TFET concept in form of planar transistors fabricated on silicon-on-insulator substrates. First, the fabrication and arising fabrication issues are described and second, the electrical characteristics of these devices are discussed in detail.

4.2. Fabrication of TFETs on SOI substrates

4.2.1. Process flow

The process flow for the fabrication of TFETs has many similarities with standard CMOS fabrication, which makes the TFET a very attractive device to be adopted by semiconductor industry. The individual process steps are sketched in Figure 4.1 and are described in the following. A scanning electron micrograph of the readily fabricated device can be found in Figure 4.2.

1. ***Preparation of the substrates.*** The starting substrates are SOI substrates with material parameters summarized in Table 4.1. The original wafers are cut into two-by-two centimeter sized pieces.
2. ***Mesa insulation and mark definition (see Figure 4.1 (a)).*** To electrically isolate the individual transistors on each sample, the top Si layer is patterned into rectangles, so called mesa-structures. In this step, also a number of lines and crosses is defined that are used as optical alignment marks in the subsequent steps. The patterning consists of two steps, which are optical lithography followed by reactive ion etching (RIE) using an SF_6/Ar-plasma (recipe Poly-Si). This plasma predominantly etches silicon, but only provides a poor selectivity to the underlying SiO_2. However, this does not create a problem, because the buried oxide layer (BOX) is thick enough to withstand the overetching. The width of the mesa structure defines the gate width, which in the present case is designed as: 20 µm, 30 µm, 40 µm and 50 µm.
3. ***Gate oxidation (see Figure 4.1 (b)).*** After the device area is defined, the samples undergo an RCA-cleaning process ([47]), which prepares the surface for the gate-oxidation. Immediately after the standard clean 2 - step (HCl/H_2O_2 chemistry) of the cleaning, the samples are oxidized

in a rapid thermal oxidation processor at 800 °C for either 20 or 30 min in O_2-atmosphere to grow an SiO_2 gate dielectric of 3.5 nm or 4.5 nm, respectively.

4. ***LPCVD Poly-Si deposition.*** To minimize contamination, within 30 minutes after the oxidation, the gate material, 200 nm thick n^+-doped poly-Si, is deposited by low-pressure chemical vapor deposition (LPCVD) method.
5. ***LPCVD SiO_2 deposition.*** A sacrificial SiO_2-layer with a thickness of 100 nm is deposited by LPCVD at 710 °C as a hard mask for the following etch process.
6. ***Poly-Si activation.*** To ensure a high level of activation in the poly-Si gate, the samples are annealed at 950 °C for 1 min in pure N_2 by means of a rapid thermal processor (RTP).
7. ***Gate patterning (see Figure 4.1 (c)).*** The patterning of the gate stack starts with another optical lithography, followed by RIE dry etching using a CHF_3-plasma (recipe JCHF3), which preferentially etches SiO_2, to pattern the sacrificial oxide. After a short HF dip to remove any residual oxide, the poly-Si gate is etched in an inductive coupled plasma etching machine using HBr and O_2 (recipe HBr2step, see [48]). This plasma features a very high etch rate selectivity between SiO_2 and Si and therefore allows for an ideal etch stop on the thin gate oxide. The resulting gate lengths are 2 µm up to 12 µm in 2 µm steps.
8. ***Source/drain implantations (see Figure 4.1 (d)).*** After the gate is defined, another optical lithography step using UV6 photoresist and hard-bake is utilized to create a mask for the source implantation which overlaps with the gate stack for self-aligned implantation. The implanta-

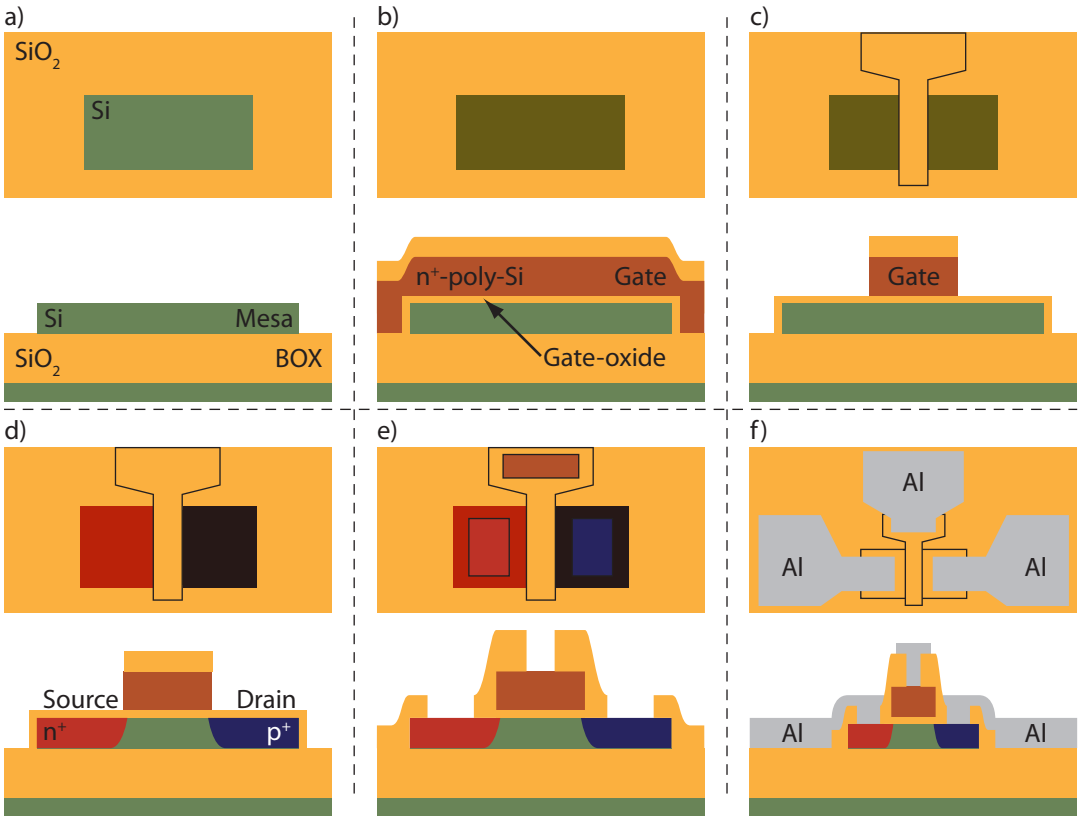


Figure 4.1. Process flow for SOI based TunnelFETs. For each step, the top view and a cross-section along transport direction are shown.

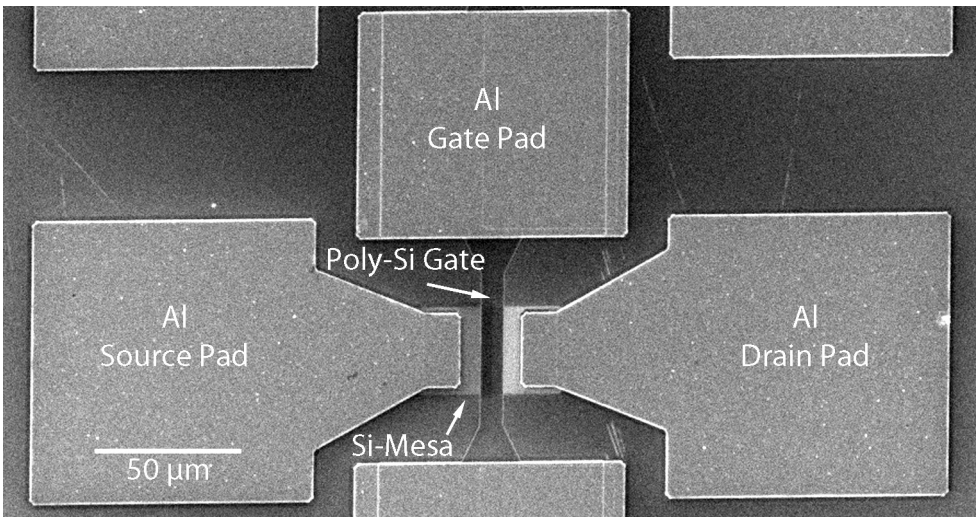


Figure 4.2. Scanning electron micrograph of a readily fabricated SOI-TFET with a gate length of $L_G = 8\,\mu m$. The silicon mesa structure is surrounded by aluminum contact pads.

tion conditions are again listed in Table 4.1. During the implantation step, the photoresist heats up the sample and the photo-resist molecules cross-link completely. This renders the removal of the photoresist by solvents impossible. Thus, to remove the photoresist, the samples are exposed to an O_2-plasma in a barrel reactor for three hours at 600 W. The barrel reactor uses a FARADAY-cage to create a field-free space around the sample that reduces the surface damage otherwise done by highly accelerated ions. The three steps, optical lithography, implantation and resist removal, are repeated for the drain side.

9. ***Annealing and activation.*** To incorporate the implanted ions in the crystal lattice and cure lattice damage created during implantation, a 1000 °C spike anneal in pure N_2 is carried out in a rapid thermal processor. These activation conditions have been selected as a trade-off between maximum activation and minimal dopant diffusion.
10. ***Isolation oxide and opening of vias (see Figure 4.1 (e)).*** To prevent surface leakage paths and provide protection from ambient air, a 100 nm thick layer of SiO_2 is deposited by physically enhanced chemical vapor deposition (PECVD). Compared to the LPCVD method that was used previously, the PECVD-oxide is deposited at much lower temperatures (350 °C - 400 °C), at which no additional dopant diffusion occurs. After the deposition, a via on top of the gate stack is defined by optical lithography and buffered HF etching. In contrast to diluted HF, the buffered HF causes less damage to the photoresist and therefore provides a better etch control. To prepare the sample for metallization, another set of vias on top of the source and drain regions is patterned using a similar process, but using a different etching time due to the different oxide thicknesses on top of both areas.
11. ***Metallization (see Figure 4.1 (f)).*** To create large contact pads that can be easily contacted by a probe tip, first a negative optical lithography is done. Immediately before the following deposition of 200 nm aluminum by electron beam evaporation, the sample is dipped in HF to remove the native oxide layer that could deteriorate the electrical contact between silicon and aluminum. The aluminum is then patterned by lift-off in an acetone bath.

12. ***Post-metal anneal.*** To further improve the metal/silicon interfaces, as final step, the samples are exposed to a ten minute long forming gas (N_2 90%, H_2 10%) anneal step at 400 °C which alloys the aluminum and the silicon.

	I203	I202	I201	I204	It202	It204	It201	I101	It101
Material	SOI	SOI	SOI	SOI	SOI	SOI	SOI	SOI	SOI
t_{Si} [nm]	20	20	20	20	20	20	20	10	10
t_{BOX} [nm]	100	100	100	100	100	100	100	100	100
t_{ox} [nm]	4.5	4.5	4.5	4.5	3.5	3.5	3.5	4.5	3.5
D_{As} [1/cm²]	5E14	1E15	3E15	6E15	5E14	1E15	3E15	3E15	3E15
E_{Source} [keV]	5	5	5	5	3	3	3	3.5	3
D_B [1/cm²]	5E14	1E15	3E15	6E15	3E15	3E15	3E15	3E15	3E15
E_{Source} [keV]	2	2	2	2	1.5	1.5	1.5	2	1.5

Table 4.1. Process parameters for the different fabricated samples. t_{Si}: Si body thickness, t_{BOX}: buried oxide thickness, t_{ox}: gate oxide thickness, D_{As}/E_{Source}: implantation dose/energy source, D_B/E_{Drain}: implantation dose/energy drain

4.2.2. Implantation & annealing

The choice of implantation conditions is very important for a successful TFET fabrication. The optimal doping profile for the source would be constant in the vertical direction and a step function in the transport direction, with a large concentration in the source and no concentration in the channel. For the drain, the requirements are similar, with the exception that the steepness of the step function can be relaxed. Two parameters determine the properties of the ion implantation:

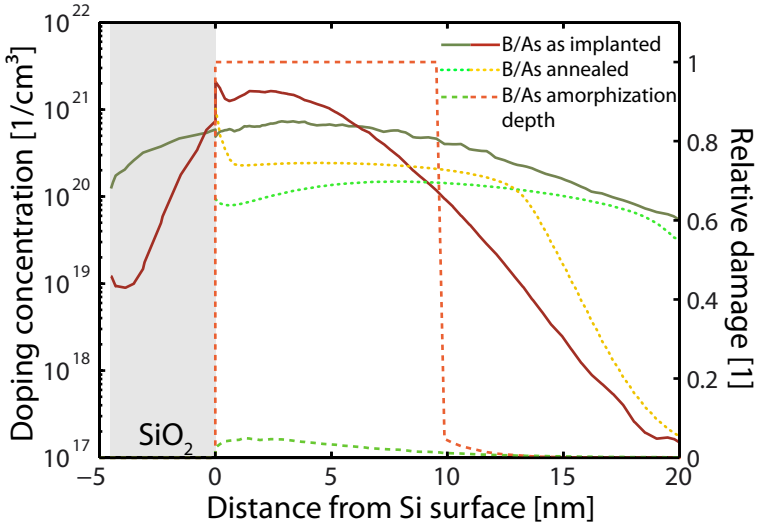


Figure 4.3. Simulated vertical doping profile for the sample I202. The doping concentration is plotted as implanted and after the 1000 °C spike anneal. The relative implantation damage is normalized to the threshold of amorphization. A value of '1' means complete amorphization.

1. ***Energy.*** The implantation energy determines the ion range and the lateral straggle of the ions. The ion range is responsible for the vertical implantation profile, which is chosen to minimize

the number of ions lost in the top or bottom oxides. Moreover, also the amorphization of the silicon layer has to be accounted for, because during the subsequent annealing step, an undamaged seed layer is needed to recrystallize the amorphous upper part. The lateral straggle also depends on the ion energy, which means it cannot be optimized separately. Nevertheless, the straggle determines the initial (i.e. before thermal treatments) steepness of the doping profile in transport direction.

2. ***Dose.*** The implantation dose determines the initial concentration of dopant ions in the semiconductor. Larger doses usually lead to higher doping concentrations after activation. Only for very high doses, when the solubility limit is exceeded, a saturation occurs and the doping concentration becomes independent of the dose. If the dose is increased further, large numbers of dopant atom clusters start to form that can even lead to a decrease in doping concentration.

After the implantation, a high temperature annealing step is done which serves the purpose of recrystallization and dopant activation. In the case of the TFET, the annealing has to be optimized for high dopant activation and small diffusion at the same time. To achieve favorable degenerate doping levels, as proposed in Chapter 3, usually a high temperature annealing (> 1000 °C) for moderate times (1 min to 10 min) is performed. However, these conditions lead to very strong diffusion, which would counteract all optimizations for steep potential profiles at the tunnel junction. Hence, the activation time is reduced to < 1 s (spike anneal) and the temperature is fixed at 1000 °C, to suppress the dopant diffusion while maintaining a reasonable activation.

Figure 4.3 shows an example of an implantation simulation carried out with Synopsys Sentaurus TCAD using TRIM [49] monte-carlo simulations for the implantation process for the sample I202 (see Table 4.1). On the abscissa the distance from the silicon surface is plotted, the left ordinate displays the doping concentration, which is plotted for As and B implantation prior and after the annealing step.

We note that due to the larger mass of arsenic, the As implantation is shallower and even after annealing only in 2/3 of the Si layer thickness, the doping level exceeds $10^{20}/cm^3$. The reason why the energy is not increased to fully dope the layer is plotted on the right ordinate. The implantation

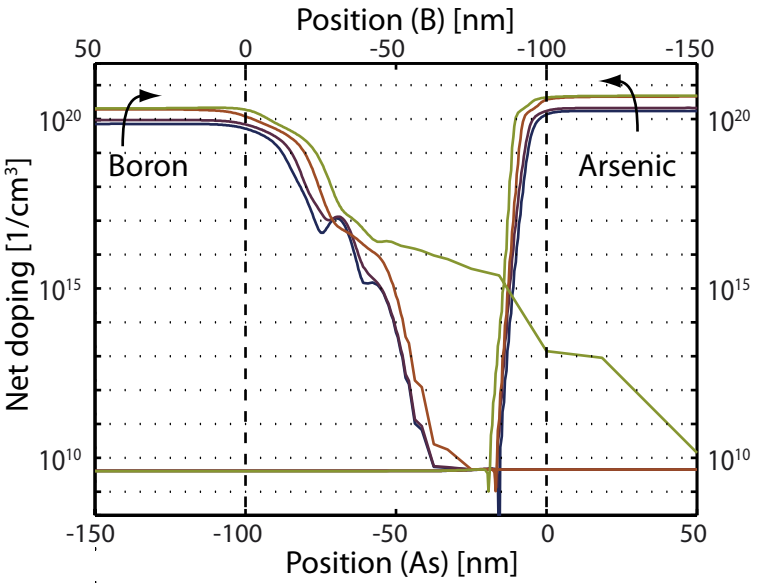


Figure 4.4. Simulated doping profile in transport direction after annealing for four different implantation doses ($5{\cdot}10^{14}$ 1/cm², $1{\cdot}10^{15}$ 1/cm², $3{\cdot}10^{15}$ 1/cm², $6{\cdot}10^{15}$ 1/cm²). Note the different abscissae for boron an arsenic. While the arsenic junction has a steepness about 1 nm/dec, the boron junction features a steepness of about 5 nm/dec. The bumpiness of the boron curves at 10^{17}/cm² is probably due to numerical artifacts. The dashed line marks the position of the gate edge.

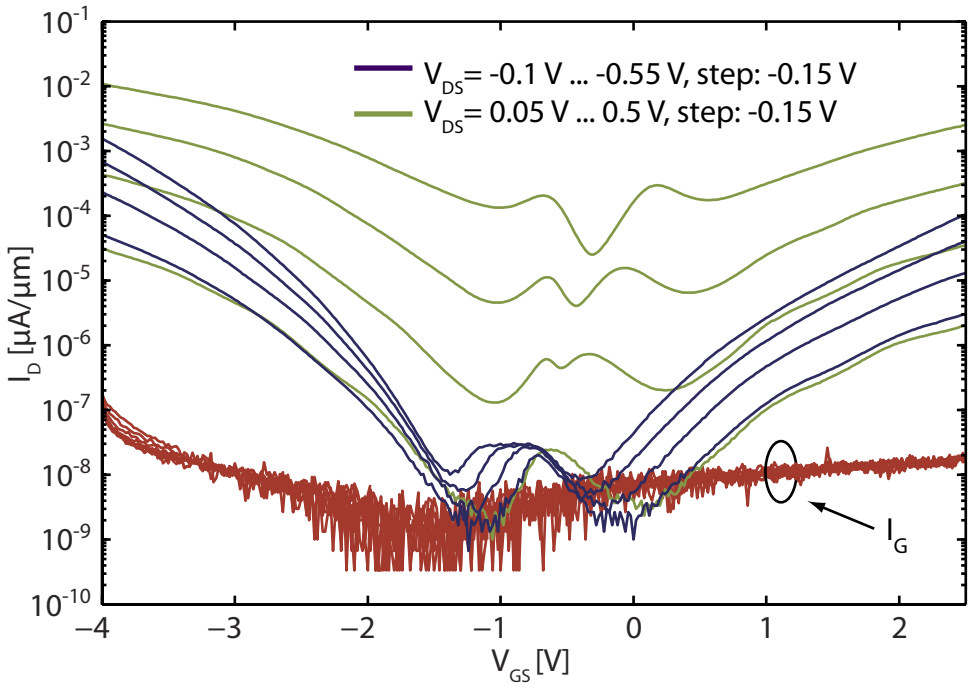


Figure 4.5. Subthreshold characteristics of an SOI-TFET with a gate length of 2 µm and a width of 30 µm on sample I201. The V_{DS} range includes positive and negative values. The minimal subthreshold slope is 320 mV/dec.

damage normalized to the threshold of amorphization shows that in case of the As, the upper half of the silicon layer is fully amorphized. To maintain a proper seed layer for recrystallization, the implantation energy was fixed at 5 keV for the sample I202. The smaller boron ions cause less damage to the silicon lattice, therefore the energy for the boron implantation is optimized to yield a homogeneous doping concentration after activation. The post-annealing profile for boron shows a property of boron, which is out-diffusion. The doping concentration close to the SiO_2-interface is reduced because boron tends to diffuse to the oxide interfaces and become electrically inactive. Even more important than the vertical doping profile is the doping profile in transport direction which is simulated for four different implantation doses by Synopsys Sentaurus TCAD and displayed in Figure 4.4. The doping profiles for boron are shown on the left side of the plot and the edge of the gate stack corresponds to the left vertical dashed line with the top abscissa. While the doping concentration is fairly constant about $1{\cdot}10^{20}$ 1/cm² on the left of the gate edge and only depends on the implantation dose, in the shadow of the gate stack, the doping concentration decreases at a rate of about 5 nm/dec. Below a dose of $1{\cdot}10^{17}$ 1/cm², the curves appear bumpy, which is probably a numeric artifact. Anyway, electrically the most important slope concerns the doping levels between $1{\cdot}10^{20}$ 1/cm² and $1{\cdot}10^{17}$ 1/cm².
In the case of arsenic, the simulated doping profile has a much steeper slope of about 1 nm/dec. Therefore, we can expect a much larger tunneling probability at the *n-i* junction than at the *p-i* junction, which in turn leads a superior performance in the p-type mode of operation where tunneling occurs at the *n-i* junction. At this point, it should be noted that the real slope might be worse than 1 nm/dec, because the diffusion simulation is not as accurate as the monte-carlo

based implantation simulation. However, qualitatively the simulations gives a correct estimation when comparing As and B.

4.3. Characterization results and discussion

The readily fabricated devices are characterized using an HP4155B semiconductor parameter analyzer which is capable of measuring I_D-V_{GS} and I_D-V_{DS} characteristics automatically with high accuracy. In the following, the different measurement results are presented and discussed. The starting point is a detailed discussion of a 'reference'-device to which all other variations will be compared.

4.3.1. Reference device - I_D-V_{GS}

Figure 4.5 shows measured subthreshold characteristics of an SOI TFET on a 20 nm thick Si body, implantation doses of $3{\cdot}10^{15}/cm^3$ for source and drain, 2 µm channel length and 30 µm channel width. The n-doped region is used as source, the potential reference. The gate leakage current I_G is plotted and with a small exception around V_{GS}=0 V, it never exceeds the drain current, which means that the characteristics are not governed by gate leakage currents.

Reverse bias - Negative V_{DS}

For negative V_{DS} (blue curves), the characteristics meet the shape predicted by the simulations. For -4 V<V_{GS}<-1.5 V, the drain current increases as a result of tunneling at the *n-i*-junction. The minimum point slope of about 320 mV/dec is measured at V_{GS}=-1.7 V. For 0 V<V_{GS}<2.5 V, tunneling occurs at the *p-i*-junction with a worse minimum slope of about 500 mV/dec at V_{GS}=0.2 V. Furthermore, the current of the n-branch is about one order of magnitude smaller than that of the p-branch. The reason for the different slopes and currents at the two junction is easily identified as the different steepnesses of the doping profiles of As versus B (see Figure 4.4). The broader profile at the *p-i*-junction leads to a larger λ_{dop}, which causes a smaller tunneling transmission, that reduces the current and the subthreshold slope compared to the situation at the *n-i*-junction.

The body thickness of 20 nm makes the SOI TFET effectively a three-dimensional device. Therefore, it is impossible to match the measurement results quantitatively with a one- or two-dimensional model and no absolute comparison of the measurement results and the models of Chapter 2 and Chapter 3 is presented. Nevertheless, in later sections, the qualitative trends predicted from the simulations are compared to the experimental parameter dependencies.

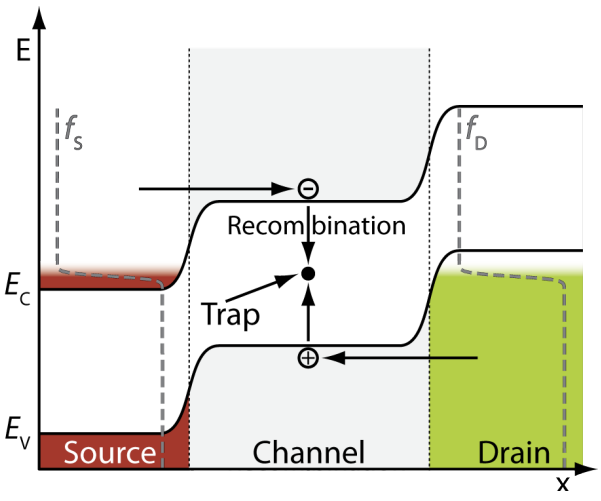


Figure 4.6. Band profile showing a possible reason for the 'hump' in the off-state of the TFET. If both, electron- and hole-densities in the channel are equal, trap-assisted recombination can occur either at the channel/oxide-surfaces or in the channel volume.

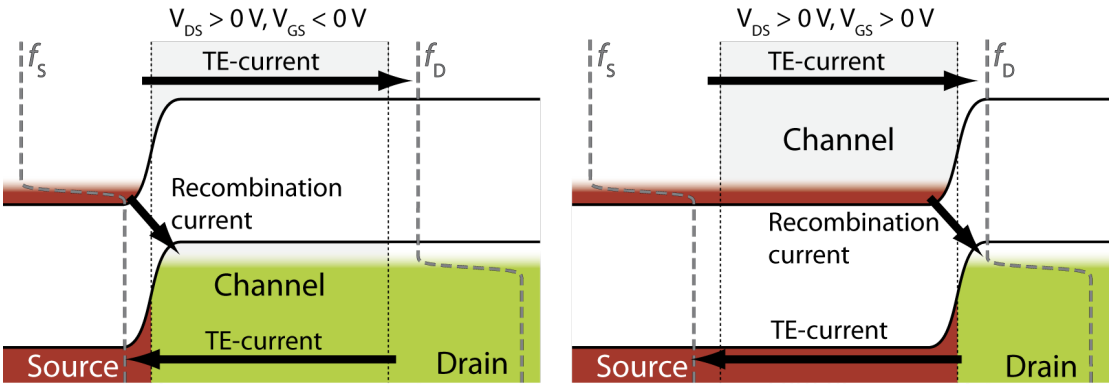


Figure 4.7. Band profiles illustrating the different components of the forward bias currents for negative (left) and positive (right) V_{GS}. For moderate V_{DS} ($<E_G$) as shown, the thermal emission currents (TE-current) are smaller than the field-enhanced recombination currents.

Recombination currents

For -1.5 V$<V_{GS}<$0 V, the characteristics reveal a hump whose width is proportional to V_{DS} which is not expected from the simulations. When looking at the band profile in this voltage range (see Figure 4.6), we find that it is the only voltage regime in which both electrons from the source and holes from the drain are present in the channel. Therefore, in the intrinsic channel recombination occurs between the two species and a recombination current starts to flow from source to drain. The effect is studied in more detail in Chapter 7.6.1 of reference [50] and in reference [51].

Forward bias - Positive V_{DS}

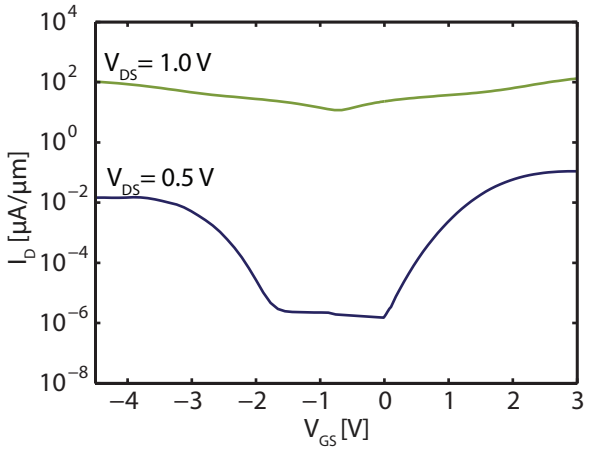


Figure 4.8. Simulated forward bias characteristics which takes into account Shockley-Read-Hall recombination including a field-enhancement factor introduced by Schenk. The simulation is done with the commercial device simulator Synopsys Sentaurus Device.

The general shape of the forward bias characteristics is similar to the reverse bias case, with a global minimum for medium gate-voltage levels and currents increasing for both larger negative and positive V_{GS}. The global minimum itself is composed of a sequence of two local maxima and three local minima that form an "M"-shape.

The existence of the global minimum cannot be explained with the WKB- or NEGF-calculations of Chapter 2 and Chapter 3. The measurements of Figure 4.5 correspond to a *p-i-n* diode at small forward bias ($V_{DS}< E_G$). In a ballistic model which does not account for regeneration and generation, the only allowed currents are majority carrier currents that overcome the built-in potential due to the doping profile (thermal emission currents). The built-in potential V_{bi} ($\approx E_G$ for

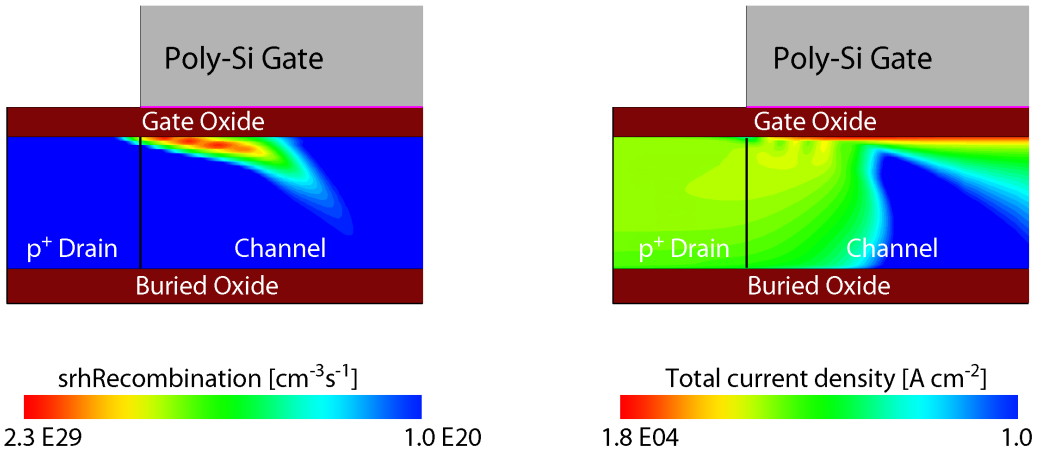


Figure 4.9. Left: 2D color plot of the Shockley-Read-Hall recombination rate at $V_{DS} = 0.5\,V$ and $V_{GS} = 3V$. Large recombination rates occur in the region of the highest electric fields at the drain/channel junction. *Right*: 2D color plot of the total current density under the same conditions. No current is flowing along the back oxide interface. The dominant current contribution is due to recombination.

degenerate doping) is compensated by V_{DS}, depending on V_{DS} either partially ($V_{DS} < V_{bi}$) or fully ($V_{DS} > V_{bi}$).
For the V_{GS} dependence, the ballistic model predicts that at large positive or negative V_{GS} one of the two majority currents becomes suppressed by the potential barrier in the channel. The other majority current can flow unobstructedly. For medium V_{GS}, the channel potential prevents neither of the two current flows. Hence, both currents are flowing and add up resulting in a global maximum.
Obviously, the measurement results are not consistent with the ballistic model. The missing factor is again a model for recombination and generation especially at the junctions. In *p-i-n* diodes that do not have a gate terminal, recombination currents are dominating the characteristics in the low-bias regime [38]. Recombination occurs all along the channel but the large electric fields at the junctions lead to an increased recombination rate. In the three-terminal TFET, a large V_{GS} leads to an increased electric field at one junction while the electric field at the other junction is reduced. Since the recombination rate depends higher than linearly on the electric field, the increase of the recombination current at one junction is not compensated by the decrease at the other junction. Therefore, I_D is expected to assume a global minimum at medium V_{GS} values, where the electric fields at the junctions are smallest and rise symmetrically for larger positive and negative V_{GS}.
Hence, the general trends of the measured curves are consistent with theory when assuming field-enhanced recombination as a dominating effect. For medium V_{GS}, recombination/generation does not occur at the source/channel or drain/channel junctions but in the depleted channel. The recombination/generation rates at the top and the floating bottom interfaces as well as in the channel volume are different, which is probably the reason for the complicated 'M'-shape.
To verify the general model, the output of the Sentaurus TCAD process simulation (see Figure 4.3 and Figure 4.4) is used as input for a Sentaurus device simulation. Transport is modeled by simultaneous solution of the Boltzmann transport equation in combination with the Poisson

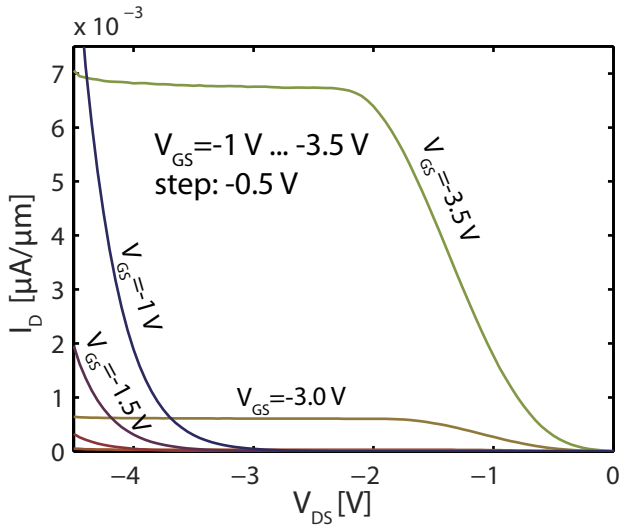


Figure 4.10. Output characteristics of an SOI-TFET with a gate length of 2 µm and a width of 30 µm on sample I201. The characteristics represent the p-mode operation where tunneling occurs at the *n-i*-junction, which features the steeper lateral doping profile.

equation. Additionally, a SHOCKLEY-READ-HALL recombination/generation model is included (see reference [52]) which assumes field-dependent carrier lifetimes according to a trap-assisted tunneling model by A. SCHENK (see reference [53]). Figure 4.8 shows forward bias characteristics for $V_{DS} = 0.5$ V and $V_{DS} = 1.0$ V. When comparing the $V_{DS} = 0.5$ V simulation with the experimental result, we find a good qualitative agreement. The 'M'-shape is not reproduced, but this is expected because the simulation does not account for surface recombination at the oxide interfaces.
Figure 4.9 shows on the left side a 2D plot of the recombination rate at $V_{GS} = 3$ V and $V_{DS} = 0.5$ V. The recombination rate reaches a value of $2.3 \cdot 10^{29}$ /(cm^3·s) in the region of the highest electric field (electric field is not shown). On the right side of Figure 4.9, the total current density is plotted and it confirms that there is no parasitic conduction at the buried oxide interface, but all conduction occurs via recombination.

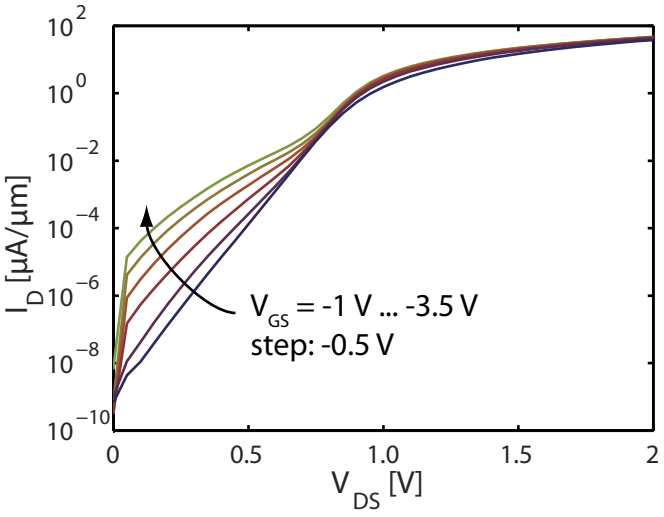


Figure 4.11. Forward bias I_D-V_{DS} characteristics resemble those of a *p-i-n* diode with a tunable recombination contribution.

4.3.2. Reference device - I_D-V_{DS}

Reverse bias - Negative V_{DS}
Figure 4.10 shows the output characteristics for p-mode operation of the reference device. For $V_{GS} = -3$ V and $V_{GS} = -3.5$ V, we observe an exponential onset followed by a linear rise and saturation around $V_{DS} \approx V_{GS} - 1.5$ V. The curve for $V_{GS} = -2.5$ V is not visible on this linear scale plot. For even smaller negative V_{GS}, an exponential rise is notable for large negative V_{DS}. This exponential rise is the signature of the ambipolar behavior of the TFET.

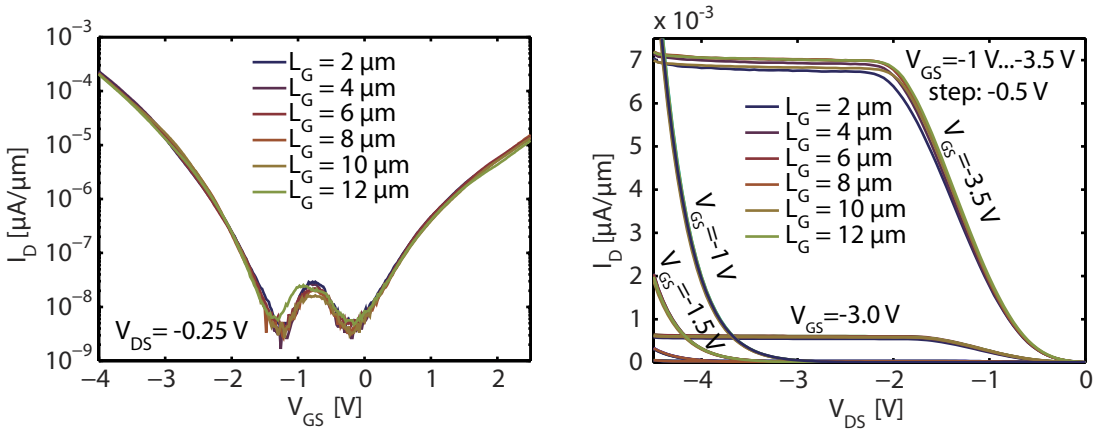


Figure 4.12. Left: Subthreshold characteristics for six different gate lengths L_G at V_{DS} = -0.25 V for the sample I201. No dependence on gate length is apparent. *Right:* Output characteristics for the same gate lengths at a V_{GS} ranging from -1 V to -3.5 V for the sample I201. On the linear scale of this plot, no systematic gate length dependence is visible. The variations are probably due to process variabilities.

As soon as the valence band in the channel is lifted above the conduction band in the source, band-to-band tunneling at the channel/drain-junction occurs which accounts for the exponentially rising current.

Forward bias - Positive V_{DS}

Figure 4.11 presents the forward bias I_D-V_{DS} characteristics of the reference TFET which are very similar to what is expected for a *p-i-n* diode [38]. For $V_{DS} < 0.7$ V, the curve at V_{GS} = -1 V shows nearly ideal diode behavior, which is determined by the majority carriers that overcome the built-in potential (thermal emission currents). At larger negative V_{GS}, the stronger electric field at the *n-i*-junction leads to an increased recombination current whose relative contribution becomes smaller for larger V_{DS}.

Above V_{DS} = 0.7 V, a small "ideal diode" region is visible. For larger V_{DS}, the channel resistance and the series resistance of source and drain $R_{S/D}$ start to limit the current flow. Assuming that the current flow at V_{DS} = 2 V is only limited by series resistance, a value of $R_{S/D}$ = 42 kΩ·µm can be extracted. However, in the following section, we will see that the conduction in the channel is also limited by scattering in the channel and a more accurate value for $R_{S/D}$ will be extracted.

4.3.3. Channel length dependence

One important parameter of the device geometry of MOSFETs is the gate length L_G. Due to diffusion-limited transport in the channel, the current in a MOSFET shows a $1/L_G$ dependence. For the TFET, the situation is different. Although scattering is present in the channel like in a MOSFET, usually the transmission of the tunnel barrier in the TFET is limiting the current. According to reference [4], the total transmission T_{tot} of a system with two scattering events is:

$$T_{tot} = \frac{T_{BTBT} T_{scat}}{T_{BTBT} + T_{scat} - T_{BTBT} T_{scat}}. \quad (4.1)$$

Here, T_{BTBT} is the band-to-band tunneling probability which forms the first scattering event and T_{scat} is the transmission that is related to scattering in the channel:

$$T_{scat} = \frac{l_{mfp}}{l_{mfp} + L_G}, \tag{4.2}$$

which is a function of the mean free path l_{mfp} and the gate length L_G. For a typical channel length in the micrometer regime, and a mean free path length of typically 10 nm, T_{scat} is of the order of 10^{-3}. If the tunneling transmission T_{BTBT} is much smaller than the scattering transmission, the total transmission is dominated by T_{BTBT}. In this case, we can expect the current of the TFET to be independent of the gate length.

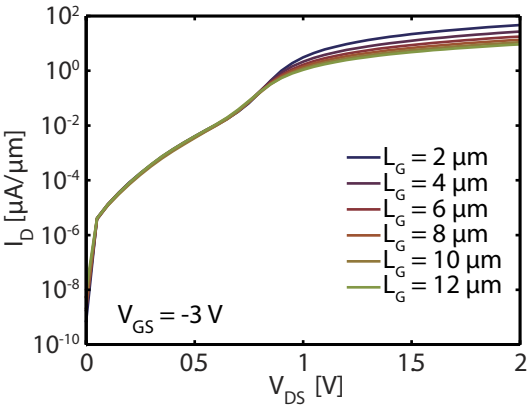


Figure 4.13. Channel length dependence of the forward bias TFET characteristics for the sample I201.

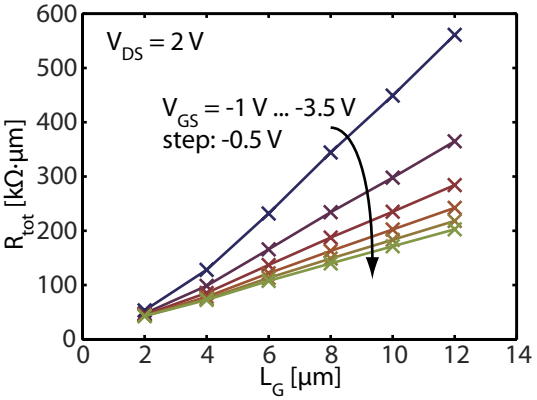


Figure 4.14. Total resistance of the TFET at $V_{DS} = 2$ V as a function of the gate length. The extrapolated value for $L_G = 0$ µm is the source/drain resistance $R_{S/D}$ and has a value of 9.1 kΩ·µm.

Subthreshold characteristics

The measured gate length dependence of the SOI TFET is demonstrated on sample I201 in Figure 4.12. On the left side, exemplary subthreshold characteristics are plotted which show no systematic dependence of the currents on the gate length L_G which varies from 2 µm to 12 µm in steps of 2 µm.

Output characteristics

On the right side of Figure 4.12, output characteristics are plotted for the same transistors with varying gate length of sample I201. In the output characteristics there is no systematic dependence on gate length visible. Neither the p-branch nor the ambipolar n-branch shows variations beyond those caused by process variability.

Forward bias - positive V_{DS}

Figure 4.13 shows the channel length dependence of the forward biased TFET for a single $V_{GS} = -3$ V. At $V_{DS} < 0.7$ V, the channel length has no impact on the characteristics, which is expected because this region is dominated by regeneration at the source/channel junction, an effect which does not depend on channel length. Also in the 'ideal' diode region at $0.7\ \text{V} < V_{DS} < 0.8$ V, the current does not depend on the channel length as expected. However, for $V_{DS} > 0.8$ V we observe a $1/L_G$ dependence that implies transport which is limited by scattering in the channel. Since at these large V_{DS} voltages, the built-in potential is compensated, the main

current contribution stems from majority carrier thermal emission currents which are diffusing across the low-field channel.
In Figure 4.14, the total device resistance R_{tot} is plotted as a function of gate length. For different V_{GS}, R_{tot} depends linearly on L_G, however, the device with $L_G = 2\,\mu m$ does not line up perfectly with the other devices. This is a consequence of the optical lithography, which for the smallest channel length is less accurate than for the longer channels.
To extract the source/drain resistance, a line is fitted to the $V_{GS} = -3.5\,V$ data and the resistance is extrapolated to zero gate length. Doing so yields $R_{S/D} = 9.1\,k\Omega{\cdot}\mu m$. Although $R_{S/D}$ is about two orders of magnitude higher than the requirement stated by the International Technology Roadmap for Semiconductors [5], which is $200\,\Omega{\cdot}\mu m$, the series resistance does not affect the performance of the TFET operation. For example, in the on-state (see Figure 4.10), the maximum current is $I_D = 7{\cdot}10^{-3}\,\mu A/\mu m$. The voltage drop caused by this current is $V_{RS/D} = 64\,\mu V$, which compared to a V_{DS} of -4 V is negligible.

Conclusion
As discussed before, the absence of any gate length dependence is a strong indication for the presence of band-to-band tunneling at both the *n-i-* and *p-i-*junctions. Furthermore, it is possible to deduce an upper limit for the band-to-band tunneling probability from eq. (4.1) and eq. (4.2). Since the variation of the gate current between the $L_G = 2\,\mu m$ and $L_G = 10\,\mu m$ is less than 1%, T_{BTBT} is smaller than 10^{-5}.

4.3.4. Impact of gate oxide thicknesses

Since the thickness of the gate oxide has direct impact on the electrostatic screening length Λ (see eq. (3.4)) at the tunnel junctions, two sets of samples with gate oxide thicknesses of 4.5 nm and 3.5 nm are fabricated. The thicknesses are measured by ellipsometry, therefore the associated measurement error is of the order of 3 Å-5 Å. The Λ-values as calculated from eq. (3.4) are

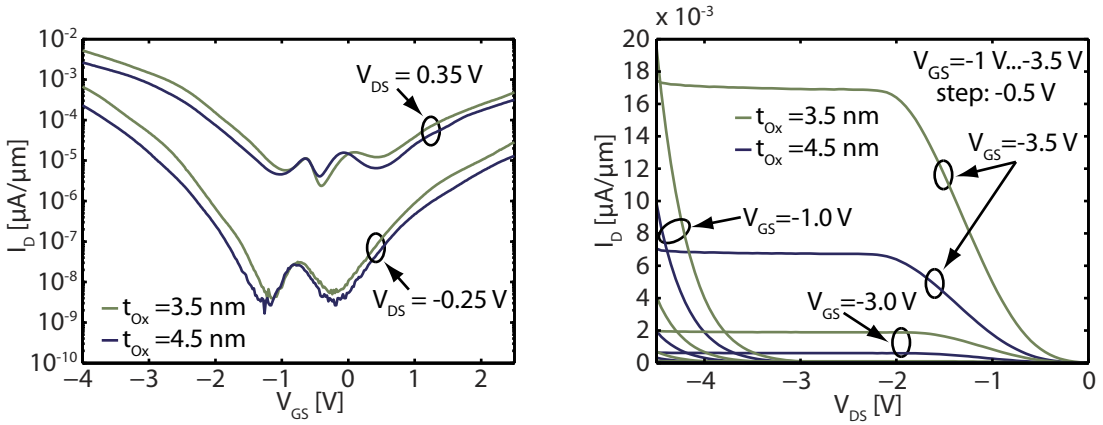


Figure 4.15. Left: Comparison of the subthreshold characteristics for two different gate oxide thicknesses t_{Ox}on samples I201 and It201. An increase of the current for the thinner dielectric is clearly visible. However, the subthreshold slope does not change significantly. *Right:* Comparison of output characteristics for different t_{Ox}. The on-state saturation current increases by a factor of 2.5 if t_{Ox} is reduced by 1 nm.

16.8 nm for the thicker oxide and 14.8 nm for the thinner oxide. A comparison of the measured subthreshold and output characteristics is presented in Figure 4.15.

Subthreshold characteristics
On the left side of Figure 4.15, the subthreshold characteristics illustrate that indeed a reduced gate oxide thickness t_{Ox} leads to improved drain currents for both, positive and negative V_{DS}. Depending on V_{GS}, the current increases by a factor of three to four. The one-dimensional WKB-model (see eq. (2.9)) predicts an increase of the current at the given bias conditions of a factor of five to six, which for such an idealizing model is in good agreement with the measurements.
For the subthreshold swing, an improvement of about 10% is expected from the WKB-model. However, the measurements do neither confirm nor disprove this hypothesis because the measurement accuracy is not sufficient to extract such a small change.
The variation of the gate oxide thickness enables the review of the explanation for the hump that occurs in the off-state of the device around $V_{GS} \approx -0.8$ V. The gate leakage current is about a factor of 2-3 larger in the off-state of the thin oxide devices (not shown), but the shape and the maximum value of the hump do not reflect this change. This observation confirms the initial analysis that gate leakage is not responsible for the parasitic effect.
The forward bias ($V_{DS} > 0$) current depends in a similar way on t_{Ox} as the reverse bias current, which is again an indication for a field enhancement of the recombination/generation rates at the junctions.

Output characteristics
The impact of the gate oxide thickness on the output characteristics is shown on the right side of Figure 4.15. Both sets of curves show distinct saturation and similar to the subthreshold characteristics, the improvement in the on-currents is clearly noticeable. The onset is exponential in both cases and saturation occurs at similar saturation voltages.

4.3.5. Impact of doping concentration and doping profile

As we have seen in Section 3.4.2, the doping concentration is an important parameter which determines the TFET performance by two means. First, the built-in voltage V_{bi} increases with a larger doping concentration and second, the potential screening length in the source and drain λ_{dop} reduces with larger doping concentration.
The preferred method to approximate λ_{dop} is to use the lateral doping profile as input for a device simulation that yields the potential profile of the structure and then extract λ_{dop}. However, as a first approximation which does not include the doping profile but the doping concentrations only, the depletion widths in a *p*-*n*-junction can be used:

$$d_{n/p} = \sqrt{\frac{2\epsilon\epsilon_0}{q} \frac{N_{A/D}}{N_{D/A}(N_{D/A} + N_{A/D})} V_{dep}}\,. \tag{4.3}$$

Here, $d_{n/p}$ is the depletion width in the respective *n*- or *p*-region with a donor or acceptor concentration of $N_{D/A}$ and a voltage across the junction of $V_{dep} = q\phi_f^0 + V_{bi}$, which corresponds to the potential drop at the source/channel junction including the effect of the gate voltage. Assuming an abrupt doping profile, the depletion width on the n-side can be calculated for $V_{dep} = -4$ V with a source donor concentration of $N_D = 2 \cdot 10^{20}/\text{cm}^2$ (see Figure 4.3) and a channel carrier concentration of about one order of magnitude less $N_A \approx 2 \cdot 10^{19}/\text{cm}^2$ (approximated value from 2D NEGF-

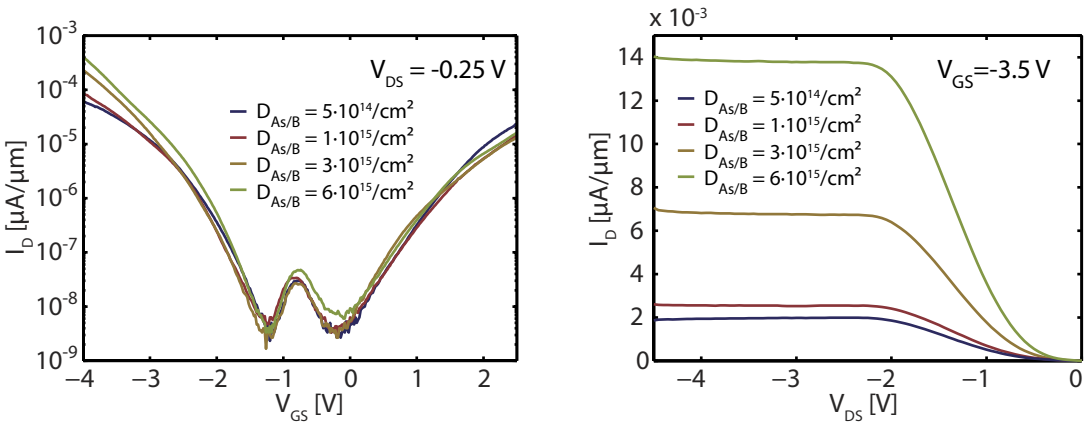


Figure 4.16. Left: Impact of the implantation dose on the transfer characteristic of the TFET samples I201, I202, I203, I204. *Right:* p-mode output characteristics plotted for the same samples.

simulations, see Figure 3.21) to a value of $d_n \approx 1.5$ nm. The depletion widths and the activation levels for different doping concentrations, extracted from the process simulations of Section 4.2.2, are summarized in Table 4.2.

	I201	I202	I203	I204
$D_{As/B}$	$3 \cdot 10^{15}$ /cm²	$1 \cdot 10^{15}$ /cm²	$5 \cdot 10^{14}$ /cm²	$6 \cdot 10^{15}$ /cm²
N_D(As)	$4.00 \cdot 10^{20}$ /cm³	$2.38 \cdot 10^{20}$ /cm³	$1.94 \cdot 10^{20}$ /cm³	$4.50 \cdot 10^{20}$ /cm³
d_n	0.78 nm	1.30 nm	1.58 nm	0.70 nm

Table 4.2. Implantation doses $D_{As/B}$, simulated doping concentrations N_D(As) and calculated depletion width d_n for the As junction.

Subthreshold characteristics

Experimental subthreshold characteristics for four different samples (see Table 4.2) are plotted in Figure 4.16 on the left side. The comparison shows a clear trend for the p-branch at negative V_{GS}, namely a larger doping concentrations leads to higher device currents. This observation meets the expectations from the simulation of Section 3.4.2. The minimum inverse subthreshold slopes do not show a doping dependence. However, the average inverse subthreshold slope improves for larger doping concentration. This means that there is no indication for a too large doping concentration which could deteriorate the subthreshold slopes as predicted in Section 3.4.2.

For the n-branch at positive V_{GS}, there is no meaningful trend, probably because of activation issues in the case of Boron. Especially for large doses, effects like dopant cluster formation or out-diffusion can lead to an unexpected dependence of the active doping concentration on the implantation dose, which is not included in process simulations.

Comparing the *n*- and *p*-branches, we note that the variation of doping concentration cannot explain the substantial difference between the inverse subthreshold slopes of both branches. Therefore, the most probable reason for the different slopes is the broader lateral doping profile at the boron-junction (see Figure 4.4), which increases λ_{dop} drastically. This observation leads

to the conclusion that of all studied parameters, the steepness of the lateral doping profile is the dominant factor which limits the TFET performance.

Output characteristics
The corresponding p-mode output characteristics are shown in Figure 4.16 on the right side. For all doses, saturation is distinct and it occurs at the same V_{DS}, which also shows that the threshold voltage variation between the samples is small. The current levels line up according to the implantation doses as expected.

4.4. Conclusion

P-i-n Tunnel FETs on SOI-substrates have successfully been fabricated and electrically characterized. The electrical characterization of a reference device reveals an inverse subthreshold slope of $S \approx 320\,\mathrm{mV/dec}$ and on-currents of the order of $I_D \approx 10^{-2}\,\mu\mathrm{A}/\mu\mathrm{m}$. These results are similar to what has been reported in [27] and demonstrate that there is a need for optimization of the SOI TFET in order to compete with MOSFETs, which feature inverse subthreshold slopes close to the thermal limit ($S = 60\,\mathrm{mV/dec}$) and on-currents of the order of $I_D \approx 10^{3}\,\mu\mathrm{A}/\mu\mathrm{m}$.
Yet, the studied variations of gate dielectric thickness and doping concentration show paths for further optimization.
A reduction of the gate dielectric thickness of about 1 nm leads to an increase of the current by a factor 2.5. The gate dielectric thickness can be seen as an indication of how the scaling of Λ affects the device performance, which is in good agreement with the simulation results and confirms that Λ should be minimized.
The experimental results on the impact of the doping concentration on device performance confirm the simulation results which predict increasing currents for larger doping concentrations. However, with the present data it is not possible to verify the impact of the doping concentration on the subthreshold slope which is predicted to deteriorate for larger doping concentrations. Another important observation concerns tunneling at the *p-i*-junction compared with tunneling at the *n-i*-junction. The minimum inverse subthreshold slope at the boron junction is with $S \approx 500\,\mathrm{mV/dec}$ much worse than the slope for tunneling at the arsenic junction. Since the doping concentration does not affect the slope substantially, the worse inverse subthreshold slope can be attributed to the broad lateral doping profile at the boron junction. A key to optimize both S and I_D is therefore the suppression of dopant diffusion.
In summary, it is still unknown whether ultra-thin body SOI TFETs could overcome the 60 mV/dec limit at current levels that are usable for digital switching applications, but the presented results give indications on what aspects the technological development should focus in the future.

Chapter 5
Si Nanowire TunnelFETs

5.1. Introduction

The results on planar SOI TFETs presented in Chapter 4 suggest that there are two main parameters which limit the tunneling performance of the devices. The first is the sub-optimal channel electrostatics and the second is the spreading of the doping profiles at the source/channel junction. Both issues are addressed in the following chapter.
To optimize the channel electrostatics, the planar SOI geometry is replaced by nanowires which are fabricated using a top-down approach. Subsequently, the steepness of the doping profile is improved by suppressing dopant diffusion during the annealing step. This suppression is implemented by replacing the spike annealing step by laser annealing.
This chapter consists of four parts: First, the fabrication of silicon nanowires is described and second, nanowires which are fabricated using spike anneal are characterized and discussed. In the third part, laser annealing is introduced and laser-annealed devices are compared to the spike-annealed reference device. In the fourth and last part, the devices with the best performance are characterized at temperatures as low as 4.2 K in order to improve the understanding of the underlying device physics.

5.2. Fabrication

5.2.1. Process flow

Although the basic structure of the process flow for the nanowire TFETs is identical to the planar SOI case, many steps have been changed to be compatible with e-beam lithography that provides a superior resolution

1. ***Preparation of the substrates.*** The starting substrates are either SOI substrates with a body thickness of 20 nm and a buried oxide thickness of 100 nm or sSOI substrates with a body thickness of 15 nm and a buried oxide thickness of 140 nm. The original wafers are cut into two-by-two centimeter sized pieces.
2. ***Mark definition.*** In this step, marks for optical and e-beam lithography are defined. The e-beam marks are 700 nm deep, 15x15 μm^2 large, quadratic pits in the substrate which are

patterned by optical lithography and etched by reactive ion etching using SF_6. This design of the e-beam marks is adopted from S. Feste [54], but the number of optical lithography steps is reduced from two to one, because the first lithography step is intended to eliminate material contrasts and to avoid etching problems due to poor selectivity between substrate and resist. However, after optimization of the reactive ion etching recipe, the selectivity is large enough to skip the first lithography step and instead directly pattern the substrate using a two-step etching recipe. The first etching step employs a CHF_3 plasma to etch the top Si layer and the buried oxide. In the second step, the gas is switched to SF_6 to perform a deep Si etch for about 500 nm.

3. ***Wire definition (see Figure 5.1 (a)).*** With marks present on the substrate (not shown), a 70 nm thick layer of Polymethylmethacrylat (PMMA) is applied to the surface followed by a pre-bake to evaporate the solvent component of the PMMA. Then, the nanowire pattern is written on the substrate using a small e-beam current of I_B=100 pA, a beam energy of 50 keV and a spatial resolution of 2 nm. The wires have a length of 1 µm and three different widths, approximately 12 nm, 20 nm and 35 nm. The large anchor regions which act as source and drain are defined by writing a 2 µm wide isolation trench around the anchors, which serves the purpose of isolating neighboring transistors. Compared to the planar case, this step is optimized, because the silicon anchor regions are now larger in order to accommodate the

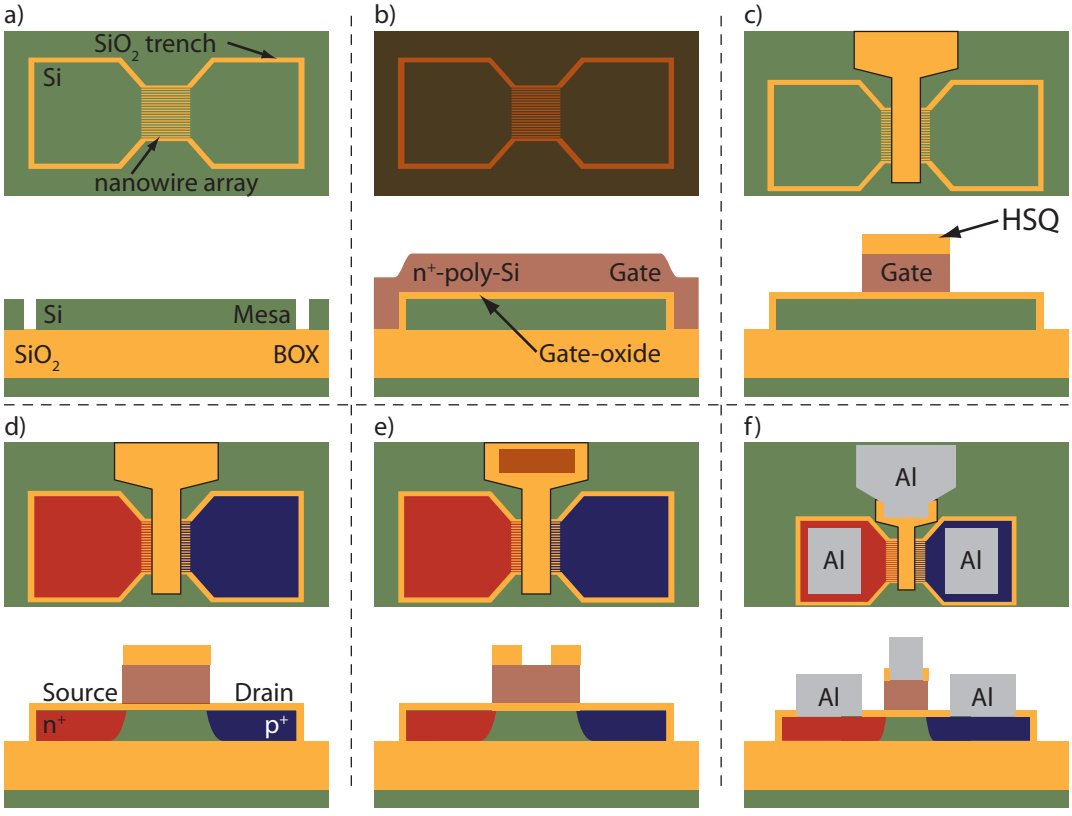


Figure 5.1. Process flow for nanowire based TunnelFETs. For each step, the top view and a cross-section along transport direction are shown. For clarity, the native Si oxide and the gate oxide are not shown in the top-view.

large metal pads that are needed for later contacting. This design avoids metal passing from the top silicon down to the buried oxide, which otherwise can create leakage paths. In the planar design of Chapter 4, this problem is solved by depositing a final isolation oxide, a step that now is omitted. Following the exposure, the PMMA is developed and the structures are transferred to the native oxide layer by a short HF dip. Directly after the HF dip, the sample is loaded into the ICP reactive ion etching machine and the top silicon layer is etched at -150 °C in an SF_6-plasma.

4. ***Gate oxidation (see Figure 5.1 (b)).*** The patterned wires undergo an RCA cleaning procedure to prepare the surface for gate oxidation (see [47]). Immediately after the standard clean 2 - step (HCl/H_2O_2 chemistry) of the cleaning, the samples are oxidized in a rapid thermal oxidation processor at 800 °C for 30 min in O_2-atmosphere to grow an SiO_2 gate dielectric of about 4.2 nm, as measured by ellipsometry.
5. ***LPCVD Poly-Si deposition.*** To minimize contamination, within 30 minutes after the oxidation, the gate material, 200 nm thick n^+-doped poly-Si is deposited by a low-pressure chemical vapor deposition (LPCVD) method.
6. ***Poly-Si activation.*** To ensure a high level of activation in the poly-Si gate, the samples are annealed at 950 °C for 1 min in pure N_2 by means of a rapid thermal processor.
7. ***Gate patterning (see Figure 5.1 (c)).*** Because the length of the nanowires is about 1 µm, the alignment of the gate cannot be done by optical lithography. Instead, e-beam lithography is used again, but this time the negative resist Hydrogen Silsesquioxane (HSQ) is employed because HSQ can directly be used as an etch mask in the subsequent etch. After developing the HSQ, the gate is etched in the ICP reactive ion etching machine using pure HBr for a short time to break through the native oxide. Then a small fraction of O_2 is mixed with the HBr to increase the selectivity and stop on the gate oxide (see [48]).
8. ***Source/drain implantations (see Figure 5.1 (d)).*** With the gate in place, self-aligned implantations for source and drain are performed with a dose of $1{\cdot}10^{15}/cm^2$ for arsenic at an energy of 5 keV and for boron at an energy of 1.5 keV. Since apart from the small width, the nanowire structure does not differ from the SOI TFETs of Chapter 4, the implantation conditions have been adopted from the SOI TFETs. Again, the optical lithography of the planar case is replaced by e-beam lithography for sake of resolution. As any organic resist that is exposed to ion implantation is heavily damaged, a 650 nm thick PMMA is used as etch mask. This large thickness ensures that underneath a severely damaged layer of resist, enough resist stays undamaged to lift-off the damaged top part. While this lift-off works perfectly well for the wires, large areas do not lift-off without leaving resist residues. Therefore, after the lift-off, the sample is exposed to a high-power, low-bias O_2 plasma in the ICP etcher which removes all resist residues.
9. ***Annealing and activation.*** To incorporate the implanted ions in the crystal lattice and cure lattice damage created during implantation while minimizing dopant diffusion, three different annealing techniques are compared. A 1000 °C spike anneal in pure N_2 serves as reference, while the other two methods utilize short-pulse laser annealing at a wavelength of λ=248 nm and a pulse duration of t=28 ns. Each sample consists of four quadrants which are annealed at four different energies in order to find the optimum pulse energy. One sample is only laser-annealed at energies of 45 mJ/cm², 50 mJ/cm², 55 mJ/cm² and 60 mJ/cm². A second sample is laser-annealed at energies of 60 mJ/cm², 65 mJ/cm², 70 mJ/cm² and 75 mJ/cm² followed by a rapid thermal anneal at 650 °C for 5 min. More details on the laser annealing process are given in Section 5.2.1.

10. ***Opening of gate via (see Figure 5.1 (e)).*** On top of the gate stack, the protecting HSQ layer is still in place and needs to be removed in order to guarantee a good electrical contact to the poly-silicon underneath. The lithography for this process step is done optically, because the alignment accuracy is less important. The HSQ is etched using a highly selective CHF_3 plasma at 0 °C in the ICP etcher.
11. ***Metallization (see Figure 5.1 (f)).*** Due to the lack of an appropriate optical mask, the lithography for the large source/drain pads is done with e-beam lithography and 650 nm thick PMMA as resist. Immediately before the deposition of 200 nm aluminum by electron beam evaporation, the samples are dipped in HF to remove the native SiO_2 layer on the silicon surface and thus ensure a good electrical contact. After the deposition, the metal is patterned by acetone lift-off.

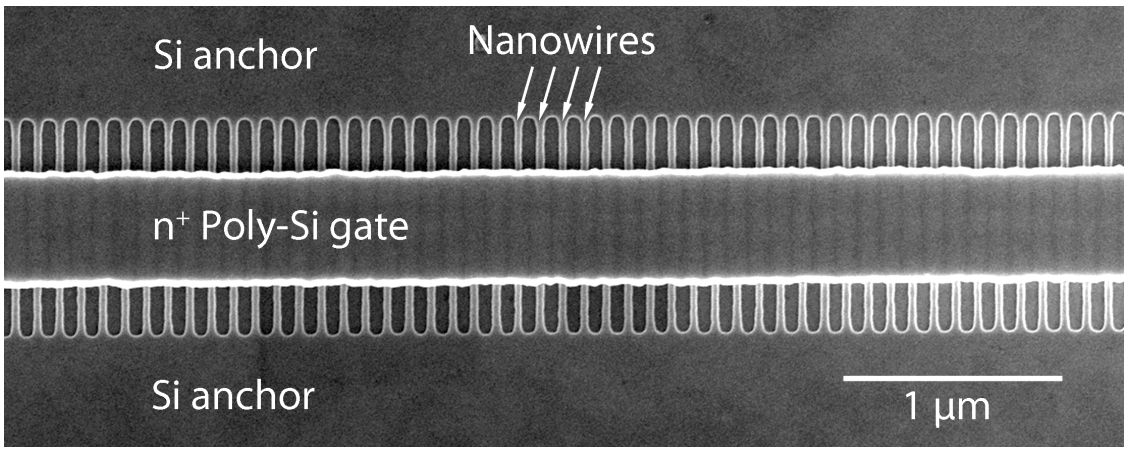


Figure 5.2. Scanning electron micrograph: Detail of a readily fabricated nanowire array transistor with 400 parallel wires. The nominal wire width is 20 nm and the pitch is 100 nm.

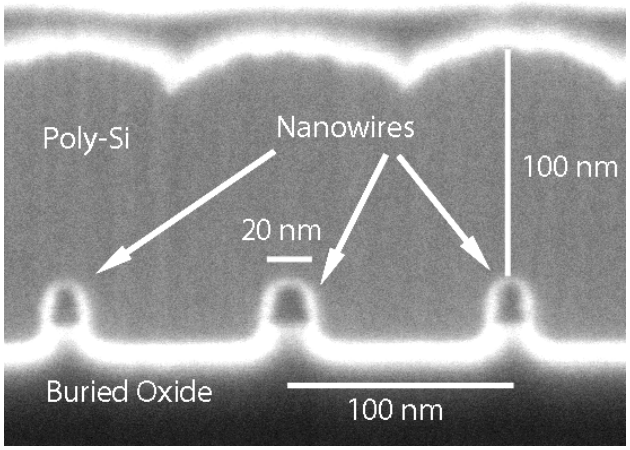


Figure 5.3. Cross-section scanning electron micrograph of a lamella prepared using a focused ion beam technique. Three nanowires which are conformally covered with Poly-Si.

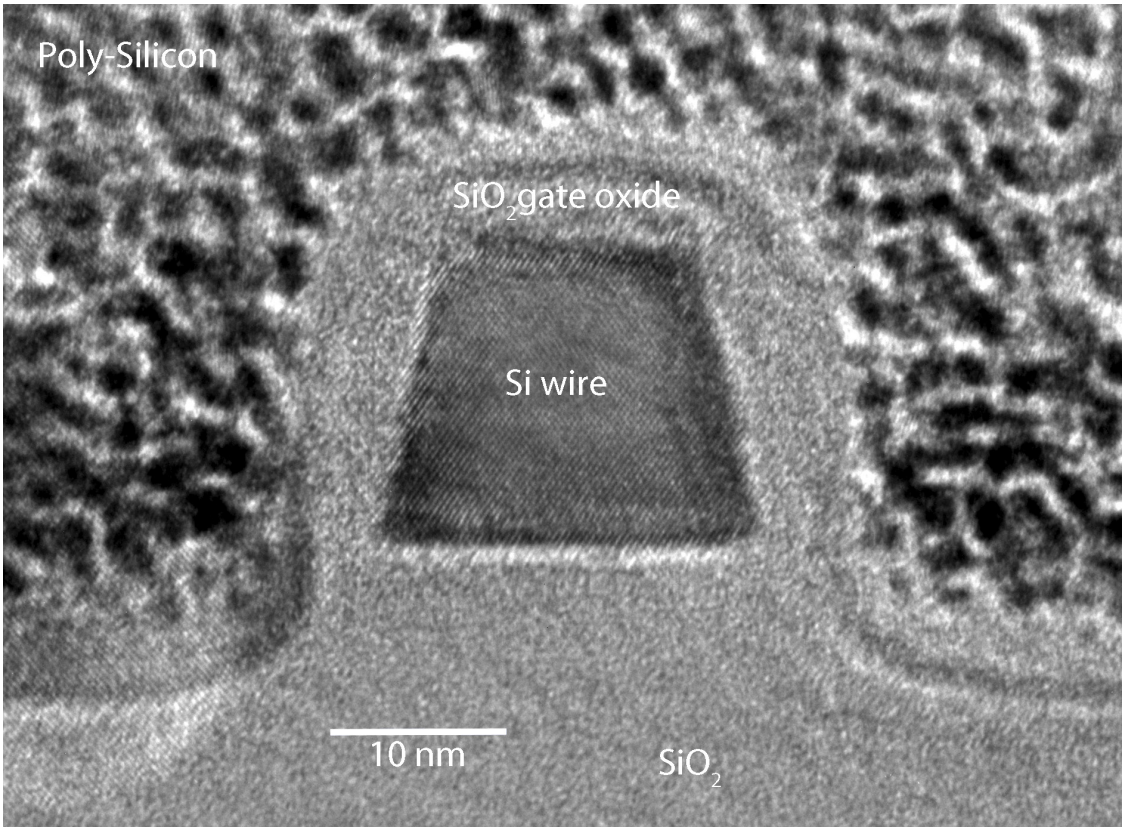


Figure 5.4. Transmission electron micrograph of a single nanowire. The Si wire is surrounded by an SiO_2 gate oxide and the Poly-Si gate.

Figure 5.2 shows a top view scanning electron micrograph of the readily fabricated structure. Figure 5.3 is a scanning electron micrograph which is done on a lamella prepared by focused ion beam. The image gives an overview of three wires and shows that the poly-Si gate material conformally covers three sides of the wires. Figure 5.4 shows a detail of one wire which demonstrates that the width and height of the wire are about 20 nm.

5.2.2. Electron beam technology

The fabrication of nanowire arrays makes heavy use of electron-beam (e-beam) lithography to achieve the desired small dimensions. Although nowadays e-beam lithography is in a mature state, the proper choice of resist thickness, exposure parameters and subsequent etching still poses a challenge for the process developer.
The structuring of a nanowire array with wire widths of < 30 nm and a pitch of about 100 nm from (s)SOI substrates can be broken into three steps:

1. ***E-beam lithography.*** After spinning an electron sensitive resist onto a substrate and baking it, the resist is exposed by an electron beam which exposes all areas that are defined in a mask file. Similar to optical lithography, the exposure dose is subject to optimization. Depending on the e-beam, resist and substrate properties, some parts of the pattern can be over- or under-exposed even with an optimal dose due to the 'proximity-effect'. This effect is a consequence

of the electron beam, which does not only expose the areas it is supposed to expose, but it also delivers a small fraction of the dose within a larger radius, that can be as high as 50 µm. Therefore, neighboring structures can influence the exposure dose of each other. After exposure, the resist is developed and ready for the wire etching.

2. ***Etching.*** Depending on the resist system, either the top silicon layer of the SOI directly or a hard mask and subsequently the top SOI-layer is patterned, usually by reactive ion etching. The result is a nanowire with residual resist or hard mask on top.

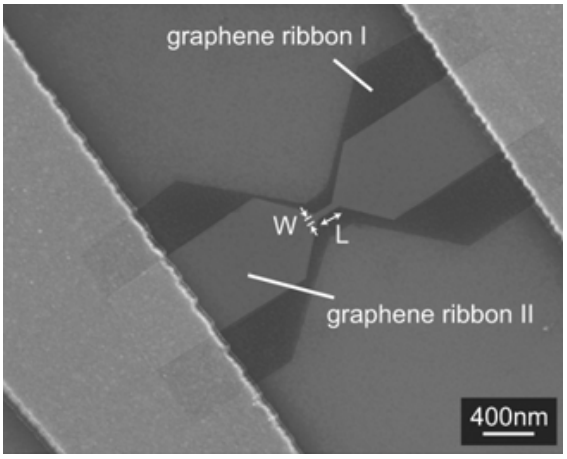


Figure 5.5. Two graphene nanoribbons that are patterned consecutively from the same graphene flake using HSQ/PMMA double-layer resist [52].

3. ***Resist removal.*** The resist or hard mask is removed to make a gate oxidation possible.

The final process which is used for the nanowire fabrication in this thesis was developed in an iterative way in the following three steps:

1. The starting point was an already existing process [54], which relies on a dual-layer resist system, composed of the optical resist AZ5206 and an HSQ top layer. HSQ is a negative e-beam resist, whose properties resemble those of SiO_2 after development. A single HSQ layer provides high resolution and stability, it can only be removed using HF, which has the disadvantage that during the removal of the resist, the buried oxide is exposed to HF, which leads to an undesired underetch of the wire. To avoid these problems, the sacrificial AZ5206 interlayer, which facilitates an acetone lift-off of the HSQ, was added. Although the additional layer step increases the sidewall roughness, in combination with an SF_6 reactive ion etch, the patterning works well for single nanowires with widths > 35 nm. For thinner wires and especially nanowire arrays, this approach meets its limitations that are given by the proximity effect. Due to the sacrificial AZ5206 layer, the resist stack reaches a height of about 300 nm. The resulting aspect ratio of > 10:1 leads to enhanced proximity effects and limited resolution. Although by means of a proximity effect correction algorithm this problem might be solved, it was not attempted because no algorithm was available.
2. In order to increase the resolution, the dual-layer resist stack was thinned to 120 nm by replacing the thick AZ5206 layer with a PMMA layer. PMMA itself is a positive high-resolution e-beam resist which is available in different dilutions that can provide layers as thin as 60 nm. The sacrificial PMMA layer provides an improved resolution compared to the AZ5206 layer for single wires. This resist system has also been adopted to pattern graphene nanoribbons down to a width of 15 nm (see Figure 5.5 and ref. [55]). For nanowire arrays with < 300 nm pitch, however, the proximity effect is still a limiting factor.
3. Since thin PMMA layers provide a high e-beam resolution, the third attempt was focused on PMMA. Yet, any plasma etching process, including the SF_6 process that works well with an HSQ mask, exhibits a very high etch-rate for PMMA, which means that the etch mask is completely etched before the nanowires are readily patterned. This problem can be solved by making use of an SiO_2 interlayer, which is patterned using PMMA and HF. Because any thick

interlayer would again deteriorate the resolution and increase proximity effects, a very thin interlayer, namely the native SiO_2 layer, present on any Si surface that has been exposed to air, is used. To make the native SiO_2 layer usable as a hard mask, the wire etch recipe needs to be highly optimized for the selectivity between Si and SiO_2. With an ICP machine (Oxford Plasmalab 100, RIE4 FZ Jülich), a receipe based on a low-temperature SF_6-process was successfully adopted to pattern the wires. As the SF_6-plasma provides a very large etch rate for PMMA, even the resist removal step after the wire etch can be skipped.

The results of the last evolution step (see Figure 5.6) show no proximity effects at all and in comparison to the other approaches very small edge roughness in combination with small wire widths. The thinnest wires that were produced reproducibly have a width of about 10 nm. Using the presented recipe, wire arrays with 200 and 400 wires are produced with a pitch of 100 nm and a width of about 20 nm. The doubling of the number of wires and therefore the total device width does neither change the e-beam dose nor the width of the resulting wires. This indicates that the e-beam recipe could scale very well to arrays with a huge number of wires.

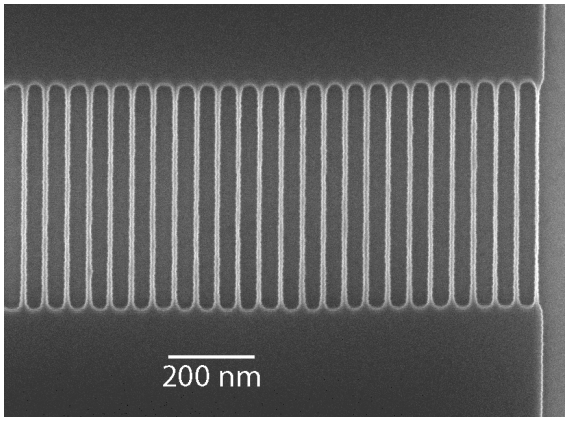


Figure 5.6. 20 nm wide silicon nanowires with a pitch of 100 nm after patterning using PMMA. No proximity effect is visible.

5.1. Room temperature characterization results and discussion

The readily fabricated nanowire devices are characterized using a Keithley 4200-SCS semiconductor parameter analyzer which is capable of measuring I_D-V_{GS} and I_D-V_{DS} characteristics automatically with high accuracy. Similar to the systematics of Chapter 4, first a reference device is discussed and than the different variations are studied in more detail.

The reference device consists of 400 nanowires with a width of about 20 nm per wire, a wire length of 1 µm and a channel length of about 450 nm. The pitch is 100 nm, however, the measured currents are normalized to the total width of the silicon wires, which is 400·20 nm = 8 µm. The implantation dose is $D_{As/B} = 1 \cdot 10^{15}/cm^2$ and the dopant activation is carried out as a spike-annealing at 1000 °C in a rapid thermal processor (RTP).

5.1.1. Reference device I_D-V_{GS}

Reverse bias - Negative V_{DS}

The reverse bias subthreshold characteristics of the reference device are shown in Figure 5.7. In this plot, the small difference, which becomes visible for small negative V_{DS} in the off-state at V_{GS} = -1 V is due to gate leakage, since $I_S+I_D+I_G$ is zero in a three-terminal setup.

The main features of the characteristics are very similar to the SOI TFETs which are presented in Chapter 4. The minimum inverse subthreshold swing is about 375 mV/dec at V_{GS} = -1.3 V, which is about 17% worse than the SOI reference device. The maximum current for V_{GS} = -4 V is I_D = 1 µA/µm, which is close to the SOI TFETs. The minimum current levels are about 10^{-8} µA/µm, which is about an order of magnitude higher than in the SOI reference device.

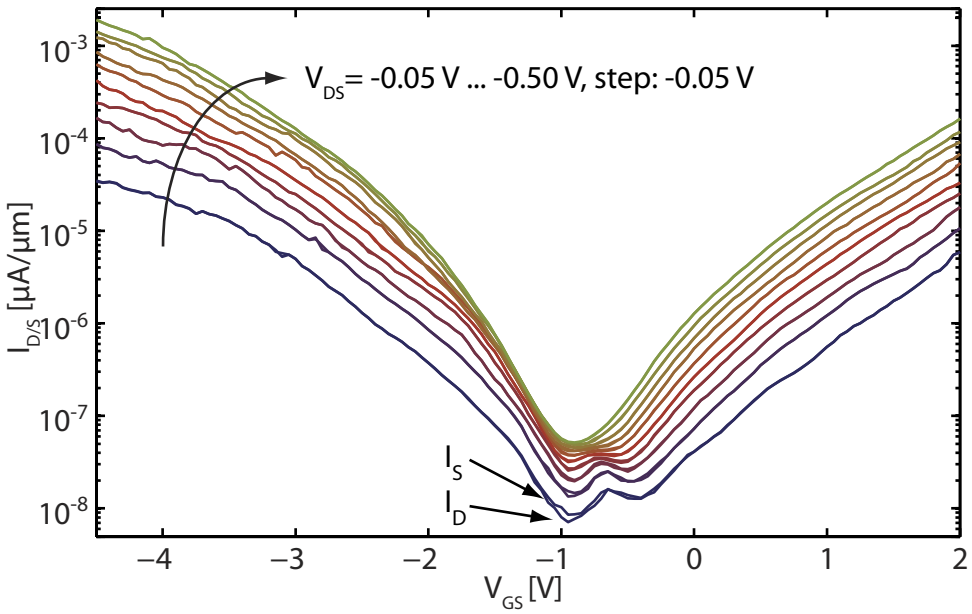


Figure 5.7. Transfer characteristics of a reference nanowire TFET consisting of 400 parallel wires with a width of 20 nm each. I_D and I_S are both plotted on top of each other in the same color. The implantation dose is $D_{As/B} = 1 \cdot 10^{15}/cm^2$ and the activation is done by spike annealing at 1000 °C.

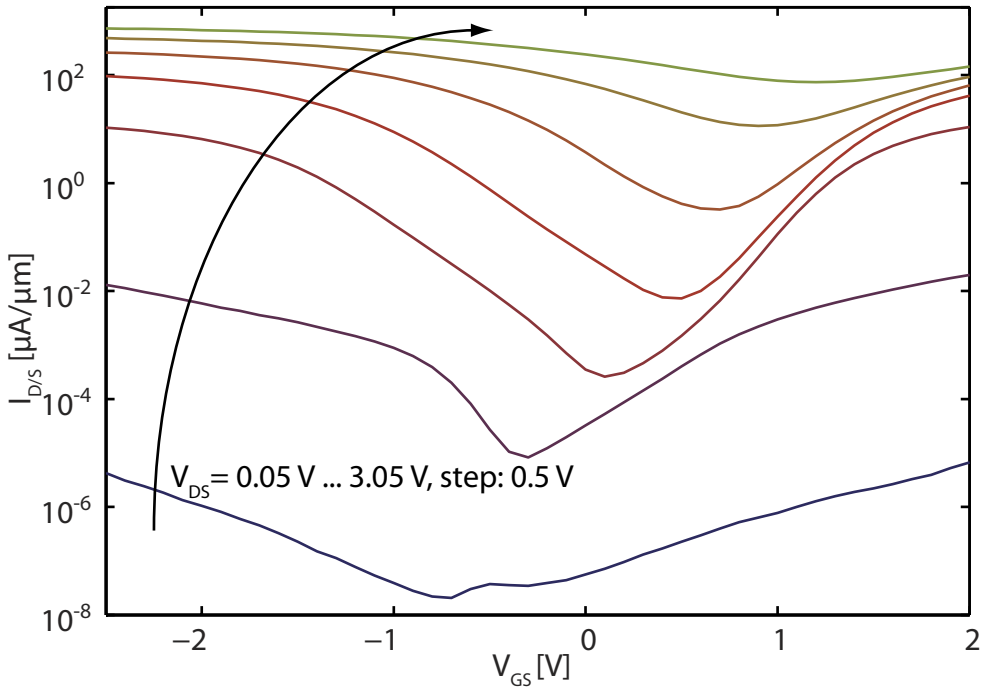


Figure 5.8. Forward bias subthreshold characteristics for the reference nanowire device. For larger V_{DS}, the current becomes more and more independent from V_{GS}.

For positive V_{GS}, the minimum inverse slope is about 475 mV/dec at V_{GS} = -0.35 V, which is an improvement of 5%. The maximum current at V_{GS} = 2 V increased by a factor of six to I_D = 1.4 µA/µm. Furthermore, we observe that the 'distance' in V_{GS} between the onset of the p-branch and n-branch reduced by about 0.6 V. This result is probably a consequence of an apparently larger tunneling probability at the *p-i* junction, which is also indicated by larger currents and a smaller inverse subthreshold slope of the *n*-branch.
The reason for an increase of the n-branch current while the p-branch current does not change compared to the SOI devices is probably a small change in the spike anneal recipe. While for the SOI devices, the spike anneal temperature profile includes a ramp from T = 650 °C to T = 1000 °C in a time of t = 10 s, the time is reduced for the nanowires to t = 4 s. The much shorter time at elevated temperatures is expected to reduce the boron diffusion significantly, which leads to a steeper junction and therefore to increased tunneling.
Moreover, it is interesting to note that the reduction of the channel length compared to the SOI devices does not affect the currents. An observation which agrees well with the channel length dependency studied in Section 4.3.3.

Parasitic currents
In contrast to the SOI reference TFET, the 'hump' in the off-state is much less visible. However, the current level at the local maximum of I_D(lmax)≈ 3 µA/µm is very close to what is observed for the SOI TFET. In the nanowire case, the local maximum is less apparent because the off-state current levels are larger due to the increased n-branch current.

Forward bias - Positive V_{DS}
The subthreshold characteristics for forward bias is plotted in Figure 5.8. For small positive V_{DS}, the curves look similar to Figure 4.5, which covers the small V_{DS} range only. For medium V_{GS} values around $V_{GS} = 0$ V, there is a minimum, from which the current rises for both positive and negative V_{GS}. As detailed in Section 4.3.1, an inverted shape with a distinct maximum for medium V_{GS} values is expected from the models used in Chapter 2 and Chapter 3. But again, the explanation for the deviation is a field-enhanced recombination which occurs at the source/channel and drain/channel junctions. Since Figure 5.8 covers a larger V_{DS}-range, we observe that the global minimum becomes less pronounced for larger V_{DS}. This behavior is understandable because for large enough V_{DS}, the built-in voltage of the *p-i-n* structure is overcome and majority carrier currents account for a larger fraction of the total current. The 'M' shape in the minimum, which is observed in the planar SOI case (Figure 4.5) is absent in the nanowires. This is probably a result of the improved gate control in the nanowire. Since the top-gate surrounds the wire on three sides, the backside interface potential does not float any more and the double peak vanishes. For V_{DS}=0.05 V, a single maximum remains.

Conclusion
The improvement which is expected due to the better gate control in nanowires is not observable in the experimental data. That means there have to be other reasons that prevent a performance improvement. Since in both cases a spike anneal was employed for dopant activation, it seems reasonable that the tunneling probability is limited by the lateral doping profiles.

5.1.2. Reference device - I_D-V_{DS}

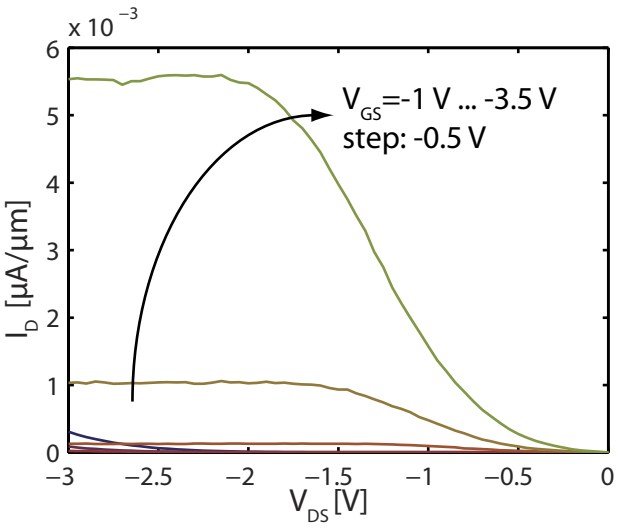


Figure 5.9. Output characteristics of a nanowire-TFET with a gate length of 450 nm and a width of 8 µm. The characteristics represent the p-mode operation where tunneling occurs at the *n-i*-junction, which features the steeper lateral doping profile.

Reverse bias - Negative V_{DS}
Figure 5.9 shows the p-mode output characteristics of the reference device. A comparison to Figure 4.10 reveals many similarities between the nanowire and the SOI devices. The maximum current level is identical within 10% and the saturation sets in at similar V_{DS} voltages for the same V_{GS}. However, the ambipolar conduction for V_{GS} = - 1 V sets in around $V_{DS} \approx$ -2.3 V, which is about 0.7 V more positive than for the SOI TFET. This again is an indication for a larger tunneling probability at the *p-i*-junction, which is probably due to the reduced annealing time for the wires.

Forward bias - Positive V_{DS}
Under forward bias, the nanowire TFET behaves again very similar to the SOI TFET. For $V_{DS} < 0.8$ V, the

otherwise ideal diode behavior is complemented by a contribution due to field-enhanced recombination at the *n-i-* and *p-i-*junctions. Despite the shortened channel length, the current for larger V_{DS} seems still limited by scattering in the channel, because the ballistic length scale for silicon is two orders of magnitude smaller than the channel length.

For $V_{GS} = -1$ V and $V_{GS} = -1.5$ V, the current for $V_{DS} > 0.7$ V seems suppressed when compared to the other gate voltages. The origin of this effect is still unclear. However, a probable explanation is that even with V_{DS} compensating the built-in voltage V_{bi}, at the junctions there are small potential barriers remaining. In the TFET device at hand, a large negative V_{GS} would therefore lower the potential barrier for holes injected from the drain into the channel. Only for small negative V_{GS}, this potential barrier would reduce the current flow, which is consistent with the measurements.

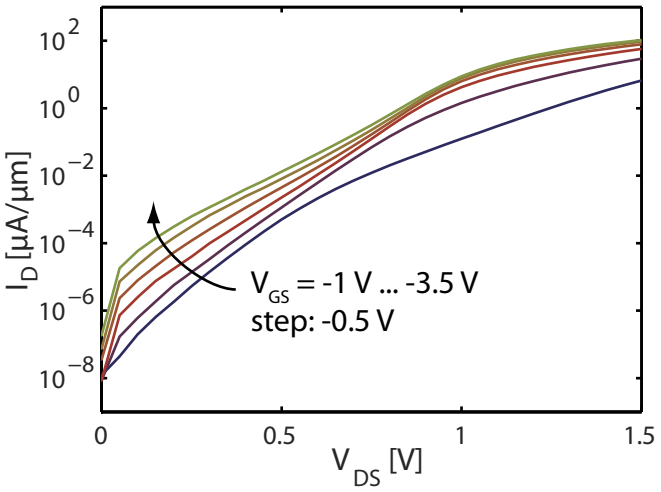


Figure 5.10. I_D-V_{DS} characteristics for positive V_{DS} which are similar to those of a forward biased p-i-n diode. For $V_{DS} < 0.8$ V, the currents are enhanced due to a recombination current which can be tuned by V_{GS}.

If the channel resistance is neglected and therefore all resistance is attributed to the source and drain parasitic series resistance, a value of $R_{S/D} = 12.5\,k\Omega{\cdot}\mu m$ can be extracted. This value is of the same order as in the SOI TFET case.

5.1.3. Strained silicon wires

According to eq. (2.9), two important material properties influence the band-to-band tunneling probability, the size of the bandgap and the effective carrier masses. In contrast to Ge and III-V semiconductors that are discussed in Section 3.4.4., strained silicon offers the opportunity to study material related effects without changing the entire fabrication process.

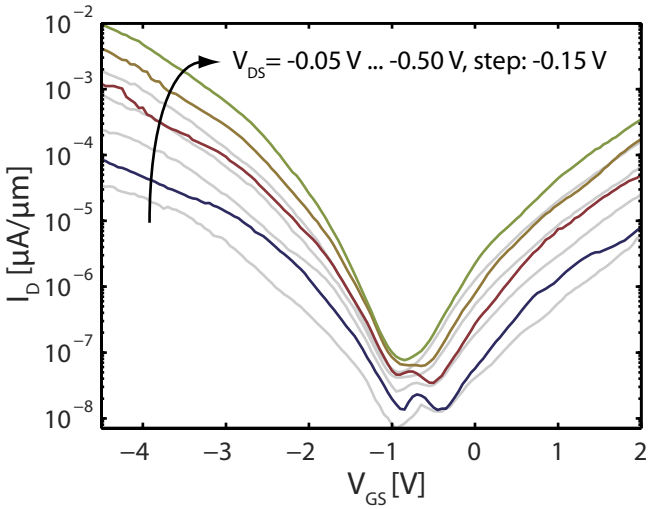


Figure 5.11. Comparison of the subthreshold characteristics of a strained silicon nanowire TFET (colored) with the reference device fabricated on normal silicon (gray).

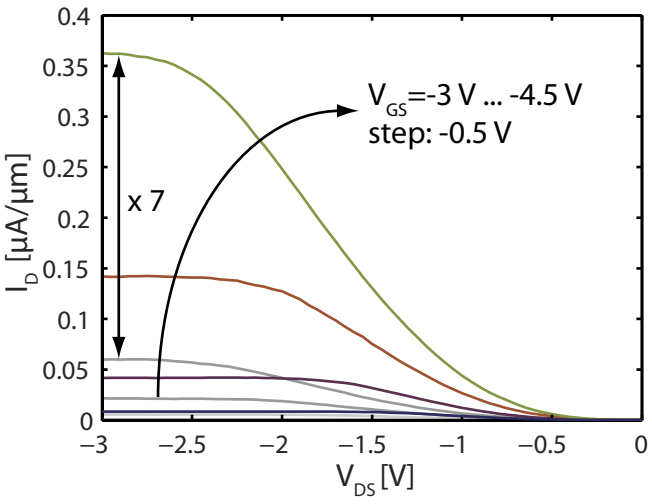


Figure 5.12. Comparison of the output characteristics of a strained silicon nanowire TFET (colored) with the reference device fabricated on normal silicon (gray).

Therefore, starting from a biaxially tensile strained SOI (sSOI) substrate with a stress of 1.2 GPa and thickness of 15 nm, nanowire TFETs are fabricated using the same process as the reference sample.

The effect of biaxial strain on the Si bandstructure is that the 6-fold degeneracy of the conduction band is lifted [56]. For example, the two conduction band valleys perpendicular to the strain direction (Δ_2) are lowered in energy while the energy of four in-plane valleys (Δ_4) is increased. Therefore, the Δ_2 valleys are preferentially occupied and since they contribute with their transverse effective masses to the in-plane conduction, the resulting in-plane effective mass is reduced [57]-[59].

When patterning nanowires, it has been shown that a biaxially strained substrate relaxes uniaxially, which leads to uniaxial tensile strain along the nanowire [60]. Similar to the biaxial case, an improvement in the mobility has been measured [61] and attributed to the preferential occupation of valleys that contribute with their transverse effective mass to transport [62].

Moreover, not only the smaller effective mass makes strained silicon an attractive material for band-to-band tunneling. The energetically lowered Δ_2 valleys also result in a reduced bandgap and, since the six-fold degeneracy is lifted, also the density of states close to the conduction band edge is reduced. The smaller density of states can be beneficial to approach the quantum capacitance limit, because it directly reduces the quantum capacitance.

Experimental results

The measured subthreshold characteristics are displayed in Figure 5.11. The colored lines represent the sSOI data, which compared to the unstrained case (gray) show an on-current that is improved by about one order of magnitude. The minimum inverse subthreshold slope of the p-branch is lowered to S = 325 mV/dec, which is an improvement of about 15%. Similar benefits are also observed for the n-mode.

Figure 5.12 shows a comparison of the output characteristics for both the strained (colored) and the unstrained (gray) wires. Again, we observe distinct saturation in both sets of curves and the exponential onset is also present in both cases. Due to strain, the saturation current increases by a factor of six, which is more than double the mobility enhancement factor which is obtained in MOSFET nanowires [63]. This is again an indication that the device performance is not limited by scattering in the channel but by other mechanisms like band-to-band tunneling.

5.2. Laser annealing

5.2.1. Motivation and setup

Since the characteristics of the nanowire reference device are very similar to what is obtained for planar SOI structures, the expected impact of Λ-scaling (see Section 3.4.3) is mitigated by a factor which limits the band-to-band tunneling rates. In Section 4.3.5 we found that the different slopes for tunneling at the source/channel and drain/channel junctions are a result of the different spreads of the corresponding lateral doping profiles. Apart from straggle effects during implantation, the main reason for the slow drop of the doping profile is diffusion during high-temperature thermal treatments.
Therefore, it seems promising to look for alternative methods of dopant activation that could provide activation comparable to a spike anneal while suppressing dopant diffusion.
There are two approaches which are popular in research, one being the flash-lamp anneal (FLA) (see references [64]-[67]) and the other one being (excimer) laser annealing (ELA) (see references [68]-[72]). Both methods rely on very short highly energetic light pulses which are produced by a flash-lamp or a laser. The main differences are the time scale and the wavelength of the light. Flash-

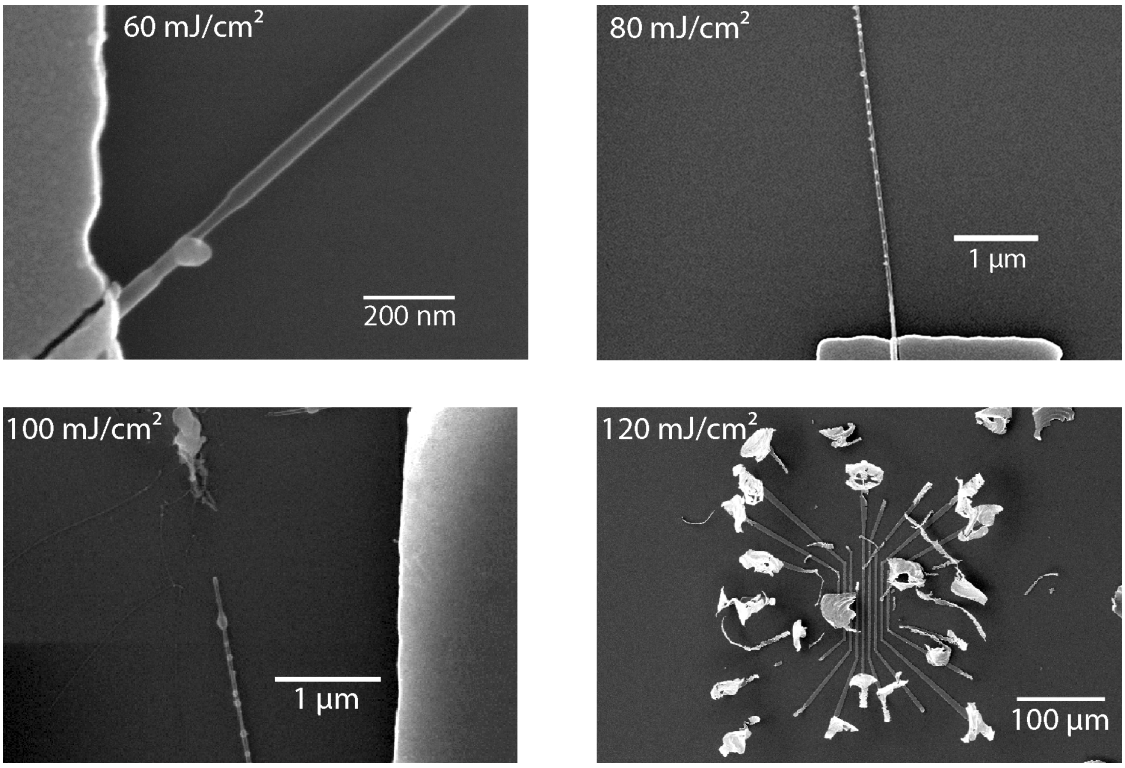


Figure 5.13. Melting study on silicon nanowires of 40 nm diameter. Each sample is annealed by 100 laser pulses of the single-pulse energy given in the images. While the E=60 mJ/cm² sample does show a single beading effect close to the Ni contacts, beading occurs all along the length of the wire for larger energies. At an energy of 120 mJ/cm², the Ni pads are destroyed.

lamp pulses usually last for milliseconds and contain the full spectrum of visible light, whereas laser pulses have a duration of nanoseconds and only consist of light of a single wavelength.
It is still not known which of the two methods yields better results, but in this work laser annealing is studied to suppress the dopant diffusion at the source/channel interface of nanowire TFETs.

Setup
The setup for the laser annealing consists of an excimer laser (λ = 248 nm), optics to widen the beam to a quadratic 7x7 mm² beam spot and a movable stage as well as a calorimeter to measure the energy density of a single pulse. The laser produces pulses of 28 ns length at a repetition rate of 1 Hz. The equipment is situated in a non-cleanroom laboratory and is exposed to air.

Choice of energy
The initial energy range was chosen according to the information from [68]-[72]. The work of Misra, et al. [71] is particularly interesting, because it focuses on nanowires and explores the energies necessary to melt the nanowires. The aim is a non-melting laser anneal to prevent structural damage on the wire or the gate dielectric which could result from melting. To verify the results for the available setup, the melting experiment has been repeated in cooperation with J. Smith (see Figure 5.13).
For this experiment, silicon nanowires with a diameter of about 40 nm, which are grown utilizing the vapor-liquid-solid (VLS) method and that are partially silicided with Ni, are exposed to 100 laser pulses of different energies of 60 mJ/cm², 80 mJ/cm², 100 mJ/cm² and 120 mJ/cm².
The wires which are annealed at E=60 mJ/cm² do not show structural damage apart from a small beading effect close to the NiSi/Si interface, which is probably induced by the NiSi. At E=80 mJ/cm² partial melting results in beading along the entire nanowire length and leads to small discontinuous breaks in the wires. At E=100 mJ/cm², the wires show severe beading and large voids, also the Ni-contacts show signs of a 'peeling'-effect. At E=120 mJ/cm² nearly all wires are missing and the Ni-contacts are completely destroyed.
These results essentially confirm the observations of [71] that the melting threshold for nanowires of the chosen diameter is around E=60 mJ/cm². Since the nanowire TFETs under study use thinner wires, the four quadrants of the first sample are annealed at energies of 45 mJ/cm², 50 mJ/cm², 55 mJ/cm² and 60 mJ/cm².

Post laser anneal regrowth
Besides the activation of dopants, the recrystallization of the regions which are amorphized by the implantation is the main purpose of the annealing step. Although the laser energy should be high enough to restore the crystal lattice, it might be helpful to do an additional low temperature regrowth anneal in an RTA-system. Especially for the large source/drain 'anchor' regions that can be treated like thin SOI, reference [68] suggests that much larger energies are needed to achieve a good activation of dopants, which probably also holds true for recrystallization. To compensate for poor activation and recrystallization in those regions, for a second sample solid phase epitaxial regrowth (SPER, see for example [73]) is carried out at 650 °C for 5 min. Since measurements of the first sample suggested that E=60 mJ/cm² does not lead to a performance degradation, the energy range for this second sample is adjusted to: 60 mJ/cm² - 75 mJ/cm² in steps of 5 mJ/cm².

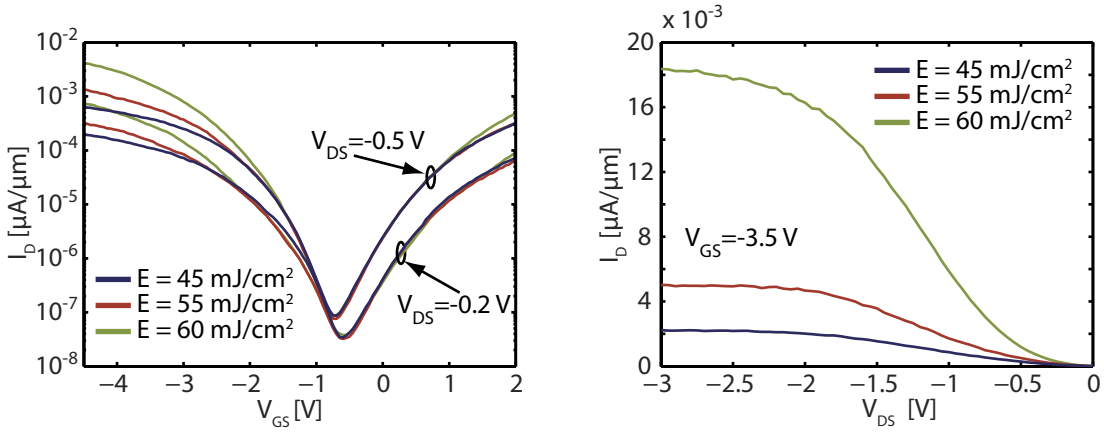


Figure 5.14. Left: Subthreshold characteristics for the laser annealed sample (ELA) at different pulse energies. *Right*: P-mode output characteristics for the same energies.

5.2.2. Results and discussion

The electrical characteristics of the two samples, one with laser-annealing only (ELA) and the other one with laser-annealing and the additional SPER step (ELA+SPER) are obtained in the same way as for the spike annealed reference sample (RTA, see Section 5.1).

Optimizing the laser energy without SPER

Before comparing the different annealing strategies, it is useful to determine which laser pulse energies leads to the best tunneling performance. Figure 5.14 compares the subthreshold characteristics for two different V_{DS} voltages and three different laser pulse energies for the sample without further thermal treatment. Clearly the on-state currents improve for higher energies which is a result of improved activation of the arsenic on the source side. The improvement of the n-branch

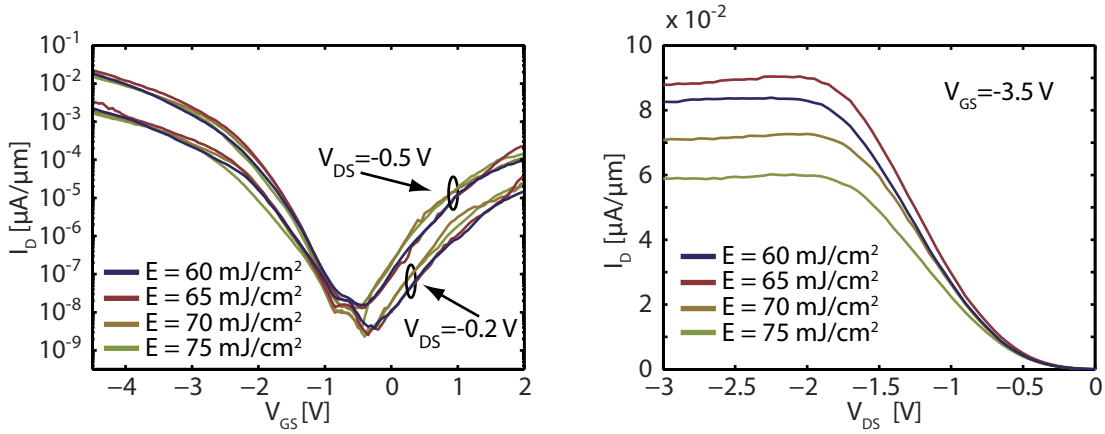


Figure 5.15. Left: Subthreshold characteristics for the laser annealed sample that is also annealed at 650 °C for 5 min (ELA+SPER) at different pulse energies. *Right*: P-mode output characteristics for the same energies.

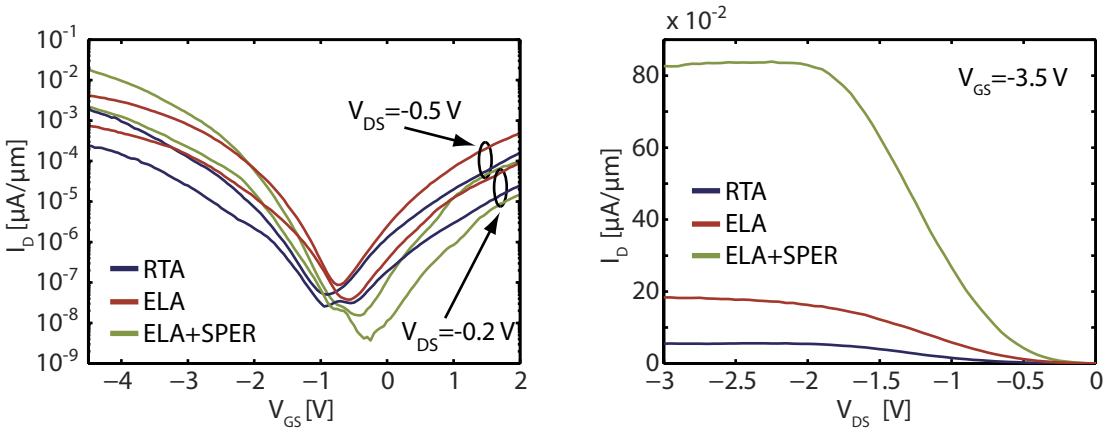


Figure 5.16. Comparison of the three different annealing stragies: 1000 °C spike anneal (RTA), laser anneal at 60 mJ/cm² (ELA) and laser anneal at 60 mJ/cm² in combination with solid-phase epitaxial recrystallization (ELA+SPER). *Left*: Subthreshold characteristics, *Right*: Output characteristics.

is less pronounced, but still visible. On the right side of Figure 5.14, the output characteristics for a single V_{DS} = -3.5 V are compared for the different annealing energies. Again, we find clear evidence that the increasing laser energies lead to larger currents. However, for all three energies the saturation currents are not constant but they increase for larger negative V_{DS}. This non-zero slope is an indication for a parasitic series resistance limiting the currents. From forward bias I_D-V_{DS} characteristics (not shown), a value for $R_{S/D}$ of 1.25 MΩ·µm is extracted under the assumption of zero channel resistance. The series resistance probably arises from the doped 'anchor' portions of the structure whose dopants are activated as good as in the nanowires for the chosen laser energies.

Optimizing the laser energy with SPER

To reduce the series resistance, a SPER annealing step is introduced for a second sample. The laser pulse energy range is adjusted to 60 mJ/cm² - 75 mJ/cm², because without SPER, there seems to be no current degradation up to 60 mJ/cm² due to possibly destroyed nanowires.

Figure 5.15 shows a comparison of transfer characteristics for the sample which was annealed by excimer laser at different laser energies and subsequently annealed at 650 °C for 5 min in an RTA-system. While the n-branch currents still increase with larger laser energy, the p-branch currents start to deteriorate for energies greater than 65 mJ/cm². This behavior is also confirmed by the p-mode output characteristics plotted on the right side of Figure 5.15. The reason for the deterioration is probably an increasing number of damaged wires which contribute less or not at all to I_D. Nevertheless, we observe distinct saturation. The increase of the n-branch is probably due to the further improvement of the boron activation for higher laser pulse energies which compensates for damaged wires. From forward bias I_D-V_{DS} characteristics (not shown), a value of $R_{S/D}$ = 65.2 kΩ·µm is extracted under the assumption of zero channel resistance. This corresponds to an improvement of a factor of about 19 in comparison to the device without SPER. Although the series resistance is still about a factor of 5 larger than in the spike-annealed device, it is small enough to not limit the on-state performance of the device at hand.

Comparison of different annealing strategies

Figure 5.16 (*Left*) compares the subthreshold characteristics for the three different annealing strategies at two different V_{DS} values (V_{DS}= -0.2 V and V_{DS}= -0.5 V). The laser pulse energy is 60 mJ/cm² in both cases which involve laser annealing. For the p-branch we observe an increase of the maximum current of about one order of magnitude when comparing the RTA and the ELA+SPER device. The behavior of the ELA curve is more complex. At small negative gate-voltage (-1.5 V < V_{GS}< -1 V), the current is larger than those of the other two devices, but for larger negative V_{GS}, falls back with respect to the ELA+SPER device. For maximum negative V_{GS}, the ELA device current is only about a factor of two larger than the current of the RTA device. The explanation for this behavior becomes apparent in the output characteristics (*Right* part of Figure 5.16). In the output characteristics, we find distinct saturation for the RTA and the ELA+SPER sample, but not so pronounced for the ELA sample. This indicates again, that the currents of the ELA device are limited by series resistance, which is probably caused by inferior dopant activation in the anchors as well as amorphization of the leads. Only by means of SPER, the series resistance is lowered and the current rises.

The analysis of the minimum inverse subthreshold slopes confirms the interpretation arising from the currents. At V_{DS} = -0.5 V, the minimum slopes are S = 375 mV/dec for the RTA, S = 315 mV/dec for the ELA and S = 290 mV/dec for the ELA-SPER. This means there is indeed an improvement of the slope because of the reduced dopant diffusion.

For the n-mode, the subthreshold situation is less obvious. In this case, the ELA sample delivers the largest I_D, followed by the RTA device and the ELA+SPER device. However, the minimum inverse subthreshold slopes do not reflect the line-up of the currents. In the n-mode, at V_{DS} = -0.5 V, the minimum slopes are S = 475 mV/dec for the RTA, S = 430 mV/dec for the ELA and S = 355 mV/dec for the ELA-SPER. Since the slope is mainly affected by the tunneling transmission, it is reasonable that the additional SPER improves the transmission similar to the case of the p-branch. The smaller current level of the ELA-SPER sample seems related to a shift of the n-branch to more positive V_{GS}. Such a shift is characteristic for a smaller active doping concentration. An explanation for a smaller active boron concentration can be found in the large surface/volume-ratio in nanowires. Boron is diffusing much faster than arsenic and during the 5 minute annealing step at 650 °C, thus the boron can diffuse to the surface and congregate at the Si/SiO_2-interface. At the interface, the boron ions become electrically inactive and do not contribute to the doping any more.

Conclusion

In summary, the use of laser annealing is beneficial for both on-currents and subthreshold slopes of the TFET. This result illustrates that indeed dopant diffusion is limiting the performance of the TFET and for ultimate device performance, the steepness of the doping profiles needs further optimization. One way to achieve steeper doping profiles is epitaxial growth of the TFET structure. In that case, dopants can be incorporated in-situ only in the source and drain regions and the doping profile can be very close to abrupt. The most promising device structure for a TFET is therefore an epitaxially grown hetero-structure with in-situ doping and different bandgaps at the source/channel and drain/channel interfaces.

5.3. Low temperature characterization

Whenever a device with a novel current modulation strategy is studied, low temperature characterization is a powerful tool to understand the underlying mechanisms. As the band-to-band tunnel-

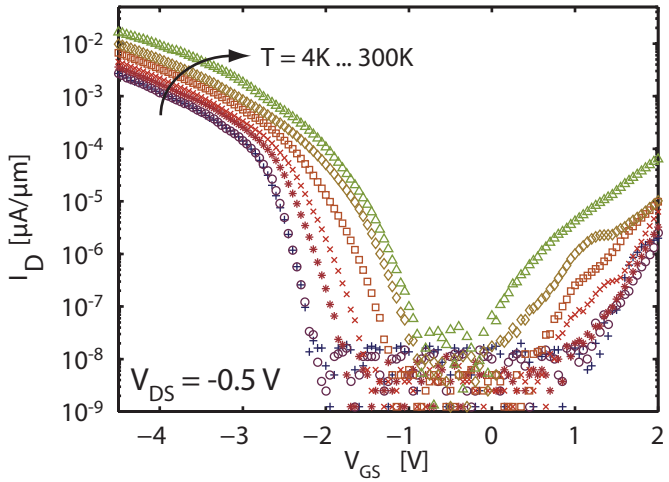


Figure 5.17. Subthreshold characteristics for the ELA-SPER sample at different temperatures and V_{DS} = -0.55 V. The subthreshold swing changes by less than a factor of two. Therefore, the conduction is not limited by thermal emission.

ing probability does not depend on temperature to first order, the temperature dependence of tunneling currents should be dramatically different from the current of MOSFETs whose transport is dominated by thermal emission. In classical MOSFETs, the subthreshold slope is proportional to the temperature (see eq. (2.7)). In the silicon TFETs, the temperature could affect the subthreshold slope by changing the available phonon density, the size of the bandgap and the carrier density in case of dopant freeze-out.

5.3.1. Experimental setup

Low-temperature measurements are performed in a Lakeshore probe-station which allows for cooling a sample using a cold finger which is cooled by liquid helium. The minimal possible temperature is therefore the evaporation temperature of helium at 4.2 K. Following the initial cool-down to 4.2 K, electrical measurements are done using an HP 4155C semiconductor parameter analyzer. After a number of devices are characterized, the temperature is slowly ramped up and additional measurements are performed at 50 K, 120 K, 150 K, 200 K, 250 K and at room temperature (300 K).

Since the sample with laser anneal and solid-phase epitaxial regrowth yields the best p-mode performance, this sample is selected for the low temperature characterization.

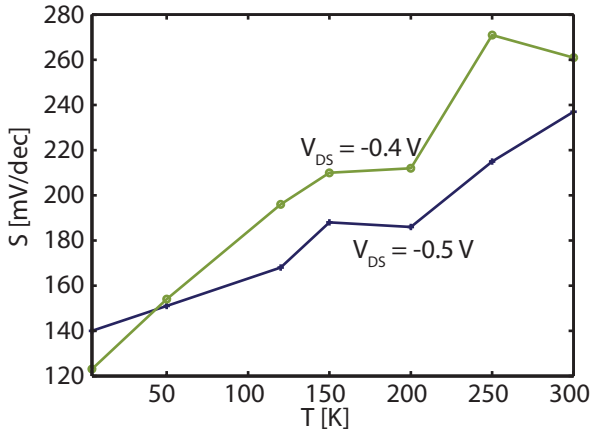


Figure 5.18. Minimum subthreshold swing of the ELA-SPER sample in p-mode for different temperatures and two different V_{DS} voltages.

5.3.2. Characterization I_D-V_{GS}

Reverse bias - negative V_{DS}

Figure 5.17 presents the subthreshold characteristics of the ELA-SPER sample at the above mentioned seven different temperatures and a constant V_{DS} of -0.5 V. By reducing the temperature, three main effects become apparent. First, the sub-

threshold swings do not change proportional to the temperature as they would for the MOSFET. Second, the distance between the p- and the n-branch widens from about 1 V to approximately 3 V. And third, the maximum currents for both branches reduce by about one order of magnitude on the p-side and two orders of magnitude on the n-side.

The first observation concerning the subthreshold slope is confirmed by Figure 5.18 which shows the temperature dependence of the minimum subthreshold slope in the p-mode for two different V_{DS} voltages. The subthreshold slope changes by about a factor of two, which is much less than the change in temperature by a factor of about 70. This leads to the conclusion that the current is not dominated by thermal emission over a potential barrier, but rather by tunneling.

The origin of the observed changes of the current and the distance between n- and p-branch is less obvious. However, there are several explanations that cannot be ruled out without further experiments and simulations:

- There are at least two other mechanism which allow for inter-band transitions. One is trap-assisted tunneling [53] , and the other one is AUGER-recombination [77] (for a good discussion on the different recombination mechanisms see reference [78]). The temperature dependence of trap-assisted tunneling is expected to be similar to phonon-assisted tunneling, since tunneling into and out of the trap both require a momentum transfer which most likely occurs by phonon-scattering. In the AUGER-recombination case, momentum is transferred to another charge carrier. Therefore, its rate mainly depends on the charge carrier concentration but not on the phonon density and therefore depends less on temperature.
- The increasing distance between n- and p-branch can partially be explained by reduced thermal broadening. At room temperature, tunneling current flows in the energy range between the band edges in source and drain. In the low temperature case, this range is further constrained by μ_S and μ_D because the FERMI-distribution exhibits sharp edges. If μ_S is situated in the band, a larger gate voltage is needed to allow tunneling. However, the shift caused by this effect should be limited to a few kT which is not sufficient to explain the observations.
- At temperatures below 150 K, dopants can 'freeze out' (see for example ref. [74]-[76]). This term describes the incomplete ionization of donor or acceptor states, which leads to a reduced carrier densities in source and drain. In the device at hand, this effect should be negligible in source and drain, because the doping concentrations should correspond to degenerate doping. In case of degenerate doping, the dopants cannot freeze out, because the impurity states form a band which overlaps with the bands of silicon.

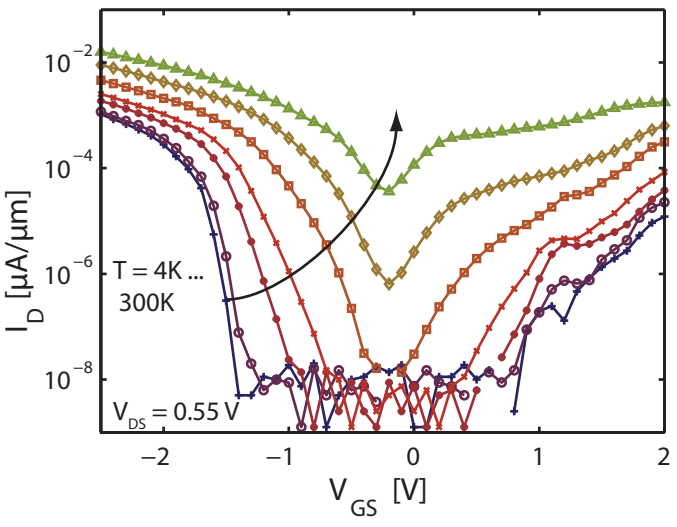


Figure 5.19. Forward bias I_D-V_{GS} characteristics of the ELASPER device at $V_{DS} = 0.55$ V.

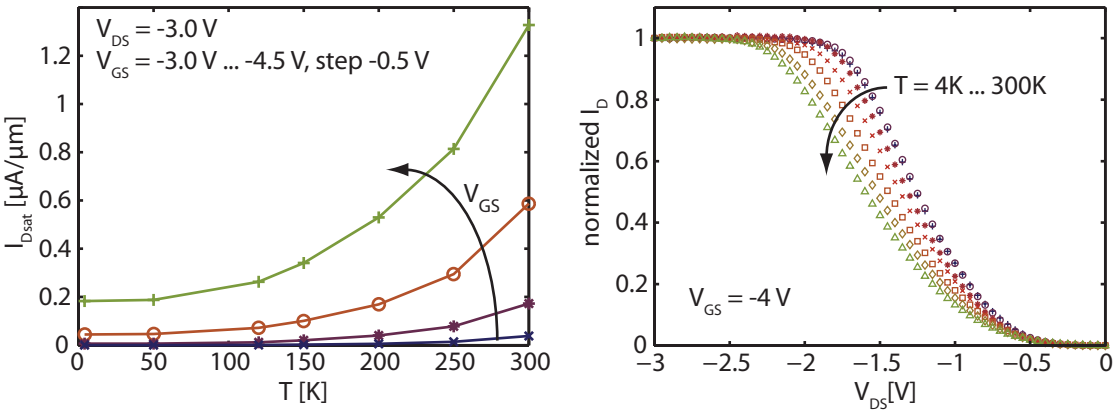


Figure 5.20. Left: Temperature dependence of output currents at V_{DS} = -3.0 V and V_{GS} ranging from -3.0 V to -4.5 V. Right: Normalized output characteristics at V_{GS} = -4.0 V for different temperatures.

- As silicon is an indirect semiconductor, all band-to-band tunneling is phonon assisted [20]. The energy and density of phonons strongly depends on temperature according to Bose-Einstein statistics. Therefore, the tunneling current is expected to reduce at low temperatures.

Since the interplay of these different contributions is very complex, they cannot be separated without further experiments and simulations.

Forward bias - positive V_{DS}

Figure 5.19 displays the measured temperature dependence of the I_D-V_{GS} characteristics of the ELA-SPER sample under positive V_{DS}. As explained before (see Section 4.3.1), for large positive and negative V_{GS}, Shockley-Read-Hall recombination in the region of highest electric field is responsible for large currents. At medium V_{GS} around 0 V, the recombination rates are small and the minimum current is given by thermal emission currents. This model is confirmed by the low temperature data. For the extreme V_{GS} voltages, there is only a moderate temperature dependence, while around V_{GS} = 0 V, the current is highly suppressed for low temperatures, which is a strong indication for thermal emission.

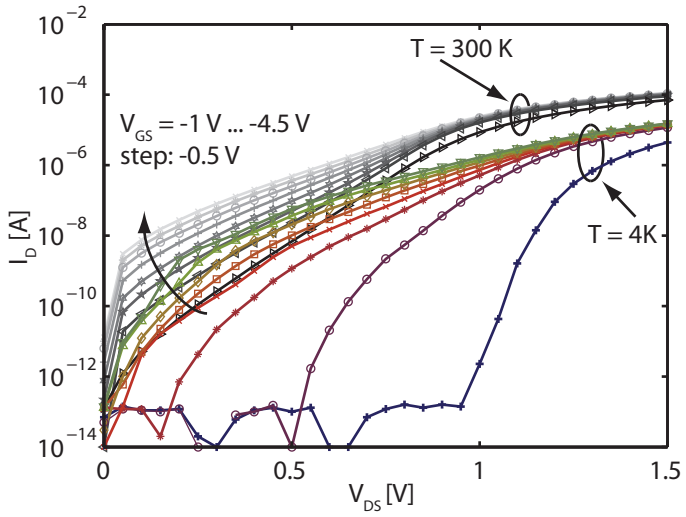


Figure 5.21. Comparison of forward bias I_D-V_{DS} characteristics at T = 4.2 K (colored) and T = 300 K (gray scale) for a V_{GS} ranging from -1 V to -4.5 V.

5.3.3. Characterization I_D-V_{DS}

Reverse bias - negative V_{DS}
The p-mode saturation currents of the ELA-SPER sample are plotted in Figure 5.20 (left) for different temperatures. In agreement with the subthreshold data, the on-state currents decrease for lower temperatures because of the interplay of the effects outlined in the section on subthreshold characteristics (Section 5.3.2). In Figure 5.20 (right), the normalized output characteristics are plotted for $V_{GS} = -4$ V. The only change for lower temperatures is a reduced saturation voltage in combination with a reduced non-linearity of the onset. One possible cause of this effect is the reduced thermal broadening of the drain FERMI-distribution. Since the high-energy tails of the FERMI-distribution are suppressed at low temperatures, saturation occurs more abruptly and therefore at lower V_{DS}.

Forward bias - positive V_{DS}
Figure 5.21 shows a comparison of forward bias I_D-V_{DS} characteristics of the ELA-SPER device at $T = 4.2$ K and $T = 300$ K. As a second parameter, V_{GS} is varied from -1 V to -4.5 V. Similar to the room temperature data, the low temperature characteristics reveal two operating regions. For $V_{DS} < 1$ V, the current is dominated by recombination at the source/channel junctions and for larger V_{DS}, thermal emission currents start to contribute to the total current. This behavior is easily understood, because at low temperatures, the thermal broadening of the FERMI-distribution is very small. Therefore, thermal emission currents are only possible if V_{DS} exceeds the built-in potential of the *p-i-n* structure. For smaller V_{DS}, the current is blocked by the built-in potential barrier. In contrast to thermal emission currents, the (trap-assisted) recombination at the source/channel junctions is nearly independent of temperature and therefore recombination becomes dominant for small V_{DS} voltages.

5.4. Conclusion

In this chapter, the first silicon nanowire TFET arrays are presented. With a width and height of less than 20 nm per wire and a pitch of 100 nm, they form the smallest nanowire TFETs that have been fabricated to date.
In order to make the fabrication at such small dimensions possible, the electron beam lithography process has been refined and complemented by an improved dry etch process.
With the implantation and activation conditions which are adopted from the planar SOI process of Chapter 4, the electrical characteristics of the nanowire TFETs are very similar to the planar SOI samples, which is again comparable to the planar Si device of reference [27] but much improved compared to other results obtained in nanowire TFETs (see reference [33]).
Since the simulations of Chapter 3 predict an improvement, dopant diffusion is identified as a limiting factor for the tunneling performance. In order to overcome this limitation, in a second step, laser annealing is introduced to suppress dopant diffusion.
To determine the laser annealing energy, melting experiments are carried out and a target energy of 60 mJ/cm² is identified as a starting point for further optimizations. Therefore, a laser annealed sample has been fabricated and characterized. The results of electrical characterizations show an improved performance compared to the initial sample.
However, series resistances in source and drain are limiting the device current and an additional thermal annealing step at a moderate temperature of 650 °C is introduced to reduce the series

resistances of a second sample. This optimization results in a boost of the on-state current of about a factor of ten and at the same time the subthreshold swing is reduced by about 25 % with respect to the starting sample.
These optimized devices are further characterized by means of low temperature measurements. At low temperatures, it becomes evident that the devices at hand do not rely on thermal emission for current modulation. Reviewing all the available information on the structure, band-to-band tunneling is the most probable explanation for device operation. However, although the low temperature characterization is a powerful tool to distinguish different injection mechanisms, it is not possible to separate all effects which could explain the observed temperature dependence.

Even with the improved electrostatics a nanowire provides and activation by laser annealing, the on-current levels of the devices are three to four orders of magnitude lower than those of MOSFETs. Furthermore, the inverse subthreshold slopes are still worse than the thermal limit, which should ultimately be overcome.
Therefore, further improvements of the device structure become necessary:

- Gate electrostatics: The SiO_2 gate dielectric should be replaced by a high-κ material in order to achieve an effective oxide thickness of < 1 nm.
- Substrate material: The bandgap of Si is too large to allow for sufficient tunneling currents. Germanium or III-V semiconductors can lead to much larger tunneling rates and therefore shift the region of very steep slopes to much larger current levels.
- Heterostructure: In order to suppress the ambipolarity of the TFET efficiently, a heterostructure with a small bandgap at the source/channel and a large bandgap at the drain/channel junction should be used.
- Doping: Although the presented laser annealing approach can lead to relatively steep dopant profiles, the profiles are limited by straggle during the implantation. In order to further improve the steepness of the profiles, the junctions should be created by growth of in-situ doped material.
- Parasitic resistance: For TFETs with large tunneling currents, the parasitic resistance will become as important as for recent MOSFETs. Therefore, source/drain engineering concepts like silicidation or epitaxial overgrowth should be exploited.

Chapter 6
InSb nanowire MOSFETs

6.1. Introduction

With the smallest band gap and effective carrier mass of known bulk semiconductors and an electron mobility at room-temperature in excess of 50,000 cm^2/Vs [79], Indium-Antimonide (InSb) holds the potential for low-power, high-speed device applications [80],[81]. Moreover, these properties predestine InSb for the use in TFETs, because very large band-to-band tunneling rates can be expected. Since before the following study, there were no published results available on InSb nanowire transistors at all, back-gated Schottky-MOS transistors are fabricated, characterized and compared with the simple one-dimensional, ballistic transport model that is introduced in Chapter 2. The work of this chapter has been performed in the group of Prof. J. Appenzeller at Purdue University, IN, USA.

6.2. Nanowire Growth and Device Fabrication

1. ***Nanowire growth.*** As a first step, single-crystalline nanowires are synthesized via a vapor-liquid-solid (VLS) process [82] with Au catalysts on InSb substrates (technique similar to [83]). Before the actual wire growth, the native oxide is removed from the growth substrate by an HF dip prior to the dispersion of the gold particles. The nanowire growth is then performed at 540 °C at a partial pressure of 50 mTorr using hydrogen as the carrier gas. High-resolution transmission electron microscopy (see Figure 6.1) confirms that the growth is occurring in the <110> direction and that a native oxide shell of approximately 10 Å thickness surrounds the nanowire core. In contrast to the transmission electron micrograph, the target diameter for device fabrication is 60 nm-100 nm.

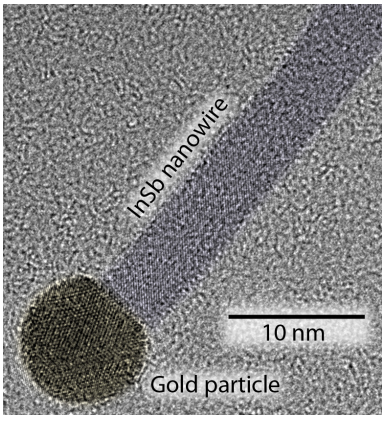


Figure 6.1. High-resolution transmission electron micrograph, showing an InSb nanowire attached to a gold growth catalyst.

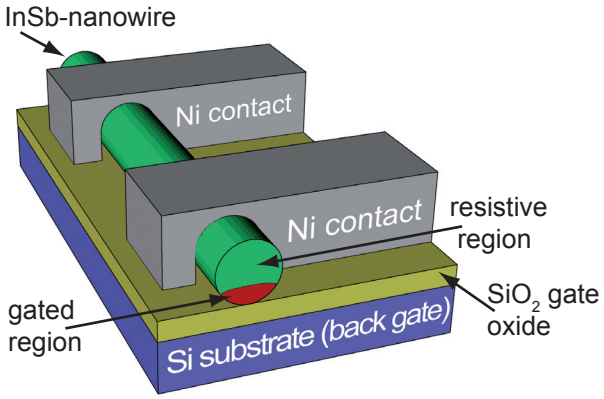


Figure 6.2. Schematic overview of the InSb back-gated nanowire devices under study.

2. ***Transfer wires to device substrate.*** For device fabrication, the wires are released from the substrate by sonication of the sample while it is immersed in a 2-Propanol solution. Afterwards, the resulting solution is dispersed on a Si substrate with a 20 nm thick SiO_2-layer on top and the sample is spinned to equally distribute the wires and remove dirt particles using a photoresist spinner. In combination with the substrate, the SiO_2-layer serves as a back gate.
3. ***Deposit source/drain contacts.*** The substrate is inspected using optical microscopy for suitable nanowires and a device layout is designed for the source/drain lithography. Metal source and drain contacts are then patterned using e-beam lithography, e-beam evaporation of 80 nm of

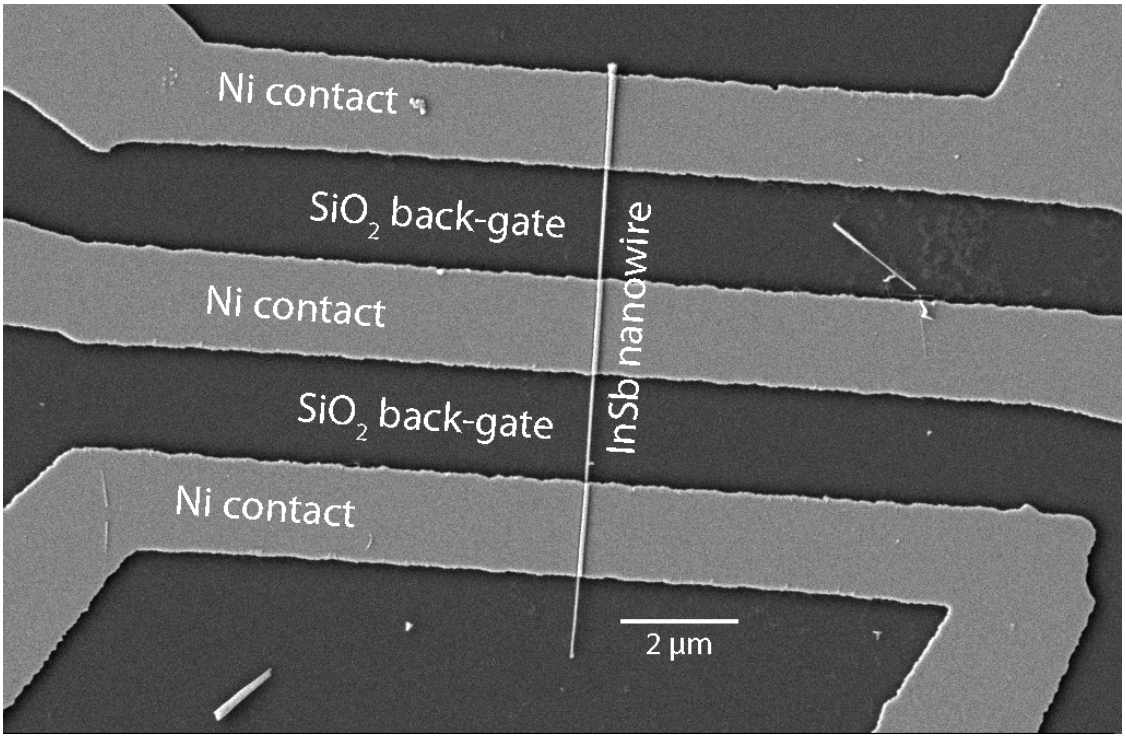


Figure 6.3. Scanning electron micrograph of an 11 µm long InSb-nanowire and three Ni fingers that are used as source/drain contacts. The nanowire is placed on top of a SiO_2 backgate oxide.

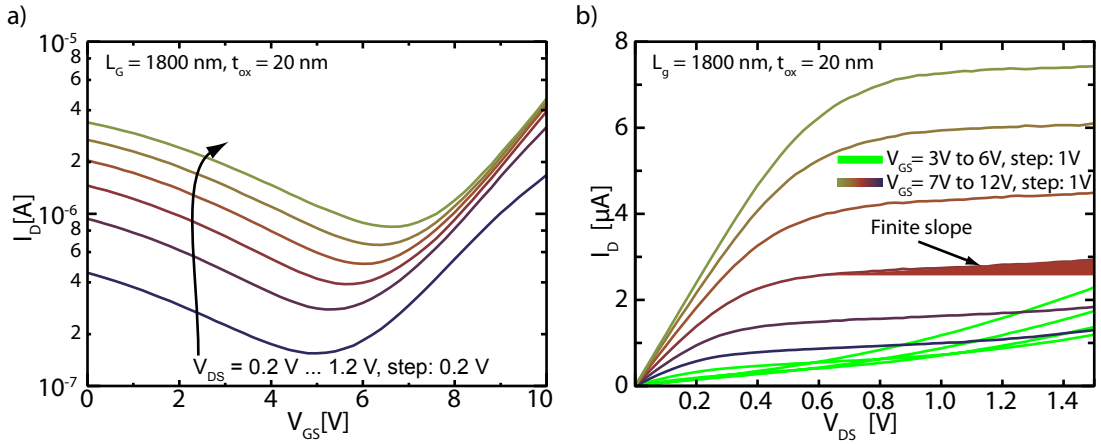


Figure 6.4. a) Transfer characteristics of an InSb-nanowire MOSFET with a gate length of L_G=1.8 µm and a back oxide thickness of t_{ox}=20 nm. b) Output characteristics of the same device, the red wedge illustrates the finite slope in saturation.

Nickel, and an acetone lift-off process. The channel lengths of the devices are designed between 1 µm and 30 µm. Figure 6.2 sketches the full device, while Figure 6.3 shows an SEM image of a 60 nm diameter nanowire FET with multiple metal contacts.

6.3. Results and Discussion

The first task at hand when characterizing a novel nanomaterial is to check the electrical device characteristics against the expected performance assuming bulk-type material properties. The most relevant parameter in this context is the band gap. InSb is expected to exhibit an energy gap of $E_g \approx 170$ meV. The transfer characteristics of a typical device, measured using an HP 4155C semiconductor parameter analyzer and shown in Figure 6.4 (a), provides initial evidence of the material quality - in particular indicating that indeed a band gap of just a few kT prevails in the nanowire device. A strong ambipolar behavior and a small I_{on}/I_{off}-ratio of around 10 are in good agreement with a small E_g-value.

To estimate the expected I_{on}/I_{off}-ratio, the Landauer equation (see eq. (2.5)) together with the assumption of a one-to-one band movement in the channel with V_{GS} is used. As before, the transmission is approximated by a step function which is one above the channel conduction band edge and zero below. The chemical potential of the source metal is chosen to align midgap with the InSb bandgap. With these assumptions, i.e. ballistic, one-dimension-

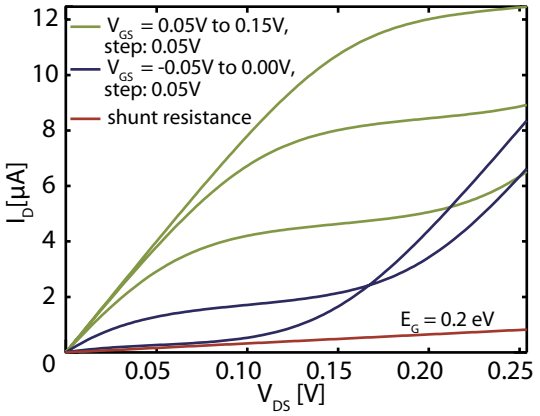


Figure 6.5. Modeled output characteristics for an InSb one-dimensional ballistic nanowire in the quantum capacitance limit. A shunt-resistance in form of a second parallel InSb nanowire which is not gated is include in the calculation.

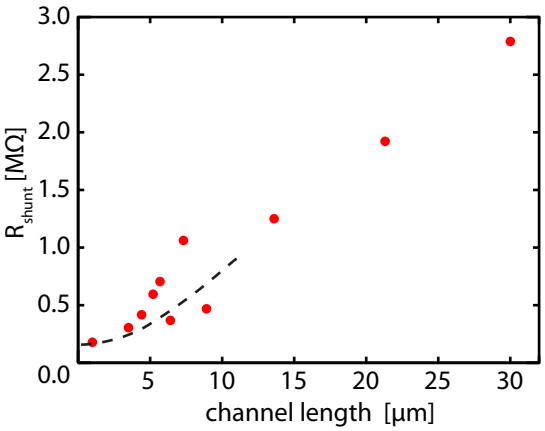


Figure 6.6. Measured shunt resistance R_{shunt} as a function of the channel length. The data suggest a non-zero minimal value that corresponds to a bandgap energy of 160 meV to 180 meV, which agrees very well with the expectations for InSb.

al transport in the quantum capacitance limit, the I_{on}/I_{off}-ratio becomes:

$$\frac{I_{on}}{I_{off}} = \frac{1}{2}\exp\left(\frac{E_g}{2kT}\right). \tag{6.1}$$

Evaluating this equation for the experimental data results in a bandgap of E_g=155 meV, which is consistent with the expected bulk value. While in the presented long-channel devices the existence of scattering and non-ideal gate control may seem to make the simple model inapplicable, it is the one-dimensional nature of a small gap semiconducting nanowire that ensures a fair description for small drain voltages in the present case. Typical output characteristics for the n-branch are shown in Figure 6.4 (b) and exhibit a smooth linear region, clear evidence of current saturation at $I_{on\text{-}max}\approx$7.5 μA for V_{GS} = 12 V, and ambipolar characteristics. Unlike silicon MOSFETs, the output characteristics of the InSb MOSFET all show a finite slope independently of the applied V_{DS} and V_{GS} (red wedge). The fact that this finite slope is also present at small V_{DS} in the off-state of the device indicates that the upper part of the nanowire is not affected by the gate voltage and thus acts as parallel shunt resistance.

To check whether the assumption of a shunt resistance is consistent with the measurements, again the one-dimensional ballistic model is employed.

In the off-state, it is assumed that the entire nanowire acts as a shunt resistor R_{shunt}. In that case, a single one-dimensional mode contributes in both the conduction and valence band. This assumption is justified when considering the mode spacing in a material with an effective mass around 0.014 m_0 [84].

In the on-state, two nanowire regions can be distinguished as shown in Figure 6.2; the gated side's current is modulated as a transistor while the shunted current remains the same as in the device off-state.

This model explains indeed the experimental observations in Figure 6.4 (b) qualitatively for the entire curve-set and even quantitatively for small enough V_{DS} values in the device off-state. Figure 6.5 shows the calculated I_D-V_{DS} characteristics including the shunt resistance. A finite slope in the saturation region of the I_D-V_{DS} characteristics as well as a substantial residual current level in the device off-state are clearly reproduced.

Since the agreement between the model and the experimental data-set is expected to be best for small V_{DS} in the device off-state, this region is studied in more detail for nanowire FETs with different channel lengths. Figure 6.6 shows the extracted R_{shunt} for in total 11 different nanowire FETs. We note that R_{shunt} reaches a minimum value of around 150 kΩ-200 kΩ for decreasing channel length. Since decreasing the channel length is a means to reach the ballistic transport limit, the value of R_{shunt} is compared to the expected ballistic shunt resistance

$$R_{shunt} \propto \frac{h}{4e^2}\exp\left(\frac{E_g}{2kT}\right) \tag{6.2}$$

from the model. A band gap of 160 meV to 180 meV can be extracted for the measured R_{shunt}, which is again in good agreement with the expectations and the on/off-current analysis.

6.4. Conclusion

In summary, within the framework of this thesis, the fabrication and characterization of InSb NWFETs are demonstrated for the first time. Clear evidence of a small band gap semiconductor with device characteristics consistent with an E_G-value of around 160 meV to 180 meV is found. The data can consistently be explained assuming two distinct regions in the device with one being a gate-independent shunt resistance. The presented device characteristics are a starting point to develop an understanding of InSb nanowire FETs that is of particular interest in the framework of TFET development, since the small bandgap promises very large on-currents. The main drawback of small bandgap materials, the strong ambipolar behavior, is also prominent in the present nanowire FETs. Together with the results of Section 3.4.4., this observation leads to the conclusion that only a TFET based on a heterostructure can lead to the desire off- and on-state performance at the same time. In the context of heterostructures, InSb provides the best tunneling performance and therefore it is the material of choice at the source/channel junction.

Chapter 7
Conclusion

The tunneling field effect transistor (TFET) has been proposed recently to overcome one of the fundamental limitations of MOSFET operation, the limited switching slope at room temperature. The minimal inverse subthreshold slope of 60 mV/dec has the potential to become a major obstacle for MOSFET scaling in the future. It prevents the down-scaling of the operation voltage, which is one of the key parameters to reduce the power consumption of integrated circuits. Therefore, overcoming the 60 mV/dec-limit could not only lead to superior mobile devices, but at the same time lead to a reduced power consumption of information technologies.

In this thesis, the concept of the TFET is evaluated in ultra-thin silicon films and silicon nanowires. The reason for choosing silicon as a starting point is the mature and widely available device technology in silicon which would permit the semiconductor industry to easily adopt the TFET for their products.

To evaluate the potential of the TFET, non-equilibrium-Green's-function (NEGF) simulations are presented, which confirm the trends that are expected from simple WKB-calculations. The simulations predict that state-of-the art silicon technology is not expected to provide TFET devices which overcome the 60 mV/dec-limit at on-current levels which are usable for digital logic applications. Furthermore, the NEGF simulations have been used to optimize typical device parameters like doping concentration, gate dielectric and body thicknesses. The doping concentration is subject to a trade-off between large currents and a good subthreshold slope. For the gate-dielectric thickness, a clear trend has been found. The tunneling performance greatly benefits from thinner gate dielectrics. For the body thickness, a similar trend is found, but for extremely thin bodies, quantization induced bandgap widening starts to deteriorate the tunneling currents. A comparison of a 1D nanowire and a 2D ultra-thin-body TFET shows that a 1D nanowire yields greatly improved channel electrostatics which results in a much better tunneling performance.

In summary, the parametric studies show optimization strategies which should be confirmed experimentally. However, even with the optimal set of parameters, Si TFETs are unlikely to deliver steep subthreshold slopes and large on-currents at the same time.

Therefore, TFETs from materials with smaller bandgap (Ge and InSb) are compared to silicon TFETs and they are found to deliver much larger on-currents at largely reduced voltages. At the same time, the small bandgaps lead to a deteriorated off-state due to an increased ambipolar behavior. To suppress this behavior, heterostructures with different bandgaps at the source/channel

and drain/channel junctions are proposed. Due to the smaller effective masses of Ge and InSb, quantization induced bandgap widening occurs at larger dimensions than in Si. Therefore, the body thickness of Ge and InSb TFETs cannot be scaled as aggressively as for Si TFETs which has to be compensated by using a high-κ gate-dielectric.

To complement the one- and two-dimensional simulations, three-dimensional silicon TFETs have been realized experimentally. First, TFETs on ultra-thin-body silicon-on-insulator (SOI) have been fabricated. The observed minimal inverse subthreshold slope of about 320 mV/dec is much larger than the corresponding values of modern MOSFETs. Nevertheless, these devices demonstrate that band-to-band tunneling can be exploited in transistors. The studied parameter dependencies show that the optimizations proposed by the simulations indeed change the device performance in the predicted way. Moreover, important parasitic conductions paths due to electron/hole-recombination at the junction as well as in the channel volume and at the oxide interfaces have been identified for the first time in TFETs. In future devices, they should be monitored carefully, because they have the potential to deteriorate TFET performance dramatically.
Since the simulations suggest that the improved electrostatics of a nanowire architecture should lead to improved TFET performance, as a next step, silicon nanowire TFETs were fabricated. To date, these devices are the smallest nanowire TFET devices available. The characterization of the nanowire TFETs reveals that the expected performance boost compared to the SOI-TFETs is missing. The reason for this behavior is identified as the poor slope of the doping profiles at the source/channel and drain/channel junctions. To further confirm this finding, laser annealing has been employed for the dopant activation of selected nanowires. This technique is known to suppress dopant diffusion, which is the main reason for poor doping profiles. The electrical characterization of laser annealed nanowire TFETs agrees well with the expectations. The minimum slope is reduced to less than 290 mV/dec and the on-currents are improved by one order of magnitude.
To examine the impact of the bandgap and of the effective carrier mass on TFET performance, a nanowire device based on strained silicon was fabricated. An improvement of the on-currents by a factor of six, which is well above the expected improvement due to mobility enhancement, is found. Furthermore, the subthreshold slope is reduced by 15% which confirms that the smaller bandgap and effective carrier masses in strained Si indeed lead to improved tunneling rates.
Finally, one of the laser annealed sample has been characterized at low temperatures (4 K). The impact of temperature on the subthreshold swing of the TFET is about a factor of 40 times smaller than on normal MOSFETs. This observation is the final prove that conduction in the TFET is based on band-to-band tunneling rather than on thermal emission.
All results obtained in Si TFETs confirm the simulation results. However, even the most improved device structure does not yield sub-60 mV/dec inverse subthreshold slopes at measurable on-current levels. Consequently, the next step in the development of TFETs is to move to heterostructure nanowires with much smaller bandgaps at the source/channel junction.
As a first step in this direction, InSb nanowire MOSFETs were fabricated in the framework of this thesis. InSb is the III-V-semiconductor with the smallest bandgap and transistors on bulk material are known for years. However, working InSb nanowire MOSFETs are reported for the first time in this thesis. Since the size of the bandgap is the most important parameter for TFETs, the bandgap of the InSb nanowire transistors is extracted and it is found to be in good agreement with the value for bulk InSb. Naturally, the small bandgap leads to strong ambipolar behavior, but in a heterostructure TFET, InSb could be the ideal material at the source/channel junction. Unfortunately, the fabrication of a full InSb TFET is well beyond the scope of this thesis because suitable dopant

materials and scalable gate dielectrics are still under development. Nevertheless, the demonstration of MOSFET operation in InSb nanowires is a big step towards a high-performance TFET. In synopsys, all simulations and experimental results of this thesis lead to the conclusion that the optimal TFET device will be a nanowire array heterostructure with a small-bandgap material at the source/channel junction and a large-bandgap material at the drain/channel junction, surrounded by a high-κ gate-dielectric.

Appendix A
Bibliography

[1] Statistics by www.comscore.com
[2] R. H. Dennard, F. H.Gaensslen, H. N. Yu, V. L. Rideout, E. Bassous, A.R. LeBlanc, "*Design of ion-implanted MOSFETs with very small physical dimensions*", IEEE Journal of Solid-State Circuits vol. **SC-9**, 1974, pp. 256-268
[3] J. Knoch, W. Riess, and J. Appenzeller, "*Outperforming the Conventional Scaling Rules in the Quantum-Capacitance Limit,*" IEEE Electron Device Letters, vol. **29**, 2008, pp. 372-374.
[4] S. Datta, "*Electronic Transport in Mesoscopic Systems*", Cambridge University Press, Cambridge 1995.
[5] International Technology Roadmap for Semiconductors, "*2008 Update*", http://www.itrs.net, 2008.
[6] M. Lundstrom, "*Elementary scattering theory of the Si MOSFET,*" IEEE Electron Device Letters, vol. **18**, 1997, pp. 361-363.
[7] L. Esaki, "*New Phenomenon in Narrow Germanium p - n Junctions*", Physical Review, vol. **109**, 1958, pp. 603-604
[8] J. Quinn, G. Kawamoto, and B. McCombe, "*Subband spectroscopy by surface channel tunneling*" Surface Science, vol. **73**, 1978, pp. 190-196.
[9] S. Banerjee, W. Richardson, J. Coleman and A. Chatterjee, "*A new three-terminal tunnel device,*" IEEE IEEE Electron Device Letters, vol. **8**, 1987, pp. 347-349.
[10] E. Takeda, H. Matsuoka, Y. Igura, S. Asai, "*A band to band tunneling MOS device (B2T-MOSFET)*" 1988 IEEE International Electron Devices Meeting, vol. **1**, 1988, pp. 402-405.
[11] T. Baba, "*Proposal for Surface Tunnel Transistors,*" Japanese Journal of Applied Physics, vol. **31**, 1992, pp. L455-L457.
[12] W.M. Reddick and G.A. Amaratunga, "*Silicon surface tunnel transistor,*" Applied Physics Letters, vol. **67**, 1995, p. 494.
[13] J. Koga and A. Toriumi, "*Negative differential conductance in three-terminal silicon tunneling device,*" Applied Physics Letters, vol. **69**, 1996, p. 1435.
[14] W. Hansch, "*A vertical MOS-gated Esaki tunneling transistor in silicon,*" Thin Solid Films, vol. **369**, 2000, pp. 387-389.

[15] C. Aydin, A. Zaslavsky, S. Luryi, S. Cristoloveanu, D. Mariolle, D. Fraboulet, and S. Deleonibus, "*Lateral interband tunneling transistor in silicon-on-insulator,*" Applied Physics Letters, vol. **84**, 2004, p. 1780.
[16] W. Hansch, C. Fink, J. Schulze and I. Eisele, "*A vertical MOS-gated Esaki tunneling transistor in silicon,*" Thin Solid Films, vol. **369**, 2000, pp. 387-389.
[17] J. Appenzeller, Y. Lin, J. Knoch, and P. Avouris, "*Band-to-Band Tunneling in Carbon Nanotube Field-Effect Transistors,*" Physical Review Letters, vol. **93**, 2004, pp. 1-4.
[18] J. Knoch, S. Mantl, and J. Appenzeller, "*Impact of the dimensionality on the performance of tunneling FETs: Bulk versus one-dimensional devices,*" Solid-State Electronics, vol. **51**, 2007, pp. 572-578.
[19] S.O. Koswatta, D. E. Nikonov, and M. Lundstrom, "*Computational study of carbon nanotube pin tunnel FETs,*" IEEE International Electron Devices Meeting, vol. **1** ,2005. IEDM Technical Digest, 2005, p. 518–521.
[20] S.O. Koswatta, M.S. Lundstrom, and D.E. Nikonov, "*Influence of phonon scattering on the performance of p-i-n band-to-band tunneling transistors,*" Applied Physics Letters, vol. **92**, 2008, p. 43125.
[21] M. Born, K. Bhuwalka, M. Schindler, U. Aelein, M. Schmidt, T. Sulima, and I. Eisele, "*Tunnel FET: A CMOS Device for high Temperature Applications,*" 2006 25th International Conference on Microelectronics, vol. **1**, 2006, pp. 124-127.
[22] A.S. Verhulst, W.G. Vandenberghe, K. Maex, and G. Groeseneken, "*Tunnel field-effect transistor without gate-drain overlap,*" Applied Physics Letters, vol. **91**, 2007, p. 053102.
[23] K. Boucart and A. Ionescu, "*A new definition of threshold voltage in Tunnel FETs,*" Solid-State Electronics, vol. **52**, 2008, pp. 1318-1323.
[24] S.O. Koswatta, M.S. Lundstrom, and D.E. Nikonov, "*Performance Comparison Between p-i-n Tunneling Transistors and Conventional MOSFETs,*" IEEE Transactions on Electron Devices, vol. **56**, 2009, pp. 456-465.
[25] A.S. Verhulst, W.G. Vandenberghe, K. Maex, S. De Gendt, M.M. Heyns, and G. Groeseneken, "*Complementary Silicon-Based Heterostructure Tunnel-FETs With High Tunnel Rates,*" IEEE Electron Device Letters, vol. **29**, 2008, pp. 1398-1401.
[26] E. Toh, G.H. Wang, L. Chan, G. Samudra, and Y. Yeo, "*Device physics and guiding principles for the design of double-gate tunneling field effect transistor with silicon-germanium source heterojunction,*" Applied Physics Letters, vol. **91**, 2007, p. 243505.
[27] T. Krishnamohan, D. Kim, S. Raghunathan, and K. Saraswat, "*Double-Gate Strained-Ge Heterostructure Tunneling FET (TFET) with record high drive currents and 60mV/dec subthreshold slope,*" 2008 IEEE International Electron Devices Meeting, vol. **67**, 2008, pp. 1-3.
[28] A. Bowonder and P. Patel, "*Low-voltage green transistor using ultra shallow junction and hetero-tunneling,*" Extended Abstracts - 2008 8th International Workshop on Junction Technology (IWJT '08), vol. **1**, 2008, pp. 93-96.
[29] K. Boucart, W. Riess, and A.M. Ionescu, "*Lateral Strain Profile as Key Technology Booster for All-Silicon Tunnel FETs,*" IEEE Electron Device Letters, vol. **30**, 2009, pp. 656-658.
[30] Q. Zhang, W. Zhao, and A. Seabaugh, "*Low-subthreshold-swing tunnel transistors,*" IEEE Electron Device Letters, vol. **27**, 2006, p. 297.
[31] K.K. Bhuwalka, M. Born, M. Schindler, M. Schmidt, T. Sulima, and I. Eisele, "*P-Channel Tunnel Field-Effect Transistors down to Sub-50 nm Channel Lengths,*" Japanese Journal of Applied Physics, vol. **45**, 2006, pp. 3106-3109.

[32] W.Y. Choi, B. Park, and J.D. Lee, "*Tunneling Field-Effect Transistors (TFETs) With Subthreshold Swing (SS) Less Than 60 mV/dec,*" Young, vol. **28**, 2007, pp. 743-745.
[33] M.T. Björk, J. Knoch, H. Schmid, H. Riel, and W. Riess, "*Silicon nanowire tunneling field-effect transistors,*" Applied Physics Letters, vol. **92**, 2008, p. 193504.
[34] F. Mayer, C. Le Royer, J. Damlencourt, K. Romanjek, F. Andrieu, C. Tabone, B. Previtali, and S. Deleonibus, "*Impact of SOI, $Si_{1-x}Ge_xOI$ and GeOI substrates on CMOS compatible Tunnel FET performance,*" 2008 IEEE International Electron Devices Meeting, vol. **4**, 2008, pp. 1-5.
[35] D. Kazazis, P. Jannaty, A. Zaslavsky, C. Le Royer, C. Tabone, L. Clavelier, and S. Cristoloveanu, "*Tunneling field-effect transistor with epitaxial junction in thin germanium-on-insulator,*" Applied Physics Letters, vol. **94**, 2009, p. 263508.
[36] C. Sandow, J. Knoch, C. Urban, Q. Zhao, and S. Mantl, "*Impact of electrostatics and doping concentration on the performance of silicon tunnel field-effect transistors,*" Solid-State Electronics, vol. **53**, 2009, pp. 1126-1129.
[37] A. Bowonder and P. Patel, "*Low-voltage green transistor using ultra shallow junction and hetero-tunneling,*" Extended Abstracts - 2008 8th International Workshop on Junction Technology (IWJT '08), IEEE, 2008, pp. 93-96.
[38] S. M. Sze, "*Physics of Semiconductor Devices*", 3rd edition, John Wiley & Sons, New York 2006.
[39] P. J. Price, "*Monte Carlo calculation of electron transport in solids*", Semiconductors and Semimetals, 1979
[40] M. Fischetti and S. Laux, "*Monte carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects,*" Physical Review B, vol. **38**, 1988, pp. 9721-9745.
[41] S. Datta, "*A simple kinetic equation for steady-state quantum transport,*" Journal of Physics: Condensed Matter, vol. **2**, 1990, p. 8023–8052.
[42] S. Datta, "*Nanoscale device modeling: the Green's function method,*" Superlattices and Microstructures, vol. **28**, 2000, p. 253-278.
[43] C. Auth and J. Plummer, "*Scaling theory for cylindrical, fully-depleted, surrounding-gateMOSFET's,*", IEEE Electron Device Letters, vol. **18**, 1997, p. 74–76.
[44] R. Lake, G. Klimeck, R.C. Bowen, and D. Jovanovic, "*Single and multiband modeling of quantum electron transport through layered semiconductor devices,*" Journal of Applied Physics, vol. **81**, 1997, p. 7845.
[45] H. Flietner, "*The E(k) Relation for a Two-Band Scheme of Semiconductors and the Application to the Metal-Semiconductor Contact,*" Physica Status Solidi (b), vol. **54**, 1972, pp. 201-208.
[46] E. Hatta, J. Nagao, and K. Mukasa, "*Tunneling through a narrow-gap semiconductor with different conduction- and valence-band effective masses,*" Journal of Applied Physics, vol. **79**, 1996, p. 1511.
[47] W. Kern and D. Puotinen, "*Cleaning Solutions Based on Hydrogen Peroxide for Use in Silicon Semiconductor Technology*", RCA Review, vol. **31**, 1970, p. 187–206.
[48] M. Lemme, T. Mollenhauer, H. Gottlob, W. Henschel, J. Efavi, C. Welch, and H. Kurz, "*Highly selective HBr etch process for fabrication of triple-gate nano-scale SOI-MOSFETs,*" Microelectronic Engineering, vol. **73**, 2004, p. 346–350.
[49] M. Posselt, "*Crystal-trim and its application to investigations on channeling effects during ion implantation,*" Radiation Effects and Defects in Solids, vol. **130**, 1994, pp. 87-119.
[50] D. K. Schroder, "*Semiconductor Material and Device Characterization*", 3rd edition, John Wiley & Sons, Hoboken, 2006.

[51] T. Ernst, S. Cristoloveanu, a. Vandooren, T. Rudenko, and J. Colinge, "*Recombination current modeling and carrier lifetime extraction in dual-gate fully-depleted SOI devices,*" IEEE Transactions on Electron Devices, vol. **46**, 1999, p. 1503-1509.
[52] "*Sentaurus Device User Manual*", version. **C-2009-06**, 2009, pp. 309-310
[53] A. Schenk, "*A model for the field and temperature dependence of Shockley-Read-Hall lifetimes in silicon,*" Solid-State Electronics, vol. **35**, 1992, p. 1585-1596.
[54] S. Feste, "*Physical Investigations of novel Materials and Structures for Nano-MOSFETs*", Dissertation, RWTH Aachen, Germany, 2009.
[55] Y. Sui and J. Appenzeller, "*Screening and interlayer coupling in multilayer graphene field-effect transistors,*" Nano Letters, vol. **9**, No. 8, 2009.
[56] Q. Ma, K. Wang, and J. Schulman, "*Band structure and symmetry analysis of coherently grown $Si_{1-x}Ge_x$ alloys on oriented substrates,*" Physical Review B, vol. **47**, 1993, pp. 1936-1953.
[57] Y. Sun, S.E. Thompson, and T. Nishida, "*Physics of strain effects in semiconductors and metal-oxide-semiconductor field-effect transistors,*" Journal of Applied Physics, vol. **101**, 2007, p. 104503.
[58] H. Irie, K. Kita, K. Kyuno, A. Toriumi, "*In-plane Mobility Anisotropy and Universality Under Uni-axial Strains in n-and p-MOS Inversion Layers on (100), (110), and (111) Si*", IEDM 2004 Technical Digest, p. 225-228, 2004.
[59] M.V. Fischetti and S.E. Laux, "*Band structure, deformation potentials, and carrier mobility in strained Si, Ge, and SiGe alloys,*" Journal of Applied Physics, vol. **80**, 1996, p. 2234.
[60] S. Feste, J. Knoch, D. Buca, and S. Mantl, "*Fabrication of uniaxially strained silicon nanowires,*" Thin Solid Films, vol. **517**, 2008, pp. 320-322.
[61] S. Thompson and T. Nishida, "*Uniaxial-process-induced strained-Si: extending the CMOS roadmap,*" IEEE Transactions on Electron Devices, vol. **53**, 2006, pp. 1010-1020.
[62] K. Uchida, T. Krishnamohan, K.C. Saraswat, Y. Nishi, "*Physical Mechanism of Electron Mobility Enhancement in Uniaxial Stressed MOSFETs and Impact of Uniaxial Stress Engineering in Ballistic Regime*", IEDM 2004 Technical Digest, p. 129, 2005.
[63] S. Feste, S. Habicht, Q. Zhao, D. Buca, and S. Mantl, "*Strained and unstrained Si nanowire FETs*", 39th European Solid-State Device Reasearch Conference ESSDERC 2009 (IEEE), Athens, Greece, 2009.
[64] T. Gebel, "*Flash lamp annealing with millisecond pulses for ultra-shallow boron profiles in silicon,*" Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms, vol. **186**, 2002, pp. 287-291.
[65] W. Yoo and K. Kang, "*Electrical activation of ultra-shallow B and BF_2 implanted silicon by flash anneal,*" Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms, vol. **237**, 2005, pp. 12-17.
[66] W. Skorupa, R. Yankov, W. Anwand, M. Voelskow, T. Gebel, D. Downey, and E. Arevalo, "*Ultra-shallow junctions produced by plasma doping and flash lamp annealing,*" Materials Science and Engineering B, vol. **114-115**, 2004, pp. 358-361.
[67] F. Lanzerath, D. Buca, H. Trinkaus, M. Goryll, S. Mantl, J. Knoch, U. Breuer, W. Skorupa, and B. Ghyselen, "*Boron activation and diffusion in silicon and strained silicon-on-insulator by rapid thermal and flash lamp annealings,*" Journal of Applied Physics, vol. **104**, 2008, p. 044908.
[68] Florakis, N. Mira, C. Grigoropoulos, K. Giannakopoulos, A. Halimaoui, and D. Tsoukalas, "*Non-melt laser annealing of Plasma Implanted Boron for ultra shallow junctions in Silicon,*" Materials Science and Engineering: B, vol. **154-155**, 2008, pp. 39-42.
[69] A. Florakis, D. Tsoukalas, I. Zergioti, K. Giannakopoulos, P. Dimitrakis, D. Papazoglou, G. Bennassayag, H. Bourdon, and A. Halimaoui, "*Laser annealing of plasma implanted boron for*

ultra-shallow junctions in Silicon," Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms, vol. **253**, 2006, pp. 13-17.
[70] N. Misra, C.P. Grigoropoulos, D.P. Stumbo, and J.N. Miller, "*Laser activation of dopants for nanowire devices on glass and plastic,"* Applied Physics Letters, vol. **93**, 2008, p. 121116.
[71] N. Misra, L. Xu, Y. Pan, N. Cheung, and C.P. Grigoropoulos, "*Excimer laser annealing of silicon nanowires,"* Applied Physics Letters, vol. **90**, 2007, p. 111111.
[72] N. Misra, L. Xu, M.S. Rogers, S.H. Ko, and C.P. Grigoropoulos, "*Pulsed laser annealing of semiconductor structures for functional devices,"* Physica Status Solidi (c), vol. **5**, 2008, pp. 3264-3270.
[73] B.J. Pawlak, R. Duffy, T. Janssens, W. Vandervorst, K. Maex, A.J. Smith, N.E. Cowern, T. Dao, and Y. Tamminga, "*Influence of the anneal conditions on arsenic activation during solid-phase epitaxial regrowth,"* Applied Physics Letters, vol. **87**, 2005, p. 031915.
[74] R. Ray and H. Fan, "*Impurity Conduction in Silicon,"* Physical Review, vol. **121**, 1961, pp. 768-779.
[75] R. Pires, R. Dickstein, S. Titcomb, and R. Anderson, "*Carrier freezeout in silicon,"* Cryogenics Dec 1990, vol.**30**, no.12, pp. 1064-8.
[76] F. Balestra and G. Ghibaudo, "*Brief review of the MOS device physics for low temperature electronics,"* Solid-State Electronics, vol. **37**, 1994, pp. 1967-1975.
[77] M.J. Kerr and A. Cuevas, "*General parameterization of Auger recombination in crystalline silicon,"* Journal of Applied Physics, vol. **91**, 2002, p. 2473.
[78] R. Van Overstraeten and R. Mertens, "*Heavy doping effects in silicon,"* Solid-State Electronics, vol. **30**, 1987, pp. 1077-1087.
[79] Litwin-Staszewska, W. Szymanska, and P. Piotrzkowski, "*The Electron Mobility and Thermoelectric Power in InSb at Atmospheric and Hydrostatic Pressures*", Physica Status Solidi (b), vol. **106(2)**,1981, pp. 551-559.
[80] T. Ashley, A. Dean, C. Elliott, R. Jefferies, F. Khaleque, and T. Phillips, "*High-speed, low-power InSb transistors,"* International Electron Devices Meeting 1997, IEDM Technical Digest, IEEE, pp. 751-754.
[81] S. Datta, T. Ashley, J. Brask, L. Buckle, M. Doczy, M. Emeny, D. Hayes, K. Hilton, R. Jefferies, T. Martin, T. Phillips, D. Wallis, P. Wilding, and R. Chau, "*85nm gate length enhancement and depletion mode InSb quantum well transistors for ultra high speed and very low power digital logic applications,"* IEEE InternationalElectron Devices Meeting, 2005, IEDM Technical Digest., IEEE, pp. 763-766.
[82] R.S. Wagner and W.C. Ellis, "*Vapor-Liquid-Solid Mechanism Of Single Crystal Growth*", Applied Physics Letters, vol. **4**, 1964, p. 89.
[83] H. Park, S. Prokes, M. Twigg, Y. Ding, and Z. Wang, "*Growth of high quality, epitaxial InSb nanowires,"* Journal of Crystal Growth, vol. **304**, 2007, pp. 399-401.
[84] Zwerdling, Solomon, B. Lax, and L.M. Roth, "*Oscillatory Magneto-Absorption in Semiconductors*", Physical Review, vol. **108(6)**, 1957, pp. 1402-1408

Appendix B
Conference Proceedings and Journal Papers

[1] C. Sandow, J. Knoch, C. Urban, Q. Zhao, and S. Mantl, "*Impact of electrostatics and doping concentration on the performance of silicon tunnel field-effect transistors*", Solid-State Electronics, vol. **53**, 2009, pp. 1126-1129.

[2] D. Candebat, Y. Zhao, C. Sandow, B. Koshel, C. Yang, J. Appenzeller, "*InSb nanowire field-effect transistors - electrical characterization and material analysis*", Proceedings of Device Research Conference, DRC 2009, Penn State University

[3] C. Urban, M. Emam, C. Sandow, Q.-T. Zhao, A. Fox, J.-P. Raskin, S. Mantl, "*High-Frequency Performance of dopant-segregated NiSi S/D SOI SB-MOSFETs.*", Proceedings of the 39th European Solid-State Device Research Conference, ESSDERC 2009, Athens, pp 149-52.

[4] M. Roeckerath, J.M. Lopes, E. Durğun Özben, C. Sandow, S. Lenk, T. Heeg, J. Schubert, and S. Mantl, "*Gadolinium scandate as an alternative gate dielectric in field effect transistors on conventional and strained silicon,*" Applied Physics A, vol. **94**, 2008, pp. 521-524.

[5] C. Urban, C. Sandow, Q. Zhao, and S. Mantl, "*High performance Schottky barrier MOSFETs on UTB SOI,*", Proceedings of 10th International Conference on Ultimate Integration of Silicon, IEEE, 2009, pp. **65-68.**

[6] Q. Zhao, S. B. Mi, C. L. Jia, C. Urban, C. Sandow, S. Habicht, S. Mantl, "*Epitaxial growth of $NiSi_2$ induced by sulfur segregation at the $NiSi_2$ /Si(100) interface,*" Journal of Materials Research, vol. **24**, 2009, pp. 135-139.

[7] M. Mueller, Q. Zhao, C. Urban, C. Sandow, D. Buca, S. Lenk, S. Estevez, and S. Mantl, "*Schottky-barrier height tuning of NiGe/n-Ge contacts using As and P segregation,*" Materials Science and Engineering: B, vol. **154-155**, 2008, pp. 168-171.

[8] C. Sandow, J. Knoch, C. Urban, and S. Mantl, "*Improving the performance of band-to-band tunneling transistors by tuning the gate oxide and the dopant concentration,*" Proceedings of Device Research Conference 2008, DRC 2008, Santa Barbara, pp. **79-80**

[9] M. Mueller, Q.T. Zhao, C. Urban, C. Sandow, U. Breuer, and S. Mantl, "*Schottky-barrier height tuning of Ni and Pt germanide/n-Ge contacts using dopant segregation,*" Proceedings of 9th International Conference on Solid-State and Integrated-Circuit Technology, IEEE, 2008, pp. **153-156.**

[10] C. Urban, Q. T. Zhao, C. Sandow, M. Müller, U. Breuer, S. Mantl, "*Schottky barrier height modulation by arsenic dopant segregation*", Proceedings of the 9th International Conference on Ultimate Integration of Silicon, 2008., Udine, pp. 151-154.

Appendix C
Curriculum Vitae

Personal Information

Name	Christian Philipp Sandow
Birthday	31 March, 1979
Place	Göttingen
Nationality	German

Career History

April 2006 - February 2010

Forschungszentrum Jülich
PhD-student working on the NEGF-simulation and top-down fabrication of silicon nanowire tunnel Field Effect Transistors (TFETs).

August 2008 - April 2009

Birck Nanotechnology Center, Purdue University, IN
Visiting research scholar with Prof. J. Appenzeller on the fabrication of silicon nanowire TFETs and InSb nanowire FETs.

April 2005 - March 2006

Rheinische Friedrich-Wilhelms Universität Bonn
Graduate assistant: Study of possible uses of silicon pixel-detectors in synchroton radiation imaging at free electron lasers.

Education

October 1999 - April 2005

Rheinische Friedrich-Wilhelms Universität Bonn
Studies of physics with focus on computational methods, particle accelerators and detectors.
Minors: computer science and chemistry.

April 13, 2005

Achieved the Diplom (equiv. Masters) with grade 1.0 "sehr gut"
Thesis title: "Studies on the reset mechanism and charge carrier collection of DEPFET single-pixels and matrices"

September 1991 - July 1998

High school "Gymnasium Corvinianum", Northeim, Germany
Abitur (high school degree) with an average grade of 1.8 "gut".
Major subjects: mathematics, physics

Appendix D
Acknowledgement

I would like to express my dearest thanks to the following people who helped and supported me in all imaginable ways:

- Prof. Dr. Siegfried Mantl who advised me well and always came up with good ideas, when I was temporarily stuck.
- Prof. Dr. L. Feld for undertaking the 'Korreferat' of this thesis.
- Prof. Dr. Joachim Knoch, who introduced me to the concept of TFETs and inspired my NEGF-simulations. Joachim taught me that experimental work can greatly benefit from theoretical input and that a good understanding of a device needs both, fabrication and simulation.
- Prof. Dr. Jörg Appenzeller, who hosted me at Purdue University and who is a critical mind. He taught me to ask the right questions. Furthermore, Jörg showed me how good 'science' is made and to how to brake down complex tasks into viable steps.
- Christoph Urban, who became a good friend and with whom I shared ups and downs during device fabrication. Our fruitful discussions were always the foundation of our motivation during difficult times. It is still mysterious to me why we share so many hobbies.
- Josh Smith, who was a constant source of inspiration and whose enthusiasm will hopefully be rewarded soon. Thank you Josh for the warm welcome at Purdue and all the unrewarded help you gave me.
- Martin Röckerath, who introduced me to the art of e-beam-evaporation and vacuum science. I have learned a lot from Martins practical skills.
- Ali Razavieh who became a friend a Purdue and who I admire because of his will to achieve his goals despite personal sacrifices.
- Yang Sui, Caitlin Burger, Drew Candebat, Hong-Yan Chen, Isaac Wildeson, Sahil M. Bansal ans Saptarshi Das who kindly accepted me as one of them and who gave me the feeling of being home although there were thousand of miles in between.
- Marcel Müller and Matthias Schmidt, who were pleasant office mates.
- Dr. Qing-Tai Zhao for introducing me to fabrication technology and his constant optimism.
- Dr. J. Schubert, Dr. J. Lopes, Dr. T. Heeg, E. Özben, S. Habicht, W. Heiermann, F. Lanzerath, N. Wilk, R. Minamisawa, Dr. R. Luptak, L. Knoll, M. Schnee, Dr. B. Holländer and A. Tiedemann for the pleasant and fair atmosphere in our group.

- Prof. Dr. M. Indlekofer for interesting discussions about 2D NEGF calculations.
- B. Rowe for teaching me e-beam lithography and how to operate Raith and Leica-Systems.
- S. Trellenkamp, who operated the e-beam writer masterfully and who always was faster than anticipated.
- M. Nonn for fabricating all the lithography masks and quick turn around times.
- W. Michelsen, S. Lenk, H. Schwallbach, W. Zander, A. Dahmen and K. Palmen for frequent and timely implantations, TEM images and discussions on LEGO Mindstorm.
- The cleanroom team at Jülich: A. Steffen, J. Müller, H. Wingens, K.-H. Deussen, S. Bunte, B. Hermanns and J. Mohr for e-beam-evaporations, oxidations, LPCVD and PECVD depositions, RIE&ICP etching and maintaining high standards in friendliness and cleanliness.
- The cleanroom team at the Birck Nanotechnology Center: S. Jurss, J. Coy, D. Hosler, G. Gardner, J. Grau, D. Lubelski, A. Mahmood, T. Miller and I. Young, who allowed me a large amount of freedom in the cleanroom and helped me to develop new RIE, oxidation and development recipes.
- Thomas Mollenhauer from AMO GmbH for ICP etchings that saved months.
- Brigitte Modolo and Mirjam Seller, who were great in administrative things.
- Mary Jo Totten and Deborah Starewich for my warm welcome at Birck.
- The Deutscher Akademischer Austausch Dienst for the nine month research scholarship which financed my stay at Purdue University.
- Daniel Hörnig for laughing about administrative nonsense.
- My parents for encouraging me, continuous support and their enthusiasm about my work.
- My wonderful wife Iris for her love and our beautiful daughter Miriam.