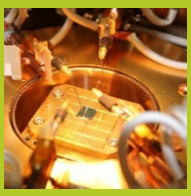
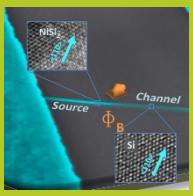
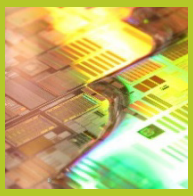
Research at

namlab







Untersuchung des elektronischen Transports an 28 nm MOSFETs und an Schottky-Barrieren FETs aus Silizium-Nanodrähten

Jürgen Beister

Technische Universität Dresden

Untersuchung des elektronischen Transports an 28nm MOSFETs und an Schottky-Barrieren FETs aus Silizium-Nanodrähten

Dipl.-Phys. Jürgen Beister

von der Fakultät Elektrotechnik und Informationstechnik
der Technischen Universität Dresden

zur Erlangung des akademischen Grades eines

Doktoringenieurs

(Dr.-Ing.)

genehmigte Dissertation

Vorsitzender:	Prof. Dr.-Ing. Dr. h.c. Karlheinz Bock
Gutachter:	Prof. Dr.-Ing. Thomas Mikolajick
	Prof. Dr.-Ing. habil. Jörg Schulze

Tag der Einreichung:	01. 11. 2016
Tag der Verteidigung:	02. 02. 2018

Vollkommenheit der Mittel
und Verworrenheit der Ziele
scheinen mir unsere Zeit zu charakterisieren.

— *Albert Einstein* [1]

Abstract

As modern microelectronics advances, enormous challenges have to be overcome in order to further increase device performance, enabling high-speed and ultra-low-power applications. State of the art Si MOSFETs with gate lengths below 32 nm use hafnium-based high-k dielectrics to reach electrical gate oxide thicknesses equivalent to ~ 1 nm SiO_2 or less. This leads to very high vertical fields in the MV/cm range, pulling the mobile channel charge closer to the gate oxide interface. Simultaneously, channel doping is increased. Both measures are needed in order to maintain gate control over the channel at reduced gate length. At the same time however, they lead to higher charge carrier scattering. For this reason, charge carrier mobility drops significantly and has become a critical parameter over the last years. Present technology nodes make use of strain engineering to partially recover this mobility loss. Even though carrier mobility is a crucial parameter for present technology nodes, it can not be determined accurately by methods typically available in industrial environments.

A major objective of this work is to study the magnetoresistance mobility μ_{MR} of strained VLSI devices based on a 28 nm ground rule. This technique allows for a more direct access to the carrier mobility, compared with the conventional current/ voltage and capacitance/ voltage mobility derivation methods like the effective mobility' μ_{eff}, in which series resistance, inversion charge density and effective channel length are necessary to extract the mobility values of the short channel devices. Aside from providing an anchor for accurate μ_{eff} measurements in linear operation conditions, μ_{MR} opens the possibility for investigations of the saturation region of the device, which cannot be accessed by μ_{eff}. Electron and hole mobilities of nFET and pFET devices with various gate lengths are studied from linear to saturation region. In

addition, the interplay between mobility enhancement due to strain improvement, and mobility degradation due to short channel effects with decreasing channel length is analyzed.

As a concept device for future nanoelectronic building blocks, silicon nanowire Schottky field-effect transistors are investigated in the second part of this work. These devices exhibit an ambipolar behavior, which gives the opportunity to measure both, electron and hole transport on a single device. The temperature dependence of the source/drain current for specific gate and drain voltages is analyzed within the framework of voltage dependent effective barrier heights. As a result, the temperature dependence of the tunnelling current is not only important for the subthreshold region, but also plays a significant role in the transistor "on-state". In addition, the effective barrier heights for electrons and holes tend towards the natural Schottky barriers of the $NiSi_2$-Si interface, if the applied external fields generate the case of flat band condition at the injection Schottky barrier, i.e. in the deep "off-state" region.

Inhaltsverzeichnis

Symbol- und Abkürzungsverzeichnis vii

1 Einleitung **1**

2 Theoretische Grundlagen **5**
2.1 Planare Feldeffekt-Transistoren 5
2.1.1 Allgemeines Schaltverhalten 5
2.1.2 Technologische Hebel zur Kurzkanalkontrolle . . 11
2.1.3 Ladungsträgerbeweglichkeit 12
2.2 Silizium-Nanodraht-Transistoren 28
2.2.1 Aufbau 28
2.2.2 Schaltverhalten 30

3 Charakterisierungsmethoden **41**
3.1 Ladungsträgerbeweglichkeit 41
3.1.1 Feldeffektbeweglichkeit 42
3.1.2 Effektive Beweglichkeit 43
3.1.3 Magnetowiderstandsbeweglichkeit 45
3.2 Serienwiderstand 50

4 Messaufbau **55**
4.1 Elektrische Charakterisierung ohne Magnetfeld: Vakuum-Kryomessplatz 56
4.2 Elektrische Charakterisierung mit Magnetfeld: PPMS Magnetfeldkryostat 59
4.2.1 Pucksystem zur Transistorcharakterisierung in hohen Magnetfeldern 61
4.2.2 Vorkehrungen zum Schutz vor elektrostatischer Entladung 66
4.2.3 Messablauf 67

5 Ergebnisse der Untersuchungen an MOSFETs 69
5.1 Struktureller Aufbau 69
5.2 Schaltverhalten 71
5.2.1 Transfer- und Ausgangskennlinien 71
5.2.2 Kapazitäts-Spannungskurven 74
5.2.3 Elektrische Bestimmung des Serienwiderstands . 75
5.3 Ladungsträgerbeweglichkeit 77
5.3.1 Magnetowiderstands-Methode 77
5.3.2 Split-CV-Methode 105
5.3.3 Magnetowiderstands-Streukoeffizient 112
5.4 Zusammenfassung und Diskussion der verschiedenen Methoden zur Bestimmung der Beweglichkeit 113

6 Ergebnisse der Untersuchungen an SiNW Transistoren 119
6.1 Struktureller Aufbau 119
6.2 Schaltverhalten 120
6.2.1 Hysterese 120
6.2.2 Transfer- und Ausgangskennlinien 122
6.2.3 Mikroskopisches Modell zum Verständnis der Temperaturabhängigkeit des Drainstroms 126
6.3 Extraktion der Schottkybarriere 130
6.3.1 Temperaturabhängigkeit der Transferkennlinien . 130
6.3.2 Einfluss externer Felder auf die effektive Barriere 132
6.4 Zusammenfassung der Ergebnisse an SiNW Transistoren 136

7 Zusammenfassung 137

A Anhang 141
A.1 Banddiagramme der MOS-Struktur 142
A.2 Messroutine Magnetowiderstandsbeweglichkeit 144
A.2.1 Datenaufbereitung 144
A.2.2 Fitroutine 145

Literaturverzeichnis 149
Eigene Publikationen 167
Danksagungen 169

Symbole

Symbol	Einheit	Beschreibung
A^*	$\mathrm{A{\cdot}K^{-2}{\cdot}m^{-2}}$	Richardsonkonstante
A_{ges}	$\mathrm{m^2}$	Fläche
B	T	magnetische Feldstärke
$\vec{B}$	T	Magnetfeldvektor
C_{do}	F	direkte Überlappkapazität
C_{ga}	$\mathrm{F{\cdot}m^{-2}}$	Kapazität zw. Gate und S/D/B
C_{gb}	$\mathrm{F{\cdot}m^{-2}}$	Kapazität zw. Gate und Bulk
C_{gc}	$\mathrm{F{\cdot}m^{-2}}$	Kapazität zw. Gate und S/D
C_{if}	F	innere Streukapazität
C_{of}	F	äußere Streukapazität
C_{ov}	F	Gesamt-Überlappkapazität
C_{ox}	F	Oxidkapazität, starke Inversion
C'_{ox}	$\mathrm{F{\cdot}m^{-2}}$	Oxidkapazität in starker Inversion, flächennormiert
C_{RLZ}	F	Kapazität der Raumladungszone
d_{ox}	m	Dicke des Gateoxids
$\vec{E}$	$\mathrm{V{\cdot}m^{-1}}$	Vektor des elektrischen Feldes
E_{c}	$\mathrm{V{\cdot}m^{-1}}$	kritisches elektrisches Feld
$E_{\mathrm{eff,lat}}$	$\mathrm{V{\cdot}m^{-1}}$	effektives laterales el. Feld (D $\rightarrow$ S)
$E_{\mathrm{eff,vert}}$	$\mathrm{V{\cdot}m^{-1}}$	effektives vertikales el. Feld (G $\rightarrow$ B)
ϵ	J	Energie
ϵ_0	J	energetisches Bandminimum
ϵ_{g}	J	Bandlücke
ϵ_{f}	J	Fermienergie
ϵ_{i}	J	intrinsische Fermienergie
ϵ_{c}	J	Leitungsbandminimum
ϵ_{v}	J	Valenzbandmaximum
$\vec{F}$	N	Kraft
f	Hz	Frequenz
$f_{(\epsilon)}$	1	Fermi-Dirac-Funktion

g_d	$\mathrm{A{\cdot}V^{-1}}$	Drainleitwert
g_d^*	$\mathrm{A{\cdot}V^{-1}}$	Kanalleitwert
g_m	$\mathrm{A{\cdot}V^{-1}}$	Transferleitwert
$g_\mathrm{m,i}$	$\mathrm{A{\cdot}V^{-1}}$	intrinsischer Transferleitwert
I_0	A	Drainstrom ohne Magnetfeld
I_ds	A	Drainstrom
$I_\mathrm{ds,lin}$	A	Drainstrom im linearen Fall
$I_\mathrm{ds,off}$	A	Drainstrom im Auszustand
$I_\mathrm{ds,sat}$	A	Drainstrom im Sättigungsfall
I_{V_th}	A	V_th-Stromwertkriterium
$\vec{j}$	$\mathrm{A{\cdot}m^{-2}}$	Stromdichte
j_therm	$\mathrm{A{\cdot}m^{-2}}$	Stromdichte, thermion. Emission
L	m	Länge (Transistorbezeichnung)
L_design	m	designte Gatelänge
L_eff	m	effektive Kanallänge
δL	m	Differenz zwischen L_design und L_eff
m^*	kg	effektive Masse
m_e^*	kg	effektive Elektronenmasse
m_h^*	kg	effektive Löchermasse
n	$\mathrm{m^{-3}}$	Elektronenkonzentration
$N_{(\epsilon)}$	$\mathrm{J^{-1}m^{-2}}$	Zustandsdichte (2D-Fall)
$N_{(\epsilon)}$	$\mathrm{J^{-1}m^{-3}}$	Zustandsdichte (3D-Fall)
N_depl	$\mathrm{m^{-2}}$	Ladungsträgerdichte in Verarmung
N_inv	$\mathrm{m^{-2}}$	Inversionsladungsträgerdichte
N_s	$\mathrm{m^{-2}}$	Ladungsträgerdichte (2D-Fall)
N_s	$\mathrm{m^{-3}}$	Ladungsträgerdichte (3D-Fall)
p	m^{-3}	Löcherkonzentration
$Q_\mathrm{inv}^{'}$	$\mathrm{C \cdot cm^{-2}}$	Inversionsladung, flächennormiert
R	Ω	elektrischer Widerstand
R_0	Ω	Widerstand ohne Magnetfeld
R_B	Ω	Widerstand im Magnetfeld
R_CA	Ω	S/D Zuleitungs- und Kontaktstöpselwiderstand
R_ch	Ω	Kanalwiderstand
R_cont	Ω	Kontaktwiderstand

R_{d}	Ω	drainseitiger Serienwiderstand
R_{ext}	Ω	Widerstand S/D-Extensiongebiete
R_{ges}	Ω	Gesamt-Transistorwiderstand
R_{s}	Ω	sourceseitiger Serienwiderstand
R_{sd}	Ω	Anschlusswiderstand zwischen Kontakt und Kanal
R_{SDdiff}	Ω	Widerstand S/D-Diffusionsgebiete
R_{ser}	Ω	Gesamt-Serienwiderstand
$R_{\mathrm{sil,int}}$	Ω	Silizid-Widerstand an S und D
R_{SRch}	Ω	Streuwiderstand am Kanalanschluss
R_{SRcont}	Ω	Streuwiderstand an S/D
S	m^2	Querschnittsfläche
S	mV/dec	Unterschwellstrom-Kennzahl
T	K	Temperatur
$\tilde{t}$	1	Tunnelwahrscheinlichkeit
$\vec{v}$	$\mathrm{m{\cdot}s^{-1}}$	Geschwindigkeit
V_{ac}	V	Wechselspannung
V_{dd}	V	positive Versorgungsspannung
V_{ds}	V	Drain-Source-Spannung
V_{ds}^*	V	intrinsische Drain-Source-Spannung
$V_{\mathrm{d,lin}}$	V	V_{d}-Kriterium für lineare Operation
$V_{\mathrm{d,sat}}$	V	Sättigungsspannung
v_{eff}	$\mathrm{m{\cdot}s^{-1}}$	effektive Geschwindigkeit
V_{eff}	V	effektive Spannung
V_{fb}	V	Flachbandspannung
V_{gs}	V	Gate-Source-Spannung
V_{gs}^*	V	intrinsische Gate-Source-Spannung
V_{ss}	V	neg. Versorgungsspannung (GND)
V_{th}	V	Einsatzspannung
$V_{\mathrm{th,nF}}$	V	Einsatzspannung, SiNW nFET-Ast
$V_{\mathrm{th,pF}}$	V	Einsatzspannung, SiNW pFET-Ast
W	m	Weite
β	$\mathrm{A{\cdot}V^{-2}}$	Verstärkungsfaktor
χ	eV	Elektronenaffinität
ε_{r}	1	relative Permittivität

$\varepsilon_{\mathrm{Si}}$	1	relative Permittivität von Si
$\varepsilon_{\mathrm{SiO_2}}$	1	relative Permittivität von SiO_2
ϕ_{B}	eV	Schottkybarriere
ϕ_{M}	eV	Metall-Austrittsarbeit
λ	m	Screeninglänge
γ_{eff}	1	R_{ser}-Korrekturfaktor für μ_{eff}
γ_{MR}	1	R_{ser}-Korrekturfaktor für μ_{MR}
κ	m^{-1}	Wellenzahl
μ	$\mathrm{m}^2{\cdot}\mathrm{V}^{-1}{\cdot}\mathrm{s}^{-1}$	Beweglichkeit
$\mu_{(\epsilon)}$	$\mathrm{m}^2{\cdot}\mathrm{V}^{-1}{\cdot}\mathrm{s}^{-1}$	mikroskopische Beweglichkeit
μ_0	$\mathrm{m}^2{\cdot}\mathrm{V}^{-1}{\cdot}\mathrm{s}^{-1}$	makroskop. (Drift-) Beweglichkeit
μ_{ch}	$\mathrm{m}^2{\cdot}\mathrm{V}^{-1}{\cdot}\mathrm{s}^{-1}$	Kanalbeweglichkeit
μ_{eff}	$\mathrm{m}^2{\cdot}\mathrm{V}^{-1}{\cdot}\mathrm{s}^{-1}$	effektive Beweglichkeit
μ_{FE}	$\mathrm{m}^2{\cdot}\mathrm{V}^{-1}{\cdot}\mathrm{s}^{-1}$	Feldeffektbeweglichkeit
μ_{MR}	$\mathrm{m}^2{\cdot}\mathrm{V}^{-1}{\cdot}\mathrm{s}^{-1}$	Magnetowiderstands-Beweglichkeit
η	1	Feldparameter: $1/2$ für Elektronen, $1/3$ für Löcher
$\boldsymbol{\sigma}$	$\mathrm{A}{\cdot}\mathrm{V}^{-1}$	Leitfähigkeitstensor
σ_0	$\mathrm{A}{\cdot}\mathrm{V}^{-1}$	makroskopische Leitfähigkeit
$\sigma_{(\epsilon)}$	$\mathrm{A}{\cdot}\mathrm{V}^{-1}$	mikroskopische Leitfähigkeit
τ	s	Streuzeit
Ψ_{B}	V	Abstand intrinsisches zu tatsächlichem Ferminiveau
Ψ_{s}	V	Oberflächenpotential

Konstanten

Symbol	**Wert**	**Beschreibung**
ε_0	$8,854188 \cdot 10^{-12}\,\mathrm{A{\cdot}s{\cdot}V^{-1}{\cdot}m^{-1}}$	elektr. Feldkonstante
$\hbar$	$6,582119 \cdot 10^{-16}\,\mathrm{eV \cdot s}$	reduziertes Planck'sches Wirkungsquantum
k_B	$8,617330 \cdot 10^{-5}\ \mathrm{eV{\cdot}K^{-1}}$	Boltzmannkonstante
q	$1,602189 \cdot 10^{-19}\,\mathrm{C}$	Elementarladung

Abkürzungen

B	Bulk (Substratkontakt des Transistors)
CPEN	Compressive Plasma Enhanced Nitride (druckverspannte Deckschicht)
CMOS	Complementary Metal-Oxide-Semiconductor
D	Drain
DIBL	Drain Induced Barrier Lowering
EOT	Equivalent Oxide Thickness
ESD	Electrostatic Discharge (elektrostatische Entladung)
eSiGe	epitaktisches Silizium-Germanium
G	Gate
GIDL	Gate Induced Drain Leakage
GND	Ground
LCC	Leadless Chip Carrier (Pinloses Chipgehäuse)
HKMG	High-k Metal Gate
MOSFET	Metal-Oxide-Semiconductor Field-Effect Transistor
MR	Magnetoresistance
NPEN	Neutral Plasma Enhanced Nitride (unverspannte Deckschicht)
PPMS	Physical Property Measurement System
REM	Rasterelektronenmikroskopie
RMS	Root Mean Square
RTN	Random Telegraph Noise
S	Source
SiNW	Silicon Nanowire (Silizium Nanodraht)
SMU	Source Measure Unit (Messkanal)
TEM	Transmissionselektronenmikroskopie
TPEN	Tensile Plasma Enhanced Nitride (zugverspannte Deckschicht)
VLS	Vapor-Liquid-Solid (Wachstumsprozess)
VLSI	Very Large Scale Integration

1 Einleitung

Bereits im Jahre 1938 berichteten Pohl und Hilsch von ersten erfolgreichen Arbeiten zu einem Festkörperverstärker welcher auf Kaliumbromidkristallen basierte [2]. Damit konnte zum ersten Mal gezeigt werden, dass "eine Steuerung elektrischer Ströme in festen Körpern mit Hilfe eines Steuergitters" grundsätzlich möglich ist. 1941 gelang Konrad Zuse mit der Konstruktion des elektro-mechanisch arbeitenden *Z3* der Bau des weltweit ersten programmierbaren Computers [3, 4]. Mit der Erfindung des Transistors durch John Bardeen, Walter Brattain und William Shockley [5–7] und der integrierten Schaltung durch Jack Kilby und Robert Noyce [8, 9] begann schließlich die rasante Entwicklung der Mikroelektronik welche bis heute anhält und immer noch ungebrochen an Fahrt aufnimmt [10]. Sie beeinflusst sämtliche Lebensbereiche in einer tiefgreifenden, fundamentalen Weise, wie es wahrscheinlich keine Technologie zuvor je vermochte. Immer neue Anwendungen werden erschlossen, getrieben durch höhere Komplexität, sinkende Kosten, geringeren Stromverbrauch und die zunehmende Vernetzung elektronischer Komponenten. Der zentrale Schlüssel für diese Entwicklung liegt in der fortschreitenden Miniaturisierung der Schaltkreise. Dabei folgt die Industrie dem „Moore'schen Gesetz", einer Beobachtung von Gordon Moore aus dem Jahr 1965, dass sich die Anzahl der Transistoren pro Flächeneinheit in integrierten Schaltkreisen etwa alle zwei Jahre verdoppelt [11]. Um die Jahrtausendwende überschritt der Anteil an digital gespeicherter Information den der analog gespeicherten [12] und markiert damit den Beginn eines neuen, „digitalen Zeitalters". Das Aufrechterhalten der fortlaufenden Verkleinerung der Transistoren, der Grundbausteine jedes integrierten Schaltkreises, stellt die Chipindustrie immer wieder vor enorme physikalische und

technologische Herausforderungen[1] und wird ermöglicht durch stete Innovation in lithographischen Abbildungsverfahren, neuen Materialien und Bauelementintegration [13–15]. Mit immer kleineren Transistordimensionen treten Kurzkanaleffekte in den Vordergrund und es müssen besondere Maßnahmen getroffen werden, um eine möglichst hohe Kontrolle des Gates über den Kanal aufrechtzuerhalten. Dabei kommen immer höhere vertikale Felder zwischen Gate und Kanalbereich zum Einsatz, wodurch die mobilen Ladungsträger des Kanals stärker an das Gateoxid gezogen werden und deren Streuung an der Grenzfläche zunimmt. Zusammen mit immer höheren Kanaldotierungen wird so die Ladungsträgerbeweglichkeit mit fortschreitenden Technologiegenerationen reduziert [16].

Während sich die Ladungsträgerbeweglichkeit in den letzten Jahren für den Transistor immer stärker zu einer Performance-limitierenden, kritischen Größe entwickelt hat [17], ist ihre akurate Bestimmung über konventionelle Methoden schwieriger geworden und mit immer größeren Unsicherheiten behaftet. Des Weiteren ist der für den Transistorbetrieb wesentliche Sättigungsbereich mit Standardverfahren überhaupt nicht zugänglich. Die vorliegende Arbeit widmet sich daher im ersten Teil der Untersuchung der Ladungsträgerbeweglichkeit an Feldeffekttransistoren in einer 28 nm HKMG- Technologie[2] mittels des Magnetowiderstandsverfahrens (engl.: „magnetoresistance“, MR).

Dieses Verfahren bietet – zum Preis eines höheren experimentellen Aufwands – einen direkteren Zugang zur Beweglichkeit der Ladungsträger, verglichen mit konventionellen Methoden wie der Messung der Feldeffektbeweglichkeit oder der effektiven Beweglichkeit. So ist die Kenntnis der exakten elektrisch effektiven Kanallänge nicht erforderlich und die Abhängigkeit von Anschlusswiderständen geringer. Wie in der Arbeit gezeigt wird, kann hierdurch eine deutlich höhere Aussagegenauigkeit erzielt werden. Für eine gewisse Technologie bestimmt,

[1] Bereits um 1975 hatten die Transistoren mit Abmessungen um 5 µm die Größe eines weißen Blutkörperchens erreicht, heutige Strukturen sind etwa 200 Mal kleiner.

[2] Technologie mit Metall-Gatebahnen und Gateoxiden hoher Permittivität „High-k Metal Gate“ (HKMG)

kann die Magnetowiderstands-Beweglichkeit so als Anker für die Bewertung und Gültigkeit der effektiven Beweglichkeit dienen. Darüber hinaus kann mit dem MR-Verfahren der Sättigungsbereich des Transistors bei hohen Drainspannungen untersucht werden, welcher über andere Verfahren nicht zugänglich ist. Ein abschließender Teil widmet sich unter Ausnutzung beider Messverfahren dem Einfluss von Verspannungstechnologien zur Steigerung der Ladungsträgerbeweglichkeit.

Aufgrund der Herausforderungen die sich bei weiterer Skalierung der Transistoren ergeben, wird im Hinblick auf zukünftige Technologiegenerationen auch an neuen Transistorkonzepten geforscht. Seit dem 22 nm Knoten sind sogenannte „Fin-FET“ Transistoren in kommerzieller Fertigung, bei welchen das Gate die Siliziumbahn von drei Seiten umschließt, um die Kanalkontrolle weiter zu steigern [18, 19]. Die ideale Geometrieform wäre dabei ein Si- Nanodraht, der von allen Seiten durch ein Gate umschlossen wird. Über die reine Skalierung („More Moore“) hinaus gibt es außerdem Bestrebungen, eine erweiterte Funktionalisierung der Transistoren zu erreichen („More than Moore“) [20]. In diesem Zusammenhang hat im letzten Jahrzehnt die Forschung an Silizium- Nanodrähten zunehmend an Bedeutung gewonnen. Die außergewöhnliche Kombination aus einer quasi- eindimensionalen Geometrie in Verbindung mit der hochentwickelten, siliziumbasierten Chiptechnologie verspricht neue Bauelementkonzepte jenseits des heute üblichen Standards [21–24]. Ein Si-Nanodraht-basierter Transistortyp, der Gegenstand aktueller Forschungs- und Entwicklungsbemühungen ist, ist der Schottky-Feldeffekt-Transistor (FET) mit extrem abrupten Schottkykontakten an Source und Drain. Dieser weist ein im Vergleich zum konventionellen MOSFET sehr unterschiedliches Schaltverhalten auf. So konnten Si-Nanodrähte verwendet werden, um rekonfigurierbare Transistoren zu realisieren, welche je nach angelegten Programmierspannungen wahlweise unipolares nFET- oder pFET-Verhalten in einem einzelnen Bauelement zeigen [25, 26]. Darüber hinaus lässt sich die Oberfläche von Si-Nanodrähten funktionalisieren, sodass der Transistor als Sensor für bestimmte Reagenzien verwendet werden kann [27–29]. In solchen Nanodraht-Schottkykontakt-FETs wird die

Performance – abgesehen von der Gatekopplung – vor allem durch die Injektion der Ladungsträger über die Schottky-Barriere in den Kanal begrenzt. Um zu einem tieferen Verständnis der Schaltvorgänge in diesen Transistoren beizutragen, widmet sich der zweite Teil dieser Arbeit der Untersuchung von Si-Nanodraht-FETs mit Nickelsilizidierten Source/Drain-Anschlüssen und Rückseitengate. Ein hierzu entwickeltes, einfaches phänomenologisches Modell dient dabei der Veranschaulichung des Temperaturverhaltens, welches hauptsächlich durch die Ladungsträgerinjektion durch Tunneln über die Schottkybarriere bestimmt wird. Mit Hilfe von temperaturabhängigen Messungen wird schließlich der Einfluss der von den beiden Schottkykontakten stammenden Energiebarrieren auf die Elektronen- und Löcherleitung in diesen Bauelementen untersucht.

2 Theoretische Grundlagen

2.1 Planare Feldeffekt-Transistoren

2.1.1 Allgemeines Schaltverhalten

Hauptelement des MOSFETs ist die MOS-Kondensatorstruktur: auf einem halbleitenden Substrat (S) wird ein Gateoxid (O) aufgebracht, auf welches ein metallischer Gateschichtstapel (M) folgt. Für einen n-MOS Transistor verwendet man üblicherweise leicht p-dotierte Substrate, während bei p-MOS Transistoren n-dotierte Substrate zum Einsatz kommen. Zum Aufbau eines Transistors wird nun das halbleitende Sub-

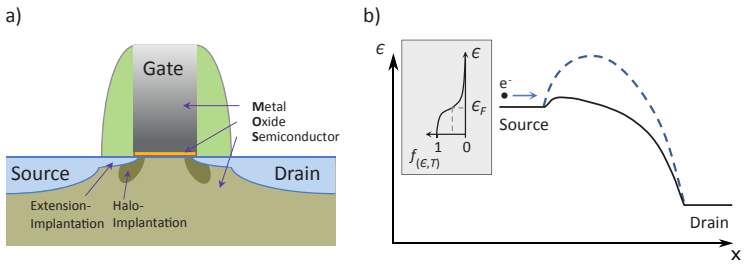


Abbildung 2.1: a) Schematischer Aufbau eines planaren MOSFETs, und b) nFET Leitungsbandkante entlang des Kanals nach [30] für $V_{\mathrm{ds}} > 0$ und mit skizzierter Fermiverteilung der Elektronen an Source. Durchgehende Linie im AN- Zustand ($V_{\mathrm{gs}} > V_{\mathrm{th}}$), gestrichelte Linie im AUS-Zustand ($V_{\mathrm{gs}} < V_{\mathrm{th}}$).

strat der MOS-Struktur um Source- und Drainanschlüsse ergänzt, welche invers zum Substrat dotiert werden, siehe Abb. 2.1a. So entsteht

im Kanalbereich eine Energiebarriere für die Majoritätsladungsträger aus den Source- und Draingebieten, wie in Abb. 2.1b am Beispiel eines nFET-Transistors gezeigt. Während Source und Drain für einen geringen Kontaktwiderstand hoch dotiert werden und räumlich etwas unter die Substratoberfläche in die Tiefe reichen, wird der eigentliche Kanalanschluss durch eine im Vergleich sehr flach eingebrachte „Extension-Implantation“ realisiert. Dies verringert das Risiko eines Kurzschlusses zwischen Source und Drain durch „tiefes Übersprechen“. Darüber hinaus werden sogenannte „Halo-Implantationen“ in den Anschlussbereichen platziert, welche lokal die Substratdotierung erhöhen. Dadurch werden die Raumladungszonen zwischen dem Bereich der Extension-Implantation und Kanal schmal gehalten, was die Herstellung sehr kurzer Gatelängen L ermöglicht.

Wird an den MOSFET eine Drain-Source-Spannung V_{ds} angelegt, während die Gate-Source-Spannung $V_{\mathrm{gs}} < V_{\mathrm{th}}$ ist, können aufgrund der Energiebarriere (gestrichelte Linie in Abb. 2.1b) nur die wenigen Source-Elektronen, die aufgrund der Fermi-Dirac-Verteilung über eine ausreichend hohe thermische Energie verfügen, zur Drain fließen. Der Transistor ist im „AUS“- Zustand und befindet sich im Unterschwellstrom- Bereich. Mit zunehmendem V_{gs} werden die Bänder im Kanalbereich nach unten verschoben und die Energiebarriere kleiner. Es können somit mehr und mehr Elektronen der Fermiverteilung im Source- Gebiet an der Leitung teilnehmen. Die thermionische Emission von Ladungsträgern von Source in den Kanal nimmt zu, und der Drainstrom I_{ds} steigt, bis sich der Transistor letztlich oberhalb der Einsatzspannung im „AN“- Zustand befindet (Abb. 2.1a und 2.2a). Für $V_{\mathrm{gs}} > V_{\mathrm{th}}$ bildet sich unter dem Gatebereich ein oberflächennaher Inversionskanal aus Minoritätsladungsträgern. Dabei wirkt sich nicht mehr die sourceseitige Injektion von Ladungsträgern limitierend auf den Stromfluss aus. Stattdessen bestimmen der Kanal- und die Anschlusswiderstände den Strom über den Transistor. Die Ladungsträgerdichte steigt mit zunehmendem V_{gs} und ist für kleine V_{ds} räumlich homogen und nahezu konstant. Für kleine V_{ds} und bei festem V_{gs} verhält sich der Kanal bezüglich V_{ds} wie ein Ohm’scher Leiter. Bei Erhöhung der Drain-Source-Spannung V_{ds} steigt I_{ds} daher zunächst linear

an (Linearbereich in Abb. 2.2b). Mit zunehmendem V_{ds} nimmt jedoch die Potentialdifferenz zwischen Gate und Drain ab, weshalb drainseitig aufgrund des geringeren Overdrives die Kanalladung Q'_{inv} reduziert wird. Dabei wird der Anstieg von $I_{\mathrm{ds}}(V_{\mathrm{ds}})$ schwächer und folgt einem parabolischen Verhalten. Bei $V_{\mathrm{ds}} = V_{\mathrm{ds,sat}}$ wird der Kanal drainseitig vollständig abgeschnürt („pinch-off“) und der Drainstrom des Langkanaltransistors erreicht sein Maximum. Mit weiterer Erhöhung von V_{ds} wandert der Abschnürpunkt in Richtung Source. Dabei bleibt das Potential am Abschnürpunkt konstant und I_{ds} nimmt praktisch nicht weiter zu (Sättigungsbereich). Mit der Verschiebung des Abschnürpunkts geht allerdings auch eine Verkürzung der elektrisch effektiven Kanallänge L_{eff} einher. Dies spielt vor allem bei kleinen Kanallängen eine wichtige Rolle, bei denen der Drainstrom jenseits von $V_{\mathrm{ds,sat}}$ nicht konstant bleibt, sondern linear ansteigt. Die in Abb. 2.2b skizzierte Ausgangskennlinie eines Langkanaltransistors zeigt leichte Anzeichen dieses Verhaltens. Weitere Details und tiefergehende Beschreibungen zu Physik und Schaltverhalten von Transistoren finden sich zum Beispiel in [30] oder [31]. Banddiagramme für verschiedene Spannungsfälle im vertikalen Schnitt durch die MOS Struktur sind in Anhang A.1 dargestellt.

Die folgenden elektrischen Kenngrößen eignen sich sehr gut, um das Schaltverhalten eines Transistors durch wenige Parameter zu beschreiben. Dabei gelten die hier gegebenen mathematischen Beschreibungen im Fall eines Langkanaltransistors. Für kurze Kanallängen sind die Formeln entsprechend anzupassen.

- Im Unterschwellstrombereich, in dem sich die MOS-Struktur in schwacher Inversion befindet, steigt der Source-Drain-Strom I_{ds} exponentiell an. Dieses Verhalten wird durch die Unterschwellstrom-Kennzahl „subthreshold swing“ oder auch „slope“ S beschrieben, welche den Anstieg von $I_{\mathrm{ds}}(V_{\mathrm{gs}})$ in Millivolt pro Dekade Strom angibt:

$$S = \frac{\mathrm{d}V_{\mathrm{gs}}}{\mathrm{d}\log_{10} I_{\mathrm{ds}}} = \phi_t(1 + C_{\mathrm{RLZ}}/C_{\mathrm{ox}}) \ln 10\,. \tag{2.1}$$

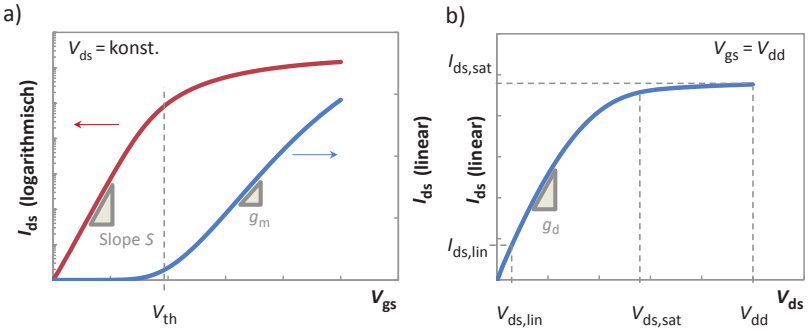


Abbildung 2.2: Übertragungskennlinie eines Langkanal- nMOSFETs: a) $I_{\mathrm{ds}}(V_{\mathrm{gs}})$ Transferkennlinien in logarithmischer und linearer Darstellung, b) $I_{\mathrm{ds}}(V_{\mathrm{ds}})$ Ausgangskennlinie mit leichtem Anzeichen einer Kanallängenmodulation, da I_{ds} oberhalb von $V_{\mathrm{ds,sat}}$ noch schwach linear ansteigt. Der Sättigungsstrom $I_{\mathrm{ds,sat}}$ wird üblicherweise bei $V_{\mathrm{ds}} = V_{\mathrm{dd}}$ abgelesen.

C_{ox} und C_{RLZ} sind die Kapazitäten des Gateoxids und der Raumladungszone (RLZ) im Substrat. Mit der Temperatur T und der Elementarladung q berechnet sich die Temperaturspannung ϕ_t zu

$$\phi_t = \frac{k_{\mathrm{B}}T}{q} \,. \tag{2.2}$$

Je steiler der Verlauf von $I_{\mathrm{ds}}(V_{\mathrm{gs}})$ im Unterschwellbereich ist, desto besser lässt sich der Transistor abschalten. Bei $T = 300\,\mathrm{K}$ beträgt $\phi_t \approx 26\,\mathrm{mV}$. Im Idealfall ist $C_{\mathrm{RLZ}}/C_{\mathrm{ox}} \approx 0$, sodass sich bei Raumtemperatur ein maximaler Anstieg von $S \approx 26\,\mathrm{mV} \cdot \ln 10 = 60\,\mathrm{mV}/\mathrm{Dekade}$ ergibt.

- Die Einsatzspannung V_{th} ist die Gate-Source-Spannung V_{gs}, bei der sich ein Inversionskanal bildet und der Transistorstrom mit steigender Gate-Source-Spannung rapide ansteigt. Dieser Punkt

ist erreicht, wenn die feldinduzierte Ladungsträgerkonzentration im Kanalbereich $n(x = 0)$ gleich der Substratdotierung N_S ist und beginnt, diese zu überschreiten. Dies ist bei einer Bandverbiegung von $q\Psi_\mathrm{s} = 2q\Psi_\mathrm{B} = 2(\epsilon_\mathrm{i} - \epsilon_\mathrm{f})$ der Fall [30], wobei ϵ_i das intrinsische und ϵ_f das tatsächliche Ferminiveau sind. Inversion setzt damit ein, wenn am nFET das Valenzband an der Grenzfläche ungefähr auf die Höhe des Ferminiveaus ϵ_F im Substrat verschoben ist. Da die Bandverbiegung nicht direkt über elektrische Messungen zugänglich ist, haben sich in der Praxis verschiedene Verfahren durchgesetzt, welche diesem Operationsbereich nahe kommen. In dieser Arbeit kamen hauptsächlich zwei weit verbreitete Methoden zur V_th- Bestimmung zum Einsatz: Bei der Konstantstrommethode ist $V_\mathrm{th} = V_\mathrm{gs}$, wenn $I_\mathrm{ds} = I_{V_\mathrm{th}} \cdot W/L$. Dabei entspricht I_{V_th} einem zuvor festgelegten Stromwertkriterium. Bei der Methode der linearen Extrapolation wird im Steigungsmaximum eine Tangente an $I_\mathrm{ds}(V_\mathrm{gs})$ angelegt, deren Schnittpunkt mit der x-Achse die Einsatzspannung darstellt [1]. Zwischen beiden Verfahren kann ein systematischer Versatz bestehen ($\leq 100\,\mathrm{mV}$). Relative V_th- Unterschiede an verschiedenen Transistortypen werden aber mit den unterschiedlichen Methoden weitestgehend gleich ermittelt ($\leq 10\,\mathrm{mV}$).

- Die äquivalente Oxiddicke EOT entspricht der Schichtdicke, die benötigt würde, um das Dielektrikum des Gateoxids bei gleicher Kapazität vollständig durch SiO_2 zu ersetzen. Sie bestimmt sich über die Formel $\mathrm{EOT} = (\varepsilon_{\mathrm{SiO_2}} \cdot d_\mathrm{ox})/\varepsilon_\mathrm{r}$, wobei d_ox die Dicke und ε_r die relative Permittivität des Gateoxids sind.

- Der Transferleitwert $g_\mathrm{m} = \partial I_\mathrm{ds}/\partial V_{\mathrm{gs}\,|V_\mathrm{ds}=\mathrm{const.}}$ entspricht der Steilheit der Transferkennlinie und gibt ein Maß für die Kontrolle über den Transistorkanal.

- Der Drainleitwert $g_\mathrm{d} = \partial I_\mathrm{ds}/\partial V_{\mathrm{ds}\,|V_\mathrm{gs}=\mathrm{const.}}$ beschreibt das Ausgangsverhalten des Transistors für ein gegebenes V_gs.

[1] Das beschriebene Verfahren gilt für den linearen Operationsbereich (kleine V_ds). Im Sättigungsbereich ergibt sich V_th aus dem Schnittpunkt der Tangente an $\sqrt{I_\mathrm{ds}(V_\mathrm{gs})}$ mit der x-Achse.

- $I_{\mathrm{ds,lin}}$ als Drainstrom I_{ds} im linearen Operationsbereich bei $V_{\mathrm{gs}} = V_{\mathrm{dd}}$ und $V_{\mathrm{ds}} = V_{\mathrm{d,lin}}$. Im linearen und parabolischen Arbeitsbereich lässt sich der Drainstrom ausdrücken als

$$I_{\mathrm{ds}} = \mu_{\mathrm{eff}} C'_{\mathrm{ox}} \frac{W}{L}[(V_{\mathrm{gs}} - V_{\mathrm{th}})V_{\mathrm{ds}} - \frac{1}{2}V_{\mathrm{ds}}^2] \tag{2.3}$$

$$= \beta[(V_{\mathrm{gs}} - V_{\mathrm{th}})V_{\mathrm{ds}} - \frac{1}{2}V_{\mathrm{ds}}^2]\,, \tag{2.4}$$

  wobei $\beta = \mu_{\mathrm{eff}} C'_{\mathrm{ox}} W/L$ als Verstärkungsfaktor des Transistors bezeichnet wird. In obigen Drainstromgleichungen wurde der Substratsteuereffekt vernachlässigt ($V_{\mathrm{bs}} = 0$). Für sehr kleine V_{ds} verliert der hintere Term ($-1/2 V_{\mathrm{ds}}^2$) an Bedeutung, sodass $I_{\mathrm{ds}}(V_{\mathrm{ds}})$ einen linearen Anstieg aufweist.

- $I_{\mathrm{ds,sat}}$ als Drainstrom im Sättigungsbereich. Die Parabelgleichung (2.4) erreicht bei $V_{\mathrm{ds}} = V_{\mathrm{gs}} - V_{\mathrm{th}}$ ihr Maximum. Der Sättigungsstrom beträgt dabei

$$I_{\mathrm{ds,sat}} = \beta(V_{\mathrm{gs}} - V_{\mathrm{th}})^2\,. \tag{2.5}$$

  In dieser Arbeit wird $I_{\mathrm{ds,sat}}$ bei $V_{\mathrm{gs}} = V_{\mathrm{dd}}$ und $V_{\mathrm{ds}} = V_{\mathrm{dd}}$ gemessen.

- $I_{\mathrm{ds,off}}$ als Drainreststrom des Transistors im Aus-Zustand, welcher im Unterschwellstrombereich der Kennlinie liegt. Die Messung erfolgt typischerweise bei $V_{\mathrm{gs}} = 0\,\mathrm{V}$ und mit $V_{\mathrm{ds}} = V_{\mathrm{dd}}$. Im Unterschwellstrombereich kann der Drainstrom näherungsweise ausgedrückt werden als [30]

$$I_{\mathrm{ds}} = \mu_{\mathrm{eff}} \frac{W}{L} \sqrt{\frac{\epsilon_{\mathrm{Si}} q N_{\mathrm{a}}}{4\Psi_{\mathrm{B}}}} \left(\frac{k_{\mathrm{B}}T}{q}\right)^2 \mathrm{e}^{q(V_{\mathrm{gs}} - V_{\mathrm{th}})/mk_{\mathrm{B}}T} \left(1 - \mathrm{e}^{-qV_{\mathrm{ds}}/k_{\mathrm{B}}T}\right) \tag{2.6}$$

  Zu diesem rein thermischen Reststrom kommen in der Praxis noch parasitäre Ströme wie Gate- und Anschlussleckströme an Source und Drain hinzu.

Für die in dieser Arbeit untersuchten nMOSFETs gilt $V_{\mathrm{ss}} = 0\,\mathrm{V}$, $V_{\mathrm{dd}} = 1\,\mathrm{V}$ und $V_{\mathrm{d,lin}} = 50\,\mathrm{mV}$; für die pFETs gelten die analogen negative Werte: $V_{\mathrm{ss}} = 0\,\mathrm{V}$, $V_{\mathrm{dd}} = -1\,\mathrm{V}$ und $V_{\mathrm{d,lin}} = -50\,\mathrm{mV}$. Dabei wird das Substrat stets auf das Sourcepotential gelegt ($V_{\mathrm{bs}} = 0\,\mathrm{V}$).

2.1.2 Technologische Hebel zur Kurzkanalkontrolle

Bei kleinen Kanallängen L treten zunehmend Abweichungen vom idealen Langkanal- Schaltverhalten auf, welche eine Beherrschung des Transistors erschweren und unter dem Begriff der Kurzkanaleffekte zusammengefasst werden. So können z.B. die Raumladungszonen, die sich natürlicherweise an den Übergängen zwischen den Source- bzw. Draingebieten und dem Substrat bilden, bereits zu signifikanten Verkürzungen der effektiven Kanallänge führen. Um dem entgegen zu wirken, kommen die bereits erwähnten Halo-Implantationen zum Einsatz, die die Raumladungszonen kanalseitig auf ein Minimum reduzieren und so einen abrupten Übergang zwischen beiden Gebieten erzeugen. Bei sehr kurzen Transistoren wird die Kanalladung stark durch das Drainfeld beeinflusst. Eine hohe Drainspannung führt dann zur Reduktion der Barriere zwischen Source und Drain (Drain Induced Barrier Lowering, DIBL). Durch diesen Effekt verringert sich zum einen die Einsatzspannung des Transistors, zum anderen verliert das Gate einen Teil der Kontrolle über die Kanalladung. Dem kann durch eine angepasste Halo- bzw. Kanalimplantation entgegen gewirkt werden. Dies führt in den Kurzkanaltransistoren aber zu einer höheren Kapazität der Raumladungszone C_{RLZ} und damit zu einer schlechteren Kanalkontrolle und höheren Unterschwellstrom-Kennzahl S. Bei weiterer Gatelängenskalierung muss daher, um die Kontrolle über die Kanalladung zu behalten, die Kapazität C_{ox} zwischen Gate und Substrat erhöht werden. Dies geschah in der Vergangenheit durch eine Reduktion der Dicke der SiO_2 Oxidschicht. Mit Unterschreiten der 3 nm-Grenze nahmen jedoch die Leckströme zwischen Gate und Substrat aufgrund von Tunnelprozessen massiv zu [32], sodass SiO_2 nach längerer Entwicklungsphase durch Hafnium-basierte Oxide (HfO_x) abgelöst wurde [33, 34]. Diese ermöglichen durch ihre hohe Permittivität („high-k“) die Verwendung höherer

physikalischer Schichtdicken bei gleichem EOT, sodass die Leckströme drastisch reduziert werden konnten. Die gleichzeitige Verwendung von metallischen Gateelektroden („Metal Gate“) verhindert Kapazitätsverluste durch Verarmungseffekte im hoch dotierten Poly-Si-Gate. Die aufgezählten Hebel, die zur technischen Realisierung heutiger Transistoren unabdingbar geworden sind, haben jedoch den Nachteil, die Beweglichkeit der Ladungsträger im Transistorkanal zu reduzieren. Einerseits erhöhen Halo-Implantationen die Substratdotierung und damit die Streuung der Ladungsträger, was die Beweglichkeit herabsetzt. Da die Halo-Implantation im Anschlussbereich des Kanals ihr Intensitätsmaximum hat, wirkt sich dieser Effekt umso stärker aus, je kürzer L ist. Die Verringerung der Schichtdicke des Gateoxids andererseits führt zu immer stärkeren vertikalen Feldern, wodurch der Transistorkanal näher an die Gateoxid- Grenzfläche gezogen wird. Typischerweise befindet sich unter dem HfO_x zwar eine dünne Lage aus SiO_2 zur Reduktion von Defekten an der Grenzfläche zum Inversionskanal im Si. Dennoch kommt es durch die Erhöhung der vertikalen Felder zur vermehrten Streuung an Oberflächendefekten sowie an entfernten Coulombstreuzentren im Hf-basierten Gatedielektrikum. Durch die daraus folgende Beweglichkeitsreduktion erhöhen sich Widerstand, Stromverbrauch und Schaltzeiten der Transistoren. Eine Möglichkeit, diesen Effekten entgegenzuwirken und die Beweglichkeit wieder anzuheben, bietet die Anwendung von Verspannungstechniken, die das Kristallgitter im Kanalbereich zug- (nFET) bzw. druckverspannen (pFET). Im Fall des pFETs eignet sich darüber hinaus die Verwendung von SiGe als Kanalmaterial mit intrinsisch höherer Löcherbeweglichkeit. Eine detailliertere Diskussion der Methoden zur Steigerung der Ladungsträgerbeweglichkeit findet sich in Abschn. 2.1.3 d).

2.1.3 Ladungsträgerbeweglichkeit

Unter dem Einfluss eines elektrischen Feldes E werden freie Ladungsträger in einem elektrisch leitenden Medium beschleunigt. Dieser Vorgang wird durch verschiedene Streuprozesse limitiert, wodurch die Ladungsträger nach einer mittleren Streuzeit τ abgebremst werden, um

dann erneut zu beschleunigen. Es stellt sich somit eine effektive Driftgeschwindigkeit

$$v_{\text{eff}} = \mu E \tag{2.7}$$

ein. Die Proportionalitätskonstante μ wird Ladungsträgerbeweglichkeit genannt. Sie steht mit der Zeit τ, die im Mittel zwischen zwei Streuprozessen vergeht und der effektiven Masse m^* sowie der Elementarladung q über

$$\mu = \frac{q\tau}{m^*} \,. \tag{2.8}$$

in Zusammenhang. Im Allgemeinen können Ladungsträger einer bestimmten Energie verschiedenen Streuprozessen unterliegen, wobei für einen speziellen Streuprozess die Streuzeit τ_i von der kinetischen Energie ϵ der Ladungsträger abhängig sein kann: $\tau_i = \tau_{i(\epsilon)}$. Dies ist zum Beispiel bei der Coulombstreuung an geladenen Störstellen der Fall. Überlagern sich verschiedene Streuprozesse, setzt sich die resultierende Streuzeit τ bei einer Energie ϵ anhand der Matthiessen'schen Regel

$$\frac{1}{\tau_{(\epsilon)}} = \sum_i \frac{1}{\tau_{i(\epsilon)}} \tag{2.9}$$

zusammen. Bei einem System, in dem nur ein Subband besetzt ist, kann die Matthiessen'sche Regel direkt auf die Beweglichkeit angewendet werden.

$$\frac{1}{\mu_{(\epsilon)}} = \sum_i \frac{1}{\mu_{i(\epsilon)}} \tag{2.10}$$

Streng genommen gilt obige Gleichung nur für Ladungsträger einer bestimmten Energie. In realen Systemen entspricht dies dem Grenzfall $k_{\text{B}}T \ll \epsilon_F - \epsilon_0$, für schwach veränderliche $\tau_i(\epsilon)$. Hierbei sind ϵ_F und ϵ_0 die Fermienergie und das energetische Bandminimum. Diese vereinfachende Betrachtung ist im Inversionskanal eines MOSFETs zumeist gültig. In einem späteren Abschnitt wird auf den allgemeineren Fall für eine nicht zu vernachlässigende Energieverteilung der zum Stromtransport beitragenden Ladungsträger näher eingegangen.

Im Inversionskanal des MOSFETs wird die Beweglichkeit durch zusätzliche Streumechanismen begrenzt, weshalb hier kleinere Werte als in moderat dotierten Silizium- Volumenkristallen erreicht werden. Sabnis und Clemens [35] fanden eine universale Beziehung zwischen der Beweglichkeit und dem effektiven vertikalen Feld $E_{\mathrm{eff,vert}}$ eines Transistors. $E_{\mathrm{eff,vert}}$ bezeichnet hier das vertikale Feld, dass im Bereich des MOSFET-Kanals herrscht und durch Anlegen einer Spannung zwischen Gate und Substrat hervorgerufen wird. Der Zusammenhang zwischen Feld und Beweglichkeit wurde von Takagi et al. eingehend charakterisiert [16]. Durch die Variation von $E_{\mathrm{eff,vert}}$ wird die Inversionsladungsträgerdichte im MOSFET-Kanal moduliert. So herrscht bei niedrigen vertikalen Feldern unterhalb der Einsatzspannung V_{th} eine geringe Ladungsträgerdichte im Kanal und die Coulomb-Streuung an Dotierstoffatomen und Gitterdefekten dominiert. Mit wachsendem Feld nimmt die Ladungsträgerkonzentration zu. Dies führt zu einer besseren Abschirmung von Streuzentren durch mobile Ladungsträger. Der Einfluss der Coulombstreuung nimmt ab und die Beweglichkeit zu, vgl. „Coulomb-Beitrag“ in Abb. 2.3. Bei sehr hohen Feldern in starker Inversion wird der Kanal mit steigendem Feld immer näher an die Grenzfläche zum Gateoxid gezogen, wodurch sich die Beweglichkeit aufgrund zunehmender Streubeiträge vermindert, siehe „Grenzflächen-Streuungsbeitrag“ in Abb. 2.3. Auch wenn nicht alle hierbei beteiligten Mechanismen vollständig verstanden sind, wird maßgeblich die Streuung durch Oberflächenrauigkeit und die indirekte Streuung durch Ladungen im Gateoxid („remote Coulomb scattering“) für die Reduktion der Beweglichkeit verantwortlich gemacht [36, 37]. Sämtliche Streubeiträge, die durch die Nähe des Elektronenkanals zur Gateoxid-Grenzfläche zustande kommen, werden in dieser Arbeit unter dem Begriff Grenzflächenstreuung („surface scattering“) zusammengefasst. Der Bereich zwischen den Regimen von Coulomb- und Grenzflächenstreuung wird von Phononenstreuung dominiert und ist daher stark temperaturabhängig. Der aus den verschiedenen Beiträgen resultierende Verlauf der Beweglichkeit kann nach (2.10) insgesamt beschrieben werden

als

$$\frac{1}{\mu} = \frac{1}{\mu_{\text{Coulomb}}} + \frac{1}{\mu_{\text{phonon}}} + \frac{1}{\mu_{\text{surface}}} \,, \tag{2.11}$$

wobei die Einzelbeiträge aus theoretischen Überlegungen berechnet werden können. Aus (2.11) wird ersichtlich, dass der Streumechanismus mit der kleinsten Beweglichkeit die Gesamtbeweglichkeit dominiert.

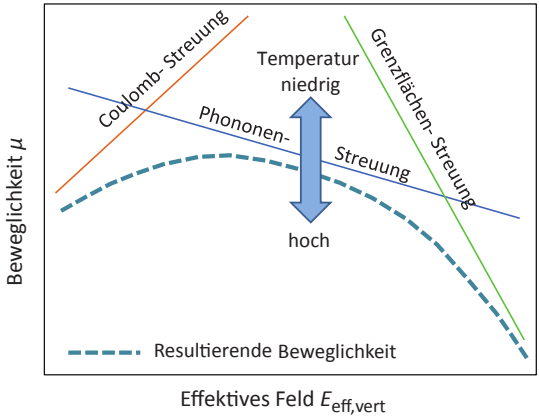


Abbildung 2.3: Kanalbeweglichkeit als Funktion des vertikalen elektrischen Feldes, nach [16]. Die Streumechanismen, welche die Beweglichkeit limitieren, folgen Exponentialgesetzen und können daher in der doppelt logarithmischen Darstellung als Geraden skizziert werden. Der Anteil an Phononenstreuung lässt sich über die Temperatur verändern. Bei Variation von $E_{\text{eff,vert}}$ wird gleichzeitig die Dichte an Inversionsladungsträgern im Kanal des MOSFETs moduliert.

a) Zusammenhang verschiedener Beweglichkeitsbegriffe

In der Literatur wird je nach verwendeter Messmethode zwischen verschiedenen Arten von Beweglichkeit unterschieden [38]:

- Die „mikroskopische Beweglichkeit“ $\mu_{(\epsilon)}$ ist eine theoretische Größe, die die Beweglichkeit einzelner Ladungsträger in ihrem jeweiligen Band mit ihrer individuellen Energie ϵ beschreibt. Sie geht nach Cohen et al. [39] in die mikroskopische Leitfähigkeit $\sigma_{(\epsilon)}$ ein, aus welcher durch Integration über alle Energien letztlich die makroskopischen, messbaren Größen Leitfähigkeit σ und Beweglichkeit μ berechnet werden können.

- Die „Driftbeweglichkeit“ μ_0 bezeichnet die makroskopische, messbare Beweglichkeit der Ladungsträger im Material. Sie ist definiert als die Proportionalitätskonstante zwischen einem angelegten elektrischen Feld E und der daraus resultierenden effektiven Ladungsträgergeschwindigkeit v gemäß $v = \mu E$. Die Leitfähigkeit einer Halbleiterprobe ist $\sigma = q(\mu_\mathrm{n} n + \mu_\mathrm{p} p)$, wobei n und p die Elektronen- und Löcherkonzentrationen und μ_n bzw. μ_p die zugehörigen Beweglichkeiten sind. Herrscht eine Ladungsträgersorte vor, kann die Driftbeweglichkeit der Majoritätsladungsträger über die Messung von Leitfähigkeit und Ladungsträgerkonzentration bestimmt werden. Im Fall eines p-dotierten Halbleiters mit $p \gg n$ ergibt sich so $\mu_\mathrm{p} = \sigma/(qp)$. Die Beweglichkeit der Minoritätsladungsträger im Volumenmaterial wird üblicherweise über Laufzeitmessungen („Time-of-flight, TOF“) zwischen zwei Kontakten bestimmt, wobei an einem Kontakt Minoritätsladungsträger durch einen Spannungspuls injiziert und an dem anderen über ein angelegtes elektrisches Feld wieder extrahiert werden. Aus der Kontaktdistanz x, der Laufzeit t und dem angelegten Feld E kann dann die Beweglichkeit mittels $\mu = v/E = x/(tE)$ bestimmt werden.

- Die „effektive Beweglichkeit“ μ_eff beschreibt die Driftbeweglichkeit der Minoritätsladungsträger im MOSFET-Kanal. Sie ist

vom theoretischen Aspekt ähnlich der Driftbeweglichkeit im Volumenmaterial, unterliegt aber bedingt durch den Probenaufbau zusätzlichen Streumechanismen wie z. B. der Oberflächen-Rauigkeitsstreuung bei hohen vertikalen Feldern. Ihre Bestimmung erfolgt über das sogenannte „Split-CV-Verfahren“ (Kap. 3.1.2), bei dem die Beweglichkeit über Strom-/Spannungs- und Kapazitäts-/Spannungsmessungen ermittelt wird. In Arbeiten, die sich ausschließlich mit Transistoren beschäftigen, werden die Begriffe „effektive Beweglichkeit“ und „Driftbeweglichkeit“ oft synonym verwendet, da eine Abgrenzung zur Beweglichkeit im Volumenmaterial meist nicht Gegenstand der Diskussion ist. Auch in der vorliegenden Arbeit wird zwischen beiden Begriffen nicht unterschieden. Sollte einmal die reine Driftbeweglichkeit im Volumenkristall gemeint sein, wird dies explizit erwähnt.

- Die „Feldeffektbeweglichkeit“ μ_{FE} (Kap. 3.1.1) stellt eine sehr grobe Abschätzung der Driftbeweglichkeit dar, welche unter vereinfachenden Annahmen mit einem schnellen, leicht durchzuführenden Messverfahren ermittelt werden kann. Sie wird aufgrund der einfachen Anwendbarkeit in der Halbleiterfertigung meist als Indikatorgröße für die Beweglichkeit verwendet.

- Die „Hallbeweglichkeit“ μ_{Hall} wird über eine Messung der Hallspannung bestimmt, welche sich bei einem stromdurchflossenen Leiter im Magnetfeld einstellt. Hierfür muss die Probengeometrie im Allgemeinen der Forderung $L \gg W$ genügen. Die Hallbeweglichkeit hängt mit der Driftbeweglichkeit über den Hall-Streufaktor r_{H} zusammen.

- Die „Magnetowiderstandsbeweglichkeit“ μ_{MR} wird über eine Messung des geometrischen Widerstandseffekts unter Einfluss eines Magnetfelds bestimmt. Im Gegensatz zur Hallmessung müssen die Proben im Allgemeinen dem hierfür notwendigen geometrischen Limit $L \ll W$ genügen. Die Magnetowiderstandsbeweglichkeit hängt, analog zur Hallbeweglichkeit, mit der Driftbeweglichkeit über den Magnetowiderstands- Streufaktor r_{MR} zusammen.

Im Folgenden wird näher auf den Zusammenhang zwischen Drift-, Hall- und Magnetowiderstandsbeweglichkeit eingegangen. Die grundlegenden Unterschiede lassen sich dabei ohne Beschränkung der Allgemeinheit bereits an einem System mit freien Ladungsträgern aus einem Band mit der gleichen effektiven Masse ableiten. Tragen mehrere Bänder zum Stromfluss bei, müssen die einzelnen Leitfähigkeiten unter Berücksichtigung der jeweiligen effektiven Massen über alle Bänder aufsummiert werden. Außerdem wurde bisher angenommen, dass nur Ladungsträger um $\epsilon \approx \epsilon_F$ zum Stromtransport beitragen. Die folgende Betrachtung berücksichtigt daher den allgemeineren Fall, dass die einzelnen Streuprozesse energieabhängig sein können.

b) Situation ohne Magnetfeld: die Driftbeweglichkeit

Die makroskopische Leitfähigkeit eines elektronischen Systems ergibt sich nach [39] im Fall eines zum Stromtransport beitragenden Subbands aus der mikroskopischen Leitfähigkeit zu

$$\sigma_0 = \int_0^{+\infty} \sigma_{(\epsilon)} \left(-\frac{\partial f_{(\epsilon)}}{\partial \epsilon} \right) \mathrm{d}\epsilon \,, \tag{2.12}$$

wobei $\sigma_{(\epsilon)}$ die mikroskopische Leitfähigkeit und ϵ die kinetische Energie bezogen auf das energetische Bandminimum ϵ_0 darstellen. Letzteres entspricht für Elektronenleitung dem Minimum des Leitungsbands und für Löcherleitung dem Maximum des Valenzbands,

$$f_{(\epsilon)} = \frac{1}{1 + e^{(\epsilon - \epsilon_\mathrm{F})/k_\mathrm{B}T}} \tag{2.13}$$

ist die Fermi-Dirac-Funktion. Um die makroskopische Leitfähigkeit zu erhalten, werden also die Leitfähigkeiten $\sigma_{(\epsilon)}$ aller Ladungsträger einer bestimmten Energie durch das Integral über alle Energien aufsummiert, wobei die partielle Ableitung $\partial f_{(\epsilon)}/\partial \epsilon$ der Fermiverteilung die Ladungsträger herausfiltert, die sich energetisch um den Bereich der Fermienergie ϵ_F befinden und tatsächlich zur Leitung beitragen können [40]. Nur in diesem Energiebereich gibt es sowohl eine ausreichende

Anzahl an Ladungsträgern, als auch eine ausreichende Anzahl an unbesetzten Zuständen, welche als Endzustände der elastischen Streuprozesse notwendig sind. Würden mehrere Bänder zum Stromfluss beitragen, wäre diese Integration für jedes Band durchzuführen und σ_0 ergäbe sich durch Summation über alle Bänder. Die mikroskopische Leitfähigkeit bei einer bestimmten Energie ϵ ist [41]

$$\sigma_{(\epsilon)} = q\epsilon\mu_{(\epsilon)}N_{(\epsilon)}\,, \tag{2.14}$$

wobei $\mu_{(\epsilon)}$ die mikroskopische Beweglichkeitsfunktion und $N_{(\epsilon)}$ die Zustandsdichte sind. $N_{(\epsilon)}$ gibt die Anzahl der Zustände pro Energie- und Volumeneinheit (3D-Fall) bzw. Flächeneinheit (2D-Fall) an. Es ergibt sich somit

$$\sigma_0 = \int_0^{+\infty} q\epsilon\mu_{(\epsilon)}N_{(\epsilon)}\left(-\frac{\partial f_{(\epsilon)}}{\partial\epsilon}\right)\mathrm{d}\epsilon\,. \tag{2.15}$$

Die makroskopische Leitfähigkeit σ_0 entspricht dem Inversen des spezifischen Widerstands $\rho_0 = 1/\sigma_0$ und hängt über

$$\sigma_0 = qN_\mathrm{s}\mu_0 \tag{2.16}$$

mit der makroskopischen Driftbeweglichkeit μ_0 zusammen. N_s ist die Dichte der gesamten im Band befindlichen Ladungsträger $N_\mathrm{s} = \int_0^{+\infty} N_{(\epsilon)}f_{(\epsilon)}\,\mathrm{d}\epsilon$. Im MOSFET-Kanal entspricht N_s der Inversionsladungsträgerdichte. Für die makroskopische Driftbeweglichkeit μ_0 ergibt sich aus (2.15) und (2.16)

$$\mu_0 = \frac{\sigma_0}{qN_\mathrm{s}} = \frac{\int_0^{+\infty}\epsilon\mu_{(\epsilon)}N_{(\epsilon)}\left(-\frac{\partial f_{(\epsilon)}}{\partial\epsilon}\right)\mathrm{d}\epsilon}{\int_0^{+\infty}N_{(\epsilon)}f_{(\epsilon)}\,\mathrm{d}\epsilon}\,. \tag{2.17}$$

Die Stromdichte kann nun über $j = \sigma_0 E = qN_\mathrm{s}\mu_0 E$ berechnet werden. Drückt man die Beweglichkeit über die mittlere Streuzeit τ und die effektive Masse m^* aus, so lautet die verallgemeinerte, energieabhängige Form von (2.8)

$$\mu_0 = \frac{\langle\mu_{(\epsilon)}\rangle}{N_\mathrm{s}} = \frac{q}{N_\mathrm{s}m^*}\langle\tau_{(\epsilon)}\rangle\,. \tag{2.18}$$

Die Klammer $\langle ... \rangle$ repräsentiert hierbei die Mittelung der durchschnittlichen Streuzeiten über die Energien der zum Stromtransport beitragenden Ladungsträger nach der Vorschrift

$$\langle x_{(\epsilon)} \rangle = \int_0^\infty x_{(\epsilon)} \epsilon N_{(\epsilon)} \frac{\partial f}{\partial \epsilon} (\epsilon - \epsilon_{\mathrm{F}}) \mathrm{d}\epsilon \,. \tag{2.19}$$

c) Situation mit Magnetfeld: Hall- und Magnetowiderstandsbeweglichkeit

Ein Ladungsträger mit einer Geschwindigkeit $\vec{v}$ in der x/y-Ebene wird im Vakuum unter einem Magnetfeld in z-Richtung $\vec{B} = \vec{B}_{\mathrm{z}}$ durch die Lorentzkraft

$$\vec{F} = q(\vec{v} \times \vec{B}) \,. \tag{2.20}$$

auf eine Kreisbahn gezwungen. Legt man nun in x-Richtung zusätzlich ein elektrisches Feld $\vec{E} = \vec{E}_{\mathrm{x}}$ an, kommt zur sogenannten „$\vec{E} \times \vec{B}$-Drift“

$$\vec{v}_{\mathrm{drift}} = \frac{\vec{E} \times \vec{B}}{B^2} \,, \tag{2.21}$$

durch die sich das Führungszentrum der Gyrationsbewegung mit $\vec{v}_{\mathrm{drift}}$ senkrecht zu $\vec{E}$ und $\vec{B}$ bewegt. Das Teilchen weicht somit auf einer spiralförmigen Trajektorie in y-Richtung aus, während im Mittel keine Strecke in x-Richtung zurückgelegt wird [42]. Im Festkörper dagegen wird diese Spiralbewegung aufgrund von Streuungseffekten immer wieder unterbrochen, wobei das Teilchen seine kinetische Energie an das Gitter abgibt. Danach wird es erneut durch das E-Feld in x-Richtung beschleunigt, während es gleichzeitig durch die Lorentzkraft bis zum nächsten Streuereignis eine immer engere Kreisbahn beschreibt. Da diese Kreisbewegung jedoch nie vollendet wird, bevor ein erneutes Streuereignis auftritt, ergibt sich im Mittel eine Bewegungskomponente in x-Richtung. Der Winkel zwischen angelegtem E-Feld und gemittelter Bewegungsrichtung heißt Hall-Winkel θ [43].

Da das Verhalten des Stromflusses unter Einwirkung eines vertikalen Magnetfelds nicht mehr als eindimensionales Problem betrachtet werden kann, wird für die Stromdichte $\vec{j} = \boldsymbol{\sigma}\vec{E}$ der Leitfähigkeitstensor $\boldsymbol{\sigma}$ relevant, der sich nach der Drude-Boltzmann-Theorie zu [44]

$$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_0 & \sigma_0\mu B \\ -\sigma_0\mu B & \sigma_0 \end{pmatrix} \frac{1}{1+\mu^2 B^2} \tag{2.22}$$

ergibt, wobei die hierin enthaltene Beweglichkeit μ im Allgemeinen ungleich der Driftbeweglichkeit ist. Die Driftbeweglichkeit geht allerdings über $\sigma_0 = N_\mathrm{s} q \mu_0$ in den Tensor ein. Die Komponenten der Stromdichte $\vec{j}$ sind

$$\begin{aligned} j_x &= \sigma_{xx(B)} E_x + \sigma_{xy(B)} E_y \\ j_y &= \sigma_{yx(B)} E_x + \sigma_{yy(B)} E_y \,, \end{aligned} \tag{2.23}$$

wobei E_x und E_y die Komponenten des elektrischen Feldes in x- bzw. y-Richtung sind. Hier gibt es zwei wichtige geometrische Grenzfälle, welche sich aus der endlichen Probengeometrie für reale Messungen ergeben:

Im Falle einer langen, aber schmalen Probe ($L \gg W$, Abb. 2.4a)) führt die Ladungsverschiebung zum Aufbau eines elektrischen Feldes E_y und damit der Hallspannung V_H, bis das Hallfeld die Lorentzkraft kompensiert und der Stromfluss in y-Richtung schließlich zum Erliegen kommt. Für den Hallwinkel gilt [45]

$$\tan\theta = \frac{E_\mathrm{y}}{E_\mathrm{x}} = \mu B \,. \tag{2.24}$$

Die so bestimmte Beweglichkeit wird Hallbeweglichkeit genannt, $\mu = \mu_\mathrm{Hall}$. Da $j_y = 0$ gilt $E_y = \mu B E_x$. Daraus folgt, dass $E_x = j_x/\sigma_0$ ist. Die Stromdichte j_x verliert somit ihre Magnetfeldabhängigkeit. An einer für den Halleffekt idealen Probe kann daher kein Magnetowiderstandseffekt beobachtet werden.

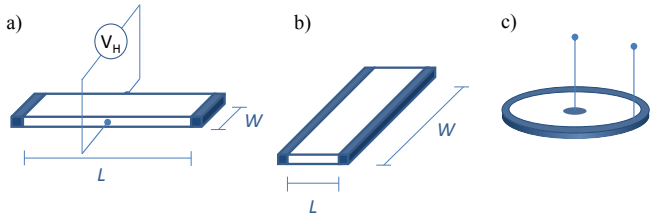


Abbildung 2.4: Probengeometrien: a) Hallprobe: $L \gg W$, b) Magnetowiderstandsprobe: $L \ll W$, c) Corbino-Scheibe: $L/W = 0$

Im entgegengesetzten Grenzfall einer kurzen, aber sehr breiten Probe ($L \ll W$, Abb. 2.4b)) kann sich hingegen das Hallfeld nicht aufbauen, da die Hallspannung über die Anschlusskontakte kurzgeschlossen wird. Das Extrem dieser geometrischen Konfiguration ist die sogenannte Corbino-Scheibe (Abb. 2.4c), bei der der Strom von einem punktförmigen Mittelkontakt zu einem ringförmig umgebenden Außenkontakt fließt. Unter solchen Umständen kann die Lorentzkraft die Ladungsträger ungehindert ablenken und die Ladungsträger legen einen größeren Weg zurück, als dies ohne Einwirkung eines magnetischen Feldes der Fall wäre. Die dadurch eintretende Erhöhung des Zweipunkt- Probenwiderstands wird geometrischer Magnetowiderstandseffekt genannt. Aus $E_y = 0$ ergibt sich $j_x = \sigma_{xx} E_x$ und damit

$$j_x = \sigma_0 E_x / (1 + \mu^2 B^2)\,, \tag{2.25}$$

wobei die so erhaltene Beweglichkeit die Magnetowiderstandsbeweglichkeit $\mu = \mu_{\mathrm{MR}}$ ist.

Durch unterschiedliche Messverfahren am gleichen Materialsystem erhält man also grundsätzlich verschiedene Beweglichkeitswerte. Wenn eine energetische Verteilung der zum Stromtransport beitragenden Ladungsträger berücksichtigt werden muss, ergeben sich für die Beweglichkeiten unterschiedliche Mittelungsvorschriften. Für die Hall- und

Magnetowiderstandsbeweglichkeit gilt [44]

$$\mu_{\mathrm{Hall}} = \frac{\int_0^\infty \epsilon \mu_{(\epsilon)}^2 N_{(\epsilon)} \left(-\frac{\partial f_{(\epsilon)}}{\partial \epsilon}\right) \mathrm{d}\epsilon}{\int_0^\infty \epsilon \mu_{(\epsilon)} N_{(\epsilon)} \left(-\frac{\partial f_{(\epsilon)}}{\partial \epsilon}\right) \mathrm{d}\epsilon} = \frac{\langle \mu_{(\epsilon)}^2 \rangle}{\langle \mu_{(\epsilon)} \rangle} = \frac{q}{m^*} \frac{\langle \tau_{(\epsilon)}^2 \rangle}{\langle \tau_{(\epsilon)} \rangle} \quad (2.26)$$

und

$$\mu_{\mathrm{MR}} = \sqrt{\frac{\int_0^\infty \epsilon \mu_{(\epsilon)}^3 N_{(\epsilon)} \left(-\frac{\partial f_{(\epsilon)}}{\partial \epsilon}\right) \mathrm{d}\epsilon}{\int_0^\infty \epsilon \mu_{(\epsilon)} N_{(\epsilon)} \left(-\frac{\partial f_{(\epsilon)}}{\partial \epsilon}\right) \mathrm{d}\epsilon}} = \sqrt{\frac{\langle \mu_{(\epsilon)}^3 \rangle}{\langle \mu_{(\epsilon)} \rangle}} = \frac{q}{m^*} \sqrt{\frac{\langle \tau_{(\epsilon)}^3 \rangle}{\langle \tau_{(\epsilon)} \rangle}} \,. \quad (2.27)$$

Die verschiedenen, methodisch unterschiedlichen Beweglichkeiten lassen sich zueinander in Beziehung setzen. In der Praxis geschieht dies durch Verhältnisbildung von μ_{Hall} bzw. μ_{MR} zu μ_{Drift}. Die Proportionalitätskonstante zwischen Hall- und Driftbeweglichkeit wird Hall-Streufaktor r_{H} genannt.

$$r_{\mathrm{H}} = \frac{\mu_{\mathrm{Hall}}}{\mu_{\mathrm{Drift}}} = \frac{\langle \tau_{(\epsilon)}^2 \rangle}{\langle \tau_{(\epsilon)} \rangle^2} \,. \quad (2.28)$$

Analog gilt für den Bezug der geometrischen Magnetowiderstandsbeweglichkeit zur Driftbeweglichkeit

$$r_{\mathrm{MR}} = \frac{\mu_{\mathrm{MR}}}{\mu_{\mathrm{Drift}}} = \frac{\langle \tau_{(\epsilon)}^3 \rangle^{1/2}}{\langle \tau_{(\epsilon)} \rangle^{3/2}} \,, \quad (2.29)$$

wobei r_{MR} den Magnetowiderstands-Streufaktor bezeichnet. Falls τ nicht von der Energie abhängt, gilt $\langle \tau \rangle = \tau$, $\langle \tau^2 \rangle = \tau^2$ und $\langle \tau^3 \rangle = \tau^3$, sodass $\mu_{\mathrm{Drift}} = \mu_{\mathrm{Hall}} = \mu_{\mathrm{MR}}$. Dies ist zum Beispiel der Fall, wenn die beteiligten Streuprozesse nicht energieabhängig sind. Für energieabhängige Streuprozesse wie z. B. der Coulombstreuung gleichen sich die unterschiedlichen Beweglichkeiten bei tiefen Temperaturen einander an, da hier praktisch nur eine einzige Energie – die Fermienergie ϵ_{F} – relevant ist. Wenn die Energie aller an der elektrischen Leitung beteiligten Ladungsträger nahe ϵ_{F} liegt, ist $\tau_{(\epsilon)} \approx \tau_{(\epsilon_{\mathrm{F}})} =$ konst. Während

die genauen Formeln zur Berechnung der einzelnen Streumechanismen sehr kompliziert sind, lassen sich die Abhängigkeiten in erster Näherung mit einem exponentiellen Ansatz der Form $\mu_{(\epsilon)} \sim \tau_{(\epsilon)} \sim \epsilon^{\eta}$ gut beschreiben. Der Exponent η ist dabei charakteristisch für die jeweilige Art der Streuung. Abb. 2.5 zeigt die Beziehung, die sich zwischen dem Energieexponent η und den Streufaktoren r_{H} und r_{MR} ergibt. Im Fall vollständiger Energieunabhängigkeit der Streuprozesse gilt $\eta = 0$ und somit $\mu_{\mathrm{MR}} = \mu_{\mathrm{Hall}} = \mu_{\mathrm{eff}}$. Dies führt in Abb. 2.5 zum Schnittpunkt der dargestellten Kurven bei (0/0).

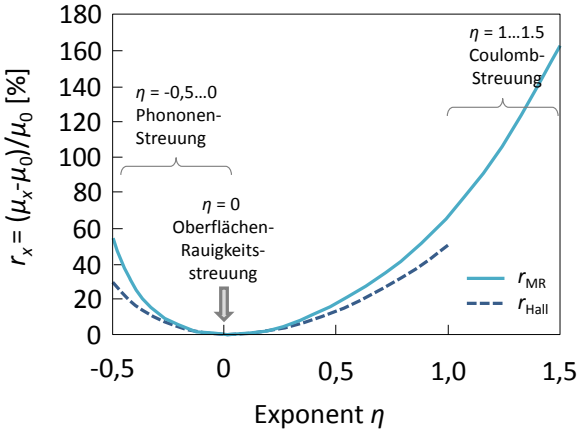


Abbildung 2.5: Zusammenhang der Streufaktoren r_{H} und r_{MR} mit dem Energieexponent η im Ansatz $\mu_{(\epsilon)} \sim \epsilon^{\eta}$ für den Fall eines einzelnen beitragenden Subbands, $N_{\mathrm{inv}} = 10^{12}\,\mathrm{cm}^{-2}$, $T = 300\,\mathrm{K}$. Nach [44], [46] und [47].

d) Ladungsträgerbeweglichkeit und Kanalverspannung

Die fortschreitende Transistorskalierung führt zur Verwendung immer höherer vertikaler Felder, um die immer stärker werdenden Kurzkanaleffekte zu begrenzen und so die Kontrolle über den Kanal zu erhalten. Aus dem universalen Zusammenhang (Abb. 2.3) wird deutlich,

dass mit einer Erhöhung der vertikalen Feldstärke eine Reduktion der Beweglichkeit verbunden ist. Hinzu kommt eine Steigerung der Substratdotierung, vor allem aufgrund der sich anteilig auf den gesamten Kanal immer stärker auswirkenden Halo-Implantation, wodurch ebenfalls die Streuung erhöht und die Beweglichkeit reduziert wird. Eine Übersicht über die Abnahme der Beweglichkeit von planaren Standardtransistoren über mehrere Generationen bis zu Gatelängen von $L = 45\,\mathrm{nm}$ gibt [17]. Aufgrund dieser Entwicklung ist die Ladungsträgerbeweglichkeit eine zunehmend kritische Größe geworden, die sich limitierend auf Leistungsparameter wie Stromverbrauch und Schaltzeiten auswirkt. Einen – zumindest teilweisen – Ausweg aus dieser Situation bietet die Möglichkeit, die Ladungsträgerbeweglichkeit durch integrative Maßnahmen zu erhöhen und so den universalen Verlauf „zu verlassen“. Zum einen kann dies durch die Verwendung von Materialien mit höherer intrinsischer Beweglichkeit geschehen. So wird z. B. durch die Verwendung von Germaniumanteilen im pFET-Kanal (channel Silicon-Germanium, cSiGe) die Löcherbeweglichkeit gesteigert [48]. Zum anderen lässt sich die Beweglichkeit durch eine mechanische Verspannung des Transistorkanals erhöhen. Eine tensile (Zug-) Verspannung erhöht die Elektronenbeweglichkeit, während kompressive (Druck-) Verspannung die Löcherbeweglichkeit steigert. Im Material bewirkt die Verspannung eine Deformation des Kristallgitters. Dadurch ändern die Bandverläufe $\epsilon(\kappa)$ im reziproken Raum [49] ihre Krümmung, zu welcher die effektiven Massen über

$$m^* = \hbar^2 \left[\frac{\partial \epsilon^2_{(\kappa)}}{\partial \kappa^2} \right]^{-1} \tag{2.30}$$

invers proportional in Beziehung stehen [50]. Diese hängen wiederum über $\mu = q\tau/m^*$ (2.8) mit der Beweglichkeit zusammen, sodass eine Reduktion der effektiven Masse zur Beweglichkeitssteigerung führt. Darüber hinaus führt eine asymmetrische, also uni- oder biaxiale Verspannung zur Aufhebung der Kristallsymmetrie im k- Raum und damit zur Aufhebung der Entartung. Die zur Leitung beitragenden Energieniveaus werden energetisch aufgespalten (Abb. 2.6), wodurch Streuverluste im Mittel reduziert werden. Die daraus folgende Zunahme der

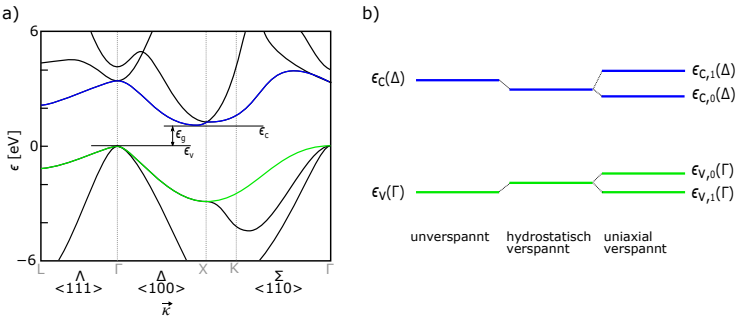


Abbildung 2.6: Bandstruktur von Silizium: a) für drei ausgewählte Richtungen im reziproken Raum im unverspannten Fall, b) Schematische Entwicklung der Energieniveaus am Valenzbandmaximum $\epsilon_{\mathrm{v}}(\Gamma)$ am Γ- Punkt und des Leitungsbandminimums $\epsilon_{\mathrm{c}}(\Delta)$ in Δ- Richtung unter Verspannung. Für Details zur Indizierung von Kristallrichtungen und der Bändertheorie siehe z. B. [49].

mittleren Streuzeiten τ trägt ebenfalls zur Steigerung der Beweglichkeit bei. Allem voran sei hier die Phononenstreuung erwähnt, welche neben der Reduktion der effektiven Masse einen Hauptteil an der Beweglichkeitssteigerung trägt [51]. Aber auch andere Streumechanismen, wie die Coulomb-, oder Grenzflächenstreuung reagieren auf Gitterverspannung [52,53]. Die detaillierte Theorie zum Einfluss verschieden gearteter Gitterverspannungen auf die Bänder im Silizium, verschiedene an der Leitung beteiligter effektiver Massen und unterschiedliche Streumechanismen ist sehr umfangreich und kann im Rahmen dieser Arbeit nicht behandelt werden. Eine sehr gute Einführung zum Thema Gitterverspannung und Beweglichkeit gibt z. B. [54].

In der hier untersuchten Transistor-Technologie wird die Gitterverspannung in den jeweiligen Gebieten durch die Verwendung von Deckschichten aus Siliziumnitrid über den Transistorgates erzeugt. Je nach Wahl der Prozessbedingungen entstehen $\mathrm{Si}_x\mathrm{N}_y$- Schichten die sich ausdehnen oder zusammenziehen, wodurch im Kanal eine kompressive

bzw. tensile Verspannung hervorgerufen wird. Die Deckschichten werden in dieser Arbeit gemäß der durch sie hervorgerufenen Kanalverspannung mit TPEN („Tensile Plasma Enhanced Nitride“, Deckschicht mit tensiler Kanalverspannung) bzw. CPEN („Compressive Plasma Enhanced Nitride“, Deckschicht mit kompressiver Kanalverspannung) gekennzeichnet. Mit diesen Schichten können sehr hohe intrinsische Verspannungswerte im Bereich von $+1,5\,\mathrm{GPa}$ (tensil) und $-3,5\,\mathrm{GPa}$ (kompressiv) erreicht werden [55]. Diese Werte gelten für planare Schichten. Wie stark der tatsächliche Übertrag auf den Transistorkanal ausfällt, ist stark geometrieabhängig und experimentell nur schwer zugänglich. Daher werden Abschätzungen über den Verspannungstransfer üblicherweise mit Hilfe von Prozesssimulationen durchgeführt. Deckschichten aus Siliziumnitrid können grundsätzlich auch ohne Verspannung abgeschieden werden („Neutral Plasma Enhanced Nitride, NPEN“). Mit fortschreitender Skalierung wird die Verspannungserzeugung mittels Deckschichten zunehmend schwieriger. Dies liegt daran, dass die immer kleiner werdenden Abstände zwischen den Transistoren in integrierten Schaltungen („Pitch“) aufgrund der Kantenbedeckung Grenzen für die maximalen Schichtdicken setzen. Erschwerend kommt hinzu, dass die kompressiven Schichten wieder vom nFET, die tensilen wieder vom pFET entfernt werden müssen, ohne die Transistoren dabei zu schädigen. Die Kantenübergänge zwischen den nFET- und pFET-Bereichen stellen integrativ eine zusätzliche Herausforderung dar.
Über die Verspannung durch kompressive Deckschichten hinaus kann die Löcherbeweglichkeit am pFET durch die Verwendung von Germanium zusätzlich weiter gesteigert werden. Dies ist insbesondere daher von Vorteil, als dass die Beweglichkeit der Löcher im Vergleich zu der der Elektronen deutlich geringer ist. Hierzu werden an den pFET Source- und Draingebieten Gräben geätzt, in denen man epitaktisch Silizium mit Germaniumanteilen aufwachsen lässt („embedded SiGe“, eSiGe). Germanium besitzt die gleiche kristalline Gitterstruktur wie Silizium. Aufgrund des größeren Atomdurchmessers von Ge ergibt sich in den eSiGe Source- und Draingebieten jedoch eine größere Gitterkonstante. Da sich der epitaktische Bereich in den umgebenden Si-Kristall einpasst, wird der Kanalbereich zwischen Source und Drain gequetscht, also kom-

pressiv verspannt. Des Weiteren ist es möglich, im Kanalbereich eine dünne Schicht SiGe („c-SiGe“, channel-SiGe) aufzuwachsen. Obwohl dies der kompressiven Verspannung wieder etwas entgegenwirkt, kann so für den Kanalbereich von der intrinsisch höheren Löcherbeweglichkeit in Germanium profitiert werden.

2.2 Silizium-Nanodraht-Transistoren

2.2.1 Aufbau

Innerhalb der NaMLab gGmbH beschäftigt sich eine Kompetenzgruppe unter der Leitung von Walter Weber mit der Untersuchung von Si-Nanodraht („Silicon Nanowire“, SiNW) Transistoren. Die im Rahmen dieser Arbeit untersuchten Einzeltransistoren basieren auf einem „bottom-up“ Fertigungsansatz. Details zu den hierfür entwickelten und verwendeten Prozessen finden sich in den Referenzen [56] (Wachstum) und [57] (Prozessierung). Im Folgenden wird eine kurze Beschreibung der wichtigsten Herstellungsschritte der hier untersuchten Si-Nanodraht-Transistoren gegeben. Bei der Probenherstellung wurden zunächst nominell undotierte SiNW über einen durch Goldpartikel katalysierten „Vapor-Liquid-Solid-“ (VLS) Prozess [58] hergestellt. Die Nanodrähte wachsen dabei epitaktisch auf Silizium in einer <110> Achsenorientierung auf, vgl. Abb. 2.7 oben. Im Anschluss erfolgte die Entfernung der Goldpartikel durch selektive Ätzung mit Flusssäure und Königswasser. Vor der weiteren Prozessierung wurden die Nanodrähte in einem Zwischenschritt in eine Suspension gebracht und auf das Zielsubstrat transferiert. Dieses weist unter einer 115 nm dicken Schicht aus SiO_2 eine sehr hohe, in einer dünnen Oberflächenschicht entartete, p-Dotierung im Silizium auf. Typischerweise enthalten die so aufgebrachten Nanodrähte eine native Oxidhülle minderer Qualität. Um eine geringere Defektdichte zu erreichen, wird diese über eine Flusssäure-Ätzung entfernt und direkt im Anschluss eine Trockenoxidation bei 875 °C für 10 min unter Sauerstoffatmosphäre durchgeführt. Bei diesem Prozess bildet sich eine koaxiale Siliziumoxidhülle, wobei intrinsisches Silizium des Nanodrahts zugunsten der wachsenden Oxidhülle

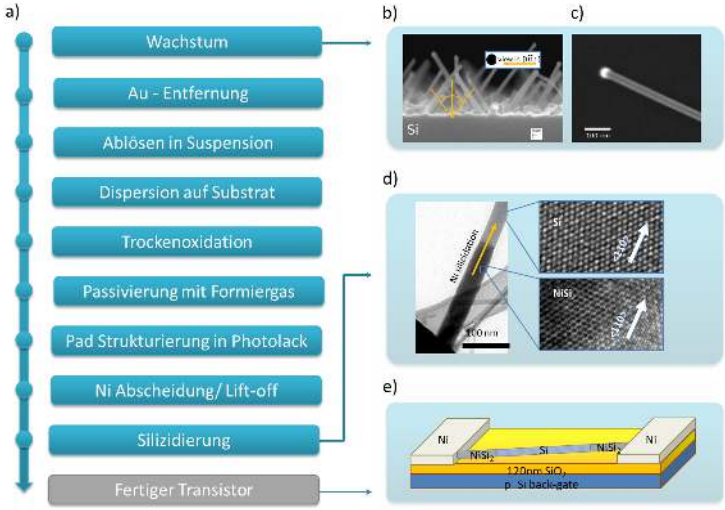


Abbildung 2.7: Wesentliche Eckpunkte der SiNW Prozessierung: a) Prozessablauf, b) REM- Aufnahme von Nanodrähten auf ihrem Wachstumssubstrat, c) einzelner Draht vor Entfernung des Goldpartikels, d) TEM Aufnahmen des Schottkykontaktbereichs, e) Skizze zum Probenaufbau; Details zum Wachstum finden sich in [56], zur Silizidierung in [59]. Die Steuerung des Transistors erfolgt über die Nutzung des Substrats als Rückseitengate.

konsumiert wird. Diese passiviert die Drahtoberfläche stark gegenüber äußeren Umwelteinflüssen. Bei Wachstum der Oxidhülle bildet sich eine radiale, kompressive Druckverspannung auf den Kern aus. Raman-Spektroskopie [60], [61] an Referenzproben mit Bündeln oxidierter Nanodrähte ergab einen mittleren kompressiven Verspannungswert von $1,3\,\mathrm{GPa}$ [57, 62]. Die Nanodrähte auf der Probe haben typischerweise $14\,\mathrm{nm}$ Si-Kerndurchmesser und eine Oxidhülle mit $8\,\mathrm{nm}$ Dicke. Der Oxidation folgte ein $10\,\mathrm{min}$ Temperschritt unter Formiergasatmosphäre

im Volumenverhältnis von 10 % H_2 zu 90 % N_2 bei 450 °C. Dies dient dazu, an der Oberfläche gegebenenfalls noch vorhandene, ungesättigte Si-Bindungen abzusättigen und die Anzahl potentieller Ladungsfallen zu reduzieren. Über einen Elektronenstrahllithografie-/Lift-off-Prozess wurden auf der Probe Felder mit Ni-Elektroden strukturiert. Diese bilden die Kontaktanschlüsse für Source und Drain und stellen darüber hinaus ein Ni-Reservoir für den nachfolgenden Silizidierungsschritt bei 450 °C zur Verfügung. Dabei diffundiert Ni aus den Kontaktgebieten lateral in die Nanodrähte ein und formt eine einkristalline, kubische $NiSi_2$ Phase [59], vgl. Abb. 2.7 Mitte. Die entstehenden Metall-Halbleiter-Schottkyübergänge sind sehr abrupt und finden teilweise auf atomare Ebene statt [63], [64]. Die Länge des Kanals, welcher sich zwischen den beiden Schottkyübergängen befindet, ist dadurch scharf definiert. Sie betrug beim hier untersuchen Transistor 500 nm, bei einem Durchmesser des Si-Kerns von etwa 14 nm. Das dotierte Siliziumsubstrat wird als Rückseitengate verwendet. Zur Verbesserung der Schalteigenschaften könnten nun noch „Top-Gates“ strukturiert werden, welche eine erhöhte elektrostatische Kopplung an den Nanodraht, sowie die individuelle Einstellung der Barrieren an Source und Drain ermöglichten [65, 66]. Dieser Schritt war für die geplanten Untersuchungen jedoch nicht erforderlich und wurde daher weggelassen. Nach Oxidation der Nanodrähte ergab sich zusammen mit dem Substratoxid eine effektive Gateoxiddicke von insgesamt 123 nm.

2.2.2 Schaltverhalten

a) Grundzüge des Schottky-Feldeffekt-Transistors

Wird ein Metall mit einem Halbleiter in Kontakt gebracht, entsteht ein sogenannter Schottkyübergang. Dieser weist grundsätzlich eine Strom-/Spannungscharakteristik ähnlich einem p-n-Übergang auf. Abb. 2.8 illustriert die Entstehung eines solchen Kontakts, wobei hierbei der Einfluss von Grenzflächenzuständen vernachlässigt wird. Sind Metall und Halbleiter noch weit auseinander, sodass keine elektrische Kopplung zwischen beiden besteht, ist die Lage der Ferminiveaus durch die Austrittsarbeit des Metalls ϕ_M und die Elektronenaffinität des Halbleiters

χ gegeben (Abb. 2.8a). Die beiden Vakuumniveaus liegen energetisch auf gleicher Höhe und dienen als Referenz. Werden die Materialien einander näher gebracht, so gleichen sich die Ferminiveaus an, wobei sich die Vakuumniveaus gegeneinander verschieben. Durch diesen Prozess kommt es im Allgemeinen zu einer Bandverbiegung im Halbleiter. Für die Elektronen im Metall ergibt sich nach Kontakt durch den Abstand des Ferminiveaus zur Leitungsbandkante im Halbleiter am Übergang eine Energiebarriere ϕ_B der Höhe

$$\phi_\mathrm{B} = \phi_\mathrm{M} - \chi\,, \tag{2.31}$$

die sogenannte Schottkybarriere (Abb. 2.8b). Im Si-Nanodraht-Transistor existieren zwei solcher Übergänge, einer auf der Sourceseite und einen an Drain. In den hier untersuchten Transistoren sind das die Übergänge zwischen dem metallischen, von den Drahtenden zur Mitte hin laufenden $NiSi_2$ und dem intrinsischen Silizium im mittleren Bereich des Drahts.

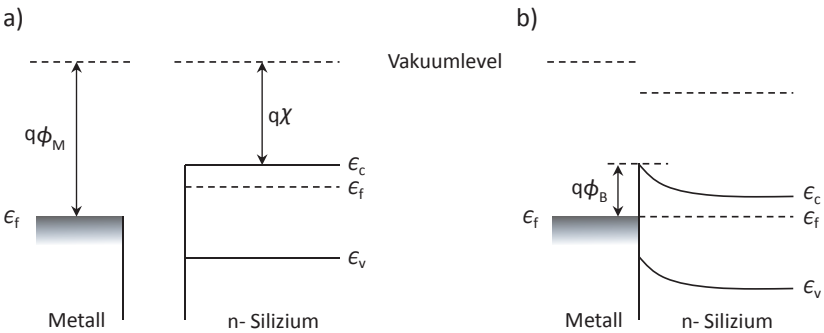


Abbildung 2.8: Formation eines Schottkykontakts ohne externe Spannungen und unter Vernachlässigung von Oberflächenzuständen: a) Metall und Halbleiter auf großer Distanz, b) Metall und Halbleiter in Kontakt; nach [67].

Der Schaltvorgang in konventionellen MOSFETs basiert auf der Erzeugung einer Inversionsladung im Kanal, sodass im AN-Zustand keine Energiebarriere zwischen Source und Drain existiert. Im Gegensatz

dazu müssen die Ladungsträger in einem SiNW-Schottkytransistor immer eine Barriere überwinden, vgl. Abb. 2.1b mit Abb. 6.10b. Dabei basiert der Schaltvorgang auf Modulation der Injektionsbarriere durch Anlegen einer Gatespannung. Hierdurch ändert sich die Fähigkeit, Ladungsträger durch die Barriere an der Sourceseite vom Metallkontakt in den Nanodraht zu injizieren [68], welche dann im weiteren Verlauf nach Drain hin abfließen können. Im Kanal selbst befindet sich die Region höchster Ladungsträgerdichte – geometrisch bedingt durch das Rückseitengate – am unteren Rand des Siliziumkerns, siehe [69] für den Fall der hier untersuchten Transistoren. Bei der Injektion in den Kanal spielen sowohl die thermionische Emission über, als auch das Tunneln durch die Schottkybarriere eine wichtige Rolle, siehe Abb. 2.9. Sofern die Ladungsträger die Barriere alleine aufgrund Ihrer durch die Fermi-Dirac-Statistik gegebene thermischen Energie überwinden ohne dabei zu tunneln, spricht man von thermionischer Emission.

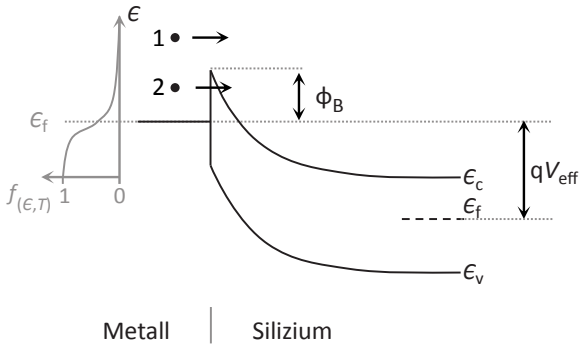


Abbildung 2.9: Schematisches Banddiagramm eines Schottkykontakts ohne Grenzflächenladung bei angelegter Spannung V_{eff}. Auf der Metallseite ist außerdem die Fermi-Dirac-Verteilung $f_{(\epsilon)}$ der Elektronen skizziert. Verfügen diese über eine ausreichende thermische Energie, können sie die Schottkybarriere ohne weiteres passieren (Pfad „1“). Elektronen mit geringerer Energie müssen durch die Barriere tunneln (Pfad „2“).

Im Fall des Tunnelprozesses wird zwischen direktem Tunneln und thermisch aktiviertem Tunneln über Phononenabsorption unterschieden. Die Form der Elektronenbarriere kann, unter positiver Vorspannung des Halbleiters und in direkter Nähe des Schottkyübergangs, in erster Näherung durch ein Dreieckspotential beschrieben werden. Die Löcherbarriere verhält sich bei inverser Spannung analog. In der Realität stellt sich je nach Ladungsdichte ein parabolischer Verlauf ein. Während bei einer reinen Schottkydiode die anliegende Spannung leicht zu bestimmen ist, ist der effektive Spannungsabfall V_{eff} über eine der beiden Schottkybarrieren im Si-Nanodraht FET nicht direkt zugänglich. Stattdessen kann V_{eff} beschrieben werden über

$$V_{\mathrm{eff}} = aV_{\mathrm{gs}} + bV_{\mathrm{ds}} + c \tag{2.32}$$

Die Koppelfaktoren a und b spiegeln dabei die Tatsache wider, dass elektrische Potentiale, die an den externen Kontakten angelegt werden, nicht unmittelbar zu einer entsprechend großen Bandverbiegung am Transistor führen. So fällt z. B. ein Teil der angelegten Spannung V_{gs} über dem Gateoxid ab, V_{ds} zu unterschiedlichen Teilen über den Kanal und die Barrieren an Source und Drain. Die Konstante c berücksichtigt Potentialverschiebungen des Flachbandfalls, welcher aufgrund unterschiedlicher Austrittsarbeiten, aber z. B. auch aufgrund von an Störstellen im Gateoxid eingefangener Ladung nicht unbedingt bei $V_{\mathrm{gs}} = V_{\mathrm{ds}} = 0\,\mathrm{V}$ liegen muss.

Im Folgenden wird auf die beiden hauptsächlichen Injektionsmechanismen, die thermionische Emission und das Tunneln, näher eingegangen. Zu diesem Zweck wird der Begriff der effektiven Barrierenhöhe ϕ_{eff} eingeführt, welche sich von der rein materialabhängigen Schottkybarriere ϕ_{B} unterscheidet. ϕ_{eff} ist die elektrisch effektive Barrierenhöhe, die sich abhängig von der angelegten Spannung einstellt und kann als die für den Stromfluss relevante Größe angesehen werden. Auf das genaue Verständnis, wie ϕ_{eff} interpretiert werden kann, wird in Abschnitt 6.2.3 näher eingegangen.

b) Bereich der thermionischen Emission

Im Bereich der thermionischen Emission ist I_{ds} gegeben durch [70]

$$I_{\mathrm{ds}} = S \underbrace{A^* T^2 \exp\left(\frac{-\phi_{\mathrm{B}}}{k_{\mathrm{B}} T}\right)}_{j_{\mathrm{therm}}} \left[\exp\left(\frac{qV_{\mathrm{eff}}}{k_{\mathrm{B}} T}\right)\right] \tag{2.33}$$

$$= S j_{\mathrm{therm}} \exp\left(\frac{qV_{\mathrm{eff}}}{k_{\mathrm{B}} T}\right), \tag{2.34}$$

wobei S die Querschnittsfläche des Nanodrahts, A^* die Richardsonkonstante, T die Temperatur, k_{B} die Boltzmannkonstante und q die Elementarladung ist. Die effektive Barriere ϕ_{eff} ist im Flachbandfall mit $V_{\mathrm{eff}} = 0\,\mathrm{V}$ zunächst gleich der Schottkybarriere. Werden die Bänder nun mit abnehmender Gatespannung lokal an der Source nach oben gezogen, reduziert sich die Injektion aufgrund von thermionischer Emission, da immer weniger Ladungsträger an Source über eine ausreichende Energie verfügen, die Barriere zu überwinden. In diesem Bereich ist die effektive Barrierenhöhe ϕ_{eff} gerade der Potentialunterschied zwischen dem Ferminiveau im Metall am Sourcekontakt und dem höchsten Punkt des Leitungsbandminimums. Sie kann dann ausgedrückt werden als Summe aus der reinen Schottkybarrierenhöhe ϕ_{B} und einem Anteil, der die durch $V_{\mathrm{eff}} < 0$ verursachte lokale Bandverbiegung in der Nähe des Schottkyübergangs berücksichtigt. Bei Langkanaltransistoren wie sie auch hier betrachtet werden und mittlerem V_{ds} ist das Gatefeld maßgeblich für die Bandverbiegung und damit die Form der Barriere verantwortlich. In halblogarithmischer Darstellung ergibt sich aus (2.34) und (2.32) im Bereich thermionischer Emission für den Drainstrom eine Geradengleichung in V_{gs} im Format

$$\log(I_{\mathrm{ds}}) = \underbrace{\log(S j_{\mathrm{therm}}) + \frac{qbV_{\mathrm{ds}}}{k_{\mathrm{B}} T} + \frac{qc}{k_{\mathrm{B}} T}}_{A_1} + \underbrace{\frac{qa}{k_{\mathrm{B}} T}}_{B_1} V_{\mathrm{gs}}. \tag{2.35}$$

Die Steigung B_1 der Geraden wird demnach durch die Kopplung a der Gateelektrode an den Nanodraht und die Temperatur festgelegt. Da

T im Experiment bekannt ist, lässt sich über die Steigung der Gatekoppelfaktor bestimmen. Die Höhe der Schottkybarriere ist dagegen im Achsenabschnitt A_1 enthalten und somit nicht direkt zugänglich. Bei langen Transistoren ist der Drainkoppelfaktor b zwar klein. Für kleine b wie auch für kleine V_{ds} sind die Auswirkungen des Drainfeldes auf die Schottkybarriere an Source nur schwach und der zweite Summand in A_1 kann vernachlässigt werden. Da eine Verschiebung der Kennlinien aufgrund von Aufladungseffekten jedoch nicht ausgeschlossen werden kann, ist c grundsätzlich unbekannt. Somit verhindert der dritte Term in A_1 eine direkte Berechnung der Schottkybarriere aus der Steigung der Transferkennlinie, selbst wenn der thermionische Bereich der Kennlinie nicht maskiert ist und daher klar identifiziert werden kann. Aus diesem Grund wurde der Ansatz verfolgt, die Schottkybarriere aus der Temperaturabhängigkeit der Transferkennlinien zu extrahieren (Abschnitt 6.3).

c) Tunnelregime

Bei einer Erhöhung der Gatespannung über die Flachbandbedingung hinaus wird die sourceseitige Barriere geschmälert (Abb. 2.10b), wodurch Tunneln durch die Barriere relevant wird. Die Barriere nimmt dabei eine parabolische Form an, welche nahe an der Metall-/ Halbleitergrenzfläche durch ein Dreieckspotential angenähert werden kann. Mit zunehmender Gatespannung nimmt die Breite der Barriere kontinuierlich ab. Die Höhe des Drainstroms hängt dabei von der energieabhängigen, (elastischen) Tunnelwahrscheinlichkeit $\tilde{t}(\epsilon, E)$ ab. In der Wentzel-Kramers-Brillouin-Näherung (WKB) kann $\tilde{t}(\epsilon, E)$ ausgedrückt werden durch [71]

$$\tilde{t}(\epsilon, E) \sim \exp\left(-K\frac{(\phi_{\mathrm{B}} - \epsilon)^{3/2}}{qE}\right), \tag{2.36}$$

wobei K eine Konstante und E das elektrische Feld am Schottkyübergang darstellen. Die Bandverbiegung, welche sich aufgrund der von außen angelegten Spannungen einstellt, geht somit über das resultierende elektrische Feld E an der Barriere in die Tunnelwahrscheinlichkeit ein.

Mit größerem V_{gs} wächst auch E, weshalb der Exponent abnimmt. Je tiefer der Transistor im AN- Zustand betrieben wird, desto weniger stark steigt der Drainstrom an. Da der Tunnelstrom nicht nur von der Tunnelwahrscheinlichkeit, sondern auch von der Energieverteilung der Ladungsträger im Sourcekontakt abhängt, ist der Tunnelstrom temperaturabhängig. In Abschnitt 6.2.3 wird ein mikroskopisches Modell vorgestellt, welches die Temperaturabhängigkeit des Tunnelstroms näher beleuchtet.

d) Transferkennlinie

Abb. 2.10a illustriert das Banddiagramm eines undotierten SiNW-Transistors mit Rückseitengate mit angelegter Drain-Source-Spannung V_{ds}. Die Gate-Source-Spannung V_{gs} ist in der Skizze als leicht negativ angenommen, sodass sich an der sourceseitigen Barriere trotz $V_{\mathrm{ds}} \neq 0\,\mathrm{V}$ gerade der Flachbandfall ergibt. Werden von hier aus die Bänder über V_{gs} nach oben gezogen, können nur thermisch angeregte Elektronen die linke Barriere überwinden, da sie für Tunnelprozesse zu breit wäre. Werden die Bänder jedoch unter den Flachbandfall gedrückt, so schmälert sich die sourceseitige Barriere und Elektronentunneln wird den Stromfluss dominieren (vgl. linke Barriere in Abb. 2.10b). Im realen Transistor müssen beide Ladungsträgersorten berücksichtigt werden, da sich der Gesamtstrom I_{ds} aus Elektronen und Löchertransport zusammensetzt. So fließt im skizzierten Fall mit $V_{\mathrm{gs}} < 0\,\mathrm{V}$ an Source in Abb. 2.10a nur ein geringer, thermionisch injizierter Elektronenstrom und die Barriere kann als „sperrend“ aufgefasst werden. Allerdings können unter diesen Bedingungen Löcher an Drain ins Valenzband tunneln; es kommt also zu einem – ebenfalls positiven – Löcherstrom. Der betrachtete Transistortyp befindet sich bei gegebenem $V_{ds} > 0\,\mathrm{V}$ sowohl für negative, als auch für positive V_{gs} im AN-Zustand. Der Kennlinienbereich hin zu negativem V_{gs} wird auch pFET-Ast genannt, der zu positivem V_{gs} nFET-Ast. Es lassen sich dementsprechend zwei Einsatzspannungen $V_{\mathrm{th,pF}}$ und $V_{\mathrm{th,nF}}$ für den Transistor definieren. Der AUS-Zustand, also der Bereich des minimalsten Stromflusses, wird um $V_{\mathrm{gs}} = 0\,\mathrm{V}$ erreicht, wo an beiden Barrieren nur eine geringe Tunnelwahrscheinlichkeit herrscht. Dieses

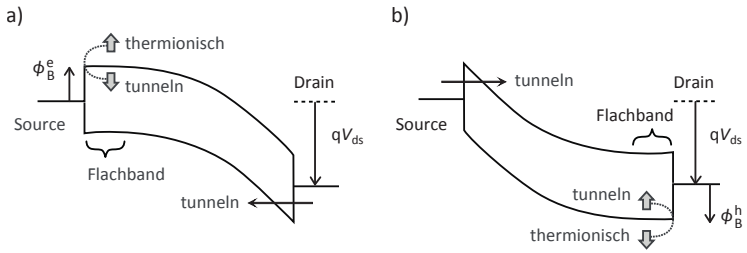


Abbildung 2.10: Banddiagramme eines SiNW Transistors mit Rückseitengate am Übergang zwischen rein thermionischer Emission und zusätzlichem Tunneln für a) Elektronentransport an Source; werden die Bänder an Source vom Flachbandfall ausgehend nach oben geschoben, dominiert thermionischer Transport, werden sie nach unten gezogen dominiert Tunneln. b) äquivalente Situation für den Löchertransport an Drain.

Verhalten nennt man „ambipolar“. Durch gezielte Strukturierung von Top-Gates über den Metall-Halbleiterübergängen im Nanodraht ist es möglich, die Bandverbiegung an beiden Barrieren, unabhängig von der Spannung am Rückseitengate, individuell anzupassen. So kann z.B. die drainseitige Barriere so eingestellt werden, dass sie das Tunneln von Löchern ins Valenzband verhindert und kein Löcherstrom zustande kommt. Die Kennlinie wird hierdurch unipolar und der Transistor verhält sich wie ein nFET. In einer Transistorsimulation können beide Äste der ambipolaren Kennlinie separat voneinander untersucht werden. Abb. 2.11 zeigt eine entsprechende Bauelementesimulation mit Hilfe des Programmpakets TCAD Sentaurus. Dabei zeigt sich im Elektronen- bzw. nFET-Ast der simulierten $\log[I_{\mathrm{ds}}]_{(V_{\mathrm{gs}})}$ Charakteristik (blau gezeichnet) ein „Knick“ bei kleinem, negativen V_{gs}. Dieser Punkt entspricht exakt der Flachbandbedingung an der sourceseitigen Barriere, wie in Abb. 2.10a skizziert. Hier geht mit wachsender Gatespannung der Injektionsmechanismus für Elektronen an der Source von reiner thermionischer Emission zu Tunneln über. Ein analoger Knick

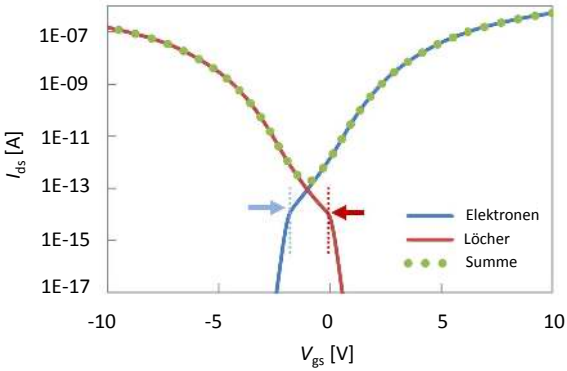


Abbildung 2.11: Simulierte $I_{\mathrm{ds}}(V_{\mathrm{gs}})$ Charakteristik eines undotierten SiNW-Transistors mit aufgewachsener Oxidhülle und Rückseitengate bei 300 K und $V_{\mathrm{ds}} = 1\,\mathrm{V}$, $L = 400\,\mathrm{nm}$. Der Si-Kerndurchmesser beträgt 14 nm und die Dicke der SiO_2-Hülle 8 nm, ferner wurden 100 nm Gateoxiddicke angenommen. Die Übergangspunkte zwischen rein thermionischer Emission und einsetzendem Tunnelbeitrag sind mit Pfeilen markiert. Details zur Simulation sind in [72] beschrieben.

findet sich im pFET-Ast (rot gezeichnet) für Löcherinjektion an der Drain. Bei abnehmendem V_{ds} werden die jeweiligen Flachbandbedingungen früher, also bei immer kleinerem $|V_{\mathrm{gs}}|$ erreicht. Dabei bewegen sich die Knicke in Abb. 2.11 aufeinander zu und treffen sich beim gleichen V_{gs}-Wert für $V_{\mathrm{ds}} = 0\,\mathrm{V}$. Da sich das Ferminiveau von $NiSi_2$ nicht exakt in der Mitte der Bandlücke von Si befindet, liegt dieser Punkt im simulierten Transistor bei leicht negativem V_{gs}, anstelle von $V_{\mathrm{gs}} = 0\,\mathrm{V}$. Auch wenn die Beiträge von Elektronen- und Löcherleitung separat berechnet werden können, wird in einem realen Transistor mit Rückseitengate stets die Summe aus beiden Ästen gemessen (gepunktete Linie in Abb. 2.11). Aus diesem Grund wird der Übergang zwischen

rein thermionischer Emission und Tunnelregime leicht durch den Tunnelbeitrag des jeweils anderen Asts maskiert. Dies ist insbesondere bei dünnen Nanodrähten wie dem hier simulierten der Fall. In real gemessenen, halblogarithmisch aufgetragenen Transferkennlinien kann daher das untere Segment des Tunnelbereichs leicht mit dem linearen Anstieg des thermionischen Abschnitts verwechselt werden.

3 Charakterisierungsmethoden

3.1 Ladungsträgerbeweglichkeit

Zur Bestimmung der Ladungsträgerbeweglichkeit (Kap. 2.1.3) gibt es verschiedene Messverfahren, die je nach Probenart in Frage kommen können. Im Folgenden wird näher auf die drei Verfahren der Feldeffektbeweglichkeit μ_{FE}, der effektiven Beweglichkeit μ_{eff} und der Magnetowiderstandsbeweglichkeit μ_{MR} eingegangen, die in dieser Arbeit zur Anwendung kamen. Allen drei Verfahren gemeinsam ist die Möglichkeit, Aussagen über die Beweglichkeit von Kurzkanaltransistoren zu machen. Die Grundlage für die Methoden der Feldeffektbeweglichkeit und der effektiven Beweglichkeit bildet die Drainstromgleichung des Feldeffekttransistors in der „charge-sheet approximation“

$$I_{\mathrm{ds}} = \frac{\mu_{\mathrm{eff}} W Q'_{\mathrm{inv}}}{L} V_{\mathrm{ds}} - W \mu_{\mathrm{eff}} \frac{k_{\mathrm{B}} T}{q} \frac{\mathrm{d}Q'_{\mathrm{inv}}}{\mathrm{d}x} \tag{3.1}$$

Der linke Term in (3.1) beschreibt den Ladungstransport unter Einfluss eines elektrischen Feldes aufgrund einer Drain-Source-Spannung V_{ds}, der sogenannten Drift. Der rechte Term berücksichtigt die Diffusion von Ladungsträgern. Diese findet statt, wenn im Material ein Gradient in den Elektron- oder Lochkonzentrationen vorherrscht. Unter der Annahme einer homogenen Ladungsverteilung im Kanal existiert zwischen Source und Drain kein Dichtegradient und der Diffusionsbeitrag kann vernachlässigt werden, da $\mathrm{d}Q'_{\mathrm{inv}}/\mathrm{d}x = 0$. Aus der so vereinfachten Gleichung (3.1) leiten sich die Methoden der Feldeffektbeweglichkeit und der effektiven Beweglichkeit (Abschnitt 3.1.2) ab. Die hierbei verwendeten Drainspannungen liegen typischerweise im Bereich $\leq 50\,\mathrm{mV}$.

3.1.1 Feldeffektbeweglichkeit

Zur Berechnung der Feldeffektbeweglichkeit wird die flächennormierte Inversionsladung Q'_{inv} im linearen Fall ($V_{\mathrm{ds}} \ll V_{\mathrm{gs}} - V_{\mathrm{th}}$) abgeschätzt durch

$$Q'_{\mathrm{inv}(V_{\mathrm{gs}})} \approx C'_{\mathrm{ox}}(V_{\mathrm{gs}} - V_{\mathrm{th}})\,, \tag{3.2}$$

mit $C'_{\mathrm{ox}} = C_{\mathrm{ox}}/A_{\mathrm{ges}}$. Die Oxidkapazität C_{ox} wird in der Regel an einer transistorähnlichen Teststruktur mit großer Kanalfläche A_{ges} bei $V_{\mathrm{gs}} = V_{\mathrm{dd}}$ gemessen. Für den Drainstrom folgt damit im linearen Operationsbereich

$$I_{\mathrm{ds}} = \frac{W}{L}\mu_{\mathrm{eff}}C'_{\mathrm{ox}}(V_{\mathrm{gs}} - V_{\mathrm{th}})V_{\mathrm{ds}}\,. \tag{3.3}$$

Der Übertragungsleitwert ist definiert als

$$g_{\mathrm{m}(V_{\mathrm{gs}},V_{\mathrm{ds}})} = \left.\frac{\partial I_{\mathrm{ds}}}{\partial V_{\mathrm{gs}}}\right|_{V_{\mathrm{ds}}}\,. \tag{3.4}$$

Aus (3.3) ergibt sich so mit (3.4) für kleine V_{ds}

$$\mu_{\mathrm{FE}} = \frac{L\, g_{\mathrm{m}(V_{\mathrm{gs}},V_{\mathrm{ds}})}}{WC'_{\mathrm{ox}}V_{\mathrm{ds}}}\,. \tag{3.5}$$

Das μ_{FE}-Verfahren birgt den geringsten Aufwand zur Bestimmung einer Ladungsträgerbeweglichkeit – es ist aber aufgrund der insbesondere nur grob abgeschätzten Inversionsladung und der Unsicherheit in L, als der tatsächlich elektrisch wirksamen Kanallänge, auch mit dem größten Fehler unter den drei hier beschriebenen Methoden verbunden. Aufgrund seiner Einfachheit findet es dennoch bevorzugt in der Halbleiterfertigung im Sinne einer Schnelltestcharakterisierung auf Waferebene häufig Verwendung. An dieser Stelle sei nochmals darauf hingewiesen, dass μ_{eff} die Driftbeweglichkeit der Inversionsladungsträger im MOSFET-Kanal beschreibt und somit von der theoretischen Beschreibung her der Driftbeweglichkeit μ_0 entspringt, siehe Kap. 2.1.3.

3.1.2 Effektive Beweglichkeit

Bei der effektiven Beweglichkeit μ_{eff} wird die Zahl der Inversionsladungsträger im Kanal N_{inv} über Kapazitäts-Spannungsmessungen (engl. 'Capacitance-Voltage', CV) ermittelt. Abb. 3.1 zeigt das typische CV-Verhalten eines n-MOSFETs. Zur Bestimmung von N_{inv} misst man die Kapazität zwischen Gate und Kanal C_{gc} in Abhängigkeit der Gatespannung und integriert von V_{fb} bis $V_{\text{g}} = V_{\text{dd}}$

$$N_{\text{inv}} = \frac{1}{q} \int_{V_{\text{fb}}}^{V_{\text{g}}} C_{\text{gc}} \, \mathrm{d}V_{\text{g}} \,, \tag{3.6}$$

wobei $Q'_{\text{inv}} = qN_{\text{inv}}/(A_{\text{ges}})$ ist. N_{inv} entspricht damit der Fläche unter $C_{\text{gc}(V_{\text{gs}})}$ in Abb. 3.1.

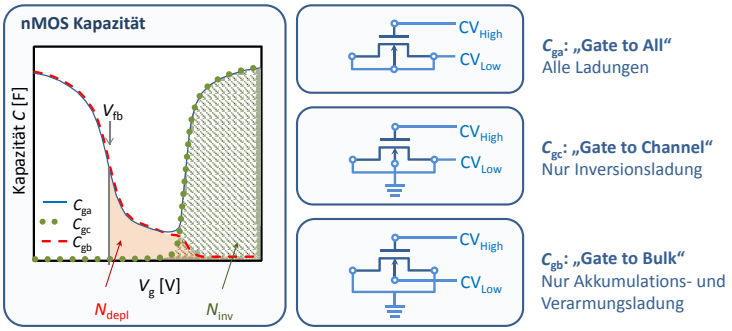


Abbildung 3.1: Anschlusskonfigurationen zur Messung verschiedener MOSFET-Kapazitäten und resultierende Messkurven am Beispiel eines n-MOS Transistors. Die Kapazität wird über eine Impedanzmessung mittels Kleinsignalmodulation bestimmt. Hierzu wird an CV_{high} auf das angelegte Potential eine kleine Wechselspannung V_{ac} der Frequenz f aufmoduliert. Zur näheren Erläuterung des Kurvenverlaufs siehe Anhang A.1.

Neben N_{inv} wird der Drainleitwert benötigt. Dieser ist definiert als

$$g_{\mathrm{d}(V_{\mathrm{gs}},V_{\mathrm{ds}})} = \left.\frac{\partial I_{\mathrm{ds}}}{\partial V_{\mathrm{ds}}}\right|_{V_{\mathrm{gs}}} . \tag{3.7}$$

Aus (3.3) und (3.7) leitet sich die effektive Beweglichkeit μ_{eff} ab:

$$\mu_{\mathrm{eff}} = \frac{L\, g_{\mathrm{d}(V_{\mathrm{gs}},V_{\mathrm{ds}})}}{W Q'_{\mathrm{inv}}} . \tag{3.8}$$

Das etwas aufwändigere μ_{eff} -Verfahren liefert aufgrund der Ladungsbestimmung durch Integration über die Gatekapazität deutlich genauere Resultate als das μ_{FE} -Verfahren. Es ist jedoch gleichermaßen auf kleine V_{ds} -Werte beschränkt. Darüber hinaus erfordert es ebenso eine genaue Kenntnis der elektrisch wirksamen Kanallänge L_{eff}, was insbesondere bei kurzen Transistoren zu großen Unsicherheiten führen kann. Weitere Unsicherheiten ergeben sich bei der Betrachtung von Kurzkanaltransistoren durch parasitäre Effekte wie z.B. Überlapp- und Millerkapazitäten, die eine akkurate Bestimmung der Inversionsladung erschweren.

Da Serienwiderstände den Drainstrom verringern und für die Kanalbeweglichkeit eigentlich der Kanalleitwert (bzw. der „intrinsische“ Drainleitwert) benötigt wird, sind diese vorher aus dem Drainleitwert herauszurechnen. Serienwiderstände führen so im μ_{eff} -Verfahren vor allem bei Kurzkanaltransistoren zu deutlich höheren Unsicherheiten als z.B. im Magnetowiderstandsverfahren, da sie über den Drainleitwert direkt in die Beweglichkeit eingehen.

Um Transistoren verschiedener Technologien miteinander vergleichen zu können, wird μ statt über V_{gs} oft über das effektive, vertikale Feld $E_{\mathrm{eff,vert}}$ aufgetragen. Hierfür wird zusätzlich die Verarmungsladung benötigt. Diese erhält man durch Messung der Kapazität zwischen Gate und Substrat, wobei Source und Drain auf Masse liegen (Abb. 3.1). Integration von V_{fb} bis $V_{\mathrm{g}} = V_{\mathrm{dd}}$ gibt die Zahl der Ladungen im Verarmungsfall

$$N_{\mathrm{depl}} = \frac{1}{q}\int_{V_{\mathrm{fb}}}^{V_{\mathrm{g}}} C_{\mathrm{gb}}\, \mathrm{d}V_{\mathrm{g}} , \tag{3.9}$$

entsprechend der Fläche unter der „Gate to Bulk“- Kurve C_{gb} in Abb. 3.1. Das effektive Feld kann dann aus (3.6) und (3.9) berechnet werden

$$E_{\mathrm{eff,vert}(V_{\mathrm{gs}})} = \frac{q}{\varepsilon_{\mathrm{Si}}\varepsilon_0}(N_{\mathrm{depl}(V_{\mathrm{gs}})} + \eta N_{\mathrm{inv}(V_{\mathrm{gs}})})\,. \tag{3.10}$$

Der hier eingeführte Parameter η ist abhängig von der Substratorientierung und trägt der Universalität der Beweglichkeit in Bezug auf $E_{\mathrm{eff,vert}}$ Rechnung [73]. Für (100) orientiertes Substrat gilt $\eta = 1/2$ für Elektronen und $\eta = 1/3$ für Löcher [74, 75].

3.1.3 Magnetowiderstandsbeweglichkeit

Im Vergleich mit den oben geschilderten Standardmethoden der Feldeffektbeweglichkeit und der effektiven Beweglichkeit bietet die Magnetowiderstandsmethode einen direkteren Zugang zur Beweglichkeit. So muss keine Annahme zur effektiven Kanallänge getroffen werden, die Sensitivität bezüglich Anschlusswiderständen ist geringer und das Verfahren ist nicht *per se* auf kleine V_{ds} begrenzt. Die Magnetowiderstandsbeweglichkeit ergibt sich aus der Drude-Boltzmann Theorie, siehe Kap. 2.1.3c). Aus (2.25) folgt

$$\frac{R_B}{R_0} = \mu_{\mathrm{MR}}^2 B^2 + 1\,, \tag{3.11}$$

wobei R_B und R_0 den Probenwiderstand mit und ohne Magnetfeld bezeichnen. Über den hier diskutierten, rein geometriebasierten Widerstandseffekt hinaus gibt es noch den sogenannten physikalischen Magnetowiderstandseffekt (PMR). Dieser kann verschiedene physikalische Ursachen haben, ist aber für die in dieser Arbeit untersuchten Si-basierten Proben wesentlich geringer als der geometrische Widerstandseffekt und kann daher vernachlässigt werden [38]. Prinzipiell wäre auch eine Bestimmung der Ladungsträgerbeweglichkeit über den Halleffekt möglich. Dies hätte zum Beispiel den Vorteil, dass die Anschlusswiderstände bei der Messung nicht betrachtet werden müssten. Hinzu käme, dass für eine Hallmessung nur vergleichsweise kleine magnetische Feldstärken vonnöten wären. Entsprechend den geometrischen Randbedingungen einer Hallmessung (mit $L \gg W$ und seitlichen, in Bezug auf die

Gesamtlänge punktförmigen Kontakten zum Abgriff der Hallspannung) wären jedoch wohl definierte Hallmessungen nicht an Kurzkanaltransistoren mit $L = 30\,\mathrm{nm}$, sondern lediglich an Langkanaltransistoren mit $L \gtrsim 300\,\mathrm{nm}$ möglich. Der Halleffekt eignet sich daher nicht, um die Beweglichkeit kürzester Transistoren zu bestimmen. Gerade diese sind jedoch aus technologischer Sicht von größtem Interesse, da sie die Performance eines hochintegrierten Chips maßgeblich bestimmen.

Im Gegensatz zum Halleffekt kann der geometrische Magnetowiderstandseffekt direkt an Standardtransistoren gemessen werden, ohne die Erfordernis einer speziellen Teststruktur. Wie bereits geschildert, tritt er komplementär zum Halleffekt auf. Mit (2.25) kann der Gesamtwiderstand R_B einer Probe im Magnetfeld geschrieben werden als

$$\begin{aligned} R_B &= R_0 + R_0 \mu_{\mathrm{MR}}^2 B^2 \\ &= R_0 + \Delta R_0(B)\,. \end{aligned} \tag{3.12}$$

Zu dieser idealen Grundgleichung gibt es zwei wichtige Korrekturen, die bei der Messung realer Proben zu berücksichtigen sind: die Geometrie- und die Serienwiderstandskorrektur. Auf diese wird in Abschnitt 5.3.1 detailliert eingegangen. Für Proben mit hohen Beweglichkeiten, wie dies z.B. oft bei III-V-Halbleitern der Fall ist, ist es ausreichend, jeweils eine Messung des Transistorstroms mit und ohne Magnetfeld durchzuführen. Aus den so erhaltenen Werten für R_B und R_0 kann dann nach Gl.(3.11) direkt die Beweglichkeit berechnet werden. Daher kann durch Messung der Ausgangs- und Transferkennlinien mit und ohne Feld die Beweglichkeit in einem großen Parameterraum sehr zeitsparend charakterisiert werden. Aufgrund des in Kap. 5.3.1 beschriebenen, durch die geringe Beweglichkeit der Proben bedingten geringen Signal-zu-Rausch-Verhältnisses konnten auf diesem Weg jedoch keine aussagekräftigen Ergebnisse erzielt werden. Aus (3.11) ergibt sich, dass die relative Änderung des Widerstands im Magnetfeld (ohne Korrekturterme)

$$\frac{R_B - R_0}{R_0} = \mu_{\mathrm{MR}}^2 B^2 \tag{3.13}$$

beträgt. Für den Kurzkanal-nFET mit $L = 30\,\mathrm{nm}$ ergibt sich so zum Beispiel im linearen Operationsbereich bei kleinen V_{ds} ein Produkt von

$\mu_{\mathrm{MR}}^2 B^2 \approx (120 \cdot 10^{-4}\,\mathrm{m^2V^{-1}s^{-1}})^2 \cdot (14\,\mathrm{V{\cdot}s{\cdot}m^{-2}})^2 = 0,028$. Dies entspricht einer geringen relativen Signalstärke von $2,8\,\%$. Durch die Ausnutzung der Abhängigkeit des Transistorstroms vom Magnetfeld $I(B)$ konnte die Rauschunsicherheit stark reduziert und damit der erreichbare Messbereich hin zu niedrigeren Beweglichkeiten deutlich erweitert werden. So wurde es möglich, die Beweglichkeit an aggressiv skalierten Kurzkanaltransistoren mit hoher Genauigkeit zu charakterisieren. Zu diesem Zweck wird der zu untersuchende Transistor bei konstanten Spannungseinstellungen betrieben, was einem bestimmten Operationspunkt auf der Transistorkennlinie entspricht. Hierbei wird kontinuierlich der Transistorstrom $I(t)$ als Funktion der Zeit gemessen (Abb. 3.2a). Nach Verstreichen einer gewissen Wartezeit von typischerweise $5-10$ Minuten, die zur Feststellung der Stärke des allgemeinen Signalrauschens sowie gegebenenfalls herrschender Drift dienten, wurde das Magnetfeld bis zur maximal zugänglichen Feldstärke von 14 T aufgebaut. Der Feldverlauf $B(t)$ wurde dabei ebenfalls als Funktion der Zeit gemessen. Aus $I(t)$ und $B(t)$ lässt sich der Stromverlauf $I(B)$ gewinnen, welcher parabolisch nach $I_B \approx I_0/(1+\mu_{\mathrm{MR}}^2 B^2)$ mit dem Magnetfeld abnimmt (Abb. 3.2b).

Durch den Fit von Gl. (5.9) unter Einbeziehung von Serienwiderstands- und Geometriekorrektur an die Messdaten kann dann die Beweglichkeit bestimmt werden. In diesem Verfahren wird somit der gesamte Verlauf $I(B)$ zur Berechnung der Beweglichkeit herangezogen, statt nur durch zwei zufällig aus dem Rauschen statistisch ausgewählte Messpunkte bei $I(B=0)$ und $I(B=B_{\max})$ gestützt zu werden. Dies hat den Vorteil, dass inkonsistente Datenbereiche leicht identifiziert und vom Fit ausgenommen werden können. Hierzu gehören z. B. Bereiche, die durch diskrete Stromsprünge zum eigentlichen Verlauf parallelverschoben erscheinen. Ein solcher Sprung ist z. B. in Abb. 3.2 b) im Bereich zwischen 8 T und 10 T zu sehen. Dieses immer wieder beobachtete Verhalten erscheint diskret und reversibel und wird daher vermutlich durch einzelne Ladungseffekte im Gateoxid oder an dessen Grenzfläche verursacht. Durch den Fit über den gesamten Datenbereich wird außerdem der Einfluss des Signalrauschens auf das Endergebnis deutlich vermindert, siehe Abschnitt 5.3.1 unter „Signalstärke“.

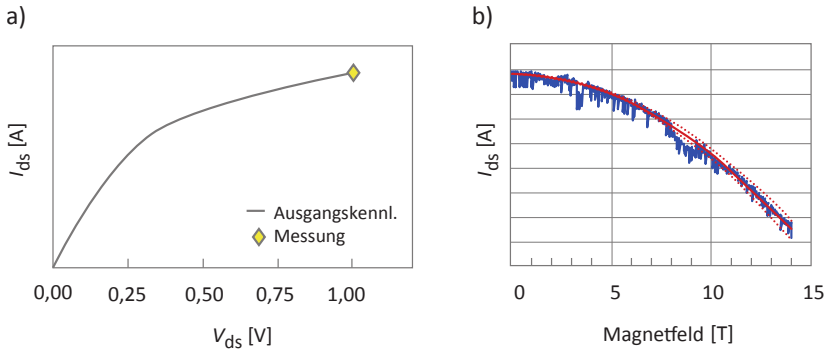


Abbildung 3.2: Bestimmung der Beweglichkeit aus $I_{\mathrm{ds}}(B)$: a) Beschränkung auf einen festen Operationspunkt des Transistors, an dem kontinuierlich der Drainstrom I_{ds} gemessen wird, b) bei Zuschalten eines Magnetfelds B kann eine parabolische Abnahme von I_{ds} beobachtet werden. Durch Fit der Daten nach Gl. (5.9) wird die Beweglichkeit erhalten.

Für eine typische Messung über den gesamten Feldstärkebereich werden für einen Durchlauf von $0 - 14\,\mathrm{T}$ ca. $30 - 45\,\mathrm{min}$ benötigt. Die Aufnahme einer Transistorkennlinie mit Beweglichkeiten bei 10 unterschiedlichen Arbeitspunkten würde somit einen Arbeitstag in Anspruch nehmen. Die erforderliche Messzeit, um die Beweglichkeit beim Gang durch die Ausgangs- und Transferkennlinien von Transistoren verschiedener Längen und Verspannungstypen bei unterschiedlichen Temperaturen zu charakterisieren wächst damit schnell ins Unermessliche. Dieses zeitliche Dilemma wurde durch die Entwicklung einer automatischen Messroutine gelöst. Diese ermöglicht es, eine vollständige Transfer- oder Ausgangskennlinie mit 6 repräsentativen Punkten während eines Magnetfelddurchlaufs zu messen. Der in Abb. 3.3 skizzierte Ablauf ist dabei wie folgt:

- Setzen der Gatespannung auf einen festen Wert
- Kontinuierliches Permutieren der angelegten Drain-Source-Spannung V_{ds}
- Speichern der Zeit t und messen des zugehörigen Stroms I_{d} für jedes V_{ds}
- Start des Magnetfeldaufbaus (0 ...14 T), dabei Speichern der aktuellen Feldstärke sowie des zugehörigen Zeitstempels in engen Intervallen.

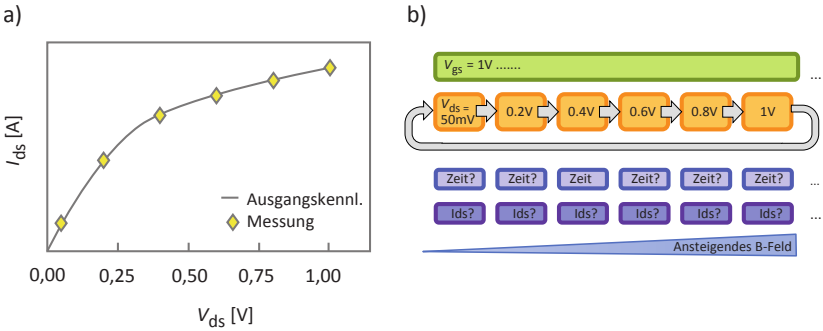


Abbildung 3.3: Illustration der automatischen Messroutine während eines Magnetfelddurchlaufs: a) Auswahl von 6 repräsentativen Datenpunkten am Beispiel einer Ausgangskennlinie $I_{\mathrm{d}}(V_{\mathrm{ds}})$ bei $V_{\mathrm{gs}} = 1\,\mathrm{V}$, b) Blockdiagramm zur Verdeutlichung der Datenaufnahme

Somit erhält man einen Verlauf $I_{\mathrm{d}}(t)$ für jeden der 6 Spannungspunkte auf der Kennlinie (Abb. 3.3a), sowie den zeitlichen Verlauf des Magnetfelds $B(t)$ während der Strommessungen. Aus der Kombination beider Informationen lässt sich für alle 6 Spannungseinstellungen $I_{\mathrm{d}}(B)$ ermitteln. Man erhält somit die Entwicklung des Drainstroms in Abhängigkeit des Magnetfelds, wodurch mit der oben beschriebenen Methode nach Gl. (5.9) die Beweglichkeit berechnet werden kann. Befindet sich der Magnet aufgrund der vorangegangenen Messung bereits

bei 14 T, kann die nachfolgende Messung analog während folgenden Feldabbaus (14 ... 0 T) erfolgen. So kann mit nur einem Magnetfeldgang zwischen den beiden Extremen $B = 0\,\mathrm{T}$ und $14\,\mathrm{T}$ der Verlauf der Beweglichkeit über die gesamte Kennlinie extrahiert werden, wodurch sich die benötigte Messzeit stark reduziert. Nähere Details zur Messroutine finden sich in Anhang A.2. Dieses Verfahren bietet verschiedene Vorteile:

- Hohe Genauigkeit für μ_{MR} gegenüber der Berechnung anhand zweier Stromwerte bei unterschiedlichen Magnetfeldern, da die Berechnung der Beweglichkeit durch Fit von $I_{\mathrm{ds}}(B)$ in einem weiten Magnetfeldbereich erfolgt und somit eine große statistische Sicherheit der Ergebnisse erreicht wird.

- Weitere Steigerung der Aussagekraft von μ_{MR} durch die Möglichkeit, inkonsistente (z.B. durch Rauschen verursachte) Bereiche in den Messdaten zu identifizieren und von der Berechnung auszunehmen.

- Automatisiertes Nachberechnen der Beweglichkeiten aller Datensätze für veränderte Randbedingungen, wie z. B. andere Anschlusswiderstände.

- Reduktion der erforderlichen Messzeit durch Permutation der elektrischen Spannungseinstellungen während eines Magnetfeldauf- bzw. -abbaus.

3.2 Serienwiderstand

Mit fortschreitender Skalierung der Transistoren nimmt die Gatelänge kontinuierlich ab, wodurch der Kanalwiderstand sinkt. Dahingegen vermindern sich die Source/Drain-Anschlusswiderstände nicht durch einfache Skalierung, sondern müssen mittels aufwändiger technologischer Hebel reduziert werden. In der Folge wird die Leistungsfähigkeit aktueller Transistorgenerationen in zunehmendem Maß durch die Anschlusswiderstände limitiert. Der Gesamt-Serienwiderstand R_{ser}

kann dabei als Summe mehrerer Teilkomponenten betrachtet werden, vgl. Abb. 5.9. Es ist zwar möglich, die Beiträge einzelner Komponenten zum Serienwiderstand wie z. B. den Schichtwiderstand der Source- und Drain-Diffusionsgebiete $R_{\mathrm{SD,diff}}$ oder auch den Metall-Halbleiterkontakt R_{cont} durch spezielle Teststrukturen wie Transferlängenstrukturen, Vier-Punkt-Kelvin-Strukturen oder Kontaktketten außerhalb des Transistors der Messung zugänglich zu machen. Die Hauptbeiträge zum Serienwiderstand liefern jedoch die Anschlusswiderstände zum Kanal, R_{ext} und $R_{\mathrm{SR,ch}}$, welche nur am integrierten Transistor zustande kommen und nicht einzeln zugänglich gemacht werden können. Um genaue Beweglichkeitswerte zu erhalten, ist sowohl für die Magnetowiderstandsmethode, als auch für das Verfahren der effektiven Beweglichkeit unbedingt eine Serienwiderstandskorrektur durchzuführen. Die nachfolgend beschriebene Methode erlaubt die hierfür erforderliche Bestimmung des Gesamt- Serienwiderstands R_{ser} am Transistor und beruht auf der Abhängigkeit des Gesamtwiderstands R_{ges} von der Transistorlänge [76, 77]. Als Ausgangspunkt hierfür dient die Drainstromgleichung für kleine V_{ds}, (3.3). Dabei wurden in der bisherigen Formulierung die Anschlusswiderstände vernachlässigt. Um sie einzubeziehen, lässt sich die Drainstromgleichung modifiziert schreiben als [38]

$$I_{\mathrm{ds}} = \frac{W}{L}\mu_{\mathrm{eff}}C'_{\mathrm{ox}}(V^*_{\mathrm{gs}} - V_{\mathrm{th}})V^*_{\mathrm{ds}}\,. \tag{3.14}$$

Dabei soll der Stern (*) die reinen, transistorinternen Beiträge bezeichnen, vgl. hierzu Abb. 3.4. Führt man nun die Anschlusswiderstände $R_{\mathrm{s}} + R_{\mathrm{d}} = R_{\mathrm{ser}}$ ein, so ergeben sich die extern zugänglichen Spannungen zu

$$V_{\mathrm{gs}} = V^*_{\mathrm{gs}} + I_{\mathrm{ds}}R_{\mathrm{s}} \;= V^*_{\mathrm{gs}} + {}^1\!/_2 I_{\mathrm{ds}}R_{\mathrm{ser}} \qquad \text{und} \tag{3.15}$$

$$V_{\mathrm{ds}} = V^*_{\mathrm{ds}} + I_{\mathrm{ds}}(R_{\mathrm{s}} + R_{\mathrm{d}}) = V^*_{\mathrm{ds}} + I_{\mathrm{ds}}R_{\mathrm{ser}} \quad , \tag{3.16}$$

wobei $R_{\mathrm{s}} = R_{\mathrm{d}} = {}^1\!/_2 R_{\mathrm{ser}}$ verwendet wurde. Diese Annahme ist aufgrund des symmetrischen Designs der hier untersuchten MOSFETs zulässig. Der Vollständigkeit halber sei angemerkt, dass in dieser quasistatischen Betrachtung aufgrund der geringen Gateleckströme kein Potentialabfall über die Gateleitung auftritt und daher $R_{\mathrm{g}} = 0$ gesetzt

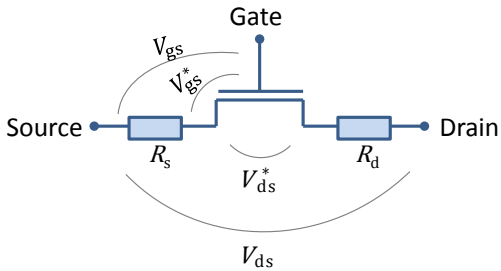


Abbildung 3.4: Ersatzschaltbild eines MOS-Feldeffekttransistors mit Anschlusswiderständen: sind $R_s, R_d \neq 0$, muss zwischen intern am Transistor abfallenden Spannungen (mit * gekennzeichnet), und den zu ihnen analogen, extern zugänglichen Größen unterschieden werden, nach [38].

werden kann. Setzt man (3.16) und (3.15) in (3.14) ein, erhält man im Fall starker Inversion ($V_{gs} - V_{th} \gg {}^1/_2 V_{ds}$)

$$I_{ds} = \frac{W_{eff}\mu_{eff}C'_{ox}(V_{gs} - V_{th})V_{ds}}{L_{eff} + W_{eff}\mu_{eff}C'_{ox}(V_{gs} - V_{th})R_{ser}} \tag{3.17}$$

L_{eff} bezeichnet die elektrisch effektive Kanallänge, welche über

$$L_{eff} = L + \delta L \tag{3.18}$$

zur strukturierten Kanallänge L in Bezug steht. Aus (3.17) und (3.18) folgt dass

$$R_{ges} = \frac{L - \delta L}{W_{eff}\mu_{eff}C'_{ox}(V_{gs} - V_{th})} + R_{ser}\,, \tag{3.19}$$

wobei $I_{ds} = V_{ds}/R_{ges}$ verwendet wurde. Bei konstantem „Gate- Overdrive“ entsprechend $V_{gs} - V_{th} =$ konst. erhält man für R_{ges} eine Geradengleichung in L:

$$R_{ges(L)} = \gamma(L - \delta L) + R_{ser} = \gamma L - \gamma\delta L + R_{ser}\,. \tag{3.20}$$

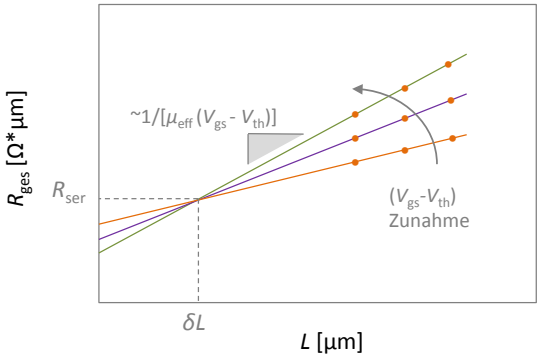


Abbildung 3.5: $R_{\text{ges}}(L)$ Plot für verschiedene Transistorlängen bei 3 unterschiedlichen Overdrive-Werten. Im Schnittpunkt der Geraden können idealerweise der Serienwiderstand und δL bestimmt werden.

Hier repräsentiert $\gamma = (W_{\text{eff}}\mu_{\text{eff}}C'_{\text{ox}}(V_{\text{gs}} - V_{\text{th}}))^{-1} =$ konst. die Steigung, welche sich invers proportional zur Ladungsträgerbeweglichkeit verhält. Messungen bei unterschiedlichem Overdrive führen somit zu unterschiedlichen Steigungen (Abb. 3.5). Allen Geraden gemeinsam ist ein Versatz in x um δL und ein Versatz in y um R_{ser}. Aus der Lage des Schnittpunkts lassen sich daher der Serienwiderstand R_{ser} und die effektive Kanallänge L_{eff} bestimmen. In der Praxis gibt es einige mögliche Abweichungen vom beschriebenen idealen, linearen Verhalten.

- So sind sowohl δL als auch R_{ser} nicht vollständig unabhängig von der Gatespannung. Bei höherem V_{gs} breitet sich der Kanal etwas weiter in die Source- und Drain Diffusionsgebiete hinein aus. Der Serienwiderstand R_{ser} nimmt damit etwas ab. Während dieser als „channel broadening“ bezeichnete Effekt bei Transistoren mit niedrig dotierten Diffusionsgebieten signifikant werden kann, ist diese Komplikation bei den in dieser Arbeit untersuchten Transistoren aufgrund der sehr hohen Dotierstoffkonzentration in Source und Drain $> 10^{20}\,\text{cm}^{-3}$ nicht zu erwarten.

- Die Beweglichkeit μ ist eine Funktion der Kanallänge. Während μ bei Langkanaltransistoren beinahe konstant ist, nimmt sie bei kleinem L aufgrund von Kurzkanaleffekten rapide ab. Dieses als „Mobility-roll-off“ bezeichnete Verhalten sollte die Steigung einer Linie in $R_{\mathrm{ges}}(L)$ graduell erhöhen und damit die Abschätzung des Serienwiderstands hin zu niedrigeren Werten verschieben.

- Der im vorigen Punkt beschriebene Effekt wird bei mechanisch verspannten Transistoren dadurch etwas reduziert, dass der Stress bei kleinerer Kanallänge besser einkoppeln kann, was den Mobility-roll- off im Vergleich zu unverspannten Technologien abmildert.

Der Serienwiderstand ist vor allem bei Kurzkanaltransistoren von Bedeutung, bei denen er einen signifikanten Anteil am Gesamtwiderstand bildet. So beträgt der Serienwiderstand beim 30 nm nFET bei $V_{\mathrm{ds}} = 50\,\mathrm{mV}$ mehr als 50 % von R_{ges}. Der sourceseitige Anteil R_{s} hat dabei einen prinzipiell größeren Einfluss auf den Drainstrom als R_{d}, da durch R_{s} zusätzlich der Overdrive von im Idealfall $(V_{\mathrm{gs}} - V_{\mathrm{th}})$ auf tatsächlich $(V^*_{\mathrm{gs}} - V_{\mathrm{th}})$ reduziert wird (3.15). Bei Langkanaltransistoren hingegen dominiert der Kanalwiderstand den Gesamtwiderstand. Der Gesamtwiderstand eines Transistors mit $L = 2\,\mu\mathrm{m}$ der in dieser Arbeit untersuchten Transistoren liegt bei kleinem V_{ds} in Inversion z. B. knapp um einen Faktor 100 über dem Anschlusswiderstand. Die Extrapolation von Messwerten bei sehr großem R_{ges} und L auf kleine R_{ser}-Werte nahe $L = 0$ unterliegt daher großen Unsicherheiten. Um den Fehler zu minimieren und zu sinnvollen Werten zu gelangen, kann die Bestimmung des Serienwiderstands mit dieser Methode nur mit Hilfe von Kurzkanaltransistoren gemacht werden. In dieser Arbeit wurden hierfür die vier kürzesten verfügbaren Kanallängen herangezogen. Zusammen mit dem Umstand, dass sich die Geraden in der Praxis nicht exakt schneiden, eignet sich die Methode aufgrund der erreichbaren Aussagegenauigkeit heutzutage nicht mehr zur Bestimmung von δL. Stattdessen wird in dieser Arbeit δL aus Messung der Überlappkapazität bestimmt und der Gesamtwiderstand über L_{eff} aufgetragen. Der Serienwiderstand kann dann bei $L_{\mathrm{eff}} = 0$ abgelesen werden, siehe Grafik 5.5.

4 Messaufbau

Die elektrische Messung der Transistoren erfolgte über ein Semiconductor Characterization System SCS4200 der Firma Keithley mit 4 Kanälen („Source Measure Unit“, SMU). Mit jeder SMU kann dabei gleichzeitig ein elektrisches Potential an einen Kontaktanschluss des Bauelements angelegt, sowie der über diesen Kontakt fließende Strom gemessen werden. Je nach Messkonfiguration können dabei Auflösungen um 50 Femtoampere erreicht werden. Als Rahmensysteme zur Probenkontaktierung und Bereitstellung der Umweltparameter wie z. B. Druck, Temperatur und Magnetfeld standen je nach Anwendungsgebiet verschiedene Messaufbauten zur Verfügung. Die Si-Nanodraht-Transistoren wurden an einem Vakuum-Kryomessplatz (Abschnitt 4.1) ohne Magnetfeld gemessen, welcher die Kontaktierung der Proben im Vakuum mittels Einzelnadeln zulässt. Die Messungen von MOSFETs auf Waferebene bei Raumtemperatur (25 °C) fanden an halbautomatischen 200 mm und 300 mm Waferprobern statt. Nach dem Herausschneiden einzelner Transistoren aus dem Wafer und Bonden in Chipgehäuse wurde die weitere Charakterisierung an PPMS („Physical Property Measurement System“) Kryostaten der Firma Quantum Design durchgeführt, an welchen für die Magnetowiderstandsmessungen Feldstärken von bis zu 14 T erzeugt werden konnten (Abschnitt 4.2). Im Folgenden werden der Vakuum-Kryomessplatz sowie das PPMS System näher vorgestellt.

4.1 Elektrische Charakterisierung ohne Magnetfeld: Vakuum-Kryomessplatz

Die in dieser Arbeit untersuchten Si-Nanodraht-Transistoren reagieren sehr sensitiv auf äußere Umwelteinflüsse wie z. B. Luftfeuchtigkeit. An der Oberfläche eines Nanodrahts adsorbierende und desorbierende polare Moleküle wie z. B. Wasser verursachen Verschiebungen der Einsatzspannung, was zu Hysteresebildung in Strom-Spannungskennlinien führt, vgl. 6.2. Während derartige Effekte für die Entwicklung neuartiger Sensorkonzepte ausgenutzt werden können [27–29], sind sie für die Charakterisierung der reinen (intrinsischen) Transistoreigenschaften eines Nanodrahts eher hinderlich. Zur Messung der Si-Nanodraht-Transistoren kam daher ein CPX-VF Kryoprober der Firma Lake Shore mit den folgenden Schlüsseleigenschaften zum Einsatz (Abb. 4.1):

- Elektrische Kontaktierung von Proben über einen Substratkontakt sowie über mikromanipulatorgeführte Einzelnadeln im Vakuum

- Videomikroskop mit bis zu 400- facher Vergrößerung zur präzisen Kontaktierung kleinster Strukturen

- Thermische Abschirmung der Probe von der Umgebung über 3 Schilde, die bei Kühlung mit flüssigem Stickstoff bzw. Helium als Kühlfallen dienen und den Kammerdruck auf $1 \cdot 10^{-7}$ mbar absenken. Durch Ausheizen der Probe unter Vakuum kann die Probenoberfläche vor der Messung effektiv von Adsorbaten, wie beispielsweise dünnen Wasserfilmen, befreit werden

- Individuell einstellbare Probentemperatur im Bereich von 400 K bis 77 K bei Kühlung mit flüssigem Stickstoff, bzw. 4 K bei Kühlung mit flüssigem Helium

- Möglichkeit der Durchführung sehr rauscharmer Messungen durch schwingungsgedämpfte Lagerung der Kammer und doppelt geschirmter Triax-Messleitungen bis zur Halterung der Messnadel

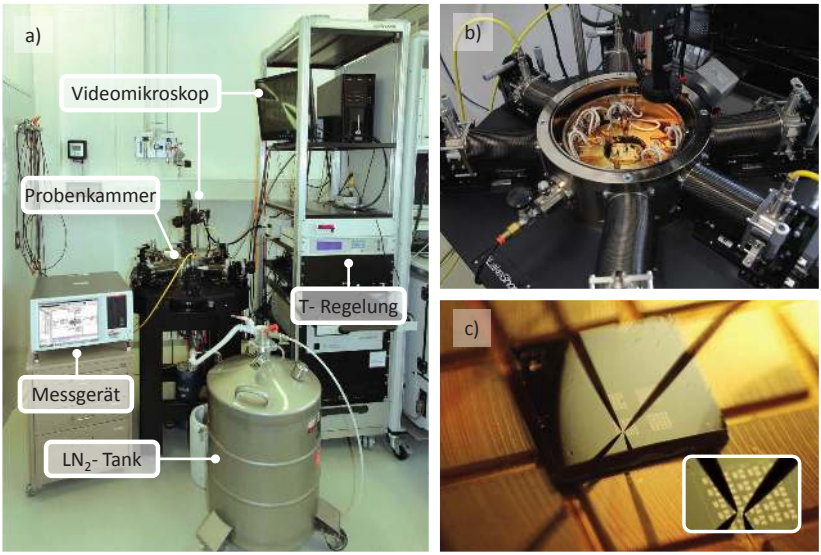


Abbildung 4.1: Aufbau des Kryomessplatzes: a) Gesamtübersicht, b) Blick in die Probenkammer mit ringförmig angeordneten Mikromanipulatoren, die die Messnadeln halten, c) Kontaktierung einer Probe. Einsatzbild: kontaktierte Nadelspitzen im Teststrukturfeld (Vergrößerung).

Der typische Aufbau für eine Messung von Nanodraht-Transistoren mit Rückseitengate ist dabei wie folgt: Der Chip wird mit Silberleitlack auf dem Probenhalter befestigt. Dieser ist elektrisch von der Umgebung isoliert und kann daher mit einem eigenen Potential beaufschlagt werden. Die Potentialleitung ist über eine Durchführung im Kammerboden von außen zugänglich und wird mit dem Messgerät über Triaxkabel verbunden. Die Kontaktierung von Source und Drain erfolgt an der Chipoberseite über die Messnadeln in der Kammer. Es stellte sich heraus, dass die eigentlich zum Prober gehörenden Spezialmessnadeln aus BeCu zu weich für eine sichere und temperaturstabile Kontaktierung der harten, mit nativem Nickeloxid überzogenen Kontaktflächen

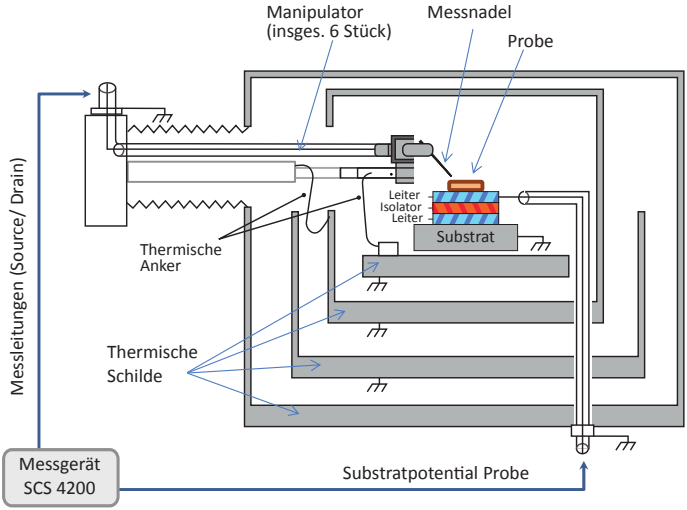


Abbildung 4.2: Schematische Zeichnung zum Aufbau der Vakuumkammer mit Kühlschilden, Manipulatoren und Probenkontaktierung, nach [78]

der Si-Nanodrahtproben waren. Aus diesem Grund wurde im Rahmen dieser Arbeit eine Aufnahme für Standard-Probernadeln aus Wolfram entwickelt und zwei der vier Manipulatoren entsprechend umgerüstet. Die Nadeln können bei geschlossener und evakuierter Kammer von außen durch Mikrometerschrauben in allen 3 Raumachsen präzise positioniert werden. Die optische Kontrolle erfolgt dabei über ein Videomikroskop, welches durch ein optisches Fenster einen vertikalen Blick auf den Probenbereich ermöglicht. Die Vergrößerung kann dabei stufenlos zwischen 50- und 400-fach eingestellt werden, sodass die Nadeln auch noch auf Kontaktflächen von z. B. $40\,\mu m \times 40\,\mu m$ präzise positioniert werden können. Die Messleitungen werden von der Nadel ausgehend in geschirmten Triax-Kabeln geführt und können an der Basis des Manipulators von außen kontaktiert werden. Die elektrische Charakterisierung erfolgte mit dem oben beschriebenen Keithley SCS 4200, dessen

Messkanäle über Triaxkabel mit den Signalleitungen am Cryo verbunden waren (Abb. 4.2). Die Temperaturregelung erfolgt durch Kühlung des Probenhalters und der Thermoschilde mittels flüssigem Stickstoff, bei Bedarf für Messungen unterhalb 77 K auch mittels flüssigem Helium. Die Durchflussmenge des Kühlmittels durch die Kapillaren des Probenhalters lässt sich dabei über ein Nadelventil gesondert einstellen. Mittels zweier Heizelemente an Probenhalter und innerem Schild kann aktiv der Kühlung entgegengeheizt und so ein stabiles thermisches Gleichgewicht bei der gewünschten Temperatur eingestellt werden. Die Messung der Ist- Temperatur und automatische Anpassung der Heizleistung zum Erreichen des Sollwerts erfolgt über einen LakeShore 366 Temperaturregler. Die hierbei erreichte Temperaturstabilität liegt üblicherweise im Bereich um $\Delta T \leq 0,1\,K$. Ein besonderer Vorteil des Vakuum-Kryomessplatzes liegt in der Kontaktierung von Bauelementen über Einzelnadeln. Hierdurch werden temperaturabhängige Messungen ermöglicht, ohne dass ein Bonden in Chipgehäuse erforderlich wird. Die ist insbesondere für sensitive, einzeln hergestellte Mikrostrukturen von Bedeutung, welche durch ein nachträgliches Bondverfahren beschädigt werden könnten.

4.2 Elektrische Charakterisierung mit Magnetfeld: PPMS Magnetfeldkryostat

Zur Bestimmung von Ladungsträgerbeweglichkeiten moderner Chiptransistoren über den Magnetowiderstandseffekt sind aufgrund der kleinen Beweglichkeitswerte inzwischen sehr hohe Magnetfelder erforderlich, wie bereits in Kapitel 3.1.3 geschildert. Die Messungen fanden daher in einem PPMS Kryostatsystem der Firma Quantum Design am Hochfeldlabor des Helmholtz-Zentrums Dresden-Rossendorf (HZDR) statt, welches mit bis zu 14 Tesla die nötigen Magnetfeldstärken erreicht. Zur Erzeugung dieser hohen Feldstärken kommt eine mit flüssigem Helium gekühlte, supraleitende Spule zum Einsatz. Diese befindet sich in einem 1,20 m hohen, zylindrischen Gehäuse mit ca. 70 cm Durchmesser, welches zur thermischen Isolation dient. Die Proben wer-

den auf einen maßgeschneiderten Probenhalter („Puck“) montiert und von oben in eine evakuierte Röhre eingebracht, an deren unterem Ende in ca. 1 m Tiefe der Puck im Zentrum der zylindrischen Spule, d. h. im Bereich des Feldmaximums auf einen Sockel gesteckt wird. Über diesen stehen 12 elektrische Kontakte zur Probencharakterisierung zur Verfügung, welche von außen über einen Lemo-Stecker zugänglich sind. Die Temperatur des Pucks und damit der Probe kann variabel im Bereich von 2 K bis 400 K eingestellt werden, wobei die Temperatur- und Magnetfeldregelung von einem Ansteuerrechner übernommen wird (Abb. 4.3b). Bedingt durch den Aufbau erfolgt die Kontaktierung der Proben nicht in der Kammer über Manipulatoren, sondern muss bereits vorher „fest“ auf dem Puck erfolgen. Zur sicheren Kontaktierung und dennoch flexiblen Einbringung von Einzeltransistoren in den PPMS Magnetfeld-Kryostaten wurde im Rahmen dieser Arbeit ein maßgeschneidertes Pucksystem entworfen, das im Folgenden näher vorgestellt wird.

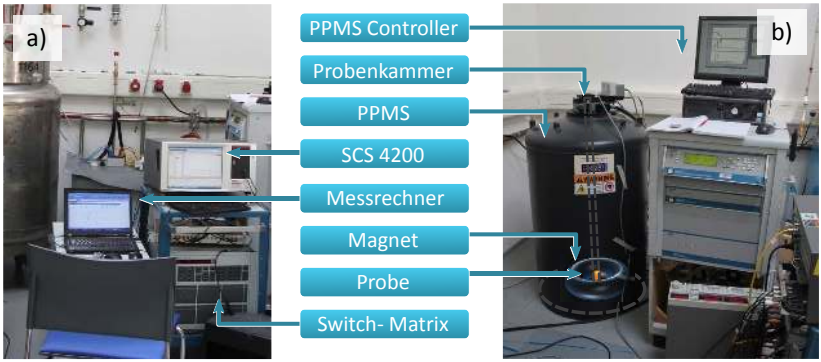


Abbildung 4.3: Aufbau zur Messung des Magnetowiderstands: a) Messgerät, Switch-Matrix und Messrechner, b) PPMS Kryostat mit Ansteuerrechner sowie skizzierter Lage von Probenkammer, Magnet und Probenpuck.

4.2.1 Pucksystem zur Transistorcharakterisierung in hohen Magnetfeldern

a) Wahl des Chipgehäuses und Probensegmentierung

Aufgrund der kleinen Anschlussflächen der Teststrukturen auf dem Wafer von nur 40 µm × 40 µm war eine direkte Kontaktierung der Transistoren durch mechanische Stempel oder handgelötete Verbindungen ausgeschlossen. Somit kam nur die Messung gebondeter Chips in Frage, für welche ein geeigneter Puck entworfen werden musste. Der im PPMS System insgesamt zur Verfügung stehende Probenraum ist auf einen Durchmesser von 2, 54 cm (1 Zoll) begrenzt. Des Weiteren war zu beachten, dass maximal 12 elektrische Kontakte über den Puck zugänglich gemacht werden können. Vor dem Entwurf des Pucksystems zur Probenhalterung stand daher zunächst die Wahl eines geeigneten Chipgehäuses, in welches die Teststrukturen gebondet werden konnten. Dieses sollte tieftemperaturbeständig und nichtmagnetisch sein, sowie mit der hierfür zu entwerfenden, nichtmagnetischen Halterung unter dem maximalen Probenraumdurchmesser bleiben. Die Wahl fiel auf ein keramisches, pinloses und in der Grundfläche quadratisches „Leadless Chip Carrier“ LCC02046- Gehäuse mit 20 Anschlusskontakten (Abb. 4.4). Die zu messenden Transistorstrukturen wurden mittels Laserschneidverfahren aus den 300 mm Wafern herausgeschnitten (Abb. 4.4a). Dies erfolgte an einem Synova LCS300 Laserschneidgerät mit einem grünen Nd:YAG Laser bei 532 nm Wellenlänge und 93 W Leistung. Zusätzlich zum Laser kommt während des Schneideprozesses ein 50 µm feiner, 175 bar starker Strahl aus entionisiertem Wasser zum Einsatz. Dieser dient der Strahlführung des Laserlichts, welches aufgrund von Totalreflexion an den Wänden stets innerhalb der Wassersäule bleibt. So wird eine Tiefenschärfe des Laserfokus im Bereich mehrerer Zentimeter erreicht. Darüber hinaus unterstützt der Wasserstrahl den Schneideprozess durch den Abtransport von herausgelösten Partikeln aus der Schneidefuge und der Probenkühlung. Hierdurch begrenzt sich die räumliche Wärmeausbreitung an der Schneidefuge auf einen Umkreis von 10 µm bis 100 µm [79]. Den kleinsten Abstand zur Sägefuge haben die ersten beiden Transistoren der zu bondenden Test-

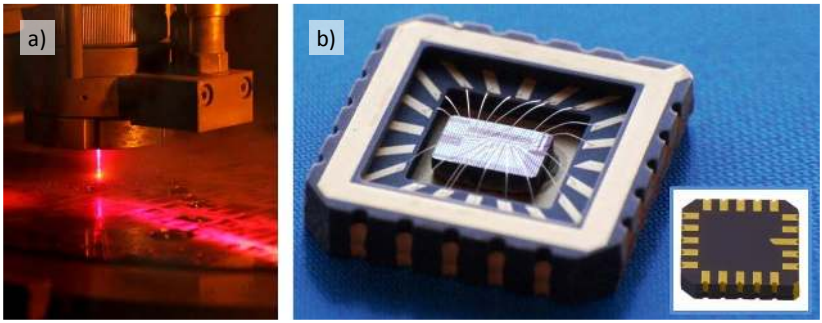


Abbildung 4.4: Probenaufbau: a) Extraktion einzelner, 2 mm × 3 mm großer Probensegmente aus einem 300 mm Wafer an einem Synova LCS300 Laserschneidegerät; die rote Färbung des Lasers ist auf den Transmissionsbereich des Sichtschutzglases zurückzuführen. b) gebondete Probe in keramischem LCC Gehäuse; kleines Bild: Kontaktflächen auf der Probenunterseite.

strukturreihe mit Abstandsmaßen von 500 µm bzw. 600 µm, für alle anderen beträgt die minimale Distanz 1 mm. Somit werden die zu untersuchenden Transistoren in der Mitte der 2 mm × 3 mm großen Probensegmente während des Schneideprozesses keinen erhöhten Temperaturen ausgesetzt. Da die Wafer bei der Herstellung eine passivierende Backend-Prozessierung erhalten haben, verursachte der Kontakt mit dem Kühlwasser oder den beim Schneiden entstandenen Partikeln keinerlei Degradation an den Proben oder dem Wafer. Im Anschluss an den Schneideprozess wurden die vereinzelten, 2 × 3 mm großen Chipsegmente in destilliertem Wasser gereinigt und getrocknet. Durch elektrische Messung der Transistoren vor und nach dem Schneideprozesses konnte verifiziert werden, dass durch das Laserschneiden keinerlei elektrische Degradation stattfand. Im Anschluss wurden die Probenstücke bei MPD („Microelectronic Packaging Dresden“) elektrisch leitend in die LCC Gehäuse eingeklebt und gemäß Verdrahtungsschema gebondet (Abb. 4.4b). Durch die Probenextraktion mittels Laserschneiden wur-

den die Wafer nicht gebrochen, sondern blieben im Ganzen erhalten und standen für weitere Messungen auf Waferprobern oder für erneute Probenextraktionen weiterhin zur Verfügung.

b) Design des Pucks

Der zu entwerfende Probenhalter musste den besonderen Anforderungen hinsichtlich der verwendeten starken Magnetfelder und den im PPMS System eingesetzten Temperaturen von 2 K - 400 K genügen. Darüber hinaus sollten die Proben leicht auswechselbar sein, wobei im gesamten Temperatur- und Magnetfeldbereich an allen Anschlüssen ein stabiler elektrischer Kontakt zur Probe zu gewährleisten war.

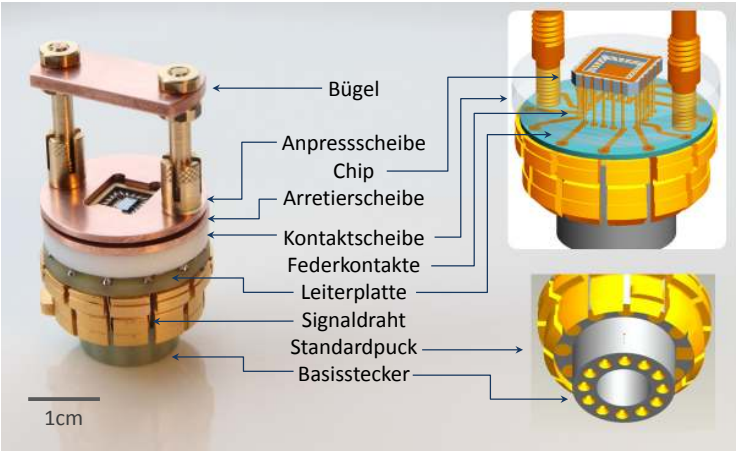


Abbildung 4.5: Aufbau des Messpucks zur Probenhalterung

Aus diesen Anforderungen heraus wurde im Rahmen dieser Arbeit eine Halterung für das LCC02046 Chipgehäuse auf dem PPMS Standardpuck von Quantum Design konzipiert und aufgebaut (Abb. 4.5). Die Kontaktierung des gebondeten Chips findet über 20 federnde Kontaktstifte aus vergoldetem BeCu mit einem Durchmesser von 0,34 mm und einer Länge von 5,9 mm statt. Diese stehen frei in den Bohrungen

einer keramischen „Kontaktscheibe“ von 5,0 mm Höhe, auf welche der Chip aufgesetzt wird. Eine Arretierscheibe sorgt für die korrekte Ausrichtung des Chips in Bezug auf die Federkontakte. Anschließend wird der Chip von oben mit einer Anpressscheibe auf die Kontaktscheibe gedrückt. Dabei werden die Federkontakte unter Spannung gesetzt, sodass eine leitende Verbindung zwischen der Leiterplatte unten unterhalb der Kontaktscheibe und dem von oben angedrückten Chip hergestellt wird. Für einen besseren Kontakt sind die Landepunkte der Kontaktstifte auf der Leiterplatte ebenfalls vergoldet. Es zeigt sich, dass das entworfene Pucksystem eine über den gesamten Temperaturbereich von 2 K - 400 K stabile elektrische Verbindung zum Chip gewährleistet. Da das PPMS nur 12 Signalleitungen zulässt, sind nicht alle Chipkontakte über die Platine tatsächlich angeschlossen. Dennoch können durch Rotation des Chips im Puck in Verbindung mit einem geeigneten Bondschema (siehe nächster Abschnitt) mehr Kontakte nutzbar gemacht werden. Von der Platine aus führen teflonummantelte Signalleitungen auf Kontaktflächen an der Unterseite des Standardpucks, welche wiederum mit dem Basisstecker verbunden sind. Beim Einsetzen des Pucks ins PPMS werden die Lamellen der Puckbasis zusammengedrückt und der Puck so arretiert. Da die Signalleitungen zwischen den Lamellen verlaufen, dient eine Teflonummantelung dem mechanischen Schutz der Drähte. Ein Sicherheitsbügel schließt den Puck nach oben hin ab. So kann sich der Puck im Probenraum des PPMS nicht verkanten und kann darüber hinaus im Notfall von oben aus dem Probenraum „herausgeangelt“ werden.

c) Probenaufbau

Die in dieser Arbeit hauptsächlich verwendete Teststruktur enthält 14 Einzeltransistoren gleichen Typs, aber unterschiedlicher Länge, welche über insgesamt 25 Kontaktflächen („Pads“) individuell angesteuert werden können. Dies ist möglich, da alle Transistoren auf einem gemeinsamen Substratkontakt liegen. Darüber hinaus teilen sich benachbarte Transistoren jeweils ihre Source- und Drainpads miteinander, wobei zwei verschiedene Gateleitungen zum Einsatz kommen. Im PPMS stehen 12 Probenanschlüsse zur Verfügung. Aufgrund der symmetrischen Ausführung des LCC Gehäuses ist es allerdings möglich, den Chip in

unterschiedlichen Orientierungen in den Puck einzusetzen. Über ein besonderes Bondschema konnten daher 17 Kontaktpads innerhalb einer Teststruktur zugänglich gemacht werden. Es standen somit auf jeder Probe nach Einbau in den Puck in der Standardorientierung „Konfiguration 1“ zunächst 5 Transistoren zur Verfügung. Durch Rotation der Probe um 90° im Halter in „Konfiguration 2“ konnten zusätzlich 4 weitere Transistoren angesprochen werden (Abb. 4.6).

Konfiguration 1

Konfiguration 2

Abbildung 4.6: Chipkonfigurationen im Probenhalter. Oben: Konfiguration mit 5 ansteuerbaren Transistoren, unten: nach Rotation des Chips um 90° mit 4 ansteuerbaren Transistoren. Der rote Kreis indiziert die Chiporientierung; jeweils kontaktierte Bonddrähte sind grün dargestellt.

4.2.2 Vorkehrungen zum Schutz vor elektrostatischer Entladung

Chips mit gebondeten Einzeltransistoren sind aufgrund fehlender Schutzschaltungen sehr empfindlich gegenüber elektrostatischer Entladung („electrostatic discharge“, ESD). Aufgrund der dünnen Gateoxide führen Potentialdifferenzen an den Anschlüssen von bereits wenigen Volt zu Degradation oder Ausfall der Transistoren. Da sich diese in den Teststrukturen aus Gründen der Platzersparnis verschiedene Anschlüsse miteinander teilen, führt die Degradation eines Transistors meist zum Ausfall des gesamten Chips. Der oben beschriebene Prozess der Probenerstellung über Laserschneiden und Bonden ist sehr zeitaufwändig und kostenintensiv, weshalb der Ausfall von Proben durch elektrostatische Schädigung unbedingt vermieden werden sollte. Es stellte zu Beginn der Arbeit eine große Herausforderung dar, ESD-sichere Prozeduren zu entwickeln. „Brennt“ ein Chip durch, so ist dies von außen nicht erkennbar, sondern wird während der elektrischen Messung im Kryostatsystem offenbar. An welchem Punkt in der Prozesskette eine Schädigung stattfand, lässt sich daher im Nachhinein nur schwer identifizieren. Der Schutz vor ESD muss während der gesamten Probenhandhabung gewährleistet sein. Prinzipielle Schutzmaßnahmen umfassen daher

- Potentialdefiniertes Bonden unter ESD Schutz
- Transport und Aufbewahrung der Proben in antistatischen ESD-Containern
- Handhabung der Proben außerhalb der ESD-Verpackung grundsätzlich nur mit durchgängiger Erdung der beteiligten Person, der verwendeten Werkzeuge und Messeinrichtungen. Dabei müssen sich aufbauende Potentiale hochohmig über einen 1 MΩ Widerstand abfließen können. Hierzu kommen unter anderem geerdete Bodenmatten, ESD Schuhe, Erdungsarmbänder, leitfähige Handschuhe und geerdete Spezialpinzetten mit hochohmiger Spitze zum Einsatz.

4.2.3 Messablauf

Zur Messung im PPMS wird ein gebondeter Chip unter ESD Schutz aus seiner antistatischen Verpackung entnommen und in der gewünschten Orientierung auf den leeren Puck gelegt. Dieser steckt dabei seinerseits in einer Tischkonsole, über die sämtliche Probenkontakte auf Masse gelegt sind. Somit liegen beim Einbau des Chips in den Puck die Federkontakte bereits vor der Berührung mit dem Chip definiert auf Masse. Anschließend wird der Chip mit der Arretierscheibe angepresst und der Puckaufbau mit dem Sicherheitsbügel abgeschlossen. Nun erfolgt der Einbau des Pucks in das PPMS System, weiterhin unter ESD Schutz und mit Hilfe eines speziellen Einbringstabs. Der Gegenstecker im PPMS („Pucksockel“) wird ebenfalls auf allen Kontakten geerdet, bevor der Puck eingesetzt wird. Die Zuweisung der 12 Probenkontakte im PPMS wahlweise auf Masse oder an einen der Kanäle des Messgeräts erfolgt über eine Switch Matrix, welche sich zwischen Messgerät und PPMS befindet (Abb. 4.7). Um das PPMS mit der Switch

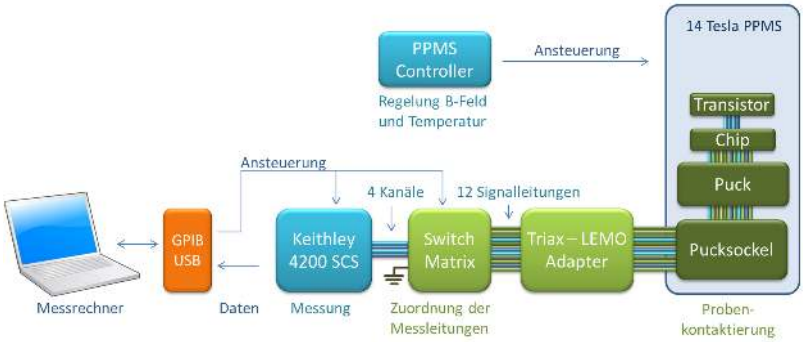


Abbildung 4.7: Verdrahtungsschema des Messaufbaus: Für das PPMS steht ein eigener Ansteuerrechner zur Verfügung. Die elektrische Messung wird über ein Keithley 4200 SCS mit 4 SMU's durchgeführt. Über eine Switch Matrix erfolgt die Zuordnung der Messleitungen zum Transistor, ungenutzte Chipkontakte werden auf Masse gelegt.

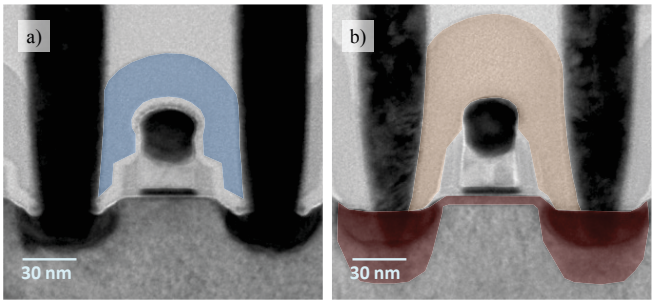
Matrix verbinden zu können, wurde ein geschirmter Adapter von LEMO auf 12 Triax-Signalleitungen entworfen. Nach Einbau des Pucks ins PPMS wird der Probenraum versiegelt und evakuiert, um der Bildung von Kondenswasser oder Eis auf der Probe vorzubeugen. Sowohl die Ansteuerung der Switch Matrix, also auch des Messgeräts erfolgt PC gestützt über den GPIB-Bus. Durch die Verwendung der Switch Matrix kann ein beliebiger (gebondeter) Transistor auf der Probe per Mausklick ausgewählt und angesteuert werden, ohne dass physisch Kabel umgesteckt werden müssen. Dadurch verringert sich das Risiko von ESD Schädigungen weiter. Darüber hinaus vereinfacht dieses Verfahren die Messung, vermeidet versehentliche Fehlzuweisungen von Kontakten und beschleunigt den Messablauf. Die Signalleitungen laufen vollständig geschirmt bis zum PPMS, wodurch die Messungen sehr rauscharm werden. Eine Beschreibung der Messroutine wird in Abschnitt 3.1.3 gegeben. Der finale Messaufbau am HZDR ist beispielhaft in Abb. 4.3 dargestellt.

Der Probenausbau im Anschluss an die Messung erfolgt analog zur Einbauprozedur und ebenfalls unter ESD Schutz. Durch elektrische Messungen der Transistoren auf Waferebene, nach dem Laserschneiden, nach dem Bonden und im PPMS konnte verifiziert werden, dass der hier beschriebene Prozess der Probenerstellung und -handhabung keinerlei elektrische Degradation verursacht.

5 Ergebnisse der Untersuchungen an CMOS-Feldeffekt-Transistoren

5.1 Struktureller Aufbau

Die in dieser Arbeit untersuchten MOSFETs wurden in einer 28 nm Technologie bei GLOBALFOUNDRIES Dresden [80] auf Entwicklungswafern mit verschiedenen Teststrukturen gefertigt. Die verwendeten Silizium-Substrate haben die (100) Kristallorientierung, wobei die Kanalorientierung der Transistoren entlang der <110> Richtung liegt. „High-k metal-gate“ (HKMG) n- und pMOS Transistoren verwenden ein Hafnium-basiertes (Hf) „high-k“ Gatedielektrikum mit $1,2$ nm bzw. $1,4$ nm EOT (Abschn. 5.2.2). Der pFET nutzt epitaktisches Silizium-Germanium (eSiGe) in Source und Drain zur kompressiven Verspannung des Transistorkanals, sowie im Kanalbereich zur Einstellung der Einsatzspannung und zur Erhöhung der intrinsischen Ladungsträgerbeweglichkeit. Darüber hinaus werden im Rahmen eines „Dual-Stress-Liner“-Integrationsschemas [81] zug- (TPEN) und druckverspannte (CPEN) Deckschichten über den n- bzw. p-Transistorgates verwendet, um die Ladungsträgerbeweglichkeit innerhalb des Kanals zu erhöhen (Kap. 2.1.3). Abbildung 5.1 zeigt beispielhaft Transmissions-Elektronen-Mikroskopieaufnahmen (TEM) des 30 nm nFETs und pFETs mit eingefärbten Verspannungsschichten. Alle hier untersuchten CMOS Transistoren haben eine Weite von $W = 1\,\mu\text{m}$ und verschiedene Gatelängen L. Die Chips wurden bis zur Passivierungsschicht nach der letzten Metallisierung prozessiert. Im Folgenden ist zu beachten, dass die Nomenklatur der Gatelängen von ei-



TEM: Alexander Würfel/ Viola Löscher, GLOBALFOUNDRIES

Abbildung 5.1: TEM Aufnahmen von einem a) nMOS Transistor mit tensiler TPEN Verspannungsschicht (blau) und b) pMOS Transistor mit kompressiv verspannendem CPEN (orange), sowie SiGe in S/D und im Kanal (rot), mit $L = 30\,\text{nm}$. Der Schnitt erfolgte so, dass die Kontakte von der ersten Metallisierungsebene auf die S/D-Gebiete sichtbar sind; die Gate-Metallelektroden auf dem high-k-Dielektrikum sind als dunkler, horizontaler Balken über dem Kanalbereich zu erkennen.

nem vorherigen Technologieknoten übernommen wurde. Die tatsächlich strukturierten Gatelängen ergeben sich über die Skalierungsformel $L_{\text{design}} = (L \cdot 0,9) + 3\,\text{nm}$, für die Weiten gilt $W_{\text{design}} = W \cdot 0,9$ (Tabelle 5.1). Dies betrifft nur die Bezeichnungen der Transistoren; in alle Berechnungen und Grafiken flossen die tatsächlichen Geometrien ein. Darüber hinaus kann die elektrisch effektive Kanallänge, z.B. durch Unterdiffusion der Source/Drain-Gebiete unter das Gate, verkürzt sein, gemäß $L_{\text{eff}} = L_{\text{design}} - \delta L$. Mit $\delta L = 6\,\text{nm}$ ergibt sich für den $L = 60\,\text{nm}$ Transistor beispielsweise eine effektive Länge von 51 nm: $L_{\text{eff}(60\,\text{nm})} = L_{\text{design}} - \delta L = L \cdot 0,9 + 3\,\text{nm} - \delta L = 51\,\text{nm}$. Der $L = 1\,\mu\text{m}$ Transistor in Abb. 5.23 ist somit bei der tatsächlichen, auf dem Wafer strukturierten Länge von $L_{\text{design}} = 903\,\text{nm}$ eingetragen. Der zugehörige Beweglichkeitswert wurde mit der elektrisch effektiven Kanallänge von $L_{\text{eff}} = L_{\text{design}} - \delta L = 897\,\text{nm}$ berechnet.

Tabelle 5.1: Beziehung zwischen Transistorbezeichnung L und tatsächlich strukturierter Gatelänge L_{design}, alle Maße in nm, die Werte sind ganzzahlig gerundet.

L	30	34	38	42	60	90	120	170	300
L_{design}	30	34	37	41	57	84	111	156	273
W_{design}	900	900	900	900	900	900	900	900	900

L	500	1000	2000
L_{design}	453	903	1803
W_{design}	900	900	900

5.2 Schaltverhalten

5.2.1 Transfer- und Ausgangskennlinien

Abbildung 5.2 und Abb. 5.3 zeigen exemplarisch Transfer- und Ausgangskennlinien von tensil verspannten nFET Transistoren für ausgewählte Gatelängen L. Dabei liegt der Strom des 1 µm Langkanaltransistors beinahe eine Größenordnung unterhalb des 30 nm Kurzkanaltransistors, da mit steigender Gatelänge der Stromfluss zunehmend durch den anwachsenden Kanalwiderstand begrenzt wird. Während der 1 µm Langkanaltransistor oberhalb von ca. $V_{\text{ds}} = 500\,\text{mV}$ beinahe vollständig sättigt, steigt der Strom des 30 nm Kurzkanaltransistors weiterhin linear an. Mit immer höherer Drainspannung wird die Kanalladung drainseitig immer geringer, bis der Kanal am Sättigungspunkt bei $V_{\text{ds}} = V_{\text{gs}} - V_{\text{th}}$ schließlich abschnürt. Mit weiterer Erhöhung von V_{ds} verschiebt sich nur der Abschnürpunkt in Richtung Source, sodass im Wesentlichen keine weitere Erhöhung des Drainstroms stattfindet. Allerdings geht mit der Verschiebung des Abschnürpunkts eine Verkürzung der effektiven Kanallänge L_{eff} und damit eine Verminderung des Kanalwiderstands einher. Dieser Effekt wird „Kanallängenmodu-

lation“ genannt und bewirkt einen Anstieg des Drainstroms jenseits des Sättigungspunkts. Während die leichte Reduktion von L_{eff} mit V_{ds} bei Langkanaltransistoren kaum ins Gewicht fällt, wird der Effekt bei kurzen Gatelängen relevant, siehe Abb. 5.3. In Bezug auf die in die-

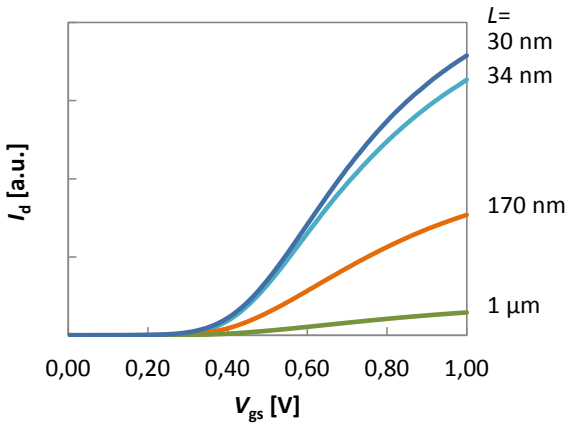


Abbildung 5.2: Transferkennlinien $I_{\mathrm{d}}(V_{\mathrm{gs}})$ für verschiedene Gatelängen L, $V_{\mathrm{ds}} = 50\,\mathrm{mV}$.

ser Arbeit zum Tragen kommenden Messbedingungen ist anzumerken, dass die nFETs stets bei positiven, die pFETs dagegen bei negativen Gate-Source- und Drain-Source-Spannungen betrieben wurden. Es gilt also $V_{\mathrm{gs}}, V_{\mathrm{ds}} \geq 0\,\mathrm{V}$ für alle nFET Messungen und $V_{\mathrm{gs}}, V_{\mathrm{ds}} \leq 0\,\mathrm{V}$ für die pFETs. Die einzige Ausnahme bildet hier die vollständige Darstellung der Kapazitäts-Spannungskurve in Abb. 5.4. Der Sperrbereich, in dem der gateinduzierte Drainleckstrom (GIDL) relevant wird, wurde in dieser Arbeit nicht untersucht. Aus Gründen der Lesbarkeit, bzw. der besseren Vergleichbarkeit mit nFET Messdaten werden daher im Folgenden für Spannungswerte an pFETs die Betragswerte verwendet. Bei Messungen an pFETs, die z. B. mit $V_{\mathrm{gs}} = 1\,\mathrm{V}, V_{\mathrm{ds}} = 50\,\mathrm{mV}$ gekennzeichnet sind, lagen also tatsächlich $V_{\mathrm{gs}} = -1\,\mathrm{V}$ und $V_{\mathrm{ds}} = -50\,\mathrm{mV}$ an. Zur Überprüfung der Transistorintegrität wurden grundsätzlich die

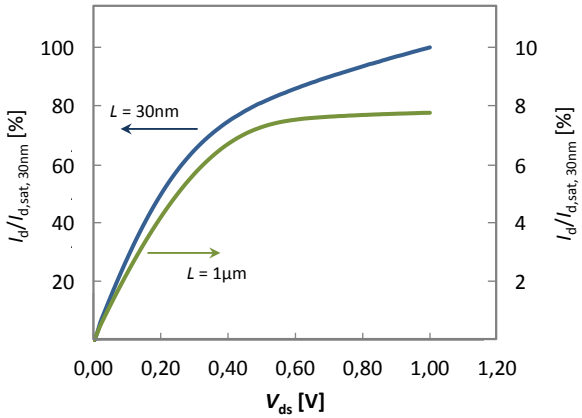


Abbildung 5.3: Ausgangskennlinien $I_\mathrm{d}(V_\mathrm{ds})$ des $L = 30\,\mathrm{nm}$ und $L = 1\,\mu\mathrm{m}$ nFET bei $V_\mathrm{gs} = 1\,\mathrm{V}$, $T = 25\,°\mathrm{C}$ relativ zum Sättigungsstrom des 30 nm Transistors. Während der Langkanaltransistor nahezu ideal in Stromsättigung geht, zeigt der Kurzkanaltransistor auch oberhalb der Sättigungsspannung einen starken linearen Anstieg des Drainstroms aufgrund von Kanallängenmodulation.

Ströme an Drain-, Source-, Substrat- und Gatekontakt aufgezeichnet. Die Gate- und Substratleckströme lagen bei den verwendeten Spannungsbedingungen für den $L = 1\,\mu\mathrm{m}$ Transistor maximal im unteren Nanoamperebereich, während die Leckströme der Kurzkanaltransistoren 1-2 Größenordnungen unter denen der Langkanaltransistoren lagen. Ferner lag oberhalb der Einsatzspannung die Differenz zwischen den an Drain ab- und an Source zufließenden Strömen mehr als 4 Größenordnungen unter dem Absolutwert des Drainstroms ($I_\mathrm{d} - I_\mathrm{s} < I_\mathrm{d}/10^4$) und damit im Bereich der Messunschärfe der verwendeten SMUs. Da der Sperrbereich in dieser Arbeit nicht untersucht wurde, konnten im weiteren Verlauf Gate- und Substratleckströme vernachlässigt werden. Der am Drainkontakt gemessene Strom I_d wird im Folgenden daher mit dem von Drain nach Source fließenden Strom I_ds synonym verwendet.

5.2.2 Kapazitäts-Spannungskurven

Zur Charakterisierung des MOS-Gateschichtstapels wurde die Kapazität großer Transistorstrukturen mit $L_{\text{design}} = 29,7\,\mu\text{m}$ und $W_{\text{design}} = 2 \times 29,7\,\mu\text{m}$ in Abhängigkeit der Gatespannung gemessen. Aus den Inversionskapazitäten bei $V_{\text{gs}} = 1\,\text{V}$ (nFET) bzw. $V_{\text{gs}} = -1\,\text{V}$ (pFET) ergaben sich die äquivalenten Gateoxiddicken zu EOT $= 1,2\,\text{nm}$ (nFET, TPEN) und $1,4\,\text{nm}$ (pFET, CPEN). Aus dem Kapazitätsverhalten

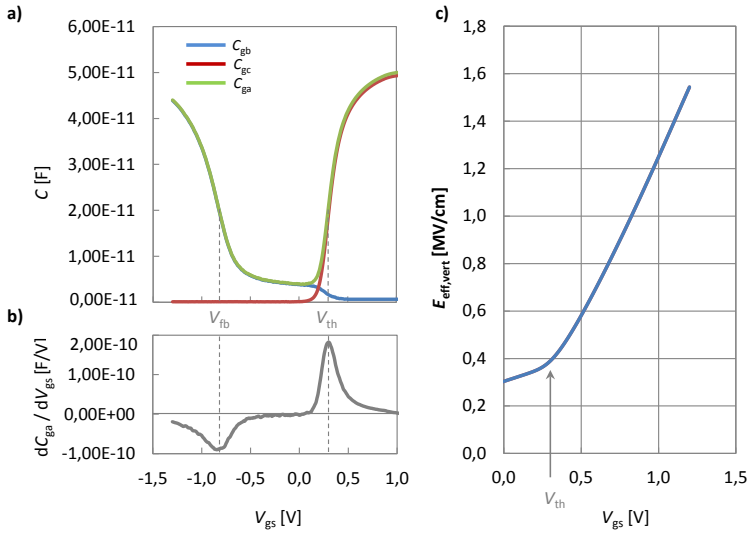


Abbildung 5.4: Messungen an einer nFET-artigen Kapazitätsstruktur mit TPEN Deckschicht und $L = 29,7\,\mu\text{m}$ sowie $W = 2 \times 29,7\,\mu\text{m}$. a) Kapazitäts-Gatespannungsverhalten bei $f = 100\,\text{kHz}$, $V_{\text{ac}} = 30\,\text{mV}$ (RMS). Aus der Inversionskapazität bei $V_{\text{gs}} = 1\,\text{V}$ von $5 \cdot 10^{-11}\,\text{F}$ ergibt sich der flächenspezifische Wert von $2,83 \cdot 10^{-2}\,\text{F/m}^2$, entsprechend $1,2\,\text{nm}$ EOT. b) Erste Ableitung von $C_{\text{ga}}(V_{\text{gs}})$ zur Ermittlung von Flachband- und Einsatzspannung, c) effektives, vertikales elektrisches Feld $E_{\text{eff,vert}}$.

konnte nach (3.10) das effektive, vertikale elektrische Feld $E_{\mathrm{eff,vert}}$ berechnet werden. Das Resultat ist in Abb. 5.4c dargestellt. Mit Einsetzen der Inversion bei $V_{\mathrm{gs}} = V_{\mathrm{th}}$ steigt die vertikale Feldstärke rapide an und nimmt linear mit V_{gs} zu. Für den nFET wird bei $V_{\mathrm{gs}} = 1\,\mathrm{V}$ ein vertikales Feld von $1,25\,\mathrm{MV/cm}$ erreicht.

5.2.3 Elektrische Bestimmung des Serienwiderstands

Zur Ermittlung der Anschlusswiderstände R_{ser} wurden die Gesamtwiderstände der Transistoren mit den drei kürzesten Gatelängen $L = 30\,\mathrm{nm}$, $34\,\mathrm{nm}$ und $38\,\mathrm{nm}$ bei konstantem Overdrive mit $V_{\mathrm{gs}} = V_{\mathrm{th}} + 0,6\,\mathrm{V}$ gemessen. Die Drain-Source-Spannung betrug dabei $V_{\mathrm{ds}} = 50\,\mathrm{mV}$ (pFET Werte negativ). Die Extrapolation $R_{\mathrm{ges}}(L)$ für $L \to 0$ erfolgte unter Berücksichtigung einer Unterdiffusion der Anschlussgebiete um $\delta L = 6\,\mathrm{nm}$ (je $3\,\mathrm{nm}$ an Source und Drain), welche aus der direkten Überlappkapazität zwischen Gate und den Source/Drain-Extensionanschlüssen abgeschätzt wurde. Für die hier verwendete Extraktion sollte R_{ser} einen signifikanten Anteil am Gesamtwiderstand ausmachen – Langkanaltransistoren kommen somit nicht in Frage, da bei diesen R_{ges} vollständig durch den Kanalwiderstand dominiert wird. Des Weiteren wird in der Methode angenommen, dass eine lineare Extrapolation auf $L = 0$ zulässig ist. Voraussetzung hierfür ist neben einer konstanten Flächenladungsdichte C'_{ox} vor allem, dass die verwendeten Transistoren die gleiche Beweglichkeit aufweisen, siehe Kapitel 3.2. Aufgrund des Beweglichkeits-roll-offs können daher nur Transistoren mit ähnlichen Gatelängen verwendet werden. Je enger diese beieinander liegen, desto sensibler reagiert allerdings die Steigung der Extrapolationsgeraden und damit R_{ser} auf kleine Schwankungen der Datenpunkte. Aus diesen Gründen dürfen die hier extrahierten Serienwiderstandswerte nur als ungefähre Richtwerte verstanden werden. Trotz seiner Unzulänglichkeiten bietet das $R_{\mathrm{ges}}(L)$-Verfahren einen Zugang zu den elektrisch effektiven Anschlusswiderständen, welche am vollständig prozessierten Transistor auch mit anderen Methoden grundsätzlich schwer zu bestimmen sind. Zwar können einzelne Beiträge wie zum Beispiel der Schichtwiderstand der Diffusionsgebiete oder die Widerstände ein-

zelner Kontaktstöpsel prinzipiell auch außerhalb des Transistors mit Hilfe spezieller Teststrukturen charakterisiert werden; wesentliche Beiträge zu R_{ser} entstehen allerdings im Bereich des Kanalanschlusses, welcher sich nur am integrierten MOSFET formiert und daher nicht isoliert zugänglich gemacht werden kann.

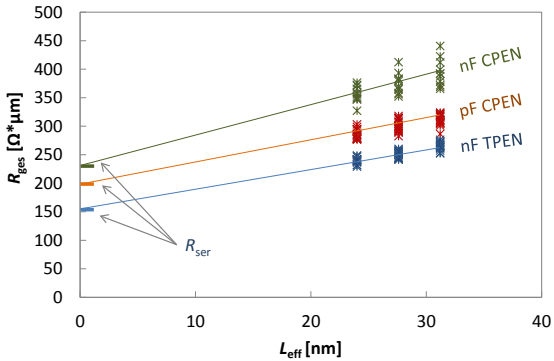


Abbildung 5.5: Bestimmung von R_{ser} durch Extrapolation der Mittelwerte der Verteilungen $R_{\mathrm{ges}}(L_{\mathrm{eff}})$ zu $L = 0$ für verschiedene Transistortypen bei festem Overdrive und $V_{\mathrm{ds}} = 50\,\mathrm{mV}$.

Wie in Abb. 5.5 ersichtlich, unterliegen die Widerstände R_{ges} verschiedener Transistoren gleicher Designlänge trotz der Overdrive-Normierung noch einer deutlichen Streuung. Dies ist verständlich, wenn man bedenkt, dass in den zusammengesetzten Parameter R_{ges} fast alle Prozessvariationen eingehen: Lithographische Schwankungen, Ätzvariationen, lokale Unterschiede in den Metall-Halbleiter-Übergängen, Implantationsschwankungen, Grenzflächenvariationen, siehe Abb. 5.9. Es ist daher ersichtlich, dass eine Bestimmung von R_{ser} anhand einiger ausgewählter Chips leicht fehlerhaft sein kann. Deshalb wurde die Extrapolation über eine große Statistik mit bis zu 20 Punkten für die jeweiligen Transistortypen durchgeführt. Um die Extrapolationsgeraden aus Abb. 5.5 zu verifizieren, wurde ein Abgleich mit

TCAD Prozess- und Bauelementesimulationen und Erfahrungswerten aus technologisch vergleichbaren Losen durchgeführt. Für die untersuchten Technologieoptionen ergaben sich somit die Werte aus Abb. 5.5 von $R_{\mathrm{ser}} = 150\,\Omega\cdot\mu\mathrm{m}$ (nFET, TPEN), $200\,\Omega\cdot\mu\mathrm{m}$ (pFET, CPEN) und $230\,\Omega\cdot\mu\mathrm{m}$ (nFET, CPEN).

5.3 Ladungsträgerbeweglichkeit

5.3.1 Magnetowiderstands-Methode

Bei den in dieser Arbeit mittels des MR Verfahrens untersuchten n- und pMOS Transistoren handelt es sich um Kurzkanaltransistoren der Längen $L = 30\,\mathrm{nm}$, $34\,\mathrm{nm}$, $38\,\mathrm{nm}$ und $42\,\mathrm{nm}$ sowie um eine Langkanalgeometrie mit $L = 1\,\mu\mathrm{m}$, jeweils mit einer konstanten Weite von $W = 1\,\mu\mathrm{m}$. Für die tatsächlichen Maße hinter diesen Bezeichnungen siehe Tabelle 5.1. Bei dieser für den Prozessormarkt entwickelten, hochintegrierten 28 nm-Technologie liegen Chip- zu Chip Variationen wichtiger Transistorkenngrößen wie z. B. Einsatzspannung, Sättigungsstrom und Unterschwellsteigung innerhalb eines Wafers bereits in einer sehr engen Verteilung. Um die Vergleichbarkeit noch weiter zu erhöhen, stammen die untersuchten Transistoren vom selben Chip und wurden damit in direkter Nähe zueinander prozessiert. Dies wurde durch das in Kap. 4.2.1 unter „Probenaufbau“ beschriebene, optimierte Bondschema ermöglicht. Es ist anzumerken, dass aus den Magnetowiderstandsdaten in Kap. 5.3.1e) folgt, dass die Schwankungen der Anschlusswiderstände R_{ser} zwischen Transistoren mit verschiedenem L innerhalb des untersuchten Chips nicht größer als $5\,\Omega\cdot\mu\mathrm{m}$ sind. Im Vergleich zur Streuung der Daten in Abb. 5.5 erscheint diese Variation nur sehr klein. Dies ist darauf zurückzuführen, dass sich die untersuchten Transistoren in direkter Nähe zueinander auf dem Wafer befinden und lokal innerhalb eines kleinen Chipbereichs eine hohe Prozesshomogenität während der Fertigung vorherrscht. Transistortypen mit unterschiedlicher Gateverspannung (tensil/ kompressiv) stammen aufgrund des speziellen Versuchssplits dabei von unterschiedlichen Wafern und wurden jeweils von im Zentrum liegenden Chipsegmenten entnommen. Aufgrund des hohen

präparativen und messtechnischen Aufwands erschien eine statistische Auswertung des gesamten Wafers nicht sinnvoll. Um dennoch aussagekräftige Vergleiche zwischen verschiedenen Wafern ziehen zu können wurden Einzelchips ausgewählt, die in ihrem Schaltverhalten hinsichtlich Einsatzspannung, Unterschwellsteigung und Sättigungsstrom bestmöglich dem jeweiligen Wafermittel entsprachen.

a) Signalstärke

Die Stärke des Magnetowiderstandseffekts skaliert in erster Näherung mit $(R_B/R_0) \approx \mu^2 B^2 + 1$, vergleiche (3.11) für ein ideales System. Bei Beweglichkeiten im Bereich um $\mu \approx 100\,\mathrm{cm^2/Vs}$ und Feldstärken von $B = 14\,\mathrm{T}$ liegt die zu erwartende Widerstandsänderung damit bei knapp zwei Prozent und sie fällt auf ein halbes Prozent für Beweglichkeiten um $50\,\mathrm{cm^2/Vs}$. Theoretisch wäre es zwar möglich, die Beweglichkeit des Transistors an jedem Operationspunkt durch Messung der Ausgangs- und Transferkennlinien mit und ohne Magnetfeld zu bestimmen. Dieser einfache und zeiteffiziente Ansatz erweist sich jedoch unter den genannten Rahmenbedingungen aufgrund des teilweise sehr schlechten Rausch- zu Signalverhältnisses von bis zu $I_{\mathrm{Noise}}/I_{\mathrm{Signal}} = \delta I_{\mathrm{d}}/\Delta I_{\mathrm{d}} = 50\,\%$ (siehe Tab. 5.2) als nicht zielführend. Dies veranschaulicht Abb. 5.6a am Beispiel eines 30 nm nFET Transistors.

Misst man den Transistorstrom z.B. bei $V_{\mathrm{gs}} = 1\,\mathrm{V}$ über der Zeit und vergrößert die Skalierung auf der y-Achse genügend, wird die Trennung zwischen den Messungen bei $B = 0\,\mathrm{T}$ und $14\,\mathrm{T}$ zwar sichtbar (Abb. 5.6b,c). Bei der Aufnahme einer I-V-Kennlinie wird jedoch bei jedem Spannungspunkt, z. B. alle $\Delta V_{\mathrm{gs}} = 10\,\mathrm{mV}$ nur eine Strommessung durchgeführt. Dabei ergibt sich jeweils zufällig einer der innerhalb des Messrauschens δI_{d} möglichen Werte. Bezogen auf die geringe Gesamtänderung des Drainstroms im Magnetfeld ΔI_{d} unterliegt eine Berechnung der Beweglichkeit basierend auf zwei Einzelwerten bei $B = 0\,\mathrm{T}$ und $14\,\mathrm{T}$ daher einer sehr großen Unsicherheit. Der Absolutstrom ist bei $V_{\mathrm{ds}} = 1\,\mathrm{V}$ in Abb. 5.6b höher als in Abb. 5.6c bei $V_{\mathrm{ds}} = 50\,\mathrm{mV}$, aber das Signal- zu Rausch-Verhältnis ist deutlich schlechter. In Abb. 5.6c bewirkt die höhere Beweglichkeit bei kleinerem V_{ds} eine größere rela-

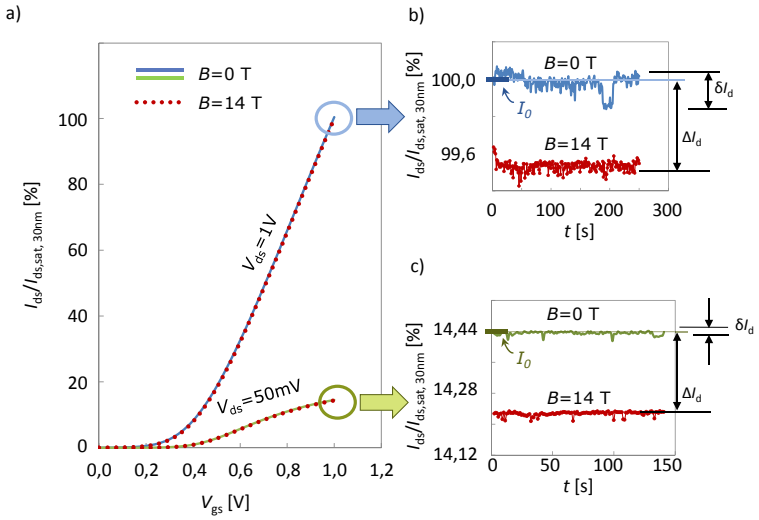


Abbildung 5.6: Einfluss des Magnetfelds auf den 30 nm nFET: a) Transferkennlinien $I_\mathrm{d}(V_\mathrm{gs})$ bei $B = 0\,\mathrm{T}$ und $B = 14\,\mathrm{T}$ können auf einer Skala, die die Kennlinie im Gesamten zeigt, kaum auseinandergehalten werden. b) und c) Strom-Zeitdiagramme bei $V_\mathrm{gs} = 1\,\mathrm{V}$ mit $V_\mathrm{ds} = 1\,\mathrm{V}$ bzw. $V_\mathrm{ds} = 50\,\mathrm{mV}$. Stromwerte sind relativ zum Sättigungswert des 30 nm-Transistors bei $B = 0\,\mathrm{T}$ angegeben.

tive Stromänderung durch das Magnetfeld. Obwohl das Rauschen des Transistorstroms z.B. im Fall des 30 nm nFET bei $V_\mathrm{ds} = 1\,\mathrm{V}$ (vgl. Abb. 5.6b und Tabelle 5.2) bei nur $0,25\,\%$ des Gesamtniveaus liegt, beläuft sich die Unsicherheit im zur Beweglichkeitsbestimmung verwendeten Stromunterschied auf $50\,\%$, da die magnetfeldbedingte Stromänderung in diesem Fall gerade mal das doppelte des Signalrauschens beträgt. Um dieses Problem zu lösen und die Genauigkeit zu erhöhen, wurde die Berechnung der Beweglichkeit über die Magnetfeldabhängigkeit des Transistorstroms untersucht.

Tabelle 5.2: Typische Signalstärken am TPEN nFET, $V_{\mathrm{gs}} = 1\,\mathrm{V}$, $T = 25°\mathrm{C}$. Das Verhältnis von Rauschen zu Signalstärke, die „Unsicherheit“ $\delta I_{\mathrm{d}}/\Delta I_{\mathrm{d}}$ führt zu den in Tabelle 5.3 angegeben Unschärfen $\Delta\mu$ in der Beweglichkeitsbestimmung über die Berechnung mit Einzelwerten.

V_{ds}	L	Rauschen $\delta I_{\mathrm{d}}/I_0$ [%]	Unsicherheit $\delta I_{\mathrm{d}}/\Delta I_{\mathrm{d}}$ [%]
50 mV	1 µm	0,03	0,8
50 mV	30 nm	0,23	16,0
1 V	1 µm	0,08	2,6
1 V	30 nm	0,25	50,0

Aus (5.8) folgt, dass

$$R_B/R_0 = 1 + \mu^2 B^2 \alpha, \quad \text{mit } \alpha = k_{\mathrm{gm}}(1 - R_{\mathrm{ser}}/R_0)\ . \tag{5.1}$$

Erwartungsgemäß sollte also das Verhältnis von R_B/R_0 aufgetragen über B^2 eine lineare Abhängigkeit zeigen. Das eindeutig lineare Verhalten von R_B/R_0 über B^2 in Abb. 5.7 bestätigt, dass sich der Transistorstrom im Magnetfeld tatsächlich entsprechend der MR Theorie entwickelt. Gut sichtbar ist auch die systematische Zunahme der Steigung mit höherem L, entsprechend einer höheren Ladungsträgerbeweglichkeit. Obwohl das Aspektverhältnis des Langkanaltransistors mit $L/W = 1$ ungünstig für den Magnetowiderstandseffekt ist, weist dieser das geringste Signalrauschen auf. Dies liegt daran, dass der Beweglichkeitswert quadratisch in den Signalunterschied eingeht und der Langkanaltransistor im gesamten Operationsbereich mit Abstand die höchste Beweglichkeit aufweist. Wie in Abb. 5.7 ersichtlich, können die MR-Signale der Kurzkanaltransistoren erst bei Feldstärken ab ca. 9 T klar voneinander getrennt werden. Für stärkere Felder steigt die Genauigkeit, mit der der Fit durchgeführt werden kann rasch an, da sich der Strom quadratisch mit dem Magnetfeld ändert.

Der Vorteil dieser Herangehensweise liegt darin, dass der gesamte

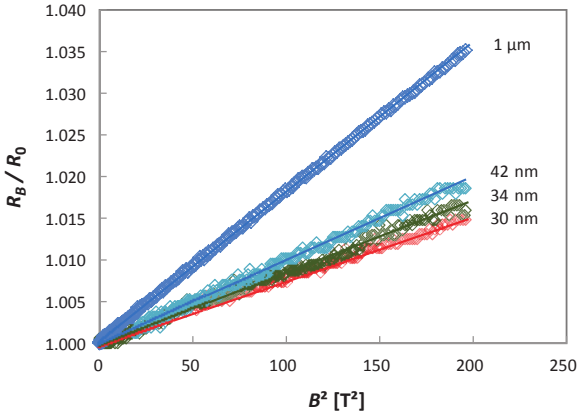


Abbildung 5.7: nFET Widerstandsverhältnis R_B/R_0 in Abhängigkeit von B^2 für verschiedene Gatelängen L bei $V_{\mathrm{gs}} = 1\,\mathrm{V}$, $V_{\mathrm{ds}} = 50\,\mathrm{mV}, T = 25\,°\mathrm{C}$.

Verlauf von $I_{\mathrm{d}}(B)$ für die Berechnung der Beweglichkeit herangezogen wird. Prinzipiell gilt, je höher die Feldstärke ist, desto klarer tritt das parabolische Verhalten hervor. Die hohe Statistik der vielen Einzelmessungen führt dabei zu einer enormen Verbesserung der Genauigkeit bezüglich des Grundrauschens. Darüber hinaus können Bereiche, die eine Drift oder Sprünge im Messsignal aufweisen leicht identifiziert und aus den zu fittenden Daten ausgenommen werden (siehe Abb. A.4 und Abb. 5.8), da sie vom erwarteten parabolischen Verhalten abweichen. Insbesondere treten diskrete Sprünge im zeitlichen Verlauf des Drainstroms auf, die wahrscheinlich auf die in [82] an MOSFETs untersuchte Signatur von „Random Telegraph Noise“ (RTN) zurückzuführen sind. Auch wenn die genauen Mechanismen von RTN noch nicht vollständig verstanden sind, handelt es sich hierbei um Umladungseffekte wie z. B. dem Einfang und die spätere Freigabe einzelner Elektronen an Gateoxid-Störstellen. Derartige Prozesse führen zu leichten Verschiebungen der Einsatzspannung und haben somit sprungartige Änderungen des Drainstroms zur Folge. Sie sind diskret und reversibel, wo-

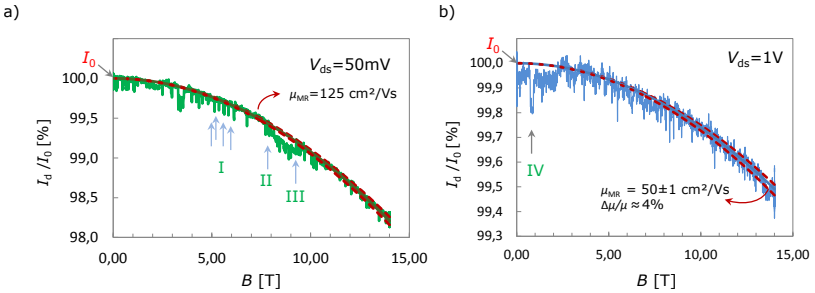


Abbildung 5.8: Bestimmung der Beweglichkeit durch Fit der $I_d(B)$-Charakteristik am Beispiel eines 30 nm nFET Transistors bei $V_{gs} = 1$V und a) $V_{ds} = 50$ mV mit gut sichtbarer RTN Signatur (Beispielpunkte I–III, siehe Haupttext), b) bei $V_{ds} = 1$ V. Durch eine vorangegangene, längere Nullfeldmessung konnte der Wert des mittleren Nullfeldstroms I_0 bestimmt werden.

bei die Lebensdauer eines solchen Zustands stark variieren kann. So gibt es in Abb. 5.8a sehr kurze I_d-Ausschläge, wie beispielhaft in Bereich I markiert. Daneben sind allerdings auch länger anhaltende Änderungen sichtbar, z. B. zwischen den Punkten II und III sowie bei IV in Abb. 5.8b. Bei II bleibt der Strom auf dem unteren Niveau der sonst meist kurzen Ausschläge liegen, was zu einer Parallelverschiebung der $I_d(B)$-Charakteristik führt. Von diesem Niveau aus finden dann weitere, kurzlebige Umladungsprozesse statt, bis der Stromfluss in III schließlich auf seine ursprüngliche Grundlinie zurückkehrt. Aus der Transfercharakteristik dieses Transistors bei $V_{ds} = 50$ mV ergibt sich, dass die Stromverschiebung zwischen II und III einer Erhöhung der Einsatzspannung um etwa $\Delta V_{th} = 1,4$ mV entspricht. Der in Abb. 5.8b für $V_{ds} = 1$ V registrierte, mit IV markierte Sprung entspricht ebenfalls einer V_{th}-Verschiebung dieser Größe. Dies legt nahe, dass in beiden Fällen höchstwahrscheinlich die gleichen mikroskopischen Umladungsprozesse beobachtet werden. Dank der Aufnahme des gesamten $I_d(B)$-Verlaufs können solche RTN-bedingten Sprünge, aber auch un-

ter Umständen vorhandene Driften im Messignal leicht identifiziert und in der Auswertung berücksichtigt werden. So beeinträchtigen derartige Störungen, aber auch das Grundrauschen des Messignals die Bestimmung der Beweglichkeit deutlich schwächer – im starken Gegensatz zu dem Verfahren, μ_{MR} aus zwei Punkten bei $I_{\mathrm{d}}(B = 0)$ und $I_{\mathrm{d}}(B = B_{\mathrm{max}})$ zu bestimmen. Die Messkurven bei höherem V_{ds} haben aufgrund der niedrigeren Beweglichkeit relativ zum Absolutstrom gesehen den kleineren Hub in $I_{\mathrm{d}}(B)$ und erscheinen somit am stärksten verrauscht (Abb. 5.8b), insbesondere im Fall der 30 nm Kurzkanaltransistoren. Hier konnte die Messunsicherheit durch das verbesserte Verfahren z. B. am nFET auf $1/10$ reduziert werden, entsprechend einer Restunsicherheit von $\Delta\mu = 2\,\mathrm{cm}^2/\mathrm{Vs}$ (Tabelle 5.3). Aufgrund der ungefähr um die Hälfte reduzierten Beweglichkeiten sind die Unsicherheiten am pFET etwa doppelt so hoch wie beim nFET.

Tabelle 5.3: Verringerung der Unsicherheit $\Delta\mu_{\mathrm{MR}}$ in der Beweglichkeitsbestimmung am Beispiel des TPEN nFET durch Fit der Feldabhängigkeit des Transistorstroms $I(B)$, anstelle der Berechnung über zwei einzelne Stromwerte I_{d} bei $B = 0$ und $B = 14\,\mathrm{T}$. $V_{\mathrm{g}} = 1\,\mathrm{V}$, $T = 25\,^{\circ}\mathrm{C}$

		Zweiwert-Methode	Berechnung über Fit $I(B)$	
V_{ds}	L	$\Delta\mu_{\mathrm{MR}}/\mu_{\mathrm{MR}}$	$\Delta\mu_{\mathrm{MR}}/\mu_{\mathrm{MR}}$	$\Delta\mu_{\mathrm{MR}}$
50 mV	1 µm	4,6 %	0,2 %	0,4 $^{\mathrm{cm}^2}/_{\mathrm{Vs}}$
50 mV	30 nm	16,3 %	1,4 %	1,7 $^{\mathrm{cm}^2}/_{\mathrm{Vs}}$
1 V	1 µm	10,9 %	0,3 %	0,5 $^{\mathrm{cm}^2}/_{\mathrm{Vs}}$
1 V	30 nm	**44,9 %**	**4,2 %**	2,0 $^{\mathrm{cm}^2}/_{\mathrm{Vs}}$

b) Geometriekorrektur

Falls die Probengeometrie das extreme geometrische Limit $L \ll W$ für den Magnetowiderstandseffekt nicht erfüllt, ist ein Korrekturfaktor k_{gm} einzufügen. Dieser berücksichtigt, dass sich bei Abweichungen

der Probengeometrie vom Idealfall bereits ein partielles Hallfeld aufbaut, das die vollständige Ausbildung des Magnetowiderstandseffekts teilweise unterdrückt. Damit erweitert sich (3.11) zu

$$R_B = R_0 + k_{\mathrm{gm}} \Delta R_0(B)\,. \tag{5.2}$$

In [83] geben Lippmann und Kuhrt eine tiefe analytische Betrachtung des Zusammenhangs zwischen Probengeometrie und Magnetowiderstandseffekt. Für den hier herrschenden Fall kleiner Hallwinkel konnten sie dabei folgende Beziehungen für k_{gm} ableiten[1]:

$$\begin{aligned} k_{\mathrm{gm}} &= 1 && \text{für } L/W = 0 \\ k_{\mathrm{gm}} &= 1 - 0,54 \cdot L/W && \text{für } 0 < L/W < 0,39 \\ k_{\mathrm{gm}} &= 0,54 \cdot W/L && \text{für } L/W \geq 1 \end{aligned} \quad . \tag{5.3}$$

Bei den in dieser Arbeit untersuchten Proben führt die Berücksichtigung von k_{gm} zu einer erheblichen Korrektur der Beweglichkeit des Langkanaltransistors mit $L/W = 1$, während die Beweglichkeiten der Kurzkanaltransistoren beinahe unbeeinflusst bleiben. Tabelle 5.4 zeigt die Geometriekorrekturfaktoren k_{gm}, die sich nach (5.3) für die untersuchten Transistorabmessungen ergeben. Die Korrektur kompensiert Abweichungen, die sich bei Probengeometrien ergeben, welche für einen reinen Magnetowiderstandseffekt ungünstig sind, da sich aufgrund eines großen L/W-Verhältnisses bereits eine Hallspannung aufzubauen beginnt. Die aufgeführten Korrekturfaktoren sind nach (5.8) bzw. (5.9) in alle Berechnungen eingeflossen, sofern nicht explizit anderweitig angegeben.

Während k_{gm} beinahe keinen Einfluss auf die Extraktion der Kurzkanalbeweglichkeit hat ($k_{\mathrm{gm}} \approx 1$), wird die Berechnung der Langkanalbeweglichkeit durch $k_{\mathrm{gm}} = 0,54$ um mehr als 40 % nach oben korrigiert. Erst nach Berücksichtigung des Korrekturfaktors liegt die Langkanalbeweglichkeit nach der Magnetowiderstandsmethode auf einem Niveau nahe der Beweglichkeit, die nach der Split-CV-Methode für kleine V_{ds}

[1] Als „klein“ gilt in [83] der Wertebereich $0° \leq \theta \leq 26°$. Der maximale Hallwinkel ergibt sich in dieser Arbeit für den nFET mit $L = 1\,\mu\mathrm{m}$ zu $\theta = \arctan(E_y/E_x) = \arctan(\mu B) \approx 15°$

Tabelle 5.4: Geometriekorrekturfaktor k_{gm} nach (5.3) auf Basis von L_{eff} und W_{design} für verschiedene nominelle Gatelängen L; $\delta L = 6\,\mathrm{nm}$

L	30 nm	34 nm	38 nm	42 nm	1 µm
L_{eff}	24 nm	27 nm	31 nm	35 nm	897 nm
W_{design}	900 nm	900 nm	900 nm	900 nm	900 nm
k_{gm}	0,986	0,984	0,981	0,979	**0,542**

ermittelt wurde. Hierbei spielt es aufgrund der untersuchten Geometrien keine Rolle, ob k_{gm} auf Basis von L oder L_{eff} berechnet wird.

c) Serienwiderstandskorrektur

Der Gesamtwiderstand eines Transistors kann in Abwesenheit magnetischer Felder als

$$R_0 = R_{\mathrm{ch}} + R_{\mathrm{ser}} \tag{5.4}$$

geschrieben werden. R_{ch} und R_{ser} bezeichnen dabei Kanal- und (Gesamt-) Serienwiderstand. Sowohl R_{ch}, als auch R_{ser} können sich prinzipiell unter dem Einfluss eines Magnetfelds verändern. Der Serienwiderstand kann weiterhin aufgeteilt werden in den Widerstand der Source- und Drainanschlüsse im Halbleitermaterial R_{sd} und dem Zuleitungs- und Kontaktwiderstand R_{cont}

$$R_{\mathrm{ser}} = R_{\mathrm{sd}} + R_{\mathrm{cont}}\,. \tag{5.5}$$

Letzterer beinhaltet die Widerstandsbeiträge aus den Zuleitungen einschließlich der Source/Drain-Kontaktstöpsel R_{CA}, dem Silizid-Übergangswiderstand $R_{\mathrm{sil,int}}$ und dem Streuwiderstand $R_{\mathrm{SR,cont}}$ auf die Source- und Drain-Diffusionsgebiete. R_{sd} wiederum setzt sich zusammen aus dem Widerstand der Diffusionsgebiete $R_{\mathrm{SD,diff}}$, den Extensiongebieten R_{ext} und letztlich dem Streuwiderstand $R_{\mathrm{SR,ch}}$ zwischen Extensiongebiet und Transistorkanal (Abb. 5.9).

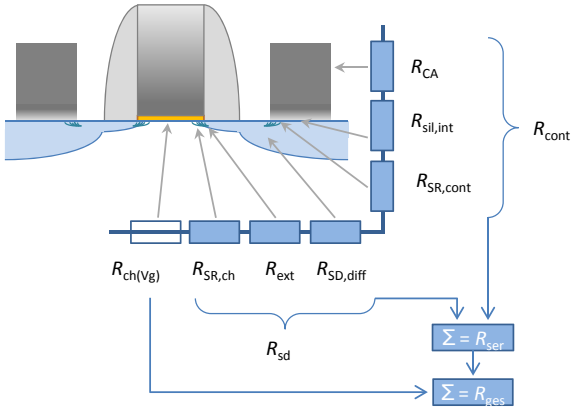


Abbildung 5.9: Einzelbeiträge zum Gesamtwiderstand eines Transistors. Außer dem Kanalwiderstand fallen alle Beiträge aufgrund des vorliegenden, symmetrischen Transistoraufbaus zweimal an, und zwar jeweils einmal an Source und einmal an Drain. Die im Bild genannten Größen repräsentieren dabei bereits die aus Source und Drain zusammengefassten Beiträge.

Die Magnetfeldabhängigkeit des vorwiegend metallischen R_{cont} ist im Vergleich zu den Halbleiteranteilen vernachlässigbar, da in isotropen Metallen beinahe kein Magnetowiderstandseffekt beobachtet werden kann [84], [85]. Hinzu kommt, dass der Strompfad in den silizidierten Kontaktstöpseln auf die Source/Drain-Gebiete nahezu parallel zu den Magnetfeldkennlinien verläuft. Aus diesen Gründen kann der Magnetfeldeinfluss auf R_{cont} vernachlässigt werden. Mit $\Delta R_{\mathrm{cont}} \simeq 0$ ergibt sich nach (3.12)

$$\begin{aligned} \Delta R_0 &= \Delta R_{\mathrm{ch}} + \Delta R_{\mathrm{sd}} \\ &= R_{\mathrm{ch}} \mu_{\mathrm{ch}}^2 B^2 + R_{\mathrm{sd}} \mu_{\mathrm{sd}}^2 B^2 \,. \end{aligned} \tag{5.6}$$

Wenn aus der Änderung des Gesamtwiderstands ΔR_0 auf die Kanalbeweglichkeit μ_{ch} geschlossen werden soll, muss daher sichergestellt werden, dass auch die Änderung ΔR_{sd} im Magnetfeld vernachlässigbar klein ist. Eine Abschätzung der Beweglichkeit in den extrem hoch dotierten Diffusions- und Extensiongebieten basierend auf den Eingangsgrößen von Serienwiderstand, Dotierstoffkonzentrationen von $2-3\cdot 10^{20}\,\mathrm{cm}^{-3}$, Anschlussgeometrien und Schichtwiderstand ergibt Obergrenzen der Beweglichkeit von 30 bzw. $20\,\mathrm{cm}^2/\mathrm{Vs}$ in den p- bzw. n-Anschlussgebieten. Somit ergibt sich ein Maximalbeitrag von $\Delta R_{\mathrm{sd}} \leq 10\,\%$, der im Folgenden vernachlässigt wird. Damit vereinfacht sich obige Relation zu

$$\begin{aligned}\Delta R_0 &\approx R_{\mathrm{ch}}\mu_{\mathrm{ch}}^2 B^2 \\ &= (R_0 - R_{\mathrm{ser}})\mu_{\mathrm{ch}}^2 B^2\,.\end{aligned} \tag{5.7}$$

Auch wenn die Änderung des Serienwiderstands im Magnetfeld zu vernachlässigen ist, so kann der absolute, konstante Gesamtserienwiderstand zu wichtigen Korrekturen in der Berechnung der Beweglichkeit führen. Mit (5.2) und (5.7) erhält man

$$\mu_{\mathrm{MR}} = \frac{1}{B}\sqrt{\frac{\dfrac{R_B(V_{\mathrm{gs}},V_{\mathrm{ds}})}{R_0(V_{\mathrm{gs}},V_{\mathrm{ds}})} - 1}{k_{\mathrm{gm}}(1-\dfrac{R_{\mathrm{ser}}}{R_0(V_{\mathrm{gs}},V_{\mathrm{ds}})})}}\,. \tag{5.8}$$

Hier wurde verwendet, dass unter Berücksichtigung der R_{ser}-Korrektur mit $\Delta R_{\mathrm{ser}} = 0$ die gemessene Beweglichkeit der Kanalbeweglichkeit entspricht, also dass $\mu_{\mathrm{MR}} := \mu_{\mathrm{ch}}$ gesetzt werden kann. Für den Drainstrom I_{d} im Magnetfeld gilt dann

$$I_{\mathrm{d}(B)} = \frac{I_0}{1 + k_{\mathrm{gm}}(1-\dfrac{R_{\mathrm{ser}}}{R_0})\mu^2 B^2}\,, \tag{5.9}$$

mit $R_0 = V_{\mathrm{ds}}/I_0$ und $I_{\mathrm{d}(B)} = V_{\mathrm{ds}}/R_B$. Im Fall kurzer und weiter Transistoren ($L/W \ll 1$, $k_{\mathrm{gm}} = 1$) und vernachlässigbarem Serienwiderstand ($R_{\mathrm{ser}}/R_0 \ll 1$) reduziert sich (5.8) zu (3.11). Andererseits erkennt man die Bedeutung der Korrekturterme, wobei die Serienwiderstandskorrektur durch das Verhältnis R_{ser}/R_0 bestimmt wird.

Die Berücksichtigung des Anschluss- oder Serienwiderstands bei der Berechnung der Ladungsträgerbeweglichkeit ist besonders dann wichtig, wenn der Kanalwiderstand klein wird und der Serienwiderstand einen signifikanten Anteil am Gesamtwiderstand bildet. Langkanaltransistoren, deren Widerstand hauptsächlich durch den Kanal bestimmt ist, werden durch die Serienwiderstandskorrektur daher kaum beeinflusst. Für sie ist in (5.8) das Verhältnis $R_{\mathrm{ser}}/R_0 \approx 0$. Für Kurzkanaltransistoren in starker Inversion bei $V_{\mathrm{gs}} = 1\,\mathrm{V}$ ist diese Korrektur hingegen sehr wichtig. Abb. 5.10 zeigt den Einfluss der Serienwiderstandskorrektur beispielhaft am 30 nm nFET für realistische Werte von R_{ser}. Für kleine V_{ds} ist der Einfluss der Serienwiderstandskorrektur am stärksten, da sich hier Kanal- und Serienwiderstand in der gleichen Größenordnung befinden. Eine Unsicherheit von $\Delta R_{\mathrm{ser}} = \pm 20\,\Omega\cdot\mu\mathrm{m}$ führt dann zu einer Abweichung $\Delta\mu = \pm 10\,\mathrm{cm}^2/\mathrm{Vs}$. Mit steigendem V_{ds} bildet sich drainseitig ein Abschnürpunkt im Kanal aus. Dabei geht der Transistor in den Sättigungsbereich, wobei der Kanalwiderstand zunimmt bis er letztlich den Gesamtwiderstand dominiert. Aus diesem Grund wird der Einfluss der Serienwiderstandskorrektur bzw. die durch R_{ser} bedingte Unsicherheit mit steigendem V_{ds} immer kleiner (vgl. Abb. 5.10). Mittels linearer Extrapolation von $R_{\mathrm{total}}(L) \to L = 0$ für die 3 kürzesten Kanallängen über mehrere Chips gemittelt und unter Abgleich mit Simulationswerten ergab sich ein Serienwiderstand von $150\,\Omega\cdot\mu\mathrm{m}$ bei den nFETs und 200 $\Omega\cdot\mu\mathrm{m}$ für die pFETs (siehe Kap. 5.2.3). Die nachfolgenden Betrachtungen beziehen diese Serienwiderstände mit ein und erfolgen unter der Annahme, dass V_{ds} keinen wesentlichen Einfluss auf R_{ser} hat.

Grundsätzlich besteht auch die Möglichkeit, über die Messung des Magnetowiderstandseffekts an zwei Transistoren mit unterschiedlichem, aber ähnlichem L einen bereits mit dem Serienwiderstand korrigierten Wert zu berechnen [86]. Voraussetzung für diese sogenann-

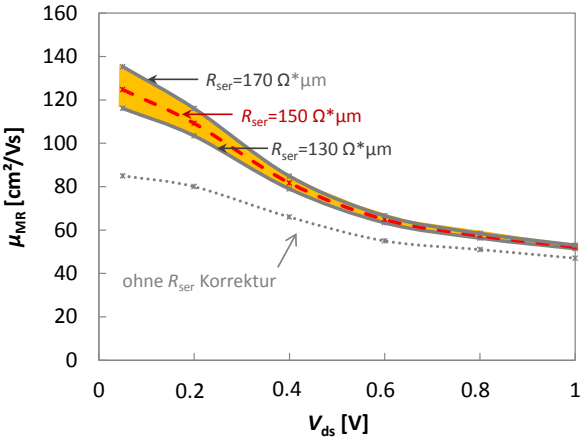


Abbildung 5.10: Beweglichkeits-Drainspannungsdiagramm für verschiedene Werte von R_{ser} am Beispiel des 30 nm NFETs. Die rot gestrichelte Linie entspricht dem nFET-Serienwiderstandswert von $R_{\mathrm{ser}} = 150\,\Omega\cdot\mu\mathrm{m}$.

te „differentielle Magnetowiderstands-Methode“ ist allerdings, dass die Transistoren in erster Näherung die gleiche Beweglichkeit und Einsatzspannung haben. Dann kann eine für beide gleichermaßen gültige, mit R_{ser} korrigierte Kanalbeweglichkeit berechnet werden. Dies erfolgt über die Beziehung $(\Delta R_{\mathrm{total,1}} - \Delta R_{\mathrm{total,2}})/(R_{\mathrm{total,1}} - R_{\mathrm{total,2}}) = \mu_{\mathrm{ch}}^2 B^2$. Hierbei bezeichnen $\Delta R_{\mathrm{total,1,2}}$ die Änderung des Gesamtwiderstands im Magnetfeld und $R_{\mathrm{total,1,2}}$ den Gesamtwiderstand des Transistors ohne Magnetfeld. Die Methode nimmt somit an, dass der Widerstandsunterschied zwischen beiden Transistoren allein aus der größeren Kanallänge herrührt. Für die in dieser Arbeit untersuchten Kurzkanaltransistoren kommt die differentielle Magnetowiderstandsmethode aufgrund des starken Beweglichkeits-roll-offs (siehe z.B. Abb. 5.23) nicht in Frage.

d) Abhängigkeit vom vertikalen elektrischen Feld

Mit der fortschreitenden Skalierung der Transistoren kommen immer höhere elektrische Felder zwischen Gate und Substrat zum Einsatz. Die dabei zunehmenden Kurzkanaleffekte machen eine immer stärkere Kontrolle der Kanalladung durch die Gatespannung erforderlich. Die bei der untersuchten 28 nm-Technologie vorhandene starke Skalierung des Gateoxids auf $1,2\,\mathrm{nm}$ EOT (nFET) bzw. $1,4\,\mathrm{nm}$ EOT (pFET) führt bei $V_{\mathrm{gs}} = 1\,\mathrm{V}$ zu hohen vertikalen Feldstärken in der Größenordnung von $E_{\mathrm{eff,vert}} \approx 1\,\mathrm{MV/cm}$, vgl. Abb. 5.4c. Dies erhöht die Streuung an der Gateoxid-Grenzfläche, wodurch die Beweglichkeit der Langkanaltransistoren im Vergleich zu früheren Technologien reduziert wird. Zusätzlich wurde die Kanaldotierung angehoben. So fallen die hier für die Langkanaltransistoren mit $L = 1\,\mu\mathrm{m}$ und $L = 2\,\mu\mathrm{m}$ ermittelten Werte beide zusammen auf einen Punkt im universalen Verlauf, den Takagi et al. bei der Untersuchung des Zusammenhangs zwischen Beweglichkeit und vertikalem elektrischen Feld in MOSFETs mit unverspanntem Silizium gefunden haben (Abb. 5.11). Dies ist ein Indikator, dass auch der kürzere Langkanaltransistor mit $L = 1\,\mu\mathrm{m}$ als mechanisch unverspannt angesehen werden kann.

Zur Charakterisierung der nFET Kurzkanalbeweglichkeit wurde μ_{MR} in Abhängigkeit von V_{gs} bei $V_{\mathrm{ds}} = 50\,\mathrm{mV}$ bestimmt und mit Hilfe der Daten aus Abb. 5.4c über $E_{\mathrm{eff,vert}}$ aufgetragen. Die Kurzkanalbeweglichkeit ist aufgrund der hohen Kanaldotierung im Vergleich zum Langkanaltransistor stark herabgesetzt, vgl. rechten Punkt in Abb 5.12 ($V_{\mathrm{gs}} = 1\,\mathrm{V}$, $V_{\mathrm{ds}} = 50\,\mathrm{mV}$) mit dem Stern in Abb. 5.11. An späterer Stelle wird gezeigt, dass hier die Kanalverspannung signifikant hilft, einen Teil des Gesamtverlustes zu kompensieren. Wie aus Abb. 5.12 weiterhin ersichtlich ist, steigt die Beweglichkeit für wachsende Feldstärken zunächst an. Dies ist ein Indiz für dominierende Coulombstreuung. Je stärker das vertikale Feld, desto mehr Ladungsträger sind im Kanal. Einzelne Coulomb-Streuzentren werden so mit wachsender Kanalladung immer besser abgeschirmt, sodass die Relevanz der Coulombstreuung anteilig abnimmt und die Beweglichkeit steigt. Für noch höhere Felder erreicht die Beweglichkeitsfunktion bei $V_{\mathrm{gs}} = 1\,\mathrm{V}$, $E_{\mathrm{eff,vert}} = 1,25\,\mathrm{MV/cm}$ ein Plateau. Dahinter muss die

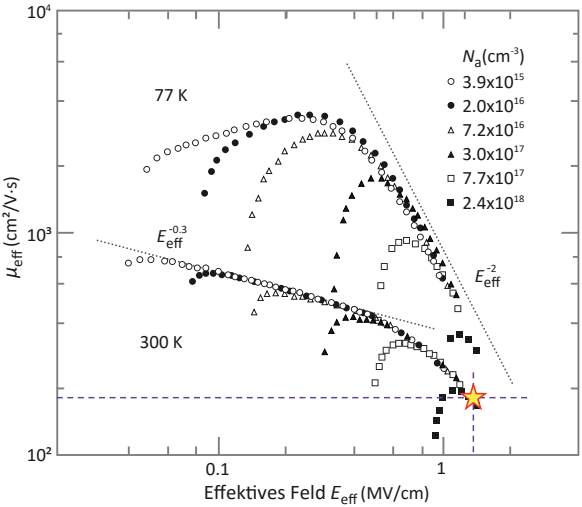


Abbildung 5.11: Beweglichkeit als Funktion des eff. vertikalen Feldes $E_{\mathrm{eff,vert}}$ aus [30], nach [16] für p-Silizium verschiedener Dotierstoffkonzentration bei $T = 77\,\mathrm{K}$ und $300\,\mathrm{K}$. Der Stern markiert die Lage $\mu_{\mathrm{MR}}(E_{\mathrm{eff,vert}})$ der in dieser Arbeit untersuchten $L = 1\,\mu\mathrm{m}$ und $L = 2\,\mu\mathrm{m}$ nFETs.

Beweglichkeitsfunktion wieder abnehmen, da der Kanal immer stärker gegen die Grenzfläche gezogen wird und so Grenzflächenstreuung der limitierende Faktor wird. Im Übergangsbereich des Plateaus, also bei $V_{\mathrm{gs}} = 1\,\mathrm{V}$, wirkt der Einfluss der Phononenstreuung am stärksten. Tatsächlich wird später über den gesamten V_{ds}-Bereich eine Abnahme der Beweglichkeit für steigende Temperaturen beobachtet (Abb. 5.19). Darüber hinaus wird aus Abb. 5.12 deutlich, dass die Kurzkanalbeweglichkeit im Bereich um $V_{\mathrm{gs}} = 1\,\mathrm{V}$ unempfindlich gegenüber leichten Verschiebungen der Einsatzspannung V_{th} ist, wie sie z. B. durch Prozessschwankungen oder DIBL verursacht werden können. [1]

[1] Eine V_{th}-Verschiebung würde in Abb. 5.12 lediglich eine kleine horizontale Verschiebung des letzten Werts bezüglich $E_{\mathrm{eff,vert}}$ im Plateaubereich bewirken

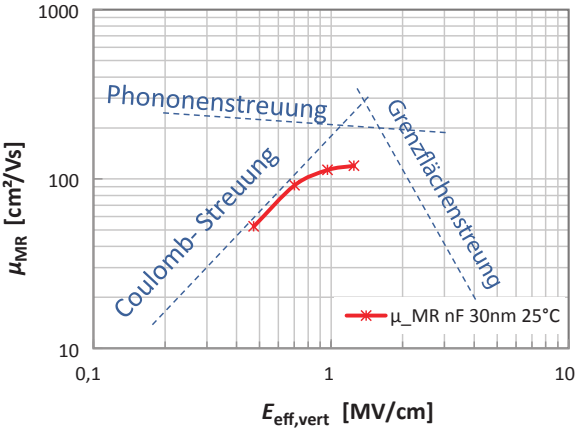


Abbildung 5.12: MR Beweglichkeit gegenüber vertikalem elektrischen Feld für den 30 nm TPEN nFET, $V_{ds} = 50\,\text{mV}$, $T = 25\,°\text{C}$. Die Datenpunkte korrespondieren zu $V_{gs} = 0,4\,\text{V}$, $0,6\,\text{V}$, $0,8\,\text{V}$ und $1\,\text{V}$. Die eingetragenen limitierenden Streumechanismen sind nicht datenbasiert sondern dienen zur groben Orientierung, vgl. Abb. 2.3.

e) Abhängigkeit vom lateralen elektrischen Feld

Eine Besonderheit der Magnetowiderstandsmethode ist die Möglichkeit, die Beweglichkeit auch für höhere V_{ds} zu untersuchen, bei denen der Transistor nicht mehr im linearen Bereich operiert. Abb. 5.13 zeigt die nFET Beweglichkeiten $\mu_{MR}(V_{ds})$ für den Fall starker Inversion bei $V_{gs} = 1\,\text{V}$. Der untersuchte Bereich reicht dabei von $V_{ds} = 50\,\text{mV}$ (linearer Bereich) bis 1 V (Sättigungsbereich). Zunächst fällt auf, dass die Beweglichkeit systematisch mit der Kanallänge zunimmt. Dies rechtfertigt die Annahme, dass bei Transistoren, die in direkter Nähe zueinander auf dem gleichen Chip hergestellt wurden, tatsächlich auch vom gleichen Serienwiderstand ausgegangen werden kann. So müssen die Unterschiede in R_{ser} zwischen den vier untersuchten Kurzkanaltransistoren kleiner $\pm 5\,\Omega\cdot\mu\text{m}$ sein, damit sich die Messungen bei $V_{ds} = 50\,\text{mV}$

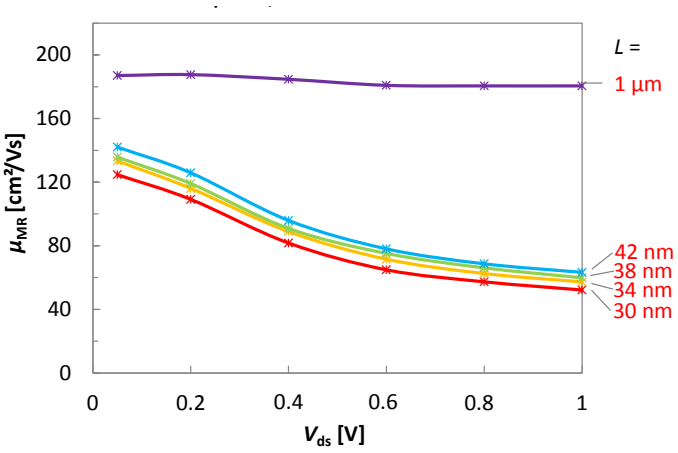


Abbildung 5.13: Entwicklung der nFET MR Beweglichkeit über V_{ds} für verschiedene Gatelängen L in Inversion bei $V_{\mathrm{gs}} = 1\,\mathrm{V}$, $T = 25\,°\mathrm{C}$, mit $R_{\mathrm{ser}} = 150\,\Omega \cdot \mu\mathrm{m}$. Jeder Punkt entspricht dem Ergebnis aus einem Fit an einen $I_{\mathrm{d}}(B)$ Datensatz bei den jeweiligen Spannungseinstellungen.

nicht schneiden; vgl. hierzu Abb. 5.10. Der $L = 1\,\mu\mathrm{m}$ Transistor weist die höchste Beweglichkeit auf. Durch die große Kanallänge ist das laterale elektrische Feld zwischen Source und Drain auch bei hohen V_{ds} klein ($E_{\mathrm{eff,lat}} \approx 0,01\,\mathrm{MV/cm}$ für $V_{\mathrm{ds}} = 1\,\mathrm{V}$), weshalb μ_{MR} mit steigendem V_{ds} kaum degradiert. Die Kurzkanaltransistoren haben dagegen eine deutlich geringere Beweglichkeit und zeigen eine starke Abhängigkeit von V_{ds}. Für kleiner werdende V_{ds} nimmt μ_{MR} zu, mit einer Tendenz, in Richtung $V_{\mathrm{ds}} = 0\,\mathrm{V}$ auf einen „Nullfeldwert“ zu sättigen. Wird V_{ds} hingegen erhöht, nimmt die Beweglichkeit zunächst immer stärker ab. Eine Abnahme der Beweglichkeit mit steigendem V_{ds} ist aus zwei Gründen zu erwarten: zum einen wird drainseitig die Dichte verringert, wodurch Coulombstreuung anteilig zunimmt – ähnlich einer leichten Reduktion von V_{gs}. Zudem ist bekannt, dass die Ladungsträger

im Bereich der Sättigungsgeschwindigkeit immer schwerer noch weiter beschleunigt werden können, weshalb gemäß $v = \mu E$ mit $v \approx v_{\mathrm{sat}}$ die Beweglichkeit selbst eine Funktion des lateralen Feldes wird und somit abnehmen muss. Diesem Sachverhalt kann über den empirischen Ansatz $\mu(E_{\mathrm{eff,lat}}) = \mu_{(E_{\mathrm{eff,lat}}=0)}/(1 + E_{\mathrm{eff,lat}}/E_{\mathrm{c}})$ Rechnung getragen werden. Dabei ist E_{c} das „kritische“ Feld, bei dem die Sättigung der Geschwindigkeit einsetzt. Ab etwa $V_{\mathrm{ds}} = 0,5\,\mathrm{V}$ schwächt sich der abfallende Trend ab. Bei $V_{\mathrm{ds}} = 1\,\mathrm{V}$ beträgt die Beweglichkeit schließlich weit weniger als die Hälfte des Werts bei niedrigem V_{ds}. Die pFETs zeigen ein ähnliches Verhalten auf der (negativen) V_{ds}-Skala (Abb.5.14), mit Absolutwerten, die ungefähr die Hälfte des nFET Niveaus betragen.

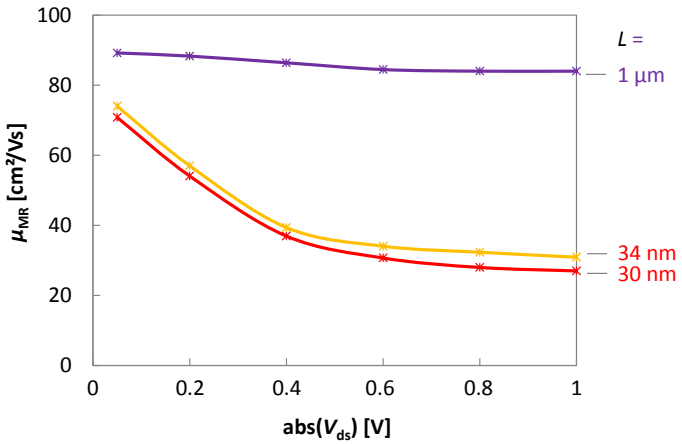


Abbildung 5.14: pFET MR Beweglichkeit über V_{ds} für verschiedene Längen L in Inversion bei $V_{\mathrm{gs}} = -1\,\mathrm{V}$, $T = 25\,^{\circ}\mathrm{C}$, $R_{\mathrm{ser}} = 200\,\Omega\cdot\mu\mathrm{m}$. Die eigentlich negative V_{ds}-Achse wurde gespiegelt, um einen leichteren Vergleich zum nFET zu ermöglichen.

Es ist anzumerken, dass die hier gezeigten Beweglichkeiten von einem technischen Standpunkt her gesehen extrahiert wurden, und zwar für den Fall der extremen Betriebsbedingung $V_{\mathrm{gs}} = 1\,\mathrm{V} = V_{\mathrm{dd}}$, $V_{\mathrm{ds}} = 0,05 \rightarrow 1\,\mathrm{V} = V_{\mathrm{dd}}$. Aus physikalischer Sicht gesehen würde die Beweglichkeit dagegen eher unter vergleichbaren Feldbedingungen extrahiert werden, also bei konstantem Overdrive (z. B. $V_{\mathrm{gs}} = 0,6\,\mathrm{V} + V_{\mathrm{th}}$). Dafür müsste für jeden Transistor zunächst an jedem Punkt in V_{ds} die Einsatzspannung bestimmt werden. Für die Messung der Inversionsbeweglichkeiten hätte dann V_{gs} für jeden Messpunkt in Abb. 5.13 und Abb. 5.14 entsprechend angepasst werden müssen. Die linearen Einsatzspannungen der untersuchten Transistoren bei $V_{\mathrm{ds}} = 50\,\mathrm{mV}$ liegen jedoch sehr eng um 400 mV. Darüber hinaus befinden sich die Beweglichkeiten der Kurzkanaltransistoren bei $V_{\mathrm{gs}} = 1\,\mathrm{V}$ auf einem Plateau (vgl. Abb. 5.12), sodass DIBL-bedingte Variationen im Inversionsfall bei $V_{\mathrm{gs}} = 1\,\mathrm{V}$ kaum Einfluss auf die Beweglichkeit haben. So zeigt sich, dass beide Herangehensweisen – Messung bei konstantem V_{gs} und Messung bei konstantem Overdrive – für die untersuchten Transistoren zur gleichen Inversionsbeweglichkeit führen.

f) Längenabhängigkeit und Kanalverspannung

Da die Langkanaltransistoren mit $L = 1\,\mu\mathrm{m}$ wie bereits oben geschildert als unverspannt angesehen werden können, ist zwischen nFET und pFET in Richtung $\langle 110 \rangle$ ein Beweglichkeitsverhältnis von etwa 2 : 1 zu erwarten, das durch die unterschiedlichen effektiven Massen von Elektronen und Löchern zustande kommt. Dieses Verhältnis kann durch Erhöhung der Ladungsträgerbeweglichkeit mittels Kanalverspannung zugunsten des pFETs verbessert werden [54]. Wie im Vergleich von Abb. 5.13 mit Abb. 5.14 gesehen werden kann, ist die Beweglichkeit des nFET Langkanaltransistors im Vergleich mit dem pFET tatsächlich doppelt so hoch. Um zu verstehen, warum die Langkanaltransistoren unverspannt zu sein scheinen, obwohl hier die gleichen Stressoren integriert wurden wie an den Kurzkanaltransistoren, hilft ein Blick in die Prozess- und Bauelementesimulation: in [87] und [88] finden Flachowsky et al. für ähnliche Stressoren, dass die eingebaute Gitterverspannung nur an Kurzkanaltransistoren effektiv wirkt, und sich mit

wachsender Kanallänge bis auf $L = 1\,\mu\text{m}$ beinahe vollständig verliert. Die hier gezeigten Magnetowiderstandsmessungen bestätigen, dass die Langkanaltransistoren mit $L = 1\,\mu\text{m}$ als unverspannte Referenzen angesehen werden können.

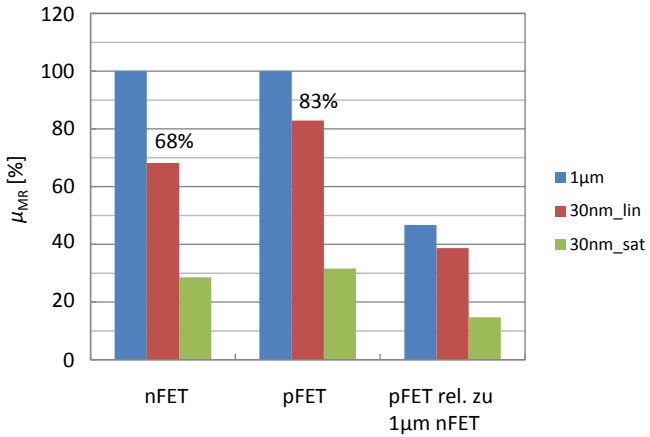


Abbildung 5.15: Vergleich der $L = 30\,\text{nm}$ Kurzkanalbeweglichkeiten mit den jeweiligen $L = 1\,\mu\text{m}$ Langkanaltransistoren. Der Index „lin“ bezieht sich auf den linearen Fall bei $V_{\text{ds}} = 50\,\text{mV}$, „sat“ auf den Sättigungsfall bei $V_{\text{ds}} = 1\,\text{V}$.

Je kürzer die Gatelänge, desto höher wird der verspannungsbedingte Beweglichkeitsgewinn. Abb. 5.15 zeigt die Beweglichkeiten der $L = 30\,\text{nm}$ Kurzkanaltransistoren relativ zu den Langkanaltransistoren mit $L = 1\,\mu\text{m}$. Während die „lineare“ Beweglichkeit ($V_{\text{ds}} = 50\,\text{mV}$) des 30 nm nFET bei 68 % des Langkanalwertes liegt, hat der 30 nm pFET 83% der Beweglichkeit „seines“ Langkanaltransistors. Dies entspricht der Erwartung, dass die Verbindung von CPEN Deckschicht und eSiGe Source/Drain einen größeren relativen Beweglichkeitsgewinn am pFET hervorruft, als die TPEN Deckschicht am nFET [55]. Wenn der Tran-

sistorstrom bei $V_{\mathrm{ds}} = 1\,\mathrm{V}$ in den Sättigungsbereich geht, sinken sowohl die n- als auch die p-Kurzkanalbeweglichkeiten unter die Hälfte des linearen Werts.

Aus den Untersuchungen zum Serienwiderstand und der MR-Beweglichkeit ergibt sich, dass der durch die Verspannung hervorgerufene relative Performancegewinn des pFETs gegenüber dem nFET durch die Kombination aus einer stärker erhöhten Kanalbeweglichkeit und einem verbesserten Serienwiderstand hervorgerufen wird. Dabei fällt der pFET Serienwiderstand relativ zum nFET mit $R_{\mathrm{ser,pF\,CPEN}}/R_{\mathrm{ser,nF\,TPEN}} = 200\Omega \cdot \mu\mathrm{m}/150\Omega \cdot \mu\mathrm{m} = 1,33$ im Vergleich zu früheren Technologien verhältnismäßig niedrig aus. So ist in „unverspannten“ Technologien eher ein Verhältnis von $R_{\mathrm{ser,pF}}/R_{\mathrm{ser,nF}} > 2$ typisch [89, 90]. Dieser geringe pFET Serienwiderstand ist zum einen auf eine durch das Germanium in den SiGe Source/Drain-Gebieten bedingte, gesteigerte Aktivierung der Bor-Dotierstoffatome zurückzuführen [91, 92], wodurch der Schichtwiderstand sinkt. Zum anderen hat auch die Verspannung einen Einfluss auf den Serienwiderstand, wie in Abb. 5.5 aus dem Vergleich der kompressiv und tensil verspannten nFETs ersichtlich wird. Diese Abhängigkeit resultiert vermutlich aus den durch Verspannung veränderten Beweglichkeiten in den Anschlusswiderständen des Kanals R_{ext} und $R_{\mathrm{SR,ch}}$, welche sich wie der Kanal im „günstig“ verspannten Bereich unter dem Gate befinden, vgl. Abb. 5.9. Da hier die Beweglichkeit beim pFET – wie dies auch im Kanal der Fall ist – stärker als beim nFET gesteigert wird, verbessert dies den Serienwiderstand des pFET stärker als den des nFET.

In Kurzkanaltransistoren nimmt die Beweglichkeit meist stark mit der Gatelänge ab. Dieser Effekt wird als „Mobility-roll-off“ bezeichnet. Je kleiner L wird, desto stärker das laterale Feld zwischen Source und Drain, dessen Größenordnung über $E_{\mathrm{eff,lat}} \approx V_{\mathrm{ds}}/L$ abgeschätzt werden kann. Eine Verkleinerung in L wirkt somit ähnlich einer Erhöhung von V_{ds} bei konstantem L. Die Ladungsverteilung im Kanal wird dabei zunehmend inhomogen. Drainseitig nimmt die Ladungsträgergeschwindigkeit zu und die Dichte ab. Es findet mehr Coulomb-Streuung statt, da die Streuzentren nicht mehr so gut von umgebender Ladung abgeschirmt werden. Hinzu kommt, dass bei einer Verkürzung von L die

Streuung in den höher dotierten Halo-Gebieten anteilig wichtiger wird. Da eine höhere Kanaldotierung die Kapazität steigert, kann dies darüber hinaus auch im Sinne eines stärkeren, effektiven vertikalen Felds verstanden werden. Insgesamt ergibt sich durch die beschriebenen Effekte so eine Verringerung der Beweglichkeit mit sinkendem L.

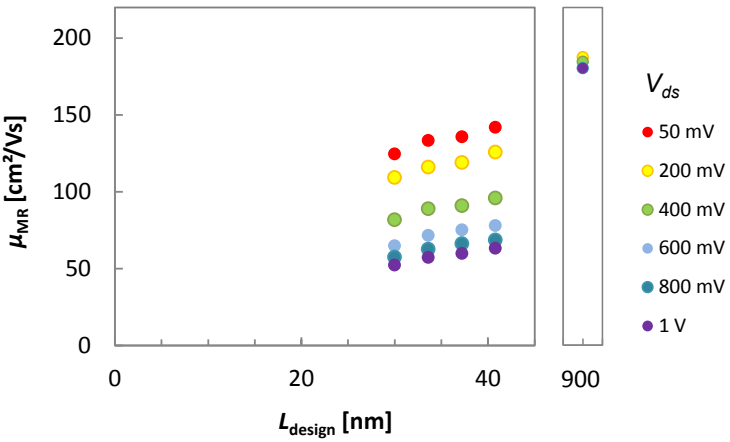


Abbildung 5.16: nFET MR Mobility-roll-off $\mu_{\mathrm{MR}}(L)$ für verschiedene V_{ds} in Inversion bei $V_{\mathrm{gs}} = 1\,\mathrm{V}$; $T = 25\,°\mathrm{C}$.

Abbildung 5.16 zeigt diese Abhängigkeit auf Basis des MR Datensatzes. Innerhalb der Messgenauigkeit ergibt sich ein Abfall von etwa $1\,\mathrm{cm}^2/(\mathrm{Vs}\cdot\mathrm{nm})$ für die kürzesten Kanallängen, mit einer Tendenz zu etwas höheren Werten bei niedrigen V_{ds}. Dies wird vermutlich dadurch begründet sein, dass eine durch den Roll-off bedingte, prozentuale Verringerung der Beweglichkeit bei den bei kleinen V_{ds} herrschenden höheren Beweglichkeiten auch zu größeren absoluten Unterschieden führt.

g) Sättigungsbereich

Den Ausgangskennlinien $I_{\mathrm{d}}(V_{\mathrm{ds}})$ kann entnommen werden, dass die hier untersuchten Transistoren bei hohen $V_{\mathrm{ds}} > 0,6\,\mathrm{V}$ im Sättigungsbereich

operieren, vgl. Abb. 5.3. Im Fall der Kurzkanaltransistoren führen die hohen lateralen Felder dazu, dass das elektrische Feld innerhalb des Kanals ortsabhängig wird und zur Drain hin zunimmt. Mit dem lateralen Feld steigt auch die Ladungsträgergeschwindigkeit, welche im Extremfall die Sättigungsgeschwindigkeit in Silizium erreichen, bzw. für kurze Zeiten sogar übertreffen kann („velocity overshoot“). Da die Stromdichte der Kontinuitätsbedingung gehorchen muss, nimmt gleichzeitig die Dichte im Kanal zur Drainseite hin ab. So wird auch die (Drift-) Beweglichkeit lokal von Punkt zu Punkt verschieden sein. Diese Situation unterscheidet sich stark vom linearen Fall bei kleinen V_{ds}, bei dem eine konstante, homogene Ladungsverteilung im Kanal herrscht. Die Beweglichkeit ist dann über den gesamten Kanal konstant und kann damit auch klar definiert angegeben werden. Im Sättigungsbereich dagegen stellt sich die berechtigte Frage, wie ein mit der MR Methode für den gesamten Kanal bestimmter Beweglichkeitswert zu verstehen ist [93]. Die folgende Analyse soll hierfür einige Anhaltspunkte liefern. Mit dem Ziel, die effektiven Ladungsträgergeschwindigkeiten [94] im Inversionsfall zu bestimmen, wurden zunächst die hierfür benötigten, intrinsischen Transferleitwerte nach [95] berechnet. Über die Messung der Ausgangskennlinie $I_{\mathrm{d}}(V_{\mathrm{ds}})$ bei zwei Gatespannungen $V_{\mathrm{gs\,1,2}}$ nahe 1 V lässt sich der Transferleitwert $g_{\mathrm{m}} = \Delta I_{\mathrm{d}}/\Delta V_{\mathrm{gs}} = [I_{\mathrm{d}(V_{\mathrm{gs2}})} - I_{\mathrm{d}(V_{\mathrm{gs1}})}]/(V_{\mathrm{gs2}} - V_{\mathrm{gs1}})$ an jedem Punkt V_{ds} bestimmen. Mit Hilfe des differentiellen Drainleitwerts $g_{\mathrm{d}} = \partial I_{\mathrm{d}}/\partial V_{\mathrm{ds}}$ kann der intrinsische Transferleitwert

$$g_{\mathrm{m,i}} = \frac{g_{\mathrm{m,0}}}{1 - R_{\mathrm{ser}} g_{\mathrm{d}}(1 + R_{\mathrm{ser}} g_{\mathrm{m,0}}/2)} \tag{5.10}$$

drainspannungsabhängig berechnet werden, wobei $g_{\mathrm{m,0}} = g_{\mathrm{m}}/(1 - R_{\mathrm{ser}} g_{\mathrm{m}}/2)$. Dabei entspricht $g_{\mathrm{m,i}}$ dem reinen Kanalleitwert, welcher sich ohne Anschlusswiderstände am Transistor ergäbe. Drückt man die Drainstromgleichung (3.3) über den Transferleitwert und mit den intrinsischen Größen aus (vgl. Abb. 3.4), erhält man

$$g_{\mathrm{m,i}} = \frac{\partial I_{\mathrm{d}}}{\partial V^*_{\mathrm{gs}}} = \mu_{\mathrm{ch}} \frac{W}{L} C'_{\mathrm{ox}} V^*_{\mathrm{ds}} \,. \tag{5.11}$$

Hieraus lässt sich eine Gleichung zur Berechnung der effektiven Geschwindigkeit der Ladungsträger v_{eff} entwickeln:

$$\frac{g_{\mathrm{m,i}}}{W} = \mu_{\mathrm{ch}} \frac{V_{\mathrm{ds}}^*}{L} C'_{\mathrm{ox}} = \mu_{\mathrm{ch}} E_{\mathrm{eff,lat}} C'_{\mathrm{ox}} = v_{\mathrm{eff}} C'_{\mathrm{ox}} \tag{5.12}$$

$$\Rightarrow v_{\mathrm{eff}(V_{\mathrm{ds}})} = \frac{g_{\mathrm{m,i}(V_{\mathrm{ds}})}}{W \cdot C'_{\mathrm{ox}}}. \tag{5.13}$$

Um des Weiteren eine Abschätzung für das drainspannungsabhängige, effektive laterale Feld auf Basis der Messdaten zu erhalten, wird die Beziehung

$$E_{\mathrm{eff,lat}(V_{\mathrm{ds}})} = v_{\mathrm{eff}(V_{\mathrm{ds}})} / \mu_{\mathrm{MR}(V_{\mathrm{ds}})} \tag{5.14}$$

verwendet. $E_{\mathrm{eff,lat}}$ wird somit letztlich bestimmt aus der Verrechnung von $I_{\mathrm{d}}(V_{\mathrm{gs}})$ bzw. $I_{\mathrm{d}}(V_{\mathrm{ds}})$ Strom-Spannungs-Messungen mit Magnetowiderstands-Beweglichkeitsdaten. Zur Berechnung von (5.13) sind sowohl $g_{\mathrm{m,i}(V_{\mathrm{ds}})}$, als auch C'_{ox} am gleichen Punkt V_{gs} zu bestimmen, im hier untersuchten Inversionsfall also bei $V_{\mathrm{gs}} = 1\,\mathrm{V}$. Dabei sollte beachtet werden, dass die Messung von C'_{ox} in der Praxis bei $V_{\mathrm{ds}} = 0$ erfolgt. Für höhere V_{ds} wird die über die Gatespannung kontrollierte Ladung insbesondere am Kurzkanaltransistor reduziert, sodass es hier zu einer Unterschätzung sowohl der Geschwindigkeit, als auch des effektiven Feldes kommt. Dennoch bietet die Methode einen guten Zugang zur qualitativen Analyse der Ladungsträgergeschwindigkeit. v_{eff} kann dabei als „effektiver“, über den Kanal gemittelter Wert aufgefasst werden.

Mit $v_{\mathrm{eff}}(V_{\mathrm{ds}})$ und $E_{\mathrm{eff,lat}}(V_{\mathrm{ds}})$ ist es nun möglich, die Beziehung zwischen Ladungsträgergeschwindigkeit und effektivem, lateralem Feld $v_{\mathrm{eff}}(E_{\mathrm{eff,lat}})$ zu untersuchen (Abb. 5.17). Wie zu erwarten, entwickeln sich bis $V_{\mathrm{ds}} = 1\,\mathrm{V}$ bei den Langkanaltransistoren nur relativ schwache elektrische Felder im unteren kV/cm-Bereich. Die Ladungsträgergeschwindigkeiten bleiben dabei bei moderaten Werten der Größenordnung 10^5cm/s. Obwohl sich die Ausgangskennlinien der Langkanaltransistoren bei $V_{\mathrm{gs}} > 1\,\mathrm{V}$ oberhalb von etwa $V_{\mathrm{ds}} > 600\,\mathrm{mV}$ aufgrund

der Kanalabschnürung („pinch-off“) bereits in der Stromsättigung befinden (vgl. Abb. 5.3), ist bezüglich der Ladungsträgergeschwindigkeit bei $V_{\mathrm{gs}} = 1\mathrm{V}$ noch keine Sättigung eingetreten. Stattdessen zeigt sich eine lineare Korrelation zwischen v_{eff} und $E_{\mathrm{eff,lat}}$ über den gesamten V_{ds}-Bereich beim n- und pFET.

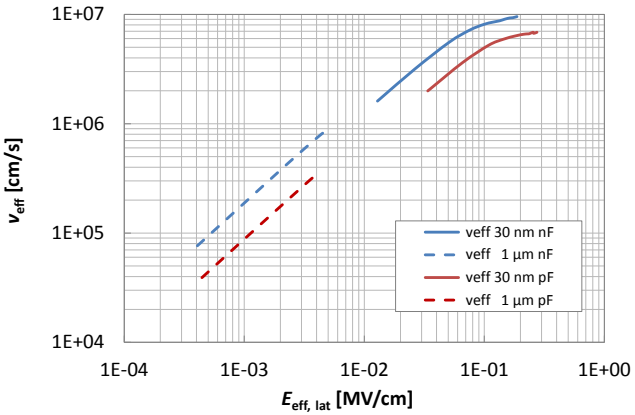


Abbildung 5.17: Effektive Ladungsträgergeschwindigkeit v_{eff} als Funktion des effektiven, lateralen elektrischen Feldes $E_{\mathrm{eff,lat}}$, für die $L = 30\,\mathrm{nm}$ und $L = 1\,\mu\mathrm{m}$ nFET und pFET Transistoren.

Im Gegensatz dazu ist das effektive laterale Feld der 30 nm Kurzkanaltransistoren auch bei niedrigen V_{ds} bereits deutlich höher und erreicht Werte in der Größenordnung um $0,2\,\mathrm{MV/cm}$ für $V_{\mathrm{ds}} = 1\,\mathrm{V}$ (Abb. 5.17). Das $v_{\mathrm{eff}}(E_{\mathrm{eff,lat}})$-Verhalten setzt dabei zunächst den linearen Trend der Langkanaltransistoren fort, sättigt aber für hohe $E_{\mathrm{eff,lat}}$ bzw. V_{ds}. Die erreichten Grenzgeschwindigkeiten liegen mit $v_{\mathrm{eff}} \approx 1{\cdot}10^7\,\mathrm{cm/s}$ (nFET) und $6,8 \cdot 10^6\,\mathrm{cm/s}$ (pFET) bei den in Silizium erreichbaren Sättigungswerten. Die Sättigung tritt ein, wenn die Ladungsträger bei höheren Feldern ihre Energie immer stärker in optische Phononen abgeben, sodass diese nicht für eine weitere Erhöhung der kinetischen

Energie, bzw. der Geschwindigkeit, zur Verfügung steht. In diesem Bereich starker lateraler elektrischer Felder hat die MR-Beweglichkeit ihren niedrigsten Wert. Hier herrscht eine inhomogene Ladungsverteilung im Kanal, weshalb die Beweglichkeit lokal unterschiedliche Werte annehmen wird. Der über die MR Methode im Sättigungsbereich extrahierte Beweglichkeitswert kann daher nicht für den gesamten Kanal gültig sein, sondern stellt eine Mittelung dar. Shahidi et al. [94] erhalten für einen 90 nm Transistor einen Wert von $v_{\text{eff}} = 1,3 \cdot 10^7\,\text{cm/s}$, bei $\mu_{\text{eff}} = 450\,\text{cm}^2/\text{Vs}$. Dies entspricht nach (5.14) einem Feld von etwa 29 kV/cm. Shahidi et al. kommen in ihrer Analyse zu dem Schluss, dass v_{eff} in ihrem Fall die Ladungsträgergeschwindigkeit an Source wiedergibt. Um zu untersuchen welcher Kanalbereich bei den 30 nm Transistoren dieser Arbeit für v_{eff} dominierend ist, wurden für die Technologie angefertigte, detaillierte 2D TCAD Prozess- und Bauelementesimulationen herangezogen.

Abbildung 5.18 zeigt die mittels Simulation berechneten lateralen Feldwerte am 30 nm nFET in Abhängigkeit von V_{ds} für verschiedene laterale Positionen im Kanal. Es zeigt sich, dass die nach (5.14) aus den Messdaten berechneten Felder auf den für den Bereich der Kanalmitte berechneten Simulationswerten liegen. Für kleine V_{ds} sind die simulierten Feldwerte beinahe unabhängig von der Kanalposition und das Feld ist konstant. Der Beweglichkeitsbegriff ist damit über den gesamten Kanal klar definiert und wird vom MR-Verfahren wiedergegeben. Da sich das Feld sourceseitig für höhere V_{ds} nur leicht erhöht („−9 nm“-Kurve in Abb. 5.18), ist zu erwarten, dass die lokale (Drift-) Beweglichkeit an Source nahe dem Wert des linearen Regimes bei niedrigen V_{ds} bleibt. Drainseitig nimmt das Feld für hohe V_{ds} dagegen sehr stark zu („+9 nm“-Kuve in Abb. 5.18). Diesem Trend folgt das aus den Messdaten berechnete Feld jedoch nur schwach. Die „durchschnittliche“ MR-Beweglichkeit wird folglich nicht von den extremen Verhältnissen an der Drainseite dominiert. Es zeigt sich, dass die MR-Beweglichkeit im Sättigungsfall einen gemittelten Wert wiedergibt, der am ehesten den lokalen Verhältnissen in der Kanalmitte entspricht. Im Gegensatz zu Shahidi's Ergebnissen liegt der Übereinstimmungspunkt zwischen gemittelten und lokalen Größen demnach nun nicht mehr rein

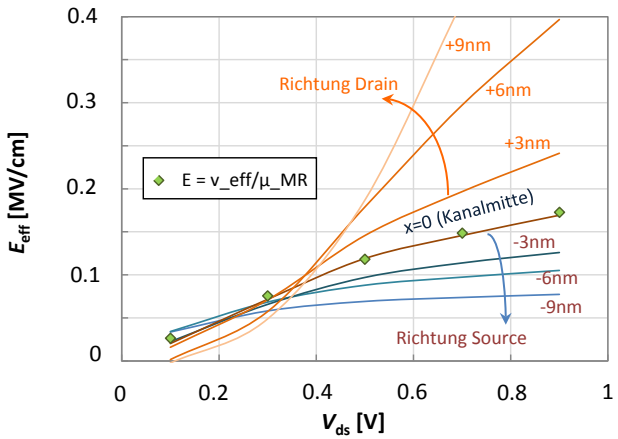


Abbildung 5.18: Simuliertes, laterales elektrisches Feld $E_{\mathrm{eff,lat}}$ als Funktion von V_{ds} für verschiedene laterale Positionen im Kanal, in konstantem Abstand von 8 Å unterhalb des Gateoxids. $x = 0$ bezeichnet die Mitte des Kanals, negative (positive) Werte eine entsprechende Verschiebung Richtung Source (Drain). Die eingetragenen Punkte entsprechen den aus den Messdaten nach $E_{\mathrm{eff,lat}} = v_{\mathrm{eff}}/\mu_{\mathrm{MR}}$ berechneten Werten.

an Source, sondern – bedingt durch die höheren lateralen Felder – in Kanalmitte. Es ist zu erwarten, dass sich dieser Trend mit fortschreitender Skalierung fortsetzt, wenn immer höhere drainseitige Felder einen immer stärkeren Einfluss auf die Mittelung ausüben.

h) Temperaturabhängigkeit

Die 30 nm und 34 nm nFETs wurden auch bei erhöhter Temperatur ($T = 85\,°\mathrm{C}$) gemessen. Dabei konnte gegenüber den Messungen bei $T = 25\,°\mathrm{C}$ keine signifikante Erhöhung des Serienwiderstands festgestellt werden. Für beide Transistorlängen ergibt sich über den gesamten V_{ds}-Bereich eine Beweglichkeitsdegradation von etwa 10%, wenn die Temperatur von 25 °C auf 85 °C erhöht wird (Abb. 5.19). Wie in Abb. 5.12 ersichtlich, befindet sich das System bei $V_{\mathrm{gs}} = 1\,\mathrm{V}$ bzw. $E_{\mathrm{eff,vert}} = 1,25\,\mathrm{MV/cm}$ in der Übergangsregion zwischen Coulomb- und Grenzflächenstreuung. So kann die Beweglichkeitsdegradation bei erhöhten Temperaturen auf eine stärkere Phononenstreuung zurückgeführt werden.

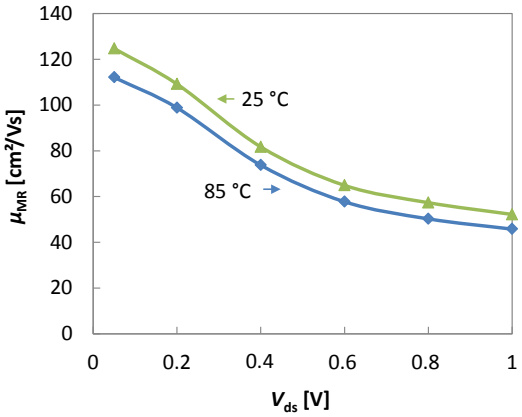


Abbildung 5.19: Ladungsträgerbeweglichkeit $\mu_{\mathrm{MR}}(V_{\mathrm{ds}})$ am 30 nm TPEN nFET: die Temperaturerhöhung $T = 25\,°\mathrm{C} \rightarrow 85\,°\mathrm{C}$ führt über den gesamten V_{ds}-Bereich zu einer etwa 10 %igen Degradation der Beweglichkeit.

5.3.2 Split-CV-Methode

Durch die Messung von $I_\mathrm{d}(V_\mathrm{gs})$ Transferkennlinien sowie $C_\mathrm{gc}(V_\mathrm{gs})$ Kapazitäts-Spannungskurven konnte nach dem in Kap. 3.1.2 beschriebenen Verfahren für kleine $V_\mathrm{ds} = 50\,\mathrm{mV}$ die effektive (Drift-) Beweglichkeit berechnet werden. Zur Bestimmung von Kurzkanalbeweglichkeiten ist dabei – wie beim Magnetowiderstandsverfahren – eine Serienwiderstandskorrektur erforderlich. Schreibt man (3.8) gemäß Abb. 3.4 um auf die internen Größen, erhält man

$$\mu_\mathrm{eff} = \frac{L\, g^*_{\mathrm{d}(V_\mathrm{gs},V_\mathrm{ds})}}{W Q'_\mathrm{inv}}. \tag{5.15}$$

mit dem reinen Kanalleitwert $g^*_\mathrm{d} = \partial I_\mathrm{d}/\partial V^*_\mathrm{ds}$. Im linearen Bereich gilt $R_\mathrm{ch} = R_\mathrm{ges} - R_\mathrm{ser}$ und es ergibt sich

$$g^*_\mathrm{d} = \frac{g_\mathrm{d}}{1 - g_\mathrm{d} R_\mathrm{ser}}\,, \tag{5.16}$$

wobei $R_\mathrm{ch} = 1/g^*_\mathrm{d}$ und $R_\mathrm{ges} = 1/g_\mathrm{d}$ verwendet wurde. Somit folgt für die serienwiderstandskorrigierte Beweglichkeit

$$\mu_\mathrm{eff} = \frac{L\, g_\mathrm{d}}{W Q'_\mathrm{inv}(1 - g_\mathrm{d} R_\mathrm{ser})}. \tag{5.17}$$

Eine weitere wichtige Korrektur für die effektive Beweglichkeit ergibt sich aus der Tatsache, dass die elektrisch relevante, effektive Kanallänge L_eff in der Regel nicht exakt der strukturierten Gatelänge entspricht. Da L direkt in μ_eff eingeht, führt z. B. eine zehnprozentige Unsicherheit in L zu einer ebenso hohen Unsicherheit in μ_eff. Durch Diffusion der Dotanden im Extensiongebiet von Source und Drain unter das Gate und hinein in den Kanalbereich ist L_eff meist etwas kürzer als L. Über die Messung des Kapazitätsverhaltens vieler parallel geschalteter Kurzkanaltransistoren konnte die Überlappkapazität C_ov bestimmt werden. Hieraus ergab sich für nFET und pFET jeweils eine Unterdiffusion von etwa 3 nm an Source und Drain unter den Gatebereich. Insgesamt wird dadurch die Gatelänge um $\delta L = 6\,\mathrm{nm}$ verkürzt. Die

Inversionsladungsdichte Q'_{inv} wurde an einer Kapazitätsstruktur mit großer Länge und Weite ($L_{\mathrm{design}} = 29,7\,\mu\mathrm{m}$, $W_{\mathrm{design}} = 2 \times 29,7\,\mu\mathrm{m}$) bestimmt und einheitlich für alle Transistorlängen verwendet. Dies trägt den folgenden Umständen Rechnung: Zum einen standen nicht für jede Gatelänge geeignete Multifet-Kapazitätsteststrukturen zur Verfügung. Zum anderen ist eine korrekte Bestimmung von Q'_{inv} an Kurzkanaltransistoren schwierig, da verschiedene parasitäre Kapazitätsbeiträge C_{ov} die eigentlich interessante Kapazität zwischen Gate und Inversionskanal maskieren. Diese betrugen in der untersuchten Technologie bis zu 30% der gemessenen Inversionskapazität und bestehen hauptsächlich aus der direkten Überlappkapazität zwischen Gate und den Source/Drain-Diffusionsgebieten (C_{do}), sowie kleineren Beiträgen aus inneren (C_{if}) und äußeren (C_{of}) Streukapazitäten: $C_{\mathrm{ov}} = C_{\mathrm{do}} + C_{\mathrm{if}} + C_{\mathrm{of}}$, vgl. Abb. 5.20.

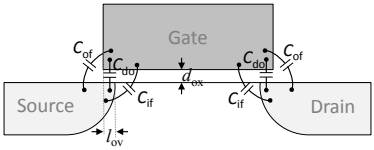


Abbildung 5.20: Parasitäre Kapazitätsbeiträge am MOSFET: direkte Überlappkapazität C_{do}, sowie innere (C_{if}) und äußere (C_{of}) Streukapazitäten, nach [30].

Abb. 5.21 zeigt die resultierende effektive Beweglichkeit in Abhängigkeit von V_{gs} für einen Lang- und einen Kurzkanaltransistor unter verschiedenen Korrekturen. Es fällt auf, dass die Anwendung der Serienwiderstandskorrektur einen fundamentalen Unterschied für die Kurzkanalbeweglichkeit macht. Bei kleinen V_{gs} dominiert zunächst der Kanal den Gesamtwiderstand und die Serienwiderstandskorrektur hat kaum Einfluss auf die extrahierte Beweglichkeit. Bei Gatespannungen oberhalb V_{th} fällt der Kanalwiderstand jedoch rapide ab, bis er schließlich in die Größenordnung des Serienwiderstands kommt. Wird in so einem Fall der gesamte Transistorwiderstand alleine dem Kanal zugeordnet, kommt es zu einer drastischen Unterschätzung der Beweglichkeit.

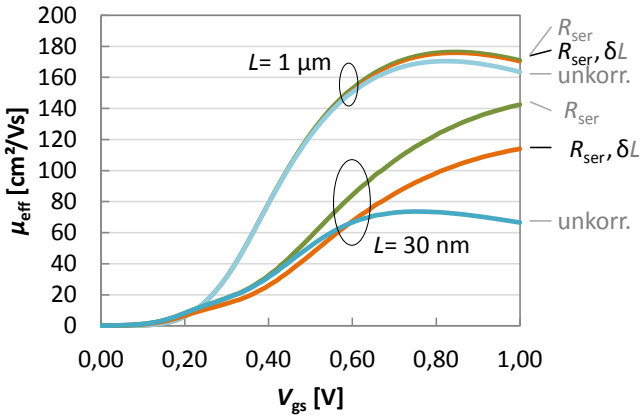


Abbildung 5.21: nFET effektive Beweglichkeit in Abhängigkeit von V_{gs} für zwei Transistorlängen $L = 30\,\text{nm}$ und $1\,\mu\text{m}$ bei $V_{ds} = 50\,\text{mV}$. Unkorrigierte Werte sind mit „unkorr“ gekennzeichnet, R_{ser} und δL stehen für $150\,\Omega \cdot \mu\text{m}$ Serienwiderstands-, bzw. 6 nm Kanallängenkorrektur.

Auch die nachfolgende Berücksichtigung der (abgeschätzten) effektiven Kanallänge hat einen deutlichen, wenngleich weniger drastischen Einfluss auf die Beweglichkeit. Da von einer um $\delta L = 6\,\text{nm}$ verkürzten Gatelänge L ausgegangen wird, reduziert sich die Beweglichkeit des „30 nm“-Transistors gegenüber dem rein R_{ser}-korrigierten Wert um 20 %. Am Langkanaltransistor fallen beide Korrekturen deutlich weniger ins Gewicht, da sein Gesamtwiderstand aufgrund der großen Länge stets durch den Kanal dominiert wird. Je kürzer der Kanal, desto kleiner der Gesamtwiderstand und desto stärker werden die Beweglichkeiten durch die Serienwiderstandskorrektur nach oben verschoben. In der unkorrigierten Messung von Abb. 5.22 liegt die Beweglichkeit des 38 nm-Transistors fast auf der des 34 nm-Transistors. Nach der Korrektur haben beide „die Plätze getauscht“. Dieses scheinbar widersprüchliche Resultat ist dem geschuldet, dass die Split-CV-Beweglichkeit sehr

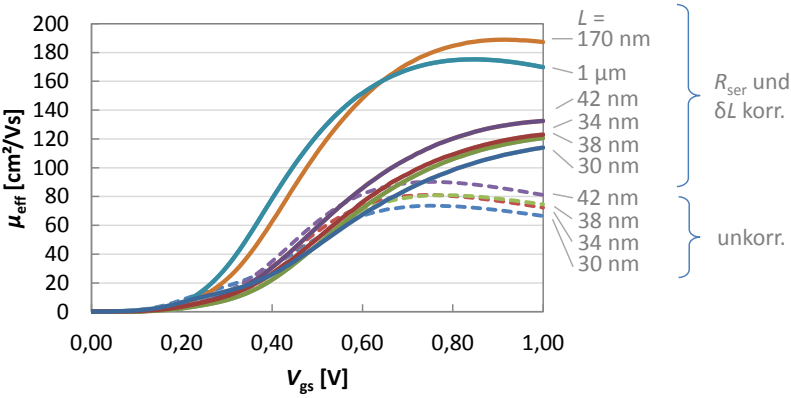


Abbildung 5.22: nFET effektive Beweglichkeit in Abhängigkeit von V_{gs} für verschiedene nominelle Transistorlängen bei $V_{\mathrm{ds}} = 50\mathrm{mV}$. Für die korrigierten Werte gilt $R_{\mathrm{ser}} = 150\,\Omega{\cdot}\mu\mathrm{m}$ und $\delta L = 6\,\mathrm{nm}$.

sensibel auf die Einflüsse von Serienwiderstand und effektiver Kanallänge reagiert, und diese Korrekturen anhand von Widerstands-Medianwerten durchgeführt wurden. Falls einzelne Transistoren nun leicht unterschiedliche Abweichungen zum Medianwert haben, kann dies zum beobachteten Verhalten führen. Es ist daher schwer, mit diesem Verfahren präzise Aussagen zu Kurzkanaltransistoren, welche sich nur wenig in der Kanallänge unterscheiden, über einzelne Messungen ohne Statistik zu treffen. Nichts desto trotz ist der erwartete Gang mit steigender Mobilität von $L = 30\,\mathrm{nm}$ bis $42\,\mathrm{nm}$ erfasst worden. Je größer der Kanal wird, desto geringer fällt der Einfluss beider Korrekturen aus. Wie in Abb. 5.22 ebenfalls ersichtlich, liegt die Beweglichkeit des 170 nm-Transistors über dem Niveau des Langkanaltransistors. Hierbei handelt es sich jedoch nicht um ein methodisches Problem. Erst durch die Serienwiderstandskorrektur kommen die effektiven Beweglichkeiten überhaupt in die richtige Größenordnung. Während kleine Unsicherheiten im Serienwiderstand R_{ser} – welcher ja für jeden Transistor in-

dividuell etwas anders ausfallen kann – bei den kürzesten Kanallängen noch leicht zu den eben beschriebenen Diskrepanzen führen können, spielen diese bei größerem L kaum noch eine Rolle. So würde schon die Annahme eines nur um $5\,\Omega\cdot\mu$m höheren R_{ser} die 38 nm-Kurve bereits wieder über die 34 nm-Kurve setzen. Im Fall des 170 nm-Transistors dagegen bewirkt eine Verschiebung in R_{ser} um $\pm 10\,\Omega\cdot\mu$m nur marginale Änderungen an der Kurve.

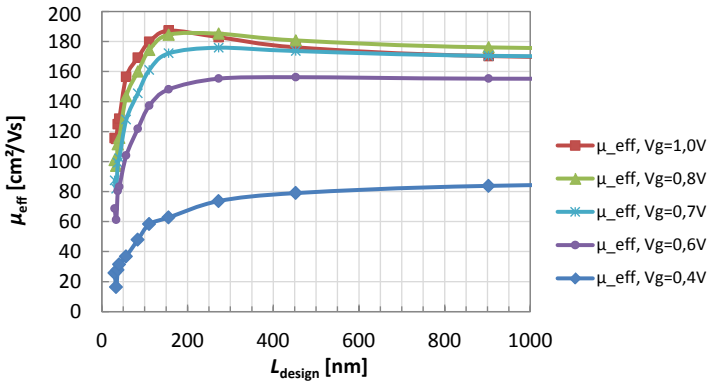


Abbildung 5.23: nFET TPEN Beweglichkeit μ_{eff} in Abhängigkeit der Kanallänge für verschiedene Gatespannungen bei $V_{\text{ds}} = 50\,\text{mV}$, für $R_{\text{ser}} = 150\,\Omega\cdot\mu$m und $\delta L = 6\,\text{nm}$.

Stattdessen handelt es sich hierbei um den sichtbaren Einfluss der Kanalverspannung auf die Beweglichkeit. Abb. 5.23 zeigt μ_{eff} in Abhängigkeit der Kanallänge für verschiedene Gatespannungen. Wenn die Kanallänge reduziert wird, kommen sich die Halo-Gebiete immer näher. Dadurch erhöht sich die Dotierstoffkonzentration des Kanals, der Effekt der Coulombstreuung und die vertikale, effektive Feldstärke nehmen zu und die Beweglichkeit ab. Dieses Verhalten wurde vorhergehend bereits als „Mobility-roll-off“ beschrieben. Darüber hinaus gibt es aber noch einen zweiten Effekt, der durch das besondere Transistordesign zustande kommt: je kürzer die Gatelänge, desto stärker kann

die TPEN Deckschicht den Kanal verspannen [96]. Die Auswirkung auf die Beweglichkeit lässt sich in Abb. 5.23 beobachten. So nimmt sie im Inversionsfall bei $V_{\mathrm{gs}} = 1\,\mathrm{V}$ beim Verkleinern der Gatelänge zunächst zu. Erst ab $L = 170\,\mathrm{nm}$ gewinnt die ebenfalls immer stärker werdende Coulombstreuung in den Halo- und Extensiongebieten die Oberhand und die Beweglichkeit fällt rapide ab. Ihr Einfluss auf den Kurzkanaltransistor wäre ohne Kanalverspannung noch drastischer. Die Verspannung hilft somit, einen Teil der unerwünschten Degradation der Kurzkanalbeweglichkeit zu kompensieren. Bei kleinen Gatespannungen setzt die Coulombstreuung bereits früher, also bei größerem L, ein, da hier die Dotierstoffatome weniger durch umgebende Kanalladungen abgeschirmt werden. Daher ist der Einfluss der Kanalverspannung bei kleinen V_{gs} nicht mehr sichtbar. Vom technologischen Standpunkt her gesehen, ist allerdings ohnehin die Beweglichkeit des Transistors im AN-Zustand entscheidend. Deshalb kommt Beweglichkeitsgewinnen im Spannungsbereich bei hohen V_{gs}-Werten eine stärkere Bedeutung zu.

Im Rahmen der Arbeit konnten auch kompressiv und damit „ungünstig“ verspannte nFETs untersucht werden. Die Prozessierung war dabei identisch zum „normalen“ nFET, nur dass die TPEN- durch eine CPEN Deckschicht ersetzt worden war. Die Extraktion des Serienwiderstands ergab beim CPEN nFET einen Wert von $R_{\mathrm{ser}} = 230\,\Omega\cdot\mu\mathrm{m}$. Hier ist – im Gegensatz zum TPEN – mit abnehmender Kanallänge keine Zunahme der Beweglichkeit zu erwarten. Stattdessen sollte die Degradation der Beweglichkeit bei abnehmender Kanallänge umso früher einsetzen. Wie in Abb. 5.24 zu sehen ist, konnte dieses Verhalten auch tatsächlich beobachtet werden. Die Beweglichkeiten der CPEN Kurzkanaltransistoren fallen dabei deutlich niedriger aus, als beim TPEN. Des Weiteren wird ersichtlich, dass die Art der Verspannungsschicht bei großen Kanallängen mit $L > 900\,\mathrm{nm}$ keine Rolle mehr spielt.

Vergleicht man die Bereiche des starken Beweglichkeits-roll-offs zwischen CPEN und TPEN, so ist der Abfall der Beweglichkeit für $L < 100\,\mathrm{nm}$ im Fall des CPEN weniger steil als beim TPEN, auch wenn die Beweglichkeitswerte insgesamt deutlich niedriger liegen. Wie bereits geschildert, bewirkt die Verspannung des Gitters eine Änderung

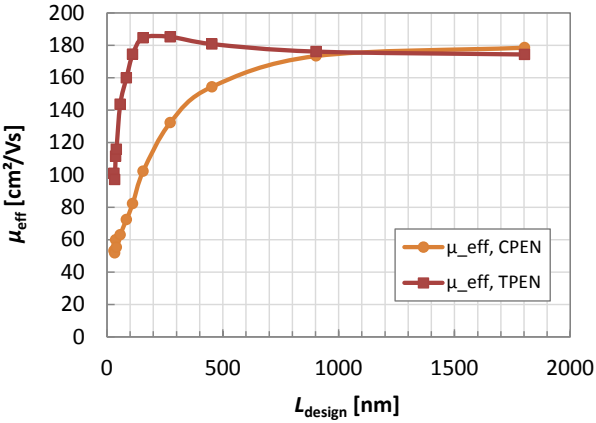


Abbildung 5.24: Vergleich der nFET Beweglichkeiten μ_{eff} für tensile (TPEN) und kompressive (CPEN) Verspannung in Abhängigkeit der Kanallänge bei $V_{\text{ds}} = 50\,\text{mV}$, $V_{\text{gs}} = 0,8\,\text{V}$, $R_{\text{ser,TPEN}} = 150\,\Omega \cdot \mu\text{m}$, $R_{\text{ser,CPEN}} = 230\,\Omega \cdot \mu\text{m}$ und $\delta L = 6\,\text{nm}$.

der effektiven Masse. Nach (2.8) ist $\mu_{(m^*)} \sim 1/m^*$. Dieses reziproke Verhalten wird in (2.8) noch mit der Streuzeit τ multipliziert. Eine Änderung der effektiven Masse bewirkt somit einen stärkeren Beweglichkeitsunterschied, je größer die ursprüngliche Beweglichkeit ist. In einer Roll-off Kurve mit konstanter (von L unabhängiger) Verspannung steigt die Beweglichkeit der Kurzkanaltransistoren mit L durch die Zunahme von τ aufgrund der geringeren Ladungsträgerstreuung an. Wird nun über eine Gitterverspannung die effektive Masse m^* verkleinert, so wird die Roll-off Kurve bei größeren Gatelängen stärker „hochgezogen“, als bei kleinen. Dieser Effekt wird beim Übergang von der CPEN-Kurve in Abb. 5.24 auf die TPEN-Kurve sichtbar, was zum steileren Anstieg des TPEN im Bereich kleiner Gatelängen führt. Die Verspannung bewirkt somit nicht einen additiven Versatz der Roll-off Kurve „nach oben“, sondern verzieht diese auf eine multiplikative Art und Weise. Da sich die Verspannung wie bereits geschildert auf großen

Abständen wieder verliert, bleiben die extremen Langkanaltransistoren hiervon jedoch unbeeinflusst.

Abschließend sei nochmals auf den großen Beweglichkeitsunterschied an kürzesten Kanallängen um den Faktor 2 hingewiesen, welcher allein durch die Variation der Deckschicht entsteht. Hieran wird deutlich, dass die kompressive Deckschicht zur Verbesserung der pFET Beweglichkeit in Strain-Technologien unbedingt vom nFET wieder entfernt werden muss, da sie die nFET-Beweglichkeit zu stark degradieren würde.

5.3.3 Magnetowiderstands-Streukoeffizient

Aus dem Verhältnis der Beweglichkeiten aus dem Magnetowiderstands- und dem Split-CV-Verfahren kann für kleine V_{ds} nach (2.29) der Magnetowiderstands-Streukoeffizient $r_{\mathrm{MR}} = \mu_{\mathrm{MR}}/\mu_{\mathrm{eff}}$ berechnet werden. Aufgrund der durch die Serienwiderstandskorrektur bedingten großen Unsicherheit am Kurzkanaltransistor vor allem für μ_{eff} aus der Split-CV-Methode konnte r_{MR} nur am Langkanaltransistor sicher bestimmt werden. Sowohl für den n-, als auch für den p-Langkanaltransistor ergibt sich ein Streufaktor von $r_{\mathrm{MR}} = 1,1$ (Tabelle 5.5). Dieser Wert impliziert, dass energieunabhängige Streumechanismen mit $\eta = 0$ dominieren, vgl. Abb. 2.5. Darüber hinaus deckt sich dies mit dem Ergebnis aus Abb. 5.12, dass sich der Kanal bezüglich $E_{\mathrm{eff,vert}}$ bei $V_{\mathrm{gs}} = 1\,\mathrm{V}$ gerade im Übergang zwischen Coulomb- und Grenzflächenstreuung befindet.

Tabelle 5.5: Magnetowiderstands-Streukoeffizient für n- und p-Langkanaltransistoren unter Berücksichtigung von Serienwiderstands- und Geometriekorrekturen.

	L	μ_{MR}	μ_{eff}	r_{MR}
nFET	1 µm	$187\,\mathrm{cm}^2/\mathrm{Vs}$	$170\,\mathrm{cm}^2/\mathrm{Vs}$	$1,10$
pFET	1 µm	$89\,\mathrm{cm}^2/\mathrm{Vs}$	$80\,\mathrm{cm}^2/\mathrm{Vs}$	$1,12$

5.4 Zusammenfassung und Diskussion der verschiedenen Methoden zur Bestimmung der Beweglichkeit

Wie in den vorherigen Abschnitten geschildert, ist sowohl im Magnetowiderstandsverfahren, als auch bei der Methode der effektiven Beweglichkeit eine Serienwiderstandskorrektur durchzuführen. Um zu untersuchen, wie sensitiv die jeweiligen Verfahren auf Unsicherheiten im Serienwiderstand reagieren, ist es sinnvoll, die Gleichungen (5.8) und (5.17) auf die Form $\mu_\mathrm{x} = \mu_{\mathrm{x}(R_\mathrm{ser}=0)} * \gamma_{\mathrm{x}(R_\mathrm{ser})}$ zu bringen. Für die Korrekturfaktoren des Serienwiderstands γ_x ergibt sich

$$\gamma_{\mathrm{eff}(R_\mathrm{ser})} = \frac{1}{1 - {}^{R_\mathrm{ser}}/_{R_\mathrm{ges}}} = \frac{R_\mathrm{ser}}{R_\mathrm{ch}} + 1 \tag{5.18}$$

$$\gamma_{\mathrm{MR}(R_\mathrm{ser})} = \frac{1}{\sqrt{1 - {}^{R_\mathrm{ser}}/_{R_\mathrm{ges}}}} = \sqrt{\frac{R_\mathrm{ser}}{R_\mathrm{ch}} + 1} \tag{5.19}$$

Das Verhalten von γ_x ist in Abb. 5.25 in Abhängigkeit von R_ser dargestellt. Dabei wurde ein für die $L = 30\,\mathrm{nm}$ Kurzkanaltransistoren bei $V_\mathrm{gs} = 1\,\mathrm{V}$ und $V_\mathrm{ds} = 50\,\mathrm{mV}$ typischer Kanalwiderstand zugrunde gelegt. Während γ_eff linear mit R_ser steigt, nimmt γ_MR nur mit der Quadratwurzel und damit deutlich schwächer zu. Der insgesamt niedrigere Korrekturfaktor zeigt, dass die unkorrigierten Beweglichkeiten im MR-Verfahren näher an den tatsächlichen Werten liegen, als im Split-CV-Verfahren. Darüber hinaus übersetzt sich der Steigungsunterschied in Abb. 5.25 in eine deutlich kleinere Unsicherheit für die MR-Beweglichkeit. So führt ein Serienwiderstand von $R_\mathrm{ser} = 200 \pm 10\,\Omega{\cdot}\mu\mathrm{m}$ im Fall der effektiven Beweglichkeit zu einer Korrektur der Beweglichkeitswerte um 150 % nach oben, bei einer Unsicherheit von 15 %. Die Magnetowiderstandsdaten werden ebenfalls signifikant, aber dennoch deutlich moderater um 60 % nach oben korrigiert. Die durch die Korrektur methodisch eingebrachte Unsicherheit beläuft sich dabei nur noch auf 5 %.

Ein weiterer wichtiger Unterschied zwischen beiden Verfahren ergibt sich bei der Berücksichtigung der Transistorgeometrien. Beim Magneto-

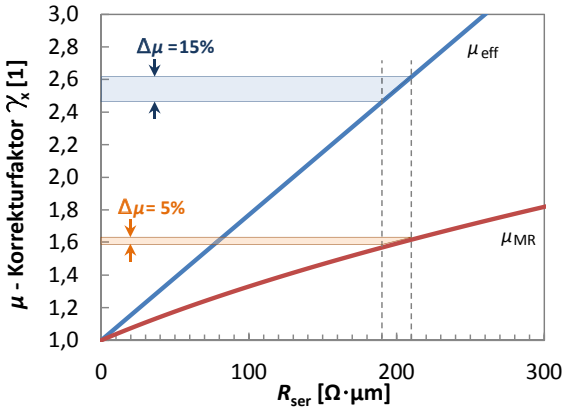


Abbildung 5.25: Serienwiderstands-Korrekturfaktor γ_x für die Extraktion der Beweglichkeit in Abhängigkeit von R_{ser} für einen typischen Kanalwiderstand am $L = 30\,\text{nm}$ Kurzkanaltransistor bei $V_{gs} = 1\,\text{V}$ und $V_{ds} = 50\,\text{mV}$. Bedingt durch die unterschiedlichen Steigungen ergeben sich aus einem Serienwiderstand von z. B. $200 \pm 10\,\Omega\cdot\mu\text{m}$ in R_{ser} je nach Verfahren verschiedene systematische Fehler für die resultierende Beweglichkeit von 5 % bzw. 15 % für μ_{MR} und μ_{eff}.

widerstandsverfahren gehen die Abmessungen des Kanals nur indirekt über den Geometriefaktor k_{gm} ein, siehe (5.3). Bei kurzen und weiten Transistoren mit $L \ll W$ ist $k_{gm} \approx 1$ und spielt faktisch kaum eine Rolle. Nur bei Langkanaltransistoren gibt k_{gm} eine wichtige Korrektur zur Magnetowiderstandsbeweglichkeit. Allerdings ist diese dann sehr genau bekannt, denn eine Unsicherheit bezüglich der effektiven Kanallänge im Bereich weniger Nanometer erzeugt bei großen L kaum Abweichungen in k_{gm}. Wichtig für das Magnetowiderstandsverfahren ist vor allem, dass zur Messung von Kurzkanalbeweglichkeiten Transistoren mit großen Weiten zur Verfügung stehen, sodass stets $L \ll W$ gewährleistet werden kann. Dies ist in der Praxis mit entsprechenden

Teststrukturen leicht zu gewährleisten. Anders liegt der Fall beim Split-CV-Verfahren. Hier stellt die Berücksichtigung der tatsächlich elektrisch effektiven Kanallänge $L_{\text{eff}} = L - \delta L$ eine wesentliche Korrektur dar, die gerade die kritische Kurzkanalbeweglichkeit beeinflusst. Allgemein ist für die Bestimmung von μ_{eff} die Kenntnis der genauen geometrischen Abmaße des Transistors erforderlich. Da diese wiederum aus statistischen oder systematischen Prozessschwankungen oder Variationen von Bauteil zu Bauteil über den Wafer variieren, können Messungen an einzelnen Transistoren stärker mit Fehlern behaftet sein. Ist δL nicht genau bekannt, können sich so schnell sehr große Fehler in μ_{eff} ergeben, da L direkt in die Beweglichkeit (3.8) eingeht. Dieser Anfälligkeit kann durch eine hohe Statistik an vielen Transistoren entgegen gewirkt werden. Die Tatsache, dass die hier angeführten Messungen an einzelnen Transistoren sehr gut mit den Magnetowiderstandsdaten übereinstimmen liegt daran, dass es sich hier um eine sehr präzise eingestellte und hinsichtlich der Prozessbedingungen in eng definierten Grenzen entwickelte Technologie handelt. Es stellt sich heraus, dass die effektive Beweglichkeit unter Berücksichtigung der genannten Korrekturen mit Annahmen in R_{ser} und δL bei der gegebenen Technologie als ein „gutes Maß“ für die Driftbeweglichkeit im linearen Operationsregime (kleine V_{ds}) gelten kann. Dass diese Aussage so getroffen werden kann wird auch mit dadurch ermöglicht, dass die messtechnisch deutlich aufwändigere Magnetowiderstandsmethode hinsichtlich der Absolutwerte einen guten Ankerpunkt liefert. Darüber hinaus ermöglicht das MR Verfahren im Gegensatz zu den anderen Methoden die Untersuchung des Sättigungsbereichs des Transistors. Abb. 5.26 stellt die Ergebnisse aus den verschiedenen Verfahren im linearen Operationsbereich mit $V_{\text{ds}} = 50\,\text{mV}$ bei $V_{\text{gs}} = 1\,\text{V}$ für den nFET gegenüber. In der linken Spalte des linken Graphs sind die MR Daten μ_{MR} gezeigt, die als Anker- bzw. Referenzpunkt zur Bewertung der anderen Verfahren dienen. In der mittleren Spalte befinden sich die Split-CV-Beweglichkeiten für verschiedene δL. Hierbei wird deutlich, dass die Kenntnis der effektiven Kanallänge eine wichtige Rolle spielt, um zu korrekten Kurzkanalbeweglichkeiten zu kommen. So werden die Daten aus unkorrigierten Kurzkanaltransistoren mit $\delta L = 0$ deutlich überschätzt. Demgegen-

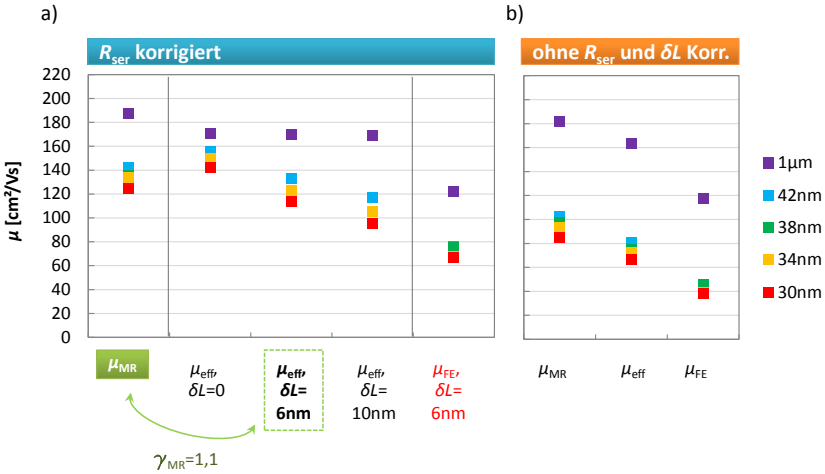


Abbildung 5.26: Beweglichkeiten aus den verschiedenen Methoden des MR-, Split-CV- und Feldeffektverfahrens am TPEN nFET unter linearen Operationsbedingungen bei $V_{ds} = 50\,\text{mV}$, $V_{gs} = 1\,\text{V}$ im Vergleich: a) mit angewandter R_{ser}-Korrektur und Einfluss der effektiven Kanallänge auf μ_{eff}, b) ohne R_{ser}- und δL-Korrektur.

über führt die Annahme von 5 nm Unterdiffusion an Source und Drain mit insgesamt $\delta L = 10\,\text{nm}$ auf 25 % zu niedrige Werte. Die korrekte Lage der Beweglichkeiten für 3 nm Unterdiffusion bzw. $\delta L = 6\,\text{nm}$ bestätigt im Vergleich mit μ_{MR}, dass sowohl δL, als letztlich auch R_{ser} hinreichend genau bestimmt werden konnten, um zu präzisen Split-CV-Beweglichkeiten μ_{eff} zu kommen. Der kleine Versatz zwischen μ_{MR} und μ_{eff} ist dabei auf den Magnetowiderstands-Streufaktor γ_{MR} zurückzuführen. Er beträgt hier, genau wie bei den Langkanaltransistoren (vgl. Tabelle 5.5) $\gamma_{MR} = 1,1$. Der Beweglichkeits-roll-off der 4 kürzesten Gatelängen fällt beim Split-CV- und beim MR-Verfahren unter diesen Bedingungen gleich hoch aus, und beträgt bei $V_{gs} = 1\,\text{V}$ und $V_{ds} = 50\,\text{mv}$ etwa $1\,\text{cm}^2/(\text{Vs} \cdot \text{nm})$. Der Vollständigkeit halber sind in der rechten

Spalte ergänzend noch die bereits R_{ser}- und δL-korrigierten Feldeffektbeweglichkeiten gezeigt. Hier ergeben sich sowohl für die Kurz-, als auch für die Langkanaltransistoren trotz der eingebrachten Korrekturen keine quantitativ aussagekräftigen Werte. μ_{FE} eignet sich daher bestenfalls als Trendindikator, um in einer Fertigungsumgebung unter minimalem Messaufwand festzustellen, ob bestimmte Prozesse einen positiven oder negativen Einfluss auf die Beweglichkeit haben.

6 Ergebnisse der Untersuchungen an SiNW Transistoren

6.1 Struktureller Aufbau

Die Si-Nanodrahtproben wurden nach erfolgter Prozessierung im Rasterelektronenmikroskop untersucht. Dies diente sowohl der Lokalisation potentiell funktionaler Transistorstrukturen, als auch der Bestimmung der zugehörigen Kanallängen. Außerdem konnte so sichergestellt werden, dass es sich um Einzeltransistoren mit beidseitig einsilizidierten Source- und Drainanschlüssen handelt und nicht um mehrere Drähte in Parallelschaltung. Abb. 6.1 zeigt exemplarisch den Transistor, an dem die Untersuchung der Schottkybarrieren durchgeführt wurde. Die links und rechts am Rand mit „Source“ und „Drain“ beschrifteten, großen Nickelgebiete führen auf ca. 60 µm breite Kontaktflächen für die elektrische Kontaktierung der Struktur über Messnadeln. Gut sichtbar ist die metallische $NiSi_2$-Phase, welche sich beim Einsilizidieren des Nickels aus den Source- und Drain-Reservoirgebieten in den Draht gebildet hat. An den Übergangspunkten zum verbleibenden intrinsischen Silizium des Kanals befinden sich die Source- und Drain-Schottkybarrieren. Die durch die räumliche Separation der Barrieren gegebene Kanallänge beträgt 500 nm.[1] Das sourceseitig am eigentlichen Transistordraht anliegende Nanodraht-Fragment hat auf die elektrischen Eigenschaften des Transistors keinen Einfluss, da es sich zum einen im Bereich des Sourcekontakts mit metallischem $NiSi_2$ Kern befindet, in dem das

[1] Berichtigter Wert gegenüber [97]. Da die Kanallänge nicht in die Auswertung eingeht, bleiben die Aussagen hiervon unberührt.

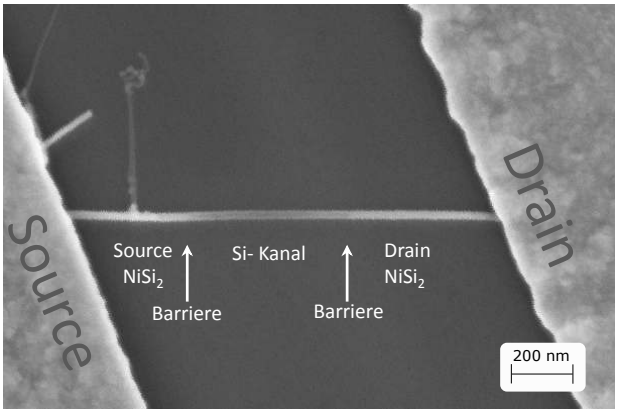


Abbildung 6.1: Rasterelektronenaufnahme des ab Kap. 6.2.2 diskutierten SiNW Transistors mit aufgewachsener Oxidhülle und Rückseitengate. Der Transistorkanal befindet sich zwischen den beiden mit Pfeilen markierten Schottkyübergängen, bzw. -barrieren.

Ferminiveau durch die Austrittsarbeit des Metalls und die angelegte Spannung festgelegt sind. Zum anderen trennen die Oxidhüllen beider Drähte die Kerne elektrisch voneinander. Sowohl der Draht, als auch die Source- und Drain Nickelgebiete liegen auf einer 115 nm dicken SiO_2-Schicht, welche als Gateoxid dient. Die Steuerung des Transistors erfolgt über das Anlegen einer Rückseitengatespannung am Substrat. Eine schematische, räumliche Darstellung des Probenaufbaus befindet sich in Abb. 2.7.

6.2 Schaltverhalten

6.2.1 Hysterese

Ein idealer Nanodraht-Transistor mit Rückseitengate würde eine Transfercharakteristik ähnlich der in Abb. 2.11 gezeigten Simulation aufweisen. In der Realität zeigen sich jedoch aufgrund von parasitären Effek-

ten Abweichungen von diesem idealen Verhalten, die im Wesentlichen auf spontanen Ladungseinfang bzw. spontane Ladungsabgabe an Störstellen im Halbleiter, an der Halbleiter/Oxid-Grenzfläche sowie an der Oberfläche des Nanodrahts zurückzuführen sind [98]. Das Schaltverhalten von Silizium-Nanodraht-Transistoren reagiert insbesondere sensitiv auf Absorption und Desorption polarer Moleküle wie z. B. Wasser an der Drahtoberfläche. Dies führt je nach den vorangegangenen Operationsbedingungen, insbesondere der jeweils angelegten Gatespannung V_{gs}, während des Betriebs zu Verschiebungen des effektiven Feldes V_{eff} an den Schottkybarrieren. Hierüber verschiebt sich stetig die Einsatzspannung, wodurch es zur Ausbildung eines Hystereseverhaltens in den Kennlinien kommt (Abb. 6.2).

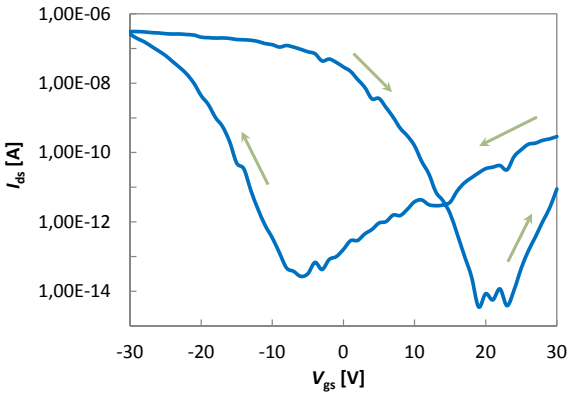


Abbildung 6.2: Strom-Spannungskennlinie $I_{\mathrm{ds}}(V_{\mathrm{gs}})$ eines unpassivierten SiNW Transistors bei 300 K. Die Oxiddicke zwischen Rückseitengate und Nanodraht beträgt $d_{\mathrm{ox}} = 300\,\mathrm{nm}$. V_{gs} wurde gemäß den Pfeilen im Schema $30\,\mathrm{V} \ldots -30\,\mathrm{V} \ldots 30\,\mathrm{V}$ durchgestimmt, $V_{\mathrm{ds}} = 500\,\mathrm{mV}$. Obwohl die Probe *in situ* ausgeheizt und gemessen wurde, tritt aufgrund der unpassivierten Oberfläche eine sehr starke Hysterese auf.

Durch kontrolliertes Aufwachsen einer definierten Oxidhülle um den Draht wird eine Passivierung erreicht, die eine starke Reduktion des Hystereseverhaltens bewirkt und dadurch einen wichtigen Schritt hin zu einem klar definiertem Schaltverhalten darstellt. Die aufgewachsene Oxidhülle sorgt für die elektrische Passivierung und trägt darüber hinaus eine kompressive Verspannung in den Nanodraht ein, welche radial auf den Kern wirkt. Nach der Oxidation erfolgt ein Temperschritt unter Formiergasatmosphäre, um ggf. noch vorhandene, freie ungesättigte Bindungen zu passivieren. Die Details zur Prozessierung sind in Kap. 2.2.1 beschrieben. Die durch die Oxidhülle eingebrachte Passivierung reduziert die Kennlinienverschiebung während einer Messung aufgrund von Hystereseeffekten gegenüber unpassivierten Transistoren um über 80 %, vgl. dazu Abb. 6.2 mit Abb. 6.3.[1] Letztere zeigt eine typische Strom-Spannungs-Kennlinie des im folgenden untersuchten Nanodraht-Transistortyps mit aufgewachsener Oxidhülle. Wird die Messung unter Umgebungsbedingungen durchgeführt, zeigt sich noch eine leichte Hysterese, die im Wesentlichen auf adsorbierte Moleküle an der Oberfläche der Hülle zurückzuführen ist [99]. Diese können durch Ausheizen der Probe im Vakuum leicht entfernt werden (Abb. 6.3).

6.2.2 Transfer- und Ausgangskennlinien

Auch wenn durch die im vorangegangenen Abschnitt beschriebenen Maßnahmen die Hysterese beinahe vollkommen eliminiert werden konnte, zeigte das Kennlinienspektrum des untersuchten SiNW-Transistors eine konstante Verschiebung um ca. $1,3\,\mathrm{V}$ hin zu negativen V_{gs}-Werten. Diese Verschiebung blieb gleich über alle verwendeten Drain-Source-Spannungen und Temperaturen. Ein solches Verhalten kann durch eine ungewollte Aufladung von Störstellen, z.B durch elektrische Messung oder Untersuchungen im Rasterelektronenmikroskop, sowie durch während der Prozessierung eingebaute, feste Ladung zustande kommen. Die Ladungen können sich dabei an der Grenzfläche zwischen Siliziumkern

[1] Obwohl es sich in beiden Grafiken um Proben mit unterschiedlicher Gateoxiddicke handelt, sind die vertikalen Felder im dargestellten V_{gs}-Bereich in etwa vergleichbar: $E_{\mathrm{gs},1} \approx V_{\mathrm{gs},1}/d_{\mathrm{ox},1} = 30\,\mathrm{V}/300\,\mathrm{nm} = 1\,\mathrm{MV/cm}$ (Abb. 6.2), $E_{\mathrm{gs},2} \approx 20\,\mathrm{V}/200\,\mathrm{nm} = 1\,\mathrm{MV/cm}$ (Abb. 6.3).

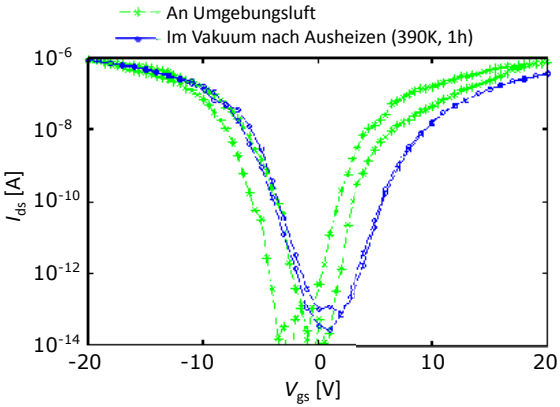


Abbildung 6.3: Transferkennlinien $I_{\mathrm{ds}}(V_{\mathrm{gs}})$ eines passivierten Si-Nanodraht-Transistors mit aufgewachsener Oxidhülle bei 300 K: *ex situ* nach Prozessierung und *in situ* nach Ausheizen für 1 h bei 390 K im Vakuum. V_{gs} wurde von $-20\,\mathrm{V} \ldots +20\,\mathrm{V} \ldots -20\,\mathrm{V}$ durchgestimmt, $d_{\mathrm{ox}} = 200\,\mathrm{nm}$. Hysterese und „Linksverschiebung“ des gesamten Spektrums konnten durch den Ausheizschritt im Wesentlichen beseitigt werden.

und Oxidhülle, innerhalb der Gateoxide, oder auf dem Draht befinden. Diesem Verhalten wird mit der Konstante c in (2.32) Rechnung getragen. Die Kennlinien wurden für die weitere Auswertung um den Versatz von $1,3\,\mathrm{V}$ zurückgeschoben, um eine Zentrierung des ambipolaren Verhaltens um $V_{\mathrm{gs}} = 0\,\mathrm{V}$ zu erreichen. Da die Anstiege der nFET- und pFET-Äste nicht gleich stark sind und auf der Probe keine Topgates zur Verfügung standen, konnte die Bestimmung des richtigen Versatzes nur ungefähr erfolgen. Es sei jedoch angemerkt, dass auch Kennlinien, die augenscheinlich „zentriert“ zu sein scheinen, leicht verschoben sein können. Eine Bestimmung der Barrierenhöhen aus dem Achsenabschnitt des thermionischen Anstiegs (2.35) ist daher grundsätzlich

schwierig, selbst wenn der thermionische Bereich eindeutig identifiziert werden kann. Die (Rück-) Verschiebung der Kennlinien beeinflusst hingegen nicht die Berechnung der effektiven Barrierenhöhen nach der hier gezeigten Methode. Die im Folgenden diskutierten Daten wurden alle an dem einzelnen SiNW Transistor mit aufgewachsener Oxidhülle aus Abb. 6.1 gemessen.

Abbildung 6.4 zeigt die Transferkennlinien bei 300 K für verschiedene V_{ds} in halblogarithmischer Darstellung, wobei der „Rauschbereich“ mit sehr geringen Strömen zwischen den pFET und nFET Kennlinienästen zugunsten der Übersichtlichkeit herausgeschnitten wurde. Es fällt auf, dass der charakteristische Knick am Übergang zwischen thermionischer Emission und Tunneln in der ambipolaren Kennlinienschar nicht sichtbar wird. Wie in der Simulation in Abb. 2.11 gezeigt, wird dieser Übergangspunkt leicht durch zum jeweils anderen Ast gehörende Tunnelströme maskiert. Der pFET zeigt ein steileres Einschaltverhalten als der nFET, da der nFET stärker durch Tunneln dominiert wird. Eine

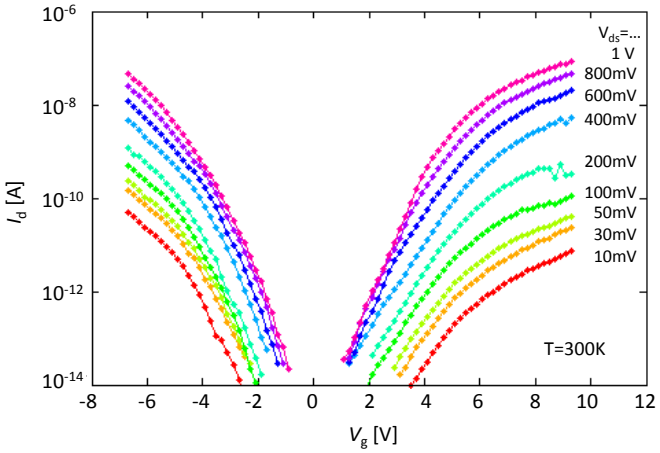


Abbildung 6.4: Strom-Spannungs-Kennlinien $I_{\mathrm{ds}}(V_{\mathrm{gs}})$ des undotierten SiNW Transistors aus Abb. 6.1 mit Oxidhülle und Rückseitengate bei 300 K für verschiedene V_{ds}

weitere, für den Transistor wichtige Charakteristik ist das in Abb. 6.5 für kleine V_{ds} dargestellte Ausgangsverhalten $I_{\mathrm{ds}}(V_{\mathrm{ds}})$ bei festem V_{gs}. Im Vergleich zur linearen MOSFET-Ausgangscharakteristik bei kleinen V_{ds} zeigt der Si-Nanodraht-Transistor grundsätzlich einen „supralinearen“, also weniger steilen Anstieg. Dieses Verhalten kann auf die Energiebarrieren an Source und Drain zurückgeführt werden, welche trotz des „geöffneten“ Gates ($V_{\mathrm{gs}} < V_{\mathrm{th,pF}}$) im Kanal vorhanden sind. Der Strompfad im MOSFET-Kanal weist hingegen keine Energiebarriere auf und ermöglicht somit im unteren Bereich der Ausgangskennlinie höhere Ströme, wobei der Anstieg über das Ohm'sche Gesetz linear erfolgt. Im Nanodraht-Transistor werden die Barrieren nach (2.32) mit steigendem V_{ds} schmäler. Wie später erläutert wird, steigt dabei die Tunnelwahrscheinlichkeit exponentiell an, wodurch es zu einem diodenartigen Anstiegsverhalten kommt. Im Hinblick auf Schaltungsentwürfe wirkt sich diese Eigenschaft ungünstig aus. Werden zwei Transistoren in Serie betrieben, vermindert die Nichtlinearität ein effektives Umladen des Potentials zwischen den Transistoren und erhöht somit die erforderlichen Schaltzeiten. Aus einem Vergleich verschiedener Nanodraht-Geometrien ergibt sich für dünnere Drähte eine stärkere Linearität im Ausgangsverhalten [100]. Eine weitere Linearisierung kann durch bessere Gatekopplung erreicht werden, z. B. durch die Verwendung von dünneren Oxiden und Topgate-Elektroden. Letztere werden nachträglich von oben z.B. über die Schottkybarrieren aufgebracht und legen sich um den Draht. Hierdurch wirkt das elektrische Feld des Gates von (fast) allen Seiten auf den Draht und nicht nur aus einer Richtung wie beim Rückseitengate. Die Verbesserung der Feldeinkopplung führt zu einer stärkeren Bandverbiegung im Silizium durch die angelegten externen Spannungen. Die räumliche Ausdehnung der Bandverbiegung spielt sich dabei im Bereich der „Screening-Länge“, auch „natürliche Länge“ λ genannt, ab [101]. Bei kleineren Werten λ wird die Breite der Energiebarrieren reduziert und das Tunneln nimmt zu. In Bereichen, in denen die Barrierenbreite (bzw. λ) die Tunnellänge von Elektronen in Si erreicht, liegt die Tunnelwahrscheinlichkeit bei Ihrem Maximalwert $\tilde{t} = 1$. Hier führt eine weitere Reduktion der Barrierenbreite, wie sie z. B. durch höheres V_{ds} erreicht werden kann, nicht zu einer größeren Transparenz der

Barriere, wodurch die Ausgangskennlinie $I_{\mathrm{ds}}(V_{\mathrm{ds}})$ linearisiert wird. Dies ist eine intrinsische Eigenschaft von quasi-eindimensionalen Strukturen mit ausreichender Gatekontrolle.

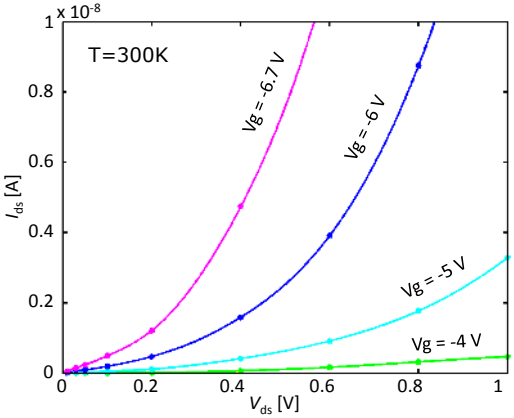


Abbildung 6.5: Strom-Spannungskennlinien $I_{\mathrm{ds}}(V_{\mathrm{ds}})$ des untersuchten SiNW Transistors aus Abb. 6.1 für verschiedene V_{gs} im pFET Bereich, $T = 300\,\mathrm{K}$.

6.2.3 Mikroskopisches Modell zum Verständnis der Temperaturabhängigkeit des Drainstroms

Da gängige Modelle wie [102] die Temperaturabhängigkeit des Drainstroms im Tunnelregime nur mit starken Einschränkungen erklären, wurde die nachfolgende Betrachtung entwickelt. Sie veranschaulicht dabei vor allem den sonst vernachlässigten Aspekt, dass der Tunnelbeitrag maßgeblich für die Temperaturabhängigkeit des Drainstroms bei festem V_{gs} und V_{ds} verantwortlich ist. Ein genaues, quantitatives Modell des Systems stellt ein sehr schwieriges und umfangreiches Unterfangen dar. Daher soll an dieser Stelle stattdessen ein einfacher, empirischer Ansatz diskutiert werden, welcher die maßgeblichen physikalischen Aspekte wiedergibt und die Temperaturabhängigkeit sowie

das Strom-Spannungsverhalten qualitativ erklärt. Hierzu ist es ausreichend, einen eindimensionalen Stromtransport über die Barriere zu betrachten. Unter den weiteren Annahmen konstanter Zustandsdichten in Metall und Halbleiter stammt die hauptsächliche Temperaturabhängigkeit des Tunnelstroms aus dem Energieintegral $\int f(\epsilon, T)\tilde{t}(\epsilon)\mathrm{d}\epsilon$, wobei $f(\epsilon, T)$ die Fermi-Dirac-Verteilung der Elektronen im Metall

$$f_{(\epsilon,T)} = \frac{1}{1 + e^{(\epsilon - \epsilon_\mathrm{F})/k_\mathrm{B}T}} \tag{6.1}$$

und

$$\tilde{t}(\epsilon, E) \sim \exp\left(-K\frac{(\phi_\mathrm{B} - \epsilon)^{3/2}}{qE}\right)$$

die Tunnelwahrscheinlichkeit aus (2.36) darstellt. Abbildung 6.6 zeigt das Ergebnis einer groben Berechnung der analytischen Ausdrücke $f(\epsilon, T)$ und $\tilde{t}(\epsilon, E)$, sowie das Produkt $f(\epsilon, T) \cdot \tilde{t}(\epsilon, E)$ für niedrige und hohe Temperaturen, a) für eine „breite“ und b) für eine „schmale“ Barriere. Das Modell berücksichtigt ebenfalls den Strombeitrag aufgrund von thermionischer Emission. Dies geschieht dadurch, dass die Tunnelwahrscheinlichkeit $\tilde{t}(\epsilon, E) = 1$ gesetzt ist für alle Elektronen, deren Energie oberhalb des (konstanten) Barrieremaximums liegt, also mit $\epsilon > \phi_\mathrm{B}$. Im Fall einer „breiten“ Barriere liefern die Elektronen mit Energien leicht unterhalb des Barrieremaximums ϕ_B den Hauptbeitrag zum Tunnel- wie auch Gesamtstrom – und zwar sowohl bei niedrigen, als auch bei hohen Temperaturen. Allerdings ergibt sich in Bezug auf den Gesamtstrom bei hohen Temperaturen ein höherer Beitrag durch thermionische Emission als bei niedrigen (vgl. hierzu in Abb. 6.6a die roten und blauen Temperaturkurven im rein thermionischen Bereich, oberhalb von $\epsilon > 0,5\,\mathrm{eV}$). Hieraus ergibt sich eine starke Temperaturabhängigkeit des Energieintegrals $\int f(\epsilon, T) \cdot \tilde{t}(\epsilon, E)\mathrm{d}\epsilon$ und damit des Gesamtstroms. Im Fall einer „schmalen“ Barriere dagegen ist diese Abnahme des Gesamtstroms mit der Temperatur weniger ausgeprägt, da das Maximum des Produkts $f(\epsilon, T) \cdot \tilde{t}(\epsilon, E)$ energetisch weit unterhalb des Barrieremaximums ϕ_B und damit näher am Ferminiveau ϵ_F liegt.

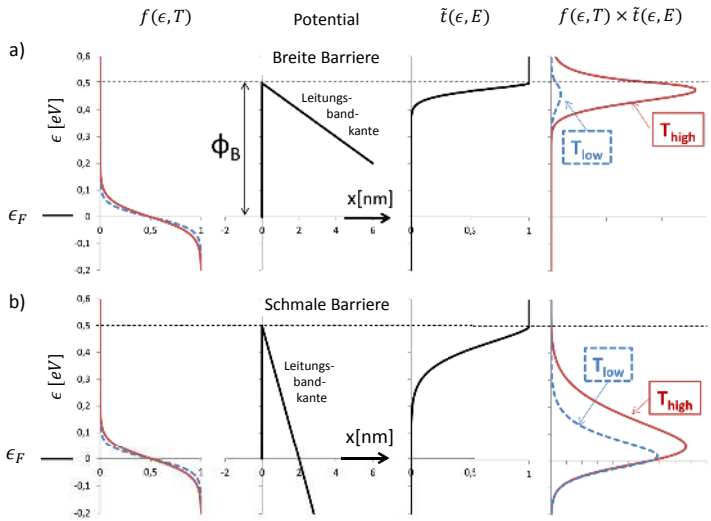


Abbildung 6.6: Einfaches, analytisches Modell zum Verständnis des Tunnelstroms durch eine dreieckige Barriere am Beispiel von Elektroneninjektion am sourceseitigen Schottkykontakt. Von links nach rechts skizziert sind die Fermi-Dirac-Verteilung $f(\epsilon,T)$ der Elektronen im Metall, das lokale, feldabhängige Barrierepotential $\phi(x)$, die energieabhängige Tunnelwahrscheinlichkeit $\tilde{t}(\epsilon,E)$ sowie das Produkt $f(\epsilon,T)\cdot\tilde{t}(\epsilon,E)$ im Fall niedriger (blau) und hoher (rot) Temperaturen, für a) eine „breite“ und b) eine „schmale“ Barriere. Die Lage des Ferminiveaus dient als Referenzpunkt auf der Energie- bzw. Potentialachse, die Höhe der Schottky Barriere für Elektronen wurde willkürlich bei $\phi_B = 0,5\,\mathrm{eV}$ gewählt.

Hier dominiert der Tunnelstrom den Gesamtstrom. Das Energieintegral über $f(\epsilon, T) \cdot \tilde{t}(\epsilon, E)$, welches den Gesamtstrom bestimmt, zeigt eine geringere Temperaturabhängigkeit als im Fall einer „breiten“ Barriere. Dieses Tunnelregime mit schwächerer Temperaturabhängigkeit wird in der Literatur häufig als Bereich der Feldemission bezeichnet, wobei der Tunnelmechanismus dem im MOSFET bekannten Fowler-Nordheim Tunneln ähnelt. Die schwache Temperaturverteilung wird durch die dünner werdende Barriere für höhere Energien verursacht. Um die Temperaturabhängigkeit des Gesamtdrainstroms zu beschreiben, kann der Ansatz

$$I_{\mathrm{ds}}(V_{\mathrm{gs}}, V_{\mathrm{ds}}) \sim T^2 \exp\left(\frac{-\phi_{\mathrm{eff}}(V_{\mathrm{gs}}, V_{\mathrm{ds}})}{k_{\mathrm{B}}T}\right) \tag{6.2}$$

im Sinne einer Energieaktivierung des Gesamtsystems verwendet werden. Hierbei stellt $\phi_{\mathrm{eff}}(V_{\mathrm{gs}}, V_{\mathrm{ds}})$ eine phänomenologische, effektive Energiebarriere dar, welche durch externe elektrische Felder moduliert wird. ϕ_{eff} kann dabei als Aktivierungsenergie des gesamten Systems aufgefasst werden. Sie soll als Maß dienen, um die verschiedenen Strommechanismen voneinander zu differenzieren. Der Vorfaktor T^2 gilt für ein dreidimensionales Modell, welches alle drei Raumkomponenten der Ladungsträgergeschwindigkeit berücksichtig. Im hier untersuchten Temperaturbereich führt dieser jedoch im Vergleich zum Exponentialterm in (6.2) nur zu einer vernachlässigbar kleinen Temperaturabhängigkeit von I_{ds}. Generell gilt, dass je größer $\phi_{\mathrm{eff}}(V_{\mathrm{gs}}, V_{\mathrm{ds}})$ wird, desto stärker fällt die Temperaturabhängigkeit von I_{ds} für eine feste Spannungseinstellung aus. Im Unterschwellstrombereich ist die Temperaturabhängigkeit von I_{ds} am stärksten.

Wie bereits erwähnt, wurden zur einfachen Beschreibung des thermisch aktivierten Tunnelstroms in Schottkybarrieren FETs in der Literatur auch alternative Modelle vorgeschlagen. So wird dieser Prozess im Modell von Knoch et al. [102] durch thermionische Emission über eine fiktive Barrierenhöhe beschrieben, die ebenfalls „effektive Barrierenhöhe“ genannt wird. Thermisch angeregte, Fermi-Dirac-verteilte Ladungsträger mit Energien über dieser effektiven Barrierenhöhe tunneln mit einer Wahrscheinlichkeit von $\tilde{t} = 1$ durch den ausgedünnten „obe-

ren“ Teil der Barriere. Die effektive Barrierenhöhe wird dabei an dem Punkt festgesetzt, an dem die Barrierenbreite $d(\epsilon)$ kleiner der Tunneldistanz in Si $d_{\mathrm{t-Si}} = 3,6\,\mathrm{nm}$ wird. Für Energien unter ϕ_{eff} wird die Tunnelwahrscheinlichkeit $\tilde{t}(\epsilon)$ diskret auf Null gesetzt. Durch diese Annahme schließt das Modell aus, dass Ladungsträger unterhalb ϕ_{eff} zur Leitung beitragen. Im Gegensatz dazu berücksichtigt das in dieser Arbeit entwickelte Modell die energetischen Beiträge unterhalb ϕ_{eff}. Im Fall einer dünnen Barriere ist die Tunnelwahrscheinlichkeit einzelner Elektronen tatsächlich $\tilde{t}(\epsilon) \ll 1$ für die stromtragenden elektronischen Zustände, aber es ist die Gewichtung mit der Fermiverteilung $f(\epsilon, T)$, die diese wichtigen Energieniveaus auswählt. Somit berücksichtigt der hier vorgestellte Ansatz die natürliche Temperaturabhängigkeit des Integrals $\int f(\epsilon, T) \cdot \tilde{t}(\epsilon)\,\mathrm{d}\epsilon$ und reduziert den Leitungsmechanismus nicht auf ein starres Übergangskriterium zwischen $\tilde{t}(\epsilon) = 1$ und $\tilde{t}(\epsilon) = 0$. Der Unterschied zwischen beiden Modellen zeigt sich dabei nicht in der phänomenologischen Beschreibung der Temperaturabhängigkeit über eine effektive Barrierenhöhe, sondern kommt vor allem in der mikroskopischen Interpretation der extrahierten Barrierenhöhen ϕ_{eff} zum Tragen.

6.3 Extraktion der Schottkybarriere

6.3.1 Temperaturabhängigkeit der Transferkennlinien

Der untersuchte SiNW Transistor wurde im Temperaturbereich zwischen 280 K und 300 K unter Vakuum gemessen. Abb. 6.7 zeigt die Temperaturabhängigkeit der Transfercharakteristik bei $V_{\mathrm{ds}} = 50\,\mathrm{mV}$ in linearer (Abb. 6.7a) und halblogarithmischer (Abb. 6.7b) Darstellung. Wie von der Simulation vorhergesagt, sind Tunnelströme dominant und der rein thermionische Bereich in den Kennlinien nicht identifizierbar. Außerdem ist sichtbar, wie höhere Temperaturen auch in diesem durch Tunnelstrom dominierten System zu einer starken Zunahme des Drainstroms führen. Dieses Verhalten kann im Kontext des in Abschnitt 6.2.3 entwickelten Modells erklärt werden. Die Bestimmung von $V_{\mathrm{th,pF}}$ und $V_{\mathrm{th,nF}}$ erfolgte durch lineare Extrapolation der Kennlinien auf $I_{\mathrm{ds}} = 0\,\mathrm{V}$ und ergab Werte um $\pm 4,5\,\mathrm{V}$ (Abb. 6.7a).

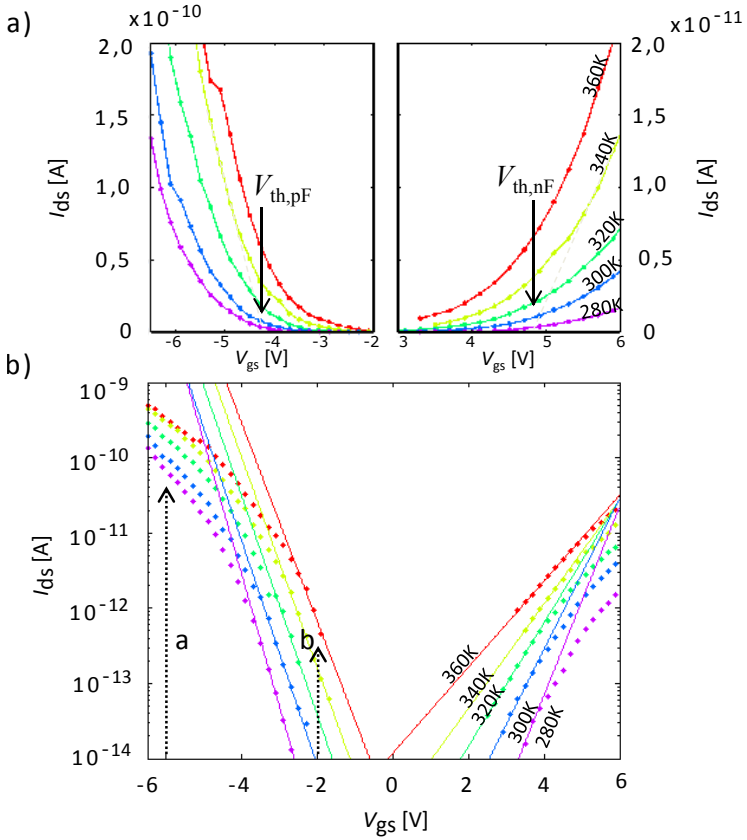


Abbildung 6.7: Drainstrom I_{ds} in Abhängigkeit der Gatespannung V_{gs} bei $V_{\mathrm{ds}} = 50\,\mathrm{mV}$ für verschiedene Temperaturen a) in linearer und b) in halblogarithmischer Darstellung. Die Pfeile 'a' und 'b' markieren die Positionen der zugehörigen Schnitte durch die Messdaten in 6.8.

6.3.2 Einfluss externer Felder auf die effektive Barrierenhöhe

Um den in Abb. 6.7 stattfindenden Prozess der thermischen Aktivierung zu analysieren und ein tieferes Verständnis des Transports durch die Schottkybarrieren des Nanodrahts zu erlangen, werden phänomenologische, effektive Barrierenhöhen ϕ_{eff} über einen weiten Bereich von V_{ds} und V_{gs} extrahiert. Hierzu dient der folgende, aus (6.2) abgeleitete Ausdruck für die effektive Barriere,

$$\phi_{\mathrm{eff}}(V_{\mathrm{gs}}, V_{\mathrm{ds}}) = -k_{\mathrm{B}} \frac{\mathrm{d}[\ln(I_{\mathrm{ds}}/T^2)]}{\mathrm{d}(1/T)}. \tag{6.3}$$

Dieser Ansatz macht die Barrierenextraktion robuster gegen sporadische, laterale Verschiebungen der zugrundeliegenden Transferkennlinien $I_{\mathrm{ds}}(V_{\mathrm{gs}})$, wie sie z. B. durch Ladungseinfang zustande kommen können. Basierend auf (6.3) können die temperaturabhängigen Drainströme halblogarithmisch als I_{ds}/T^2 über $1000/T$ aufgetragen werden. Ein solcher Arrhenius-Graph ist für die pFET-Seite des Kennlinienfelds in Abb. 6.8 dargestellt. Dabei entsprechen die Datenpunkte in den Schnitten 'a' und 'b' in Abb. 6.7 den entsprechend markierten Linien in Abb. 6.8. Durch identische Verarbeitung der temperaturabhängigen Transferkennlinien bei anderen V_{ds} ergibt sich so die Gesamtheit der Datenpunkte in Abb. 6.8. Aus den Steigungen der sich im Arrhenius-Graph ergebenden Geraden kann die Barrierenhöhe ϕ_{eff} berechnet werden. Die Analyse wird hier auf zwei besondere Gatespannungen begrenzt: einen mit 'a' gekennzeichneten AN-Zustand bei $V_{\mathrm{gs}} = -6\,\mathrm{V}$, sowie einen mit 'b' gekennzeichneten AUS-Zustand im noch nicht durch Elektronentransport dominierten pFET Unterschwellbereich bei $V_{\mathrm{gs}} = -2\,\mathrm{V}$. Wie für einen Schottky-FET zu erwarten zeigt sich, dass die Temperaturabhängigkeit im Unterschwellbereich am stärksten ausgeprägt ist (siehe untere Kurvenschar in Abb. 6.8). Nichtsdestotrotz kann auch in den meisten zum AN-Zustand gehörenden Kurven eine signifikante Temperaturabhängigkeit beobachtet werden (obere Kurvenschar in Abb. 6.8).

Darüber hinaus wird sichtbar, dass die Temperaturabhängigkeit mit zunehmendem V_{ds} abnimmt. Für die höchsten anliegenden Potentiale bei $V_{\mathrm{ds}} = 1\,\mathrm{V}$ und $V_{\mathrm{gs}} = -6\,\mathrm{V}$ ist I_{ds}/T^2 schließlich nahezu konstant.[1]

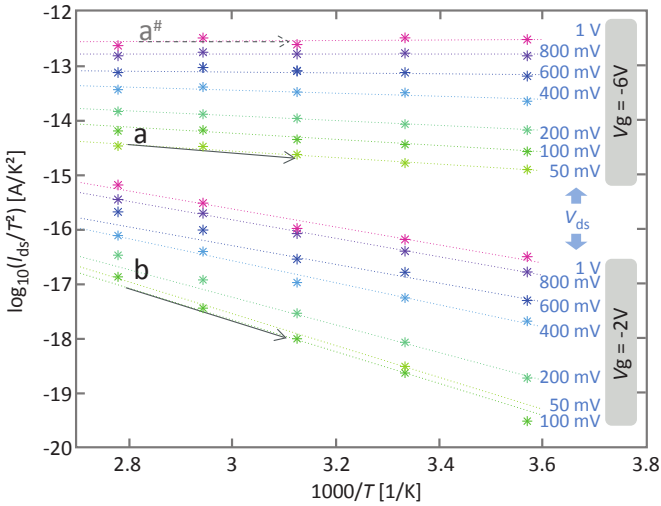


Abbildung 6.8: Arrhenius Graph (I_{ds}/T^2 über $1000/T$) für $V_{\mathrm{gs}} = -2\,\mathrm{V}$ und $V_{\mathrm{gs}} = -6\,\mathrm{V}$ auf der pFET-Seite des temperaturabhängigen $I_{\mathrm{ds}}(V_{\mathrm{gs}})$ Datensatzes. Die Punkte auf den Geraden 'a' und 'b' entsprechen den Messpunkten entlang der eingetragenen Pfeile in 6.7. Mit steigenden Werten für V_{ds} und $|V_{\mathrm{gs}}|$ reduziert sich die effektive Barriere, wie aus der Abnahme der Steigungen zwischen den Geraden b – a und a – a$^{\#}$ ersichtlich wird.

[1] Es sei angemerkt, dass aufgrund der Symmetrie der Banddiagramme für negatives und positives V_{ds} am Nanodraht-Transistor zwar die Richtung, nicht aber die Höhe von I_{ds} durch das Vorzeichen der Drainspannung beeinflusst werden. Elektronen- und Löchertransport lassen sich daher grundsätzlich mit der gleichen Drainspannung untersuchen. Dies steht in starkem Gegensatz zum MOSFET, bei dem das Vorzeichen der Drainspannung sowohl das Verhalten der Drain-Substrat-Diode, als auch – über den Overdrive – die Kanalladung beeinflusst.

Im nächsten Schritt können aus den Steigungen nach (6.3) die jeweiligen effektiven Barrierenhöhen bestimmt werden. Durch analoge Anwendung dieses Verfahrens auf der nFET-Seite der Transferkennlinien bei $V_{\mathrm{gs}} = 2\,\mathrm{V}$ und $V_{\mathrm{gs}} = 6\,\mathrm{V}$ ergibt sich das für Elektronen- und Löchertransport zusammengefasste Resultat in Abb. 6.9.

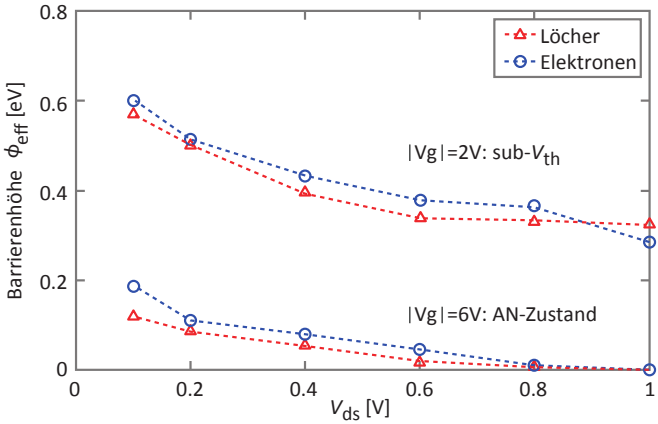


Abbildung 6.9: Effektive Barrierenhöhe ϕ_{eff} in Abhängigkeit der Drainspannung im Unterschwellbereich „sub-Vth“ bei $|V_{\mathrm{gs}}| = 2\,\mathrm{V}$ und im AN-Zustand bei $|V_{\mathrm{gs}}| = 6\,\mathrm{V}$. Während im sub-$V_{\mathrm{th}}$-Bereich thermionische Emission und Tunneln vorherrschen, ist der Injektionsmechanismus im AN-Zustand durch reines Tunneln bestimmt. Auch wenn sich der Transistor bei $|V_{\mathrm{gs}}| =$ konst. in ähnlichen Operationsbedingungen befindet, können die extrahierten Barrieren für Elektronen und Löcher aufgrund der Unsicherheiten in der Zentrierung der Strom-Spannungskennlinien nicht direkt verglichen werden.

Es zeigt sich, dass ϕ_{eff} in der Tat stark von den angelegten Spannungen abhängt. In Übereinstimmung mit dem Modell aus Abschnitt 6.2.3 ist ϕ_{eff} für Elektronen und Löcher im Unterschwellbereich (b) viel größer, als im AN-Zustand (a). Auch wenn der rein thermionische Bereich

in der ambipolaren Transferkennlinie nicht zugänglich ist, sollten die effektiven Barrierehöhen im Grenzfall weiter Tunnelbarrieren zu den physikalischen Schottkybarrieren streben, da die Ladungen, die hauptsächlich zum Stromfluss beitragen, Energien nahe dem Barrieremaximum haben, vgl. 6.6a. Im unteren Unterschwellbereich mit $|V_{\mathrm{gs}}| \ll |V_{\mathrm{th}}|$ gilt daher für kleine Drainspannungen V_{ds}

$$\phi_{\mathrm{eff}}(V_{\mathrm{gs}} \to 0, V_{\mathrm{ds}} \to 0) \approx \phi_{\mathrm{B}} \tag{6.4}$$

Im in Abb. 6.9 gezeigten Unterschwellbereich ($|V_{\mathrm{gs}}| = 2\,\mathrm{V}$) ist bei hohen V_{ds} die effektive Barriere zunächst beinahe konstant. Mit sinkendem V_{ds} nimmt ϕ_{eff} unter $V_{\mathrm{ds}} = 0,5\,\mathrm{V}$ zu. Diese Zunahme deutet darauf hin, dass mit kleinem V_{ds} die Barriere auf der Extraktionsseite relevant wird. Abb. 6.10 skizziert die Banddiagramme eines Schottkytransistors mit Rückseitengate für verschieden große V_{ds}.

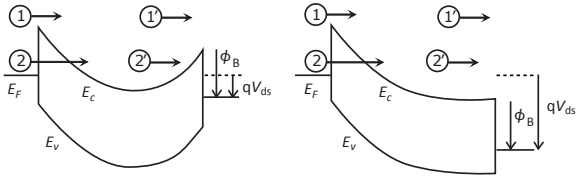


Abbildung 6.10: Schematische Banddiagramme eines SiNW Schottky-Transistors mit Rückseitengate für a) $qV_{\mathrm{ds}} < \phi_B$ und b) $qV_{\mathrm{ds}} > \phi_B$. Strompfade für thermionische Emission und Tunneln sind mit 1 bzw. 2 gekennzeichnet. Im Falle rein thermionischer Emission kann die drainseitige Barriere vernachlässigt werden, da aufgrund der Einkristallinität des Nanodrahts die sourceseitig injizieren Ladungsträger fast keinen energetischen Relaxationen bzw. Streuverlusten unterliegen.

Hierbei wird ersichtlich, dass bei Operation im Tunnelregime und mit $qV_{\mathrm{ds}} < \phi_{\mathrm{B}}$ grundsätzlich auch die Ladungsträgerextraktion über die Barriere an der Drainseite in Betracht gezogen werden muss (Strompfad ② in Abb. 6.10a). Im rein thermionischen Regime, in dem die Tunnelpfade unterdrückt werden und nur der Strompfad ① zum Tragen kommt

(AUS-Zustand), kann die Barriere auf der Extraktionsseite dagegen vernachlässigt werden. Bei hinreichend großem V_{ds} gilt dies auch für den AN-Zustand, da die durch Tunneln injizierten Ladungsträger ② an Drain keine Barriere erfahren (Abb. 6.10b). Im AN-Zustand herrscht, bedingt durch die hohe Gatespannung, eine hohe Bandverbiegung an den Barrieren, und die Tunnelwahrscheinlichkeit $\tilde{t}$ erhöht sich stark im Vergleich zum Sub-V_{th}-Fall. Bei gleichzeitig hohen V_{ds} kann kaum noch eine Barriere nachgewiesen werden (vgl. Abb. 6.9). Reduziert man nun jedoch die Drainspannung, steigt ϕ_{eff} wieder leicht an. Dies liegt daran, dass die Barriere auf der Extraktionsseite an Bedeutung gewinnt.

6.4 Zusammenfassung der Ergebnisse an SiNW Transistoren

In dieser Arbeit wurde das Schaltverhalten nominell undotierter Silizium Nanodraht Feldeffekttransistoren mit Ni-silizidierten Source/Drain-Kontakten und Rückseitengate untersucht. Diese Transistoren haben jeweils an Source und an Drain eine Schottkybarriere im Leitungskanal. Daraus ergibt sich ein ambipolares Schaltverhalten, welches die Möglichkeit eröffnet, Elektronen- und Löcherleitung am selben Bauelement zu studieren. Mit Hilfe temperaturabhängiger Messungen der Transferkennlinien unter verschiedenen Gate-Source- und Drain-Source-Spannungen konnten effektive Barrierenhöhen bestimmt werden. Diese lassen sich im Rahmen eines einfachen Modells qualitativ verstehen, welches das Tunneln von Ladungsträgern über eine Schottkybarriere beschreibt. Die effektive Barrierenhöhe kann dabei im Sinne einer Aktivierungsenergie für das gesamte System aufgefasst werden. Es zeigt sich, dass die Barrieren auch im AN-Bereich des Transistors noch eine Rolle spielen. Insbesondere ist dabei nicht nur die Barriere auf der Injektionsseite, sondern auch die Barriere an der Extraktionsseite der Ladungsträger in Betracht zu ziehen. Dies hat wichtige Implikationen für das Design zukünftiger, auf Nanodraht Schottky-FETs basierender Elektronik.

7 Zusammenfassung

Die vorliegende Arbeit widmet sich der Untersuchung verschiedener Aspekte des elektronischen Transports in State-of-the-Art MOSFETs einer planaren 28 nm HKMG Technologie, sowie von Silizium-Nanodraht- Transistoren als möglichen Bausteinen für zukünftige, neuartige Anwendungen. Dabei liegt der Schwerpunkt auf Seite der MOSFETs auf der Untersuchung der Ladungsträgerbeweglichkeit. Diese nimmt im Zuge der fortschreitenden Skalierung der Bauelemente drastisch ab und hat sich so in den letzten Jahren zunehmend zu einem für das Leistungsverhalten des Transistors limitierenden Faktor entwickelt. Aus diesem Grund wird in dieser Arbeit die auf dem geometrischen Magnetowiderstands- Effekt beruhende Magnetowiderstands- Beweglichkeit μ_{MR} untersucht. Diese Methode hat gegenüber den konventionellen, auf Kapazitäts- Spannungs- Messungen beruhenden Verfahren, verschiedene Stärken. So muss die genaue, elektrisch effektive Kanallänge nicht bekannt sein und die Abhängigkeit von Anschlusswiderständen fällt deutlich geringer aus, als z.B. bei der „effektiven Beweglichkeit“, μ_{eff}. Die Magnetowiderstands- Beweglichkeit ist damit potentiell genauer. Sie bietet somit die Möglichkeit, als Referenzpunkt für μ_{eff} zu dienen und dieser zugrunde liegende Annahmen implizit zu verifizieren. Darüber hinaus eröffnet das μ_{MR}- Verfahren die Möglichkeit, das Sättigungsregime bei hohen Drain-Source-Spannungen zu untersuchen, welches mit anderen Methoden nicht zugänglich ist. Aufgrund der geringen Beweglichkeiten heutiger Kurzkanaltransistoren mit typischen Werten unter $150\,\mathrm{cm}^2/\mathrm{Vs}$ sind für die Messung des MR Effekts jedoch sehr hohe Magnetfeldstärken um 14 T erforderlich. Die Messungen können somit nicht auf Waferebene [103] durchgeführt werden, sondern erfordern das Bonden einzelner Strukturen in Chipgehäuse, welche anschließend in

einem Hochfeld- Kryostatsystem gemessen werden.[1] Hierfür wurde im Rahmen dieser Arbeit eine nichtmagnetische und temperaturbeständige Probenhalterung entwickelt. Diese ermöglicht in Verbindung mit dem Chipgehäuse und einem angepassten Bondschema eine mehrfache Ausnutzung der im Magnetsystem begrenzten Probenkontakte. Als besondere experimentelle Herausforderung erwies sich der Aufbau einer über den gesamten Prozess der Probenerstellung und -charakterisierung durchgängig sicheren Verfahrensweise zum Schutz der empfindlichen, gebondeten Einzeltransistoren vor elektrostatischer Entladung.

Zur präzisen Bestimmung von μ_{MR} wird eine Formel für gleichzeitige Serienwiderstands- und Geometriekorrekturen hergeleitet und deren Einfluss auf die extrahierte Beweglichkeit anhand der Messdaten analysiert. Hierbei zeigt sich, dass die Geometriekorrektur nur für den Langkanaltransistor mit $W/L = 1$ relevant ist, während die Serienwiderstandskorrektur bei den Kurzkanaltransistoren wichtig ist. Es zeigt sich, dass Unsicherheiten im zugrundeliegenden Serienwiderstand auf μ_{MR} einen deutlich geringeren Einfluss haben, als auf μ_{eff}. Mit der Untersuchung von Elektronen- und Löcherbeweglichkeiten an Transistoren von bis zu 30 nm Gatelänge vom linearen Operationsbereich bis ins Sättigungsregime beleuchtet die vorliegende Arbeit einen Bereich, der in vorhergehenden Arbeiten bisher nicht zugänglich war [93, 104, 105]. Der Abgleich mit Berechnungen der Ladungsträger- Geschwindigkeiten und mit TCAD Bauelementesimulationen hilft dabei, zu einem tieferen Verständnis der Beweglichkeit im Sättigungsregime zu kommen.

Ein weiterer Abschnitt befasst sich mit dem Einfluss der Stressortechnologie auf die Beweglichkeit. Mit Hilfe von μ_{eff} wird dabei der Einfluss einer kleiner werdenden Kanallänge auf die Beweglichkeit untersucht. Dabei wird das Gegenspiel zwischen Beweglichkeitsverbesserungen aufgrund stärker einkoppelnder Kanalverspannung und Beweglichkeitsverschlechterungen durch zunehmende Kurzkanaleffekte sichtbar. Zum Abschluss zeigt ein Vergleich der verschiedenen Methoden zur

[1] Prinzipiell kann die Beweglichkeit auch über den Halleffekt bei niedrigeren Feldern gemessen werden. Dies erfordert allerdings im Gegensatz zum MR- Verfahren spezielle Teststrukturen mit einem Potentialabgriff im Kanal, wodurch eine Charakterisierung an Kurzkanaltransistoren nicht möglich ist.

Bestimmung der Ladungsträgerbeweglichkeit Limitierungen und Stärken der jeweiligen Verfahren auf.

Der zweite Teil der Arbeit beschäftigt sich mit dem elektronischen Transport an Si- Nanodraht Transistoren mit Nickel-silizidierten Source/Drain-Schottkykontakten und Rückseitengate. Diese weisen ein ambipolares Schaltverhalten auf und ermöglichen so die Untersuchung von Elektronen- und Löchertransport am gleichen Bauelement. Es zeigt sich, dass das eine unter definierten Prozessbedingungen aufgewachsene Oxidhülle um den Draht wesentlich dafür ist, ein hysteresearmes und damit präzises, reproduzierbares Schaltverhalten zu erreichen. Weitere Verbesserungen der Schalteigenschaften können durch Ausheizen und Messen der Probe im Vakuum erreicht werden. Zur Untersuchung des elektronischen Transportverhaltens wird die Temperaturabhängigkeit des Source/Drain-Stroms im Rahmen spannungsabhängiger, effektiver Energiebarrieren ϕ_{eff} betrachtet. Dabei zeigt sich, dass die Temperaturabhängigkeit des Tunnelstroms nicht nur für den Unterschwellbereich wichtig ist, sondern auch für den AN- Zustand. Darüber hinaus liegen die effektiven Barrieren für Elektronen und Löcher bei kleinen Gate- und Drain-Source-Spannungen in der Größenordnung der physikalischen Schottkybarriere des $NiSi_2$-Si Übergangs. Für kleine Drain-Source-Spannungen kann eine Zunahme von ϕ_{eff} beobachtet werden, und zwar sowohl im Unterschwellbereich, als auch im AN- Zustand des Transistors. Dies zeigt, dass an Si-Nanodraht Transistoren mit Schottkykontakten nicht nur die Barriere an der Injektionsseite der Ladungsträger, sondern auch die auf der Extraktionsseite von elementarer Bedeutung für das Schaltverhalten ist.

A Anhang

A.1 Banddiagramme der MOS-Struktur

Abb. A.1 zeigt die Banddiagramme im vertikalen Schnitt durch den Kondensatorstapel – beispielhaft an einer nMOS Struktur mit p-Silizium und SiO_2 als Gateoxid – für verschiedene Spannungsfälle. Eine negatives Gatepotential führt im Silizium zu einer Bandverbiegung „nach oben“ (Abb. A.1b). An der Grenzfläche zum SiO_2 sammeln sich daher Löcher aus dem p-dotierten Substrat. Die Kapazität der Struktur wird maximal (vgl. Abb. 3.1), da die gesamte Spannung über dem SiO_2 abfällt, ähnlich einem Plattenkondensator mit geringem Plattenabstand. Mit steigendem V_g werden die Löcher von der Grenzfläche weggeschoben (Abb. A.1c) und die Kapazität nimmt ab – das Substrat geht in die Verarmung. Wird bei sehr kleinen Frequenzen (unter 1 Hz) gemessen, so können Elektron-Loch-Paare, die sich durch spontane Generation im Halbleiter bilden, aufgrund des hohen Feldes getrennt werden, sodass es zu keiner direkten Rekombination kommt. Diese Elektronen sammeln sich mit zunehmendem V_g an der Grenzfläche, wodurch der p-dotierte Halbleiter „invertiert“ wird (Abb. A.1d). Die Kapazität nimmt wieder ihren maximalen Wert an. Im MOSFET kann dieser Ladungsanteil trotz der üblicherweise hohen verwendeten Frequenzen (10 kHz - 1 MHz) gemessen werden, da die benötigten Elektronen von den Source/Drain-Gebieten zur Verfügung gestellt werden (C_{ga} in Abb. 3.1). Misst man die Kapazität am MOSFET dagegen nur zwischen Gate und Substrat, können die wenigen, spontan generierten Ladungsträger dem Feld nicht folgen und es zeigt sich statt des erneuten Anstiegs eine weitere Abnahme der Kapazität für hohe V_g, siehe C_{gb} in Abb. 3.1 (tiefe Verarmung).

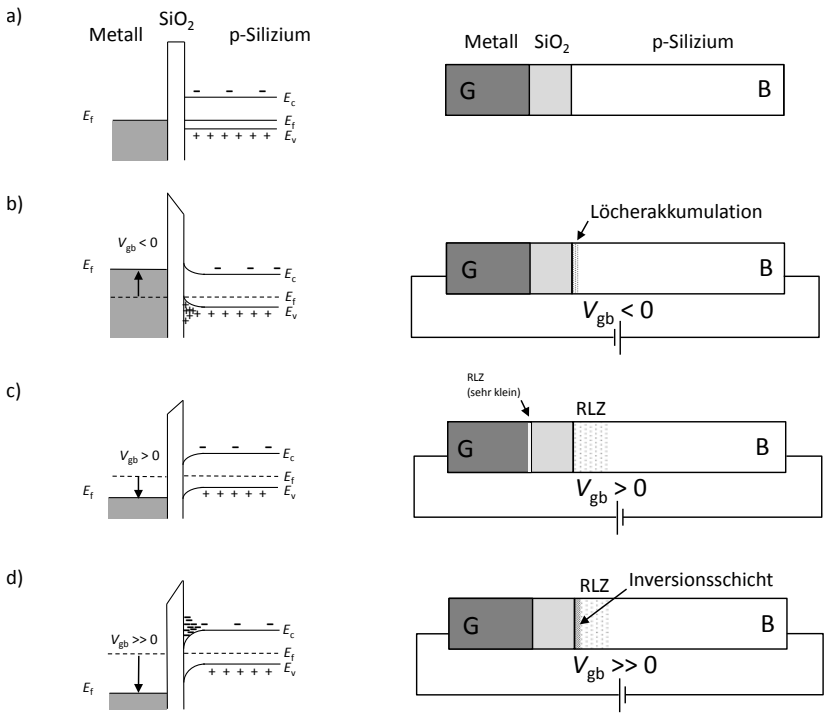


Abbildung A.1: Banddiagramme einer nFET MOS-Gatestruktur mit p-Silizium (links) und entsprechende räumliche Ladungsverteilungen (rechts) nach [30] und [31] für verschiedene Spannungsfälle: a) Flachbachfall, b) Akkumulation, c) Verarmung, d) Inversion

A.2 Messroutine Magnetowiderstandsbeweglichkeit

A.2.1 Datenaufbereitung

Die in Abschnitt 3.1.3 beschriebene Routine zur Messung der Magnetowiderstandsbeweglichkeit legt bei jeder Messung eine Textdatei mit den gemessenen Stromwerten an Source und Drain sowie den dazugehörigen Zeitstempeln an. Da sich wie in Abb. 3.3b) illustriert fortlaufend dieselben 6 Spannungseinstellungen wiederholen, gehört jeder sechste Messwert zu einem bestimmten Operationspunkt des Transistors. Die Aufbereitung der Rohdaten geschieht mittels eines MATLAB-Scripts. Zunächst wird die große, mit verschiedenen Strommessungen „durchmischte“ Ursprungsdatei in ihre 6 Einzeltabellen zerlegt, von denen jede die fortlaufenden Messwerte enthält, die zu einer konstanten Spannungseinstellung gehören (Abb. A.2).

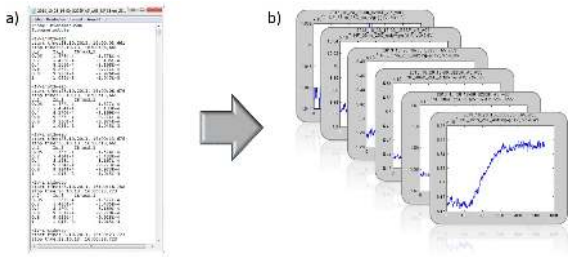


Abbildung A.2: Extraktion der Rohdaten: a) Ausschnitt aus einer Messdatei mit 4 von ca. 580 aufgenommenen Datenpaketen, von denen in jedem die gemessenen Source- und Drainströme zu 6 unterschiedlichen Spannungseinstellungen abgelegt sind und b) die 6 daraus rekonstruierten zugehörigen Einzelmessungen der Drainstromentwicklung während des Magnetfelddurchlaufs.

Dann werden die Daten an Anfang und Ende der Messung, die bei konstantem Feld aufgenommen wurden, vom Datensatz abgeschnitten. Damit bleiben nur die Daten übrig, die in die eigentliche Zeit der Magnetfeldänderung fallen. Der um Vor- und Nachmessung reduzierte Datensatz in der Form $I(t)$ entspricht dem Bereich II in Abb. A.3. Dieser wird nun anhand der Zeitstempel zusammen mit der ebenfalls vorliegenden Information $B(t)$ korreliert (siehe Abb. A.3) und als $I(B)$ in eine neue Datentabelle geschrieben, die im Anschluss an die Fitroutine übergeben wird.

In einem zweiten Schritt werden Datenbereiche identifiziert, die aufgrund von spontanen Sprüngen oder aufgrund von Drift im Signal augenscheinlich nicht zum Rest der Messung passen. Ein solches Beispiel ist der spontane Signaleinbruch im oberen Bereich von Abschnitt II in Abb. A.3. Die Zeiten vor und nach dem eigentlichen Felddurchlauf sind hier für die Bewertung sehr hilfreich. So ist z. B. im erwähnten Beispiel zu sehen, wie sich das Signal nach Abschluss der eigentlichen Messung in der Folge (Bereich III) auf das ursprünglich zu erwartende „Nullfeldsignal“ erholt hat. Bei der Auswahl der für den Fit geeigneten Messwerte sucht man die Datenbereiche, die den größten konsistenten, auf einer Parabel liegenden Abschnitt bilden. Da die Beweglichkeit in der Krümmung der Parabel steckt, ist die Absolutlage in I_d dabei nicht entscheidend. Falls der Strom z. B. aufgrund von spontanem Ladungseinfang und -abgabe am Gateoxid zwischen zwei diskreten Zuständen hin- und herspringt (vgl. Abb. 5.8), hat dies keine Auswirkung auf die ermittelte Beweglichkeit. Die Information über vom Fit auszunehmende Zeitbereiche wird ebenfalls an die Fitroutine übergeben.

A.2.2 Fitroutine

Im Anschluss an die Aufbereitung der Messdaten erfolgt die Übergabe an eine MATLAB-Routine, die die Datensätze unter Berücksichtigung von Serienwiderstands- und Geometriekorrektur fittet, um gemäß Gl. (5.9) die Beweglichkeit zu bestimmen. Hierzu werden folgende Werte an die Fitroutine übergeben:

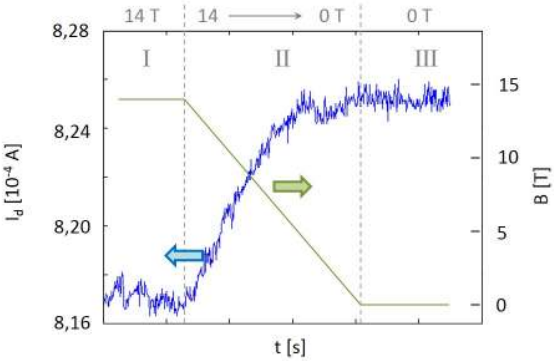


Abbildung A.3: Beispiel für einen der sechs $I_\mathrm{d}(t)$ Datensätze aus Abb. A.2; die gesamte Messung gliedert sich in 3 Bereiche: I) die Zeit vor der Magnetfeldänderung, II) die eigentliche Messzeit und III) eine Zeit der Nachmessung. Im gezeigten Beispiel handelt es sich um eine Messung während eines Feldabbaus. Dies ist daran ersichtlich, dass der Strom aufgrund eines sinkenden Magnetowiderstands mit der Zeit zunimmt.

- gemessene Datenpunkte als Funktion des Magnetfelds $I(B)$
- Richtung des Felddurchlaufs; im Falle eines Feldabbaus werden die Daten zunächst so umsortiert, dass sie der Reihenfolge nach mit steigendem Feld erscheinen. Die dient dazu, die Parabel in der richtigen Krümmung an die Daten zu fitten.
- Information über für den Fit nicht zu berücksichtigende Datenbereiche
- Länge L und Weite W des Transistors zur Anwendung der Geometriekorrektur
- Serienwiderstand R_sd zur Anwendung der Serienwiderstandskorrektur

Als variable Parameter werden dabei der Nullfeldstrom I_0 und die Beweglichkeit μ_{MR} zugelassen, als Startwert für I_0 wird der bei $B = 0\,\mathrm{T}$ gemessene Stromwert herangezogen. Der Fit erfolgt nach der Methode der kleinsten Quadrate und wird zur Erhöhung der Genauigkeit dreimal durchlaufen, wobei I_0 und μ_{MR} aus dem vorigen Durchlauf jeweils automatisch als neue Startwerte herangezogen werden. Anschließend wird die resultierende Fitkurve mit den Messdaten in einer gemeinsamen Grafik dargestellt (Abb. A.4) und für spätere Referenz als Grafikdatei gespeichert.

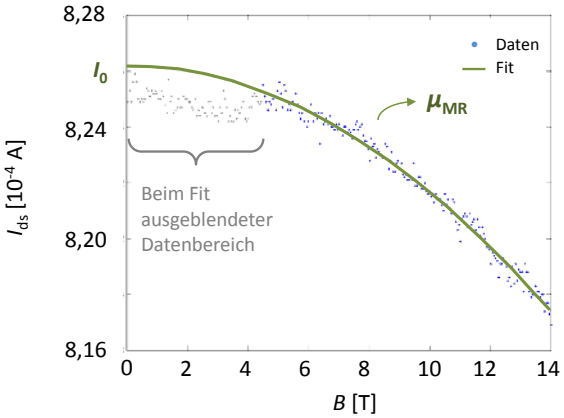


Abbildung A.4: Nach Datenaufbereitung resultierender $I_d(B)$-Graph. Die Fitroutine legt einge Funktion nach Gl. (5.9) durch die Messdaten (durchgezogene Linie), wobei ausgeblendete Bereiche (grau hinterlegte Messpunkte) nicht berücksichtigt werden. Die Beweglichkeit ergibt sich bei festgelegtem R_{sd} aus der Krümmung der Parabel.

Vom Fit ausgenommene Datenpunkte sind hierbei farblich von den anderen Punkten abgesetzt. Außerdem wird die berechnete Beweglichkeit mit ausgegeben. Die direkte grafische Ausgabe des Resultats ermöglicht eine sofortige visuelle Überprüfung, ob die sich ergebende Fit-

kurve auch zu den Daten passt, bevor mit dem eigentlichen Ergebnis, der extrahierten Beweglichkeit, die Auswertung fortgeführt wird. Nach erfolgtem Fit werden die ermittelten Beweglichkeiten zur weiteren Verarbeitung in MATLAB bzw. Excel gespeichert. Die Matlab-Scripte zum Einlesen, Aufbereiten und Fitten der Messdaten sind aufeinander abgestimmt. Sobald einmal für jeden einzulesenden Datensatz hinterlegt ist welcher Bereich gefittet werden soll, können sämtliche Beweglichkeiten aus allen Datensätzen z.B. für andere Serienwiderstände R_{ser} automatisiert neu berechnet werden. Dies ermöglicht eine schnelle und effiziente Datenanalyse.

Literaturverzeichnis

[1] Einstein, Albert: The Common Language of Science (Die übernationale Sprache der Wissenschaft). In: *Einstein Archives Online* 28-557 (1941), Sept, p. 3. `http://alberteinstein.info/vufind1/Record/EAR000020633`

[2] Pohl R. W., Hilsch R.: Steuerung von Elektrodenströmen mit einem Dreielektrodenkristall und ein Modell einer Sperrschicht. In: *Zeitschrift für Physik* 111 (1938), 399-408. `https://doi.org/10.1007/BF01342357`

[3] Zuse, K.: *Patentanmeldung Z 391 IXc/42m.* 1941

[4] Rojas, R.: Konrad Zuse's Legacy: The Architecture of the Z1 and Z3. In: *IEEE Annals of the History of Computing* 19 (1997), Nr. 2, S. 5–16

[5] Brattain, Walter H. ; John, Bardeen: *Three-electrode circuit element utilizing semiconductive materials.* Oct 1950. – US Patent 2,524,033

[6] Brattain, Walter H. ; John, Bardeen: *Three-electrode circuit element utilizing semiconductive materials.* Oct 1950. – US Patent 2,524,035

[7] William, Shockley: *Circuit element utilizing semiconductive material.* Sept 1951. – US Patent 2,569,347

[8] Kilby, Jack S.: *Miniaturized electronic circuits.* Jun 1964. – US Patent 3,138,743

[9] Noyce, Robert N.: *Semiconductor device-and-lead structure.* Apr 1961. – US Patent 2,981,877

[10] Ford, Dale: Celebrating the 50th Anniversary of Moore's Law. In: *IHS Technology Report* (2015). https://technology.ihs.com/api/binary/532884

[11] Moore, Gordon E.: Cramming more components onto integrated circuits, Reprinted from Electronics, volume 38, number 8, April 19, 1965, pp.114 ff. In: *Solid-State Circuits Society Newsletter, IEEE* 11 (2006), Sept, Nr. 5, S. 33–35. http://dx.doi.org/10.1109/N-SSC.2006.4785860. – DOI 10.1109/N–SSC.2006.4785860. – ISSN 1098–4232

[12] Hilbert, Martin ; López, Priscila: The world's technological capacity to store, communicate, and compute information. In: *science* 332 (2011), Nr. 6025, S. 60–65

[13] Totzeck, Michael ; Ulrich, Wilhelm ; Göhnermeier, Aksel ; Kaiser, Winfried: Semiconductor fabrication: Pushing deep ultraviolet lithography to its limits. In: *nature photonics* 1 (2007), Nr. 11, S. 629–631

[14] Thompson, S.E. ; Chau, R.S. ; Ghani, T. ; Mistry, K. ; Tyagi, S. ; Bohr, M.T.: In search of "Forever,"continued transistor scaling one new material at a time. In: *Semiconductor Manufacturing, IEEE Transactions on* 18 (2005), Feb, Nr. 1, S. 26–36. http://dx.doi.org/10.1109/TSM.2004.841816. – DOI 10.1109/TSM.2004.841816. – ISSN 0894–6507

[15] Mistry, Kaizad: *Transistor Scaling: The Age of Innovation.* https://nanohub.org/resources/20880. Version: Apr 2014

[16] Takagi, S. ; Toriumi, Akira ; Iwase, M. ; Tango, H.: On the universality of inversion layer mobility in Si MOSFET's: Part I-effects of substrate impurity concentration. In: *Electron Devices, IEEE Transactions on* 41 (1994), Dec, Nr. 12, S. 2357–2362. http://dx.doi.org/10.1109/16.337449. – DOI 10.1109/16.337449. – ISSN 0018–9383

[17] Thompson, S.E. ; Armstrong, M. ; Auth, C. ; Alavi, M. ; Buehler, M. ; Chau, R. ; Cea, S. ; Ghani, T. ; Glass, G. ; Hoffman, T. ; Jan, Chia-Hong ; Kenyon, C. ; Klaus, J. ; Kuhn, K. ; Ma, Zhiyong ; Mcintyre, B. ; Mistry, K. ; Murthy, A. ; Obradovic, B. ; Nagisetty, R. ; Nguyen, Phi ; Sivakumar, S. ; Shaheed, R. ; Shifren, L. ; Tufts, B. ; Tyagi, S. ; Bohr, M. ; El-Mansy, Y.: A 90-nm logic technology featuring strained-silicon. In: *Electron Devices, IEEE Transactions on* 51 (2004), Nov, Nr. 11, S. 1790–1797. `http://dx.doi.org/10.1109/TED.2004.836648`. – DOI 10.1109/TED.2004.836648. – ISSN 0018–9383

[18] Chau, R.S. ; Doyle, B.S. ; Kavalieros, J. ; Barlage, D. ; Datta, S. ; Hareland, S.A.: *Tri-gate devices and methods of fabrication*. `https://www.google.com/patents/US6858478`. Version: feb 2005. – US Patent 6,858,478

[19] Auth, C. ; Allen, C. ; Blattner, A. ; Bergstrom, D. ; Brazier, M. ; Bost, M. ; Buehler, M. ; Chikarmane, V. ; Ghani, T. ; Glassman, T. ; Grover, R. ; Han, W. ; Hanken, D. ; Hattendorf, M. ; Hentges, P. ; Heussner, R. ; Hicks, J. ; Ingerly, D. ; Jain, P. ; Jaloviar, S. ; James, R. ; Jones, D. ; Jopling, J. ; Joshi, S. ; Kenyon, C. ; Liu, H. ; McFadden, R. ; Mcintyre, B. ; Neirynck, J. ; Parker, C. ; Pipes, L. ; Post, I. ; Pradhan, S. ; Prince, M. ; Ramey, S. ; Reynolds, T. ; Roesler, J. ; Sandford, J. ; Seiple, J. ; Smith, P. ; Thomas, C. ; Towner, D. ; Troeger, T. ; Weber, C. ; Yashar, P. ; Zawadzki, K. ; Mistry, K.: A 22nm high performance and low-power CMOS technology featuring fully-depleted tri-gate transistors, self-aligned contacts and high density MIM capacitors. In: *VLSI Technology (VLSIT), 2012 Symposium on*, 2012. – ISSN 0743–1562, S. 131–132

[20] Arden, W. ; Brillouet, M. ; Cogez, P. ; Graef, M. ; Huizing, B. ; Mahnkopf, R.: *More than Moore - White Paper*. `http://www.itrs.net/papers.html`. Version: 2014

[21] Colinge, Jean-Pierre ; Lee, Chi-Woo ; Afzalian, Aryan ; Akhavan, Nima D. ; Yan, Ran ; Ferain, Isabelle ; Razavi, Pedram ; O'Neill, Brendan ; Blake, Alan ; White, Mary u. a.: Nanowire transistors without junctions. In: *Nature nanotechnology* 5 (2010), Nr. 3, S. 225–229

[22] Knoll, L ; Zhao, QT ; Nichau, A ; Richter, S ; Luong, GV ; Trellenkamp, S ; Schäfer, A ; Selmi, L ; Bourdelle, KK ; Mantl, S: Demonstration of Improved Transient Response of Inverters with Steep Slope Strained Si NW TFETs by Reduction of TAT with Pulsed IV and NW Scaling. In: *International Electron Device Meeting (IEDM)*, 2013

[23] Specht, M. ; Kommling, R. ; Hofmann, F. ; Klandzievski, V. ; Dreeskornfeld, L. ; Weber, W. ; Kretz, J. ; Landgraf, E. ; Schulz, T. ; Hartwich, J. ; Rosner, W. ; Stadele, M. ; Luyken, R.J. ; Reisinger, H. ; Graham, A. ; Hartmann, E. ; Risch, L.: Novel dual bit tri-gate charge trapping memory devices. In: *Electron Device Letters, IEEE* 25 (2004), Dec, Nr. 12, S. 810–812. `http://dx.doi.org/10.1109/LED.2004.838621`. – DOI 10.1109/LED.2004.838621. – ISSN 0741–3106

[24] Weber, Walter M. ; Geelhaar, L. ; Lamagna, L. ; Fanciulli, M. ; Kreupl, F. ; Unger, E. ; Riechert, H. ; Scarpa, G. ; Lugli, P.: Tuning the Polarity of Si-Nanowire Transistors Without the Use of Doping. In: *Nanotechnology, 2008. NANO '08. 8th IEEE Conference on*, 2008, S. 580–581

[25] Heinzig, André ; Slesazeck, Stefan ; Kreupl, Franz ; Mikolajick, Thomas ; Weber, Walter M.: Reconfigurable Silicon Nanowire Transistors. In: *Nano Letters* 12 (2012), Nr. 1, 119-124. `http://dx.doi.org/10.1021/nl203094h`. – DOI 10.1021/nl203094h. – PMID: 22111808

[26] De Marchi, Michele ; Sacchetto, Davide ; Frache, Stefano ; Zhang, Jian ; Gaillardon, P ; Leblebici, Yusuf ; De Micheli, Giovanni: Polarity control in double-gate, gate-all-around

vertically stacked silicon nanowire FETs. In: *Electron Devices Meeting (IEDM), 2012 IEEE International* IEEE, 2012, S. 8–4

[27] Cui, Yi ; Wei, Qingqiao ; Park, Hongkun ; Lieber, Charles M.: Nanowire nanosensors for highly sensitive and selective detection of biological and chemical species. In: *Science* 293 (2001), Nr. 5533, S. 1289–1292

[28] Patolsky, Fernando ; Zheng, Gengfeng ; Lieber, Charles M.: Fabrication of silicon nanowire devices for ultrasensitive, label-free, real-time detection of biological and chemical species. In: *Nature protocols* 1 (2006), Nr. 4, S. 1711–1724

[29] Pregl, Sebastian: Fabrication and characterization of a silicon nanowire based Schottky-barrier field effect transistor platform for functional electronics and biosensor applications (Dissertation). In: *Technische Universität Dresden* (2015). `http://nbn-resolving.de/urn:nbn:de:bsz:14-qucosa-171112`

[30] Taur, Yuan ; Ning, Tak H.: *Fundamentals of modern VLSI devices.* Cambridge University Press, 2009

[31] Hoffmann, Kurt: *Systemintegration: vom Transistor zur großintegrierten Schaltung.* Oldenbourg Verlag München Wien, 2006

[32] Chau, Robert ; Datta, Suman ; Doczy, Mark ; Kavalieros, Jack ; Metz, Matthew: Gate dielectric scaling for high-performance CMOS: from SiO2 to High-K. In: *International workshop on gate insulator*, 2003, S. 124–126

[33] Mistry, Kaizad ; Allen, C ; Auth, C ; Beattie, B ; Bergstrom, D ; Bost, M ; Brazier, M ; Buehler, M ; Cappellani, A ; Chau, R u. a.: A 45nm logic technology with high-k+ metal gate transistors, strained silicon, 9 Cu interconnect layers, 193nm dry patterning, and 100% Pb-free packaging. In: *Electron Devices Meeting, 2007. IEDM 2007. IEEE International* IEEE, 2007, S. 247–250

[34] Auth, Chris ; Cappellani, A ; Chun, J-S ; Dalis, A ; Davis, A ; Ghani, T ; Glass, G ; Glassman, T ; Harper, M ; Hattendorf, M u. a.: 45nm high-k+ metal gate strain-enhanced transistors. In: *VLSI Technology, 2008 Symposium on* IEEE, 2008, S. 128–129

[35] Sabnis, A. G. ; Clemens, J. T.: Characterization of the electron mobility in the inverted <100> Si surface. In: *Electron Devices Meeting, 1979 Internationa* Bd. 25, 1979, S. 18–21

[36] Ando, Tsuneya ; Fowler, Alan B. ; Stern, Frank: Electronic properties of two-dimensional systems. In: *Reviews of Modern Physics* 54 (1982), Nr. 2, S. 437

[37] Mamatrishat, Mamat ; Kouda, Miyuki ; Kawanago, Takamasa ; Kakushima, Kuniyuki ; Ahmet, P ; Aierken, Abuliemu ; Tsutsui, KAZUO ; Nishiyama, Akira ; Sugii, Nobuyuki ; Natori, Kenji u. a.: Effect of remote-surface-roughness scattering on electron mobility in MOSFETs with high-k dielectrics. In: *ECS Transactions* 33 (2010), Nr. 3, S. 249–255

[38] Schroder, Dieter K.: *Semiconductor material and device characterization.* John Wiley & Sons, 2006

[39] Cohen, Morrel H. ; Economou, E. N. ; Soukoulis, C. M.: Microscopic mobility. In: *Phys. Rev. B* 30 (1984), Oct, 4493–4500. `http://dx.doi.org/10.1103/PhysRevB.30.4493`. – DOI 10.1103/PhysRevB.30.4493

[40] Palenskis, V.: Drift Mobility, Diffusion Coefficient of Randomly Moving Charge Carriers in Metals and Other Materials with Degenerated Electron Gas. In: *World Journal of Condensed Matter Physics* 3 (2013), 73-81. `http://dx.doi.org/10.4236/wjcmp.2013.31013`. – DOI 10.4236/wjcmp.2013.31013

[41] Ghibaudo, G: Transport in the inversion layer of a MOS transistor: use of Kubo-Greenwood formalism. In: *Journal of Physics C: Solid State Physics* 19 (1986), Nr. 5, 767. `http://stacks.iop.org/0022-3719/19/i=5/a=015`

[42] Schumacher, Uwe: *Fusionsforschung: Eine Einführung.* Wiss. Buchges., 1993

[43] Weiss, H: Magnetoresistance. In: *Semiconductors and Semimetals* 1 (1966), S. 315–376

[44] Meziani, Y. M. ; Łusakowski, J. ; Knap, W. ; Dyakonova, N. ; Teppe, F. ; Romanjek, K. ; Ferrier, M. ; Clerc, R. ; Ghibaudo, G. ; Boeuf, F. ; Skotnicki, T.: Magnetoresistance characterization of nanometer Si metal-oxide-semiconductor transistors. In: *Journal of Applied Physics* 96 (2004), Nr. 10, S. 5761–5765. `http://dx.doi.org/http://dx.doi.org/10.1063/1.1806991`. – DOI http://dx.doi.org/10.1063/1.1806991

[45] Putley, Ernest H.: The Hall effect and semi-conductor physics. (1968)

[46] Cassé, M. ; Rochette, F. ; Thevenod, L. ; Bhouri, N. ; Andrieu, F. ; Reimbold, G. ; Boulanger, F. ; Mouis, M. ; Ghibaudo, G. ; Maude, D. K.: A comprehensive study of magnetoresistance mobility in short channel transistors: Application to strained and unstrained silicon-on-insulator field-effect transistors. In: *Journal of Applied Physics* 105 (2009), Nr. 8. `http://dx.doi.org/http://dx.doi.org/10.1063/1.3097764`. – DOI http://dx.doi.org/10.1063/1.3097764

[47] Donetti, L. ; Gámiz, F. ; Cristoloveanu, S.: A theoretical interpretation of magnetoresistance mobility in silicon inversion layers. In: *Journal of Applied Physics* 102 (2007), Nr. 1. `http://dx.doi.org/http://dx.doi.org/10.1063/1.2752103`. – DOI http://dx.doi.org/10.1063/1.2752103

[48] Liu, P.W. ; Lin, Y.H. ; Chiang, W.T. ; Tsai, C.H. ; Cheng, L.W. ; Tsai, C.T. ; Ma, G.H.: SiGe Channel CMOSFETs Fabricated on (110) Surfaces with TaC/HfO2 Gate Stacks. In: *VLSI Technology, Systems and Applications, 2007. VLSI-TSA 2007. International Symposium on*, 2007. – ISSN 1524–766X, S. 1–2

[49] Ashcroft, Neil W. ; Mermin, N D.: Solid State Physics. In: *Saunders College, Philadelphia* (1976)

[50] Lundstrom, Mark: *Fundamentals of carrier transport*. Cambridge University Press, 2009

[51] Sun, Y ; Thompson, SE ; Nishida, T: Physics of strain effects in semiconductors and metal-oxide-semiconductor field-effect transistors. In: *Journal of Applied Physics* 101 (2007), Nr. 10, S. 104503

[52] Hadjisavvas, G. ; Tsetseris, L. ; Pantelides, S.T.: The Origin of Electron Mobility Enhancement in Strained MOSFETs. In: *Electron Device Letters, IEEE* 28 (2007), Nov, Nr. 11, S. 1018–1020. `http://dx.doi.org/10.1109/LED.2007.906471`. – DOI 10.1109/LED.2007.906471. – ISSN 0741–3106

[53] Zhao, Yi ; Takenaka, M. ; Takagi, S.: Comprehensive understanding of surface roughness and Coulomb scattering mobility in biaxially-strained Si MOSFETs. In: *Electron Devices Meeting, 2008. IEDM 2008. IEEE International*, 2008. – ISSN 8164–2284, S. 1–4

[54] Flachowsky, Stefan: Verspannungstechniken zur Leistungssteigerung von SOI-CMOS-Transistoren (Dissertation). In: *Technische Universität Dresden* (2010). `http://www.qucosa.de/fileadmin/data/qucosa/documents/6313/StefanFlachowsky-Dissertation2010.pdf`

[55] Flachowsky, S. ; Herrmann, T. ; Hontschel, J. ; Illgen, R. ; Ong, S.Y. ; Wiatr, M. ; Baldauf, T. ; Klix, W. ; Stenzel, R.: Mobility and strain effects for <100> and <110> oriented silicon and SiGe transistor channels. In: *Ultimate Integration on Silicon (ULIS), 2012 13th International Conference on* (2012), March, S. 5–8. `http://dx.doi.org/10.1109/ULIS.2012.6193343`. – DOI 10.1109/ULIS.2012.6193343

[56] Weber, Walter M. ; Duesberg, Georg S. ; Graham, Andrew P. ; Liebau, Maik ; Unger, Eugen ; Cheze, Caroline ; Geelhaar, Lutz ; Lugli, Paolo ; Riechert, Henning ; Kreupl, Franz: Silicon nanowires: catalytic growth and electrical characterization. In: *physica status solidi (b)* 243 (2006), Nr. 13, 3340–3345. http://dx.doi.org/10.1002/pssb.200669138. – DOI 10.1002/pssb.200669138. – ISSN 1521–3951

[57] Heinzig, André ; Mikolajick, Thomas ; Trommer, Jens ; Grimm, Daniel ; Weber, Walter M.: Dually Active Silicon Nanowire Transistors and Circuits with Equal Electron and Hole Transport. In: *Nano Letters* 13 (2013), Nr. 9, 4176-4181. http://dx.doi.org/10.1021/nl401826u. – DOI 10.1021/nl401826u. – PMID: 23919720

[58] Wagner, R. S. ; Ellis, W. C.: VAPOR-LIQUID-SOLID MECHANISM OF SINGLE CRYSTAL GROWTH. In: *Applied Physics Letters* 4 (1964), Nr. 5

[59] Weber, Walter M. ; Geelhaar, Lutz ; Graham, Andrew P. ; Unger, Eugen ; Duesberg, Georg S. ; Liebau, Maik ; Pamler, Werner ; Chèze, Caroline ; Riechert, Henning ; Lugli, Paolo ; Kreupl, Franz: Silicon-Nanowire Transistors with Intruded Nickel-Silicide Contacts. In: *Nano Letters* 6 (2006), Nr. 12, 2660-2666. http://dx.doi.org/10.1021/nl0613858. – DOI 10.1021/nl0613858. – PMID: 17163684

[60] Anastassakis, E. ; Cantarero, A. ; Cardona, M.: Piezo-Raman measurements and anharmonic parameters in silicon and diamond. In: *Phys. Rev. B* 41 (1990), Apr, 7529–7535. http://dx.doi.org/10.1103/PhysRevB.41.7529. – DOI 10.1103/PhysRevB.41.7529

[61] Moselund, K.E. ; Najmzadeh, M. ; Dobrosz, P. ; Olsen, S.H. ; Bouvet, D. ; De Michielis, L. ; Pott, V. ; Ionescu, A.M.: The High-Mobility Bended n-Channel Silicon Nanowire Transistor. In: *Electron Devices, IEEE Transactions on* 57

(2010), April, Nr. 4, S. 866–876. http://dx.doi.org/10.1109/TED.2010.2040939. – DOI 10.1109/TED.2010.2040939. – ISSN 0018–9383

[62] Baldauf, T. ; Heinzig, A. ; Trommer, J. ; Mikolajick, T. ; Weber, W.M.: Stress-Dependent Performance Optimization of Reconfigurable Silicon Nanowire Transistors. In: *Electron Device Letters, IEEE* 36 (2015), Oct, Nr. 10, S. 991–993. http://dx.doi.org/10.1109/LED.2015.2471103. – DOI 10.1109/LED.2015.2471103. – ISSN 0741–3106

[63] Ogata, K ; Sutter, E ; Zhu, X ; Hofmann, S: Ni-silicide growth kinetics in Si and Si/SiO 2 core/shell nanowires. In: *Nanotechnology* 22 (2011), Nr. 36, 365305. http://stacks.iop.org/0957-4484/22/i=36/a=365305

[64] Lu, Kuo-Chang ; Tu, K. N. ; Wu, W. W. ; Chen, L. J. ; Yoo, Bong-Young ; Myung, Nosang V.: Point contact reactions between Ni and Si nanowires and reactive epitaxial growth of axial nano- NiSi/Si. In: *Applied Physics Letters* 90 (2007), Nr. 25. http://dx.doi.org/http://dx.doi.org/10.1063/1.2750530. – DOI http://dx.doi.org/10.1063/1.2750530

[65] Trommer, Jens ; Heinzig, André ; Baldauf, Tim ; Slesazeck, Stefan ; Mikolajick, Thomas ; Weber, Walter M.: Functionality-enhanced logic gate design enabled by symmetrical reconfigurable silicon nanowire transistors. In: *IEEE Transactions on Nanotechnology* 14 (2015), Nr. 4, S. 689–698

[66] Trommer, Jens ; Baldauf, Tim ; Mikolajick, Thomas ; Weber, Walter M. ; Raitza, Michael u. a.: Reconfigurable nanowire transistors with multiple independent gates for efficient and programmable combinational circuits. In: *2016 Design, Automation & Test in Europe Conference & Exhibition (DATE)* IEEE, 2016, S. 169–174

[67] Henisch, Heinz K.: *Rectifying semi-conductor contacts.* Clarendon Press, 1957

[68] Martin, Dominik ; Heinzig, Andre ; Grube, Matthias ; Geelhaar, Lutz ; Mikolajick, Thomas ; Riechert, Henning ; Weber, Walter: Direct Probing of Schottky Barriers in Si Nanowire Schottky Barrier Field Effect Transistors. In: *Phys. Rev. Lett.* 107 (2011), Nov, 216807. `http://dx.doi.org/10.1103/PhysRevLett.107.216807`. – DOI 10.1103/PhysRevLett.107.216807

[69] Heinzig, André: Entwicklung und Herstellung rekonfigurierbarer Nanodraht- Transistoren und Schaltungen (Dissertation). In: *Technische Universität Dresden* (2014)

[70] Sze, Simon M. ; Ng, Kwok K.: *Physics of semiconductor devices.* John Wiley & Sons, 2006

[71] Landau, L. D. ; Lifshitz, E. M.: Quantum mechanics: non-relativistic theory, 2nd rev. ed. In: *Course of theoretical physics* 3 (1965)

[72] Trommer, J. ; Heinzig, A. ; Slesazeck, S. ; Mikolajick, T. ; Weber, W.M.: Elementary Aspects for Circuit Implementation of Reconfigurable Nanowire Transistors. In: *Electron Device Letters, IEEE* 35 (2014), Jan, Nr. 1, S. 141–143. `http://dx.doi.org/10.1109/LED.2013.2290555`. – DOI 10.1109/LED.2013.2290555. – ISSN 0741–3106

[73] Takagi, S. ; Toriumi, Akira ; Iwase, M. ; Tango, H.: On the universality of inversion layer mobility in Si MOSFET's: Part II-effects of surface orientation. In: *Electron Devices, IEEE Transactions on* 41 (1994), Dec, Nr. 12, S. 2363–2368. `http://dx.doi.org/10.1109/16.337450`. – DOI 10.1109/16.337450. – ISSN 0018–9383

[74] Watt, JT ; Plummer, JD: Universal mobility-field curves for electrons and holes in MOS inversion layers. In: *VLSI Technology, 1987. Digest of Technical Papers. Symposium on* IEEE, 1987, S. 81–82

[75] Takagi, S. ; Iwase, M. ; Toriumi, A.: On the universality of inversion-layer mobility in n- and p-channel MOSFETs. In: *Electron Devices Meeting, 1988. IEDM '88. Technical Digest., International*, 1988. – ISSN 0163–1918, S. 398–401

[76] Terada, Kazuo ; Muta, Hiroki: A New Method to Determine Effective MOSFET Channel Length. In: *Japanese Journal of Applied Physics* 18 (1979), Nr. 5, 953. `http://stacks.iop.org/1347-4065/18/i=5/a=953`

[77] Chern, John G. ; Chang, Peter ; Motta, Richard F. ; Godinho, Norm: A new method to determine MOSFET channel length. In: *Electron Device Letters, IEEE* 1 (1980), Nr. 9, S. 170–173

[78] Inc., Lake Shore C.: *Model CPX-VF High Performance Cryogenic Probe Station*

[79] Synova S.A., CH-1024 Ecublens S. Chemin de la Dent d'Oche d. Chemin de la Dent d'Oche: LCS 300 W User's manual. (2008). `http://www.synova.ch`

[80] *GLOBALFOUNDRIES Module One LLC & Co. KG, Wilschdorfer Landstr. 101, 01109 Dresden, Deutschland*

[81] Horstmann, M. ; Wei, A. ; Kammler, T. ; Hontschel, J. ; Bierstedt, H. ; Feudel, T. ; Frohberg, K. ; Gerhardt, M. ; Hellmich, A. ; Hempel, K. ; Hohage, J. ; Javorka, P. ; Klais, J. ; Koerner, G. ; Lenski, M. ; Neu, A. ; Otterbach, R. ; Press, P. ; Reichel, C. ; Trentsch, M. ; Trui, B. ; Salz, H. ; Schaller, M. ; Engelmann, H.-J. ; Herzog, O. ; Ruelke, H. ; Hubler, P. ; Stephan, R. ; Greenlaw, D. ; Raab, M. ; Kepler, N. ; Chen, H. ; Chidambarrao, D. ; Fried, D. ; Holt, J. ; Lee, W. ; Nii, H. ; Panda, S. ; Sato, T. ; Waite, A. ; Liming, S. ; Rim, K. ; Schepis, D. ; Khare, M. ; Huang, S.F. ; Pellerin, J. ; Su, L.T.: Integration and optimization of embedded-SiGe, compressive and tensile stressed liner films, and stress memorization in advanced SOI CMOS technologies. In:

Electron Devices Meeting, 2005. IEDM Technical Digest. IEEE International, 2005, S. 233–236

[82] Abe, Kenichi ; Teramoto, Akinobu ; Sugawa, Shigetoshi ; Ohmi, Tadahiro: Understanding of traps causing random telegraph noise based on experimentally extracted time constants and amplitude. In: *Reliability Physics Symposium (IRPS), 2011 IEEE International* IEEE, 2011, S. 4A–4

[83] Lippmann, HJ ; Kuhrt, F: Der Geometrieeinfluß auf den transversalen magnetischen Widerstandseffekt bei rechteckförmigen Halbleiterplatten. In: *Zeitschrift Naturforschung Teil A* 13 (1958), S. 462

[84] Beer, AC: Galvanomagnetic Effects in Semiconductors (Academic, New York, 1963). In: *Solid State Physics, Suppl* 4

[85] Sommerfeld, A. ; Frank, N.: The Statistical theory of thermoelectric, galvano- and thermomagnetic phenomena in metals. In: *Rev. Mod. Phys.* 3 (1931), Jan, 1-42. `http://dx.doi.org/10.1103/RevModPhys.3.1`. – DOI 10.1103/RevModPhys.3.1

[86] Chaisantikulwat, W ; Mouis, M ; Ghibaudo, G ; Gallon, C ; Fenouillet-Beranger, C ; Maude, DK ; Skotnicki, T ; Cristoloveanu, S: Differential magnetoresistance technique for mobility extraction in ultra-short channel FDSOI transistors. In: *Solid-state electronics* 50 (2006), Nr. 4, S. 637–643

[87] Flachowsky, Stefan ; Illgen, Ralf ; Herrmann, Tom ; Klix, Wilfried ; Stenzel, Roland ; Ostermay, Ina ; Naumann, Andreas ; Wei, Andy ; Höntschel, Jan ; Horstmann, Manfred: Detailed simulation study of embedded SiGe and Si:C source/drain stressors in nanoscaled silicon on insulator metal oxide semiconductor field effect transistors. In: *Journal of Vacuum Science & Technology B* 28 (2010), Nr. 1, C1G12-C1G17. `http://dx.doi.org/http://dx.doi.org/10.1116/1.3258631`. – DOI http://dx.doi.org/10.1116/1.3258631

[88] Flachowsky, S. ; Wei, A. ; Herrmann, T. ; Illgen, R. ; Horstmann, M. ; Richter, R. ; Salz, H. ; Klix, W. ; Stenzel, R.: Gate length scaling trends of drive current enhancement in {CMOSFETs} with dual stress overlayers and embedded-SiGe. In: *Materials Science and Engineering: B* 154–155 (2008), Nr. 0, 98 - 101. `http://dx.doi.org/http://dx.doi.org/10.1016/j.mseb.2008.09.022`. – DOI http://dx.doi.org/10.1016/j.mseb.2008.09.022. – ISSN 0921–5107. – Front-End Junction and Contact Formation in Future Silicon/Germanium Based Devices

[89] International Roadmap for Semiconductors ITRS, Santa Clara, CA: 2001 ITRS Front End Process. (2001), Nov. 29, S. 22

[90] Flachowsky, Stefan ; Wei, Andy ; Illgen, Ralf ; Herrmann, Tom ; Höntschel, Jan ; Horstmann, Manfred ; Klix, Wilfried ; Stenzel, Roland: Understanding strain-induced drive-current enhancement in strained-silicon n-MOSFET and p-MOSFET. In: *Electron Devices, IEEE Transactions on* 57 (2010), Nr. 6, S. 1343–1354

[91] Thompson, Scott E. ; Sun, Guangyu ; Choi, Youn S. ; Nishida, Toshikazu: Uniaxial-process-induced strained-Si: extending the CMOS roadmap. In: *IEEE Transactions on Electron Devices* 53 (2006), Nr. 5, S. 1010–1020

[92] Gannavaram, Shyam ; Pesovic, Nemanja ; Ozturk, C: Low temperature (800/spl deg/C) recessed junction selective silicon-germanium source/drain technology for sub-70 nm CMOS. In: *Electron Devices Meeting, 2000. IEDM'00. Technical Digest. International* IEEE, 2000, S. 437–440

[93] Subramanian, N. ; Ghibaudo, G. ; Mouis, M. ; Maude, D.K.: Magnetoresistance mobility extraction in the saturation regime of short channel MOS devices. In: *Semiconductor Conference Dresden-Grenoble (ISCDG), 2012 International* (2012), Sept, S.

119–122. http://dx.doi.org/10.1109/ISCDG.2012.6360041. – DOI 10.1109/ISCDG.2012.6360041

[94] Shahidi, G.G. ; Antoniadis, D.A. ; Smith, Henry I.: Electron velocity overshoot at room and liquid nitrogen temperatures in silicon inversion layers. In: *Electron Device Letters, IEEE* 9 (1988), Feb, Nr. 2, S. 94–96. http://dx.doi.org/10.1109/55.2051. – DOI 10.1109/55.2051. – ISSN 0741–3106

[95] Chou, S.Y. ; Antoniadis, D.A.: Relationship between measured and intrinsic transconductances of FET's. In: *Electron Devices, IEEE Transactions on* 34 (1987), Feb, Nr. 2, S. 448–450. http://dx.doi.org/10.1109/T-ED.1987.22942. – DOI 10.1109/T–ED.1987.22942. – ISSN 0018–9383

[96] Pham-Nguyen, L. ; Fenouillet-Beranger, C. ; Ghibaudo, G. ; Skotnicki, T. ; Cristoloveanu, S.: Mobility enhancement by CESL strain in short-channel ultrathin SOI MOSFETs. In: *Solid-State Electronics* 54 (2010), Nr. 2, 123 - 130. http://dx.doi.org/http://dx.doi.org/10.1016/j.sse.2009.12.006. – DOI http://dx.doi.org/10.1016/j.sse.2009.12.006. – ISSN 0038–1101. – Selected Full-Length Extended Papers from the EURO-SOI 2009 Conference

[97] Beister, Jürgen ; Wachowiak, Andre ; Heinzig, André ; Trommer, Jens ; Mikolajick, Thomas ; Weber, Walter M.: Temperature dependent switching behaviour of nickel silicided undoped silicon nanowire devices. In: *physica status solidi (c)* 11 (2014), Nr. 11-12, 1611–1617. http://dx.doi.org/10.1002/pssc.201400055. – DOI 10.1002/pssc.201400055. – ISSN 1610–1642

[98] Fleetwood, DM ; Winokur, PS ; Reber Jr, RA ; Meisenheimer, TL ; Schwank, JR ; Shaneyfelt, MR ; Riewe, LC: Effects of oxide traps, interface traps, and "border traps"on metal-oxide-semiconductor devices. In: *Journal of applied physics* 73 (1993), Nr. 10, S. 5058–5074

[99] Olthuis, W ; Bergveld, P: On the charge storage and decay mechanism in silicon dioxide electrets. In: *Electrical Insulation, IEEE Transactions on* 27 (1992), Nr. 4, S. 691–697

[100] Weber, Walter M. ; Heinzig, Andre ; Trommer, Jens ; Grube, Matthias ; Kreupl, Franz ; Mikolajick, Thomas: Reconfigurable Nanowire Electronics-Enabling a Single CMOS Circuit Technology. In: *Nanotechnology, IEEE Transactions on* 13 (2014), Nr. 6, S. 1020–1028

[101] Ferain, Isabelle ; Colinge, Cynthia A. ; Colinge, Jean-Pierre: Multigate transistors as the future of classical metal-oxide-semiconductor field-effect transistors. In: *Nature* 479 (2011), Nr. 7373, S. 310–316

[102] Knoch, J. ; Zhang, M. ; Appenzeller, J. ; Mantl, S.: Physics of ultrathin-body silicon-on-insulator Schottky-barrier field-effect transistors. In: *Applied Physics A* 87 (2007), Nr. 3, 351-357. `http://dx.doi.org/10.1007/s00339-007-3868-1`. – DOI 10.1007/s00339–007–3868–1. – ISSN 0947–8396

[103] Yu, LC ; Cheung, KP ; Tilak, V ; Dunne, G ; Matocha, K ; Campbell, JP ; Suehle, JS ; Sheng, K: A fast, simple wafer-level Hall-mobility measurement technique. In: *Integrated Reliability Workshop Final Report, 2009. IRW'09. IEEE International* IEEE, 2009, S. 73–76

[104] Hess, Karl: Magnetoresistance of n-silicon inversion layers in the ohmic and in the hot-electron range. In: *Surface Science* 58 (1976), Nr. 1, S. 235–237

[105] Shin, Minju ; Shi, Ming ; Mouis, Mireille ; Cros, Antoine ; Josse, Emmanuel ; Mukhopadhyay, Saibal ; Piot, Benjamin ; Kim, Gyu-Tae ; Ghibaudo, Gerard: Experimental and Theoretical Investigation of Magnetoresistance From Linear Regime to Saturation in 14-nm FD-SOI MOS Devices. In: *Electron Devices, IEEE Transactions on* 62 (2015), Nr. 1, S. 3–8

Eigene Publikationen

(1) **J. Beister**, A. Wachowiak, R. Boschke, T. Herrmann, M. Uhlarz and T. Mikolajick: Mobility Investigations on Strained 30 nm High-k Metal Gate MOSFETs by Geometrical Magnetoresistance Effect, IEEE Trans. El. Dev. Vol. 62, No. 6, pp. 1819 - 1825 (2015)

(2) **J. Beister**, A. Wachowiak, A. Heinzig, J. Trommer, T. Mikolajick and W. M. Weber: Temperature dependent switching behaviour of nickel silicided undoped silicon nanowire devices, Phys. Status Solidi C 11, No. 11-12, pp. 1611-1617, November (2014), *with Journal Cover Page, Title: 'Schottky Barriers in Ni- silicided Si nanowire devices'*

(3) **J. Beister**, A. Wachowiak, A. Heinzig, J. Trommer, T. Mikolajick and W. M. Weber: Investigation of Schottky Barriers in Ni-silicided Silicon Nanowire Devices, *Conference Talk* Symposium X, eMRS Spring Meeting, Lille, France, (2014)

(4) J. Trommer, A. Heinzig, U. Mühle, M. Löffler, A. Winzer, P. M. Jordan, **J. Beister**, T.Baldauf, M. Geidel, B. Adolphi, E. Zschech, T. Mikolajick and W. Weber: Enabling Energy Efficience and Polarity Control in Germanium Nanowire Transistors by Individually Gated Nanojunctions, ACS Nano 11(2), pp 1704-1711 (2017)

Danksagungen

Auf dem Weg zu dieser Arbeit haben mich viele Leute begleitet, denen ich hier herzlich danke sagen möchte. Zunächst einmal danke ich meinem Doktorvater Prof. Thomas Mikolajick zusammen mit Dr. Stefan Jakschik für die Vergabe des interessanten Themas, sowie Prof. Jörg Schulze für die Begutachtung der Arbeit. Danke auch an das BMBF, das den Magnetowiderstandsteil dieser Arbeit im Rahmen des CoolSilicon Spitzenclusters im Projekt 'CoolRule' (16N10184) gefördert hat, sowie an die DFG, die Teile der Nanodraht-Untersuchungen im Rahmen des ReproNano Projekts (MI 1247/6-2 und WE 4853/1-2) unterstützt hat.

Dr. Maciej Wiatr und dem Team bei GLOBALFOUNDRIES danke ich für das entgegengebrachte Vertrauen und die Bereitstellung der Proben. Besonderer Dank geht dabei an Roman Boschke für die Prozessierung der Stressliner-Splitwafer, an Dr. Stefan Flachowsky für die vielen wertvollen Diskussionen der Daten, Dr. Tom Herrmann für Simulationen, Martin Drubba für die Führung durch den Parameterdschungel der Inline-Daten und Layout-Files und ebenso an John Morgan und Dr. Ricardo Mikalo für Diskussionen, Teststrukturfragen- und Probenunterstützung. Danke an Dr. Bo Bai und Dr. Juliano Razafindramora für die Unterstützung bei der Auswertung einiger Inline-Daten. Danke an Dr. Hans-Jürgen Engelmann und seinem Team für die tollen TEM-Analysen und den Zugang zum FIB-System. Ganz besonderer Dank geht an Dr. Sabine Kolodinski für die administrative Unterstützung und den Gästezugang bei GLOBALFOUNDRIES.

Danke an Prof. Wosnitza vom HZDR für den Zugang zu den Einrichtungen des Hochfeld-Magnetlabors sowie zu den 14 Tesla PPMS

Systemen. Ganz herzlich bedanken möchte ich mich an dieser Stelle nochmals bei Dr. Marc Uhlarz für seinen großen Einsatz bei der Rescourcenplanung am HZDR, die tatkräftige Hilfe im Labor, die wertvollen Diskussionen und die herzliche Gastfreundschaft. Danke auch an Danilo Bürger und Fr. Dr. Heidemarie Schmidt für den Zugang zu Ihrem 9 Tesla System in der frühen Phase des Projekts, sowie an Dr. Iida und Dr. Nenkov vom IfW Dresden für das Ermöglichen einiger Testmessungen, welche für die Entwicklung des gesamten Messsystems sehr wertvoll waren.

Bei der Werkstatt der Elektrotechnik der TU Dresden bedanke ich mich sehr für die genaue Anfertigung einiger feinmechanischer Teile für den in dieser Arbeit entwickelten PPMS Probenhalter; danke auch an Herr Günther und das Team der IfW-Institutswerkstatt für die Maßanfertigung der Kontaktplatine des Probenhalters, sowie an die Mitarbeiter der Firma Microceram für die präzise Fertigung der Keramikscheibe mit extrem feinen Bohrungen in engen Toleranzen zur Aufnahme der Federkontakte. Danke an das Team um Herrn Woldt bei MPD für den tollen Support beim Aufbau und Bonden der Chips. Danke für kurze Kommunikationswege, rasche Lieferung und die große Sorgfalt an den wertvollen Teilen trotz kleinem Auftragsvolumen und die hilfreichen Tipps zur Entwicklung einer geeigneten Prozedur zur ESD-sicheren Handhabung der Proben.

Am NaMLab danke ich neben meinem Doktorvater vor allem auch dem Geschäftsführer Dr. Alexander Ruf für das tolle Arbeitsklima und den einzigartigen Mitarbeiterumgang am Institut. Danke für viele unkomplizierte und unbürokratische Regelungen und Möglichkeiten. Danke für das entgegengebrachte Vertrauen, für herzliche Gespräche und dass Deine Tür immer offen steht! Dr. Walter Weber und Dr. André Heinzig von der Nanowire-Gruppe danke ich für die Herstellung der Nanodraht-Proben und die wertvollen fachlichen Diskussionen. Danke auch an Steve Knebel für die tatkräftige LabView-Unterstützung und Dr. Guntrade Roll für die Bereitstellung von Testproben sowie der Einführung in die Messtechnik. Claudia Richter und Terry Jay danke

ich für die meisterliche Hilfe mit dem grünen 0,1 kW Laserschwert zur Extraktion filigraner Probensegmente aus fragilen, 300 mm großen Wafern. Danke an Dr. Stefan Döring, Jan Gärtner und Dr. Dominik Martin für die Unterstützung bei interessanten SSRM- und AFM-Messungen die vor allem im Rahmen des CoolRule Projekts durchzuführen waren. Danke an Dr. Martin Krupinski und Dr. Stefan Schmult für die fachlichen wie persönlichen Gespräche und gesellige Mittagsrunden. Ein riesen DANKE! an Dr. Andre Wachowiak für die wissenschaftliche Begleitung dieser Arbeit, die Motivation in herausfordernden Zeiten und das Mitfiebern in vielen spannenden und unvergesslichen Messnächten im Labor! Danke an Susi für die große Hilfe bei der grafischen Nachbearbeitung einiger Abbildungen.

Zu guter Letzt möchte ich mich bei Freunden und Familie für die Begleitung durch diese besondere Zeit bedanken; insbesondere bei meinen Eltern und Schwiegereltern für die fantastische familiäre Unterstützung, die ihr uns gegeben habt. Danke an unsere Kinder für die quirligen und schönen Momente, Legobauen und herumtoben, wenn Papa mal wieder den Kopf frei brauchte... Ohne den Rückhalt meiner Frau Barbara hätte ich diese Arbeit wahrscheinlich nicht zu Ende gebracht. Danke für die Gespräche mit Dir und dass Du mich immer wieder motiviert und aufgebaut hast. Danke für Deine Freundschaft und dass Du mit mir Seite an Seite durch diese herausfordernde Zeit gegangen bist.

Ich liebe Dich!

Abstract

As modern microelectronics advances, enormous challenges have to be overcome in order to further increase device performance, enabling high-speed and ultra-low-power applications. With progressive scaling of Silicon MOSFETs, charge carrier mobility has dropped significantly and became a critical device parameter over the last decade. Present technology nodes make use of strain engineering to partially recover this mobility loss. Even though carrier mobility is a crucial parameter for present technology nodes, it cannot be determined accurately by methods typically available in industrial environments. A major objective of this work is to study the magnetoresistance mobility μ_{MR} of strained VLSI devices based on a 28 nm ground rule. This technique allows for a more direct access to charge carrier mobility, compared to conventional current/ voltage and capacitance/ voltage mobility derivation methods like the effective mobility μ_{eff}, in which series resistance, inversion charge density and effective channel length are necessary to extract the mobility values of the short channel devices. Aside from providing an anchor for accurate μ_{eff} measurements in linear operation conditions, μ_{MR} opens the possibility to investigate the saturation region of the device, which cannot be accessed by μ_{eff}. Electron and hole mobility of nFET and pFET devices with various gate lengths are studied from linear to saturation region. In addition, the interplay between mobility enhancement due to strain improvement, and mobility degradation due to short channel effects with decreasing channel length is analyzed.

As a concept device for future nanoelectronic building blocks, silicon nanowire Schottky field-effect transistors are investigated in the second part of this work. These devices exhibit an ambipolar behaviour, which gives the opportunity to measure both electron and hole transport on a single device. The temperature dependence of the source/drain current for specific gate and drain voltages is analyzed within the framework of voltage dependent effective barrier heights.