

Technische Universität Dresden

Integration von Multi-Gate-Transistoren auf Basis einer 22 nm-Technologie

Tim Baldauf

von der Fakultät Elektrotechnik und Informationstechnik der Technischen Universität
Dresden

zur Erlangung des akademischen Grades eines

Doktoringenieurs
(Dr.-Ing.)

genehmigte Dissertation

Vorsitzender: Prof. Dr. rer. nat. Johann W. Bartha

Gutachter: Prof. Dr.-Ing. habil. Gerald Gerlach

Prof. Dr.-Ing. habil. Roland Stenzel

Tag der Einreichung: 18.10.2013

Tag der Verteidigung: 10.01.2014

Kurzfassung

Die kontinuierliche Skalierung der planaren MOSFETs war in den vergangenen 40 Jahren der Schlüssel, um die Bauelemente immer kleiner und leistungsfähiger zu gestalten. Hinzu kamen Techniken zur mechanischen Verspannung, Verfahren zur Kurzzeitausheilung, die in-situ-dotierte Epitaxie und neue Materialien, wie das High-k-Gateoxid in Verbindung mit Titanitrid als Gatemetall. Jedoch erschwerten Kurzkanaleffekte und eine zunehmende Streuung der elektrischen Eigenschaften die Verkleinerung der planaren Transistoren erheblich. Somit gelangten die planaren MOSFETs mit der aktuellen 28 nm-Technologie teilweise an die Grenzen ihrer Funktionalität.

Diese Arbeit beschäftigt sich daher mit der Integration von Multi-Gate-Transistoren auf Basis einer 22 nm-Technologie, welche eine bessere Steuerfähigkeit des Gatekontaktes aufweisen und somit die Fortführung der Skalierung ermöglichen. Zudem standen die Anforderungen eines stabilen und kostengünstigen Herstellungsprozesses als Grundvoraussetzung zur Übernahme in die Volumenproduktion stets im Vordergrund.

Die Simulationen der Tri-Gate-Transistoren stellten dabei den ersten Schritt hin zu einer Multi-Gate-Technologie dar. Ihre Prozessabfolge unterscheidet sich von den planaren Transistoren nur durch die Formierung der Finnen und bietet damit die Möglichkeit eines hybriden 22 nm-Prozesses. Am Beispiel der Tri-Gate-Transistoren wurden zudem die Auswirkungen der Kristallorientierung, der mechanischen Verspannung und der Überlagerungseffekte des elektrischen Feldes auf die Leistungsfähigkeit von Multi-Gate-Strukturen analysiert.

Im nächsten Schritt wurden Transistoren mit vollständig verarmten Kanalgebieten untersucht. Sie weisen aufgrund einer niedrigen Kanaldotierung eine Volumeninversion, eine höhere Ladungsträgerbeweglichkeit und eine geringere Anfälligkeit gegenüber der zufälligen Dotierungsfluktuation auf, welche für leistungsfähige Multi-Gate-Transistoren entscheidende Kriterien sind. Zu den betrachteten Varianten zählen die planaren ultradünnen SOI-MOSFETs, die klassischen FinFETs mit schmalen hohen Finnen und die vertikalen Nanowire-Transistoren.

Anschließend wurden die Vor- und Nachteile der verschiedenen Transistorstrukturen für eine mittel- bis langfristige industrielle Nutzung betrachtet. Dazu erfolgte eine Analyse der statistischen Schwankungen und eine Skalierung hin zur 14 nm-Technologie.

Eine Zusammenfassung aller Ergebnisse und ein Ausblick auf die mögliche Übernahme der Konzepte in die Volumenproduktion schließen die Arbeit ab.

Abstract

Within the past 40 years the continuous scaling of planar MOSFETs was key to shrink the devices and to improve their performance. Techniques like mechanical stressing, rapid thermal annealing and in-situ doped epitaxial growing as well as novel materials, such as high-k-gate-oxide in combination with titanium nitride as metal-gate, has been introduced. However, short-channel-effects and increased scattering of electrical properties significantly complicate the scaling of planar transistors. Thus, the planar MOSFETs gradually reached their limits of functionality with the current 28 nm technology node.

For that reason, this work focuses on integration of multi-gate transistors based on a 22 nm technology, which show an improved gate control and allow a continuous scaling. Furthermore, the requirements of a stable and cost-efficient process as decisive condition for mass fabrication were always taken into account.

The simulations of the tri-gate transistors present the first step toward a multi-gate technology. The process sequence differs from the planar one solely by a fin formation and offers the possibility of a hybrid 22 nm process. Also, the impact of crystal orientation, mechanical stress and superposition of electrical fields on the efficiency of multi-gate structures were analyzed for the tri-gate transistors.

In a second step transistors with fully depleted channel regions were studied. Due to low channel doping they are showing a volume inversion, a higher carrier mobility and a lower sensitivity to random doping fluctuations, which are essential criteria for powerful multi-gate transistors. Reviewed structure variants include planar ultra-thin-body-SOI-MOSFETs, classic FinFETs with a tall, narrow fins and vertical nanowire transistors.

Then advantages and disadvantages of the considered transistor structures have been observed for a medium to long term industrial use. For this purpose, an analysis of statistical fluctuations and the scaling-down to 14 nm technology was carried out.

A summary of all results and an outlook to the transfer of concepts into mass fabrication complete this work.

Inhaltsverzeichnis

Inhaltsverzeichnis	I
Symbol- und Abkürzungsverzeichnis	V
1 Einleitung	1
2 Grundlagen und Entwicklung der CMOS-Technologie	5
2.1 Planare Transistoren	6
2.1.1 Theoretische Grundlagen von MOSFETs	6
2.1.2 Skalierung und Kurzkanalverhalten planarer Transistoren	17
2.1.3 Mechanische Verspannung von Silizium	25
2.1.4 Techniken zur mechanischen Verspannung	28
2.2 Multi-Gate-Transistoren	31
2.2.1 Multi-Gate-Strukturen	31
2.2.2 Überlagerungseffekte	34
2.2.3 Quanteneffekte	36
2.3 Stand der Technik	39
3 Grundlagen der Simulation	43
3.1 Prozesssimulation	44
3.1.1 Abscheiden und Abtragen von Schichten	44
3.1.2 Implantation	45
3.1.3 Thermische Ausheilung mit Diffusion	46
3.2 Bauelementesimulation	48
3.2.1 Grundgleichungen und Ladungsträgertransport	48
3.2.2 Bandlückenverengung	50
3.2.3 Generation und Rekombination	51
3.2.4 Ladungsträgerbeweglichkeit	53
3.2.5 Effekte der mechanischen Verspannung	56
3.2.6 Ladungsträgerquantisierung	60
3.3 Kalibrierung der Modellparameter	60
3.3.1 Prozessparameter	60
3.3.2 Modellparameter	63

4	Planare Transistoren auf Basis einer 22 nm-Technologie	67
4.1	Transistoraufbau	68
4.1.1	Replacement-Gate-Prozess	68
4.1.2	In-situ-dotierte Source-Drain-Gebiete	71
4.1.3	Haloimplantation	73
4.1.4	Elemente der mechanischen Verspannung	74
4.2	Charakterisierung des elektrischen Verhaltens	76
4.2.1	Stationäres Verhalten	76
4.2.2	Gatesteuerung und Kurzkanaleffekte.....	78
4.2.3	Dynamisches Verhalten	80
5	Tri-Gate-Transistoren	83
5.1	Prozessintegration und Transistoraufbau.....	84
5.1.1	Anforderungen an hochintegrierte Schaltkreise	84
5.1.2	Hybride CMOS-Technologie.....	84
5.1.3	Strukturierung der Finne	86
5.1.4	Geometrieabhängiges Dotierungsprofil	87
5.2	Charakterisierung des elektrischen Verhaltens	88
5.2.1	Stationäres Verhalten	88
5.2.2	Kurzkanaleffekte und Gatesteuerung.....	90
5.2.3	Eckeneffekt	91
5.2.4	Eckenimplantation	97
5.2.5	Finnengeometrie	100
5.2.6	Dynamisches Verhalten	101
5.3	Optimierung der Tri-Gate-Struktur.....	103
5.3.1	Gestaltung der epitaktischen Source-Drain-Gebiete.....	103
5.3.2	Mechanisch verspanntes Isolationsoxid.....	105
5.3.3	Substratorientierung.....	109
6	Transistoren mit vollständig verarmtem Kanal	119
6.1	Ultra-Dünne-SOI-MOSFETs.....	120
6.1.1	Prozessintegration	120
6.1.2	Charakterisierung des elektrischen Verhaltens	121
6.2	FinFETs	123
6.2.1	Prozessintegration	123
6.2.2	Charakterisierung des elektrischen Verhaltens	125
6.3	Vertikale Nanowire-MOSFETs	129
6.3.1	Prozessintegration	129
6.3.2	Strukturierung des Aktivgebiets	132
6.3.3	Charakterisierung des elektrischen Verhaltens	134
6.3.4	Asymmetrisches Dotierungsprofil	140
6.3.5	Mechanische Verspannung	142

7	Skalierung und statistische Schwankungen der Strukturen	145
7.1	Skalierung zur 14 nm-Technologie	146
7.1.1	Leistungsfähigkeit	146
7.1.2	Kurzkanalverhalten und Steuerfähigkeit	147
7.2	Statistische Schwankungen	149
7.2.1	Impedanz-Feld-Methode	149
7.2.2	Zufällige Dotierungsfluktuation	152
7.2.3	Fixe Ladungen im Oxid	154
7.2.4	Metall-Gate-Granularität	155
7.2.5	Geometrische Variationen	156
7.2.6	Kombination der Störquellen	158
8	Zusammenfassung und Ausblick	161
	Anhang	167
	Literaturverzeichnis	173
	Danksagung	185
	Acknowledgement	186
	Lebenslauf	187



Symbol- und Abkürzungsverzeichnis

Indizes	Beschreibung
n	Elektronen- bzw. n-leitend
p	Löcher- bzw. p-leitend
lin	Linearbereich der Drain-Source-Spannung
sat	Sättigungsbereich der Drain-Source-Spannung
ref	Referenzwert der entsprechenden Größe

Symbol	Einheit	Beschreibung
a	m	Gitterkonstante
C_{Box}	F	Kapazitätsanteil zwischen Wanne und Substrat
C_d	F	Kapazität der Verarmungszone
C_{GS}	F	Gate-Source-Kapazität
C'_{inv}	$F \cdot m^{-2}$	Flächennormierte Inversionskapazität
C_{ISS}	F	Eingangskapazität bei Source-Schaltung
C_j	F	Sperrschichtkapazität
C_{mill}	F	MILLERkapazität
C_{ov}	F	Überlappungskapazität
C_{ox}	F	Gateoxidkapazität
CD	m^{-2}	Dosis der Eckenimplantation
CE	1	Verteilung der Ladungsträger zwischen Eck- und Top-Gate-Bereich eines Tri-Gate-Transistors (Eckeneffekt)
D	$m^2 \cdot s^{-1}$	Diffusionskoeffizient der Ladungsträger
D_{Dot}	$m^2 \cdot s^{-1}$	Diffusionskonstante von Dotanden
$DIBL$	$mV \cdot V^{-1}$	Drain Induced Barrier Lowering
E	$V \cdot m^{-1}$	Elektrische Feldstärke
E_a	eV	Aktivierungsenergie
E_{eff}	$V \cdot m^{-1}$	Effektive elektrische Feldstärke
E_q	$V \cdot m^{-1}$	Quantisierungsfeldstärke
EI	1	Elektrostatische Integrität
EOT	m	Equivalent Oxide Thickness (äquivalente Gateoxidicke)
$E_{ }$	$V \cdot m^{-1}$	Parallele elektrische Feldstärke
E_{\perp}	$V \cdot m^{-1}$	Orthogonale elektrische Feldstärke
f_{RO}	Hz	Ringoszillatorfrequenz
f_t	Hz	Transitfrequenz

Symbol	Einheit	Beschreibung
G	$\text{m}^{-3} \cdot \text{s}^{-1}$	Generationsrate
$G_{k,s}^I, G_{k,s}^U$	-	GREENsche Funktionen der Impedanz-Feld-Methode
g_m	S	Steilheit
H_{Fin}	m	Höhe der Finne
h_{21}	1	Kleinsignalstromverstärkung
HD	m^{-2}	Dosis der Haloimplantation
I_D	A	Drainstrom
$I_{D,th}$	A	Konstantstrom (Drain) zur Bestimmung der Schwellspannung
$I_{D,off}$	A	Sperrstrom (Drain)
$I_{Ro,eff}$	A	Effektive Stromaufnahme des Ringoszillators (schwingend)
$I_{Ro,off}$	A	Stromaufnahme des Ringoszillators (ohne Rückkopplung)
J	$\text{A} \cdot \text{m}^{-2}$	Stromdichte
J_T	$\text{m}^{-2} \cdot \text{s}^{-1}$	Teilchenstromdichte
\mathbf{k}	m^{-1}	Wellenvektor
L	m	Länge des Transistors
L_{eff}	m	Effektive Gatelänge
L_{Gate}	m	Physikalische Gatelänge
L_{Pitch}	m	Mittlerer Abstand der Gatekontakte zweier Transistoren
m^*	kg	Effektive Masse
\bar{m}^*	kg	Effektive Leitfähigkeitsmasse
m_i^*, m_t^*	kg	Longitudinale/Transversale effektive Masse
m_{lh}^*, m_{hh}^*	kg	Effektive Masse des leichten/schweren Löcherbandes
m_q^*	kg	Effektive Quantisierungsmasse
MUG	1	Vierpolverstärkung ohne Rückkopplung
N	m^{-3}	Teilchendichte
N_a	m^{-3}	Teilchendichte der aktiven Dotanden
N_A	m^{-3}	Akzeptordichte
N_{Ch}	m^{-3}	Mittlere Dotierungskonzentration im Kanalgebiet
N_D	m^{-3}	Donatordichte
N_{tot}	m^{-3}	Absolute Störstellendichte
n	m^{-3}	Elektronendichte
n_i	m^{-3}	Eigenleitendichte
n_{MG}	1	Effektive Anzahl der Gateflächen
p	m^{-3}	Löcherdichte
$Pitch_{Fin}$	m	Abstand der Finnen eines Multi-Gate-Transistors
$Pitch_{NW}$	m	Abstand der Drähte eines Nanowire-Transistors
R	$\text{m}^{-3} \cdot \text{s}^{-1}$	Rekombinationsrate
R_{ges}	Ω	Gesamtwiderstand des Transistors
R_{Kanal}	Ω	Kanalwiderstand

Symbol	Einheit	Beschreibung
R_{on}	Ω	Transistorwiderstand im eingeschalteten Zustand
$R_{on,CO}$	Ω	Transistorwiderstand bei konstanter Inversion
R_{SD}	Ω	Parasitärer Source-Drain-Widerstand
\mathbf{r}	m	Ortsvektor
\mathbf{S}	$W \cdot m^{-2}$	Energiestromdichte der Ladungsträger
S_{ij}	Pa^{-3}	Elastizitätskonstante
$S_{I,kl}, S_{U,kl}$	A^2s, V^2s	Spektraldichte des Rauschstromes bzw. der Rauschspannung
SCE	V	Short Channel Effect (Kurzkanaleffekt)
SS	mV/dec	Unterschwellsteigung
T	K	Ladungsträgertemperatur
T_L	K	Gittertemperatur
t_{BOX}	m	Dicke des vergrabenen Oxids
t_D	s	Mittlere Verzögerungszeit einer Inverterstufe
t_f	s	Abfallzeit einer Inverterstufe
t_r	s	Anstiegszeit einer Inverterstufe
t_{Si}	m	Dicke des Siliziumkanals
t_{ox}	m	Dicke des Gateoxids
U_{BS}	V	Backgate-Source-Spannung
U_{DS}	V	Drain-Source-Spannung
U_{GS}	V	Gate-Source-Spannung
U_T	V	Temperaturspannung im Silizium
U_{th}	V	Schwellspannung
v	$m \cdot s^{-1}$	Ladungsträrgeschwindigkeit
v_{sat}	$m \cdot s^{-1}$	Sättigungsgeschwindigkeit der Ladungsträger
W	m	Weite des Transistors
W_A	eV	Austrittsarbeit
W_C	eV	Energie des Leitungsbandminimums
W_F	eV	Fermienergie
W_{Fin}	m	Weite der Finne
W_g	eV	Energiedifferenz der Bandlücke
W_{Gate}	m	Effektive Gateweite
W_i	eV	Intrinsische Fermienergie des Siliziums
W_{NW}	m	Drahtweite eines Nanowire-Transistors
W_{Str}	m	Strukturweite des Transistors
W_v	eV	Energie des Valenzbandmaximums
w_d	m	Weite der Verarmungszone
x_{Ge}	1	Germaniumanteil im Silizium-Germanium
x_j	m	Tiefe der Source-Drain-Gebiete
\mathbf{y}	S	Admittanzmatrix eines Vierpols

Symbol	Einheit	Beschreibung
α	1	Skalierungsparameter für Multi-Gate-Strukturen
α_{ii}	1	Ionisationsrate der Ladungsträger
$\overline{\Delta E}_C, \overline{\Delta E}_V$	eV	Verschiebung der mittleren Leitungs- bzw. Valenzbandkante
$\Delta E_{C,i}, \Delta E_{V,i}$	eV	Verschiebung der Leitungs- bzw. Valenz-Subbänder
ΔE_{Sub}	eV	Maximale Aufspaltung der Subbänder
$\Delta t_{hi}, \Delta t_{low}$	s	Verzögerungszeit zwischen Ein- und Ausgang
$\delta I_{\vartheta,k}$	A	Änderungen des Rauschstromes
$\delta U_{\vartheta,k}$	V	Änderungen der Rauschspannung
ε	1	Mechanische Deformation
ϑ_n	1	Anzahl der äquivalenten Leitungsbandtöler
κ	1	Skalierungsfaktor
κ_T	$W \cdot m^{-1} \cdot K^{-1}$	Thermische Leitfähigkeit von Elektronen bzw. Löchern
λ	m	Natürliche Länge
λ_{QP}	eV	Quantenkorrekturpotenzial
μ	$m^2 \cdot V^{-1} \cdot s^{-1}$	Ladungsträgerbeweglichkeit
μ_{AP}	$m^2 \cdot V^{-1} \cdot s^{-1}$	Ladungsträgerbeweglichkeit (Gitterschwingung)
μ_{HS}	$m^2 \cdot V^{-1} \cdot s^{-1}$	Ladungsträgerbeweglichkeit (Hochfeldsättigung)
μ_L	$m^2 \cdot V^{-1} \cdot s^{-1}$	Ladungsträgerbeweglichkeit (Gittertemperatur)
μ_{LF}	$m^2 \cdot V^{-1} \cdot s^{-1}$	Ladungsträgerbeweglichkeit bei geringer Feldstärke
μ_{Lomb}	$m^2 \cdot V^{-1} \cdot s^{-1}$	Ladungsträgerbeweglichkeit (LOMBARDI-Modell)
μ_{RCS}	$m^2 \cdot V^{-1} \cdot s^{-1}$	Ladungsträgerbeweglichkeit (ferner Ladungsstreuung)
μ_{RPS}	$m^2 \cdot V^{-1} \cdot s^{-1}$	Ladungsträgerbeweglichkeit (ferner Gitterschwingung)
μ_{SR}	$m^2 \cdot V^{-1} \cdot s^{-1}$	Ladungsträgerbeweglichkeit (Oberflächenrauheit)
σ, σ_{ij}	Pa	Mechanische Verspannung
σ_{el}	$S \cdot m^{-1}$	Elektrische Leitfähigkeit
σI_D	A	Standardabweichung des Drainstroms
σI_{komb}	A	Kombinierte Standardabweichung des Drainstroms
σU_{GS}	V	Standardabweichung der Gate-Source-Spannung
τ	s	Ladungsträgerlebensdauer
τ_0	s	Grundlebensdauer der Ladungsträger
τ_e	s	Energierelaxationszeit
τ_s	s	Mittlere Stoßzeit
$\Phi_{n,p}$	V	Quasi-Fermi-Potenzial
φ	V	Elektrisches Potenzial
$\varphi_{D,i}$	V	Intrinsisches Potenzial des Draingebiets
φ_s	V	Oberflächenpotenzial
Γ, L, K, X		Symmetriepunkte der Brillouin-Zone
Δ, Λ, Σ		Symmetrielinien der Brillouin-Zone

Konstante	Wert	Beschreibung
e	$1.602189 \cdot 10^{-19} \text{ C}$	Elementarladung
\hbar	$6.582119 \cdot 10^{-16} \text{ eV} \cdot \text{s}$	Reduzierte PLANCKSche Wirkungsquantum
k_B	$8.617343 \cdot 10^{-5} \text{ eV} \cdot \text{K}^{-1}$	BOLTZMANN-Konstante
m_0	$9.109390 \cdot 10^{-31} \text{ kg}$	Elektronenruhemasse
ϵ_0	$8.854188 \cdot 10^{-12} \text{ As} \cdot \text{V}^{-1} \cdot \text{m}^{-1}$	Permittivität des Vakuums
ϵ_{Si}	11.8	Relative Permittivität des Siliziums
ϵ_{ox}	3.9	Relative Permittivität des Siliziumoxids (SiO_2)

Abkürzung	Beschreibung
ALD	Atomic Layer Deposition
AP	Acoustic Phonon
B2B	Band to Band
BOX	Buried Oxide
BT	Body Tight
CE	Corner Effect
CMOS	Complementary Metal Oxide Semiconductor
CMP	Chemisch-Mechanisches Polieren
CO	Constant Overdrive
COL	Compressive Overlayer
CPU	Central Processor Unit
cSiGe	Channel Silicon-Germanium
CVD	Chemical Vapour Deposition
DG-MOS	Double Gate Metal Oxide Semiconductor
DIBL	Drain Induced Barrier Lowering
DSL	Dual Stress Liner
EOT	Equivalent Oxide Thickness
eSiGe	Embedded Silicon-Germanium
FB	Floating Body
FD-SOI	Fully Depleted Silicon On Insulator
FinFET	Feld-Effekt-Transistor mit schmaler hoher Finne
HDP	High Density Plasma
hh	Heavy Hole Band (schweres Löcherband)
ISS	Inter Subband Scattering
ITRS	International Technology Roadmap for Semiconductors
lh	Light Hole Band (leichtes Löcherband)
LSA	Laser Spike Annealing
MGG	Metall-Gate-Granularität
MGS	Metal Gate Stress
MLD	Molecular Layer Deposition

Abkürzung	Beschreibung
MOSFET	Metal Oxide Semiconductor Field Effect Transistor
MUG	Maximal Unilateral Gain
n-IFM	Rauschartige Impedanz-Feld-Methode (noise-like)
n-MOSFET	n-Kanal MOSFET
p-MOSFET	p-Kanal MOSFET
RCS	Remote Coulomb Scattering
RDF	Random Doping Fluctuation
RO	Ringoszillator
RPS	Remote Phonon Scattering
RTA	Rapid Thermal Annealing
SCE	Short Channel Effect
SD	Source-Drain
SiC	Silizium-Kohlenstoff
s-IFM	Statistischen Impedanz-Feld-Methode
SMT	Stress Memorization Technique
so	Spin-Orbit-Löcherband
SOI	Silicon On Insulator
SR	Surface Roughness
SS	Subthreshold Swing
sSOI	Strained Silicon On Insulator
STI	Shallow Trench Isolation
TAT	Trap Assisted Tunneling
TEM	Transmissionselektronenmikroskopie
TOL	Tensile Overlayer
UTB	Ultra Thin Body

1 Einleitung

Seit dem Beginn der industriellen Fertigung von Schaltkreisen auf Basis der CMOS-Technologie (Complementary Metal Oxide Semiconductor) in den 1960er Jahren hat sich unser Leben unter dem zunehmenden Einfluss der Rechentechnik stark verändert. Anfangs waren es nur einzelne Großrechner, die von Universitäten und Hochschulen für die Forschung eingesetzt wurden, und schrankgroße Computer, die größeren Unternehmen und Behörden zur Verarbeitung ihrer Daten dienten, doch mit der kontinuierlichen Verkleinerung (Skalierung) der planaren MOSFETs (Metal Oxide Semiconductor Field Effect Transistor) erhöhte sich auch die Funktionalität der Schaltkreise. Der erste seriengefertigte Ein-Chip-Mikroprozessor war der „4004“ von INTEL aus dem Jahr 1971 (Abbildung 1.1). Er bestand aus rund 2300 Transistoren, die mittels 10 μm -Technologie auf 2“-Wafers prozessiert wurden. Prozessoren mit immer kleineren Strukturgrößen und immer höherer Transistorzahl ermöglichten schließlich die Entwicklung von PCs und die industrielle Nutzung von Rechnern. In den vergangenen 40 Jahren sorgte die kontinuierliche Skalierung der Transistoren für einen Anstieg der Komplexität und der Schaltgeschwindigkeit hochintegrierter Schaltkreise.

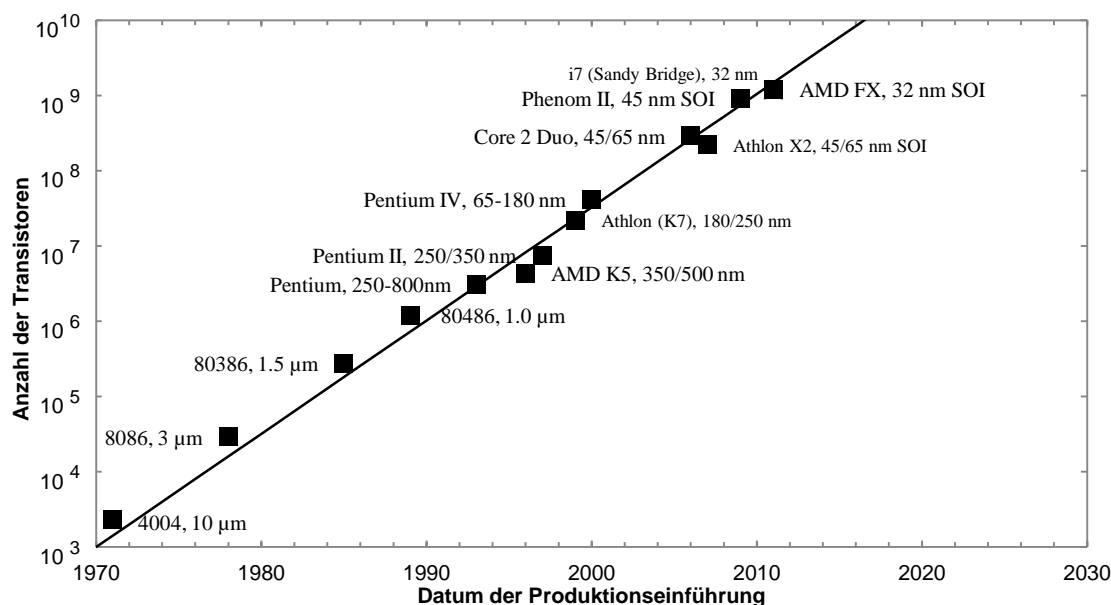


Abbildung 1.1: Transistoranzahl verschiedener Prozessorarchitekturen in Abhängigkeit von der Jahreszahl ihrer Produktionseinführung. Nach dem Gesetz von MOORE verdoppelt sich die Anzahl der integrierten Transistoren aller zwei Jahre, dargestellt durch die Trendlinie [1], [2], [3].

Die aktuellen CPUs (Central Processor Unit) erreichen Taktraten von bis zu 5 GHz und besitzen bereits über eine Milliarde Transistoren, die mittels 32 nm- bzw. 28 nm-Technologie auf 300 mm-Wafern hergestellt werden. Mit den weltweit etablierten mobilen Anwendungen (Smartphones, Tablets etc.), der Steuerung und Regelung von industriellen Maschinen und der stetig wachsenden Infrastruktur des Internets (Rechenzentren und Serverfarmen) ist die Halbleitertechnik heute ein fundamentaler Bestandteil unseres Alltags und stellt einen der führenden Industriebereiche dar.

Die Skalierungsschritte der planaren Transistoren erfolgten in der Vergangenheit aller zwei Jahre und ermöglichten die Halbierung der benötigten Fläche bzw. eine Verdoppelung der Transistorzahl auf einem Chip. Diese Tatsache ist auch als das Gesetz von MOORE bekannt, welches bereits 1975 veröffentlicht wurde [4]. Längere Zeit stellte dabei die Lithographie zur Strukturierung der Bauelemente die wesentliche Herausforderung dar, um die Verkleinerung voranzutreiben, während die physikalischen Grenzen der Transistoren noch in weiter Ferne lagen.

Parallel dazu wurde auch an neuen Materialien und Transistorkonzepten geforscht, um die elektrischen Eigenschaften der MOSFETs unabhängig von ihrer Skalierung zu verbessern [5]. Ein wesentlicher Schritt dabei war der Ansatz des SOI-Substrats (Silicon On Insulator), bei dem das Kanalgebiet durch ein vergrabenes Oxid (Buried Oxide, BOX) vom Substrat getrennt ist und der Anfang der 2000er Jahre unter anderem von IBM, AMD und FREESCALE in die Volumenproduktion übernommen wurde. Aber auch die mechanische Verspannung des Siliziums wurde genutzt, um die Beweglichkeit der Ladungsträger im Kanal zu erhöhen und somit die Leistungsfähigkeit der Transistoren zu steigern.

Mit der 90 nm-Technologie rückten jedoch Kurzkanaleffekte aufgrund der immer kleiner werdenden Kanallängen zunehmend in den Vordergrund der Entwicklung. Neuartige Dotierungsverfahren und Kurzzeitausheilungen konnten vorerst die Schalteigenschaften der Transistoren auch bei fortschreitender Skalierung sichern. Spätestens die Einführungen der 45 nm- und 32 nm- bzw. 28 nm-Technologien haben aber gezeigt, dass planare Transistoren trotz aller Innovationen nicht weiter verkleinert werden können, ohne deren Funktionalität massiv zu beeinträchtigen oder gar ganz zu verlieren.

Da diese Tatsache schon längere Zeit erwartet worden war, begann man bereits frühzeitig mit der Erarbeitung von Multi-Gate-Transistoren und Ansätzen mit vollständig verarmten Kanalgebieten. Solche Strukturen ermöglichen durch eine bessere Steuerfähigkeit des Gatekontakts eine weitere Verringerung der Gatelänge unterhalb von 30 nm und damit eine Fortführung der klassischen Skalierung. Die meisten Konzepte sind jedoch nicht für die Massenproduktion der Halbleiterhersteller geeignet, da ihre Integration trotz kürzerer Gatelänge im Vergleich zum planaren MOSFET mehr Platz beanspruchen würde bzw. mit einem sehr hohen Entwicklungsaufwand und dementsprechenden Kosten verbunden wäre.

Diese Arbeit beschäftigt sich daher mit der Integration von Multi-Gate-Transistoren auf Basis einer 22 nm-Technologie und speziell mit Konzepten, die auch den Richtlinien der ITRS (International Technology Roadmap for Semiconductors) gerecht werden, um die nötigen Voraussetzungen für die Volumenproduktion zu erfüllen. Dabei wird zunächst das grundlegende Verhalten der MOSFETs erläutert und auf die physikalischen Hintergründe der Kurzkanaleffekte, die mechanischen Verspannung sowie die Besonderheiten der Multi-Gate-Strukturen eingegangen.

Einen zentralen Punkt stellen die Tri-Gate-Transistoren dar, welche in Verbindung mit planaren MOSFETs in 22 nm-Technologie einen hybriden Prozessansatz ermöglichen, der eine einfache und kostengünstige Realisierung beider Strukturvarianten auf einem Substrat erlaubt. Die elektrischen Eigenschaften der Bauelemente wurden über eine Prozess- und Bauelementesimulation berechnet, deren Modellparameter mithilfe experimenteller Daten einer 32 nm-SOI-Technologie angepasst wurden. Die Simulationen ermöglichen dabei eine frühzeitige Abschätzung von Ergebnissen und ersparen somit zahlreiche experimentelle Versuche. Zudem lassen sich Parameter abschätzen, die mit der derzeitigen Technik nicht messbar sind.

Neben den Tri-Gate-Transistoren werden auch Strukturen mit vollständig verarmten Kanalgebieten betrachtet. Zu ihnen gehören der ultradünne SOI-Transistor (Ultra Thin Body, UTB), der Feld-Effekt-Transistor mit schmaler hoher Finne (FinFET) und der vertikale Nanowire-Transistor. Sie ermöglichen eine vollständige Steuerung ihres Kanalgebiets und besitzen aufgrund der geringen Kanaldotierung gegenüber den dotierten Strukturen eine bessere Ladungsträgerbeweglichkeit.

Mit der Skalierung zu immer kleineren Strukturen gewinnen auch die statistischen Variationen der elektrischen Eigenschaften immer mehr an Bedeutung. Sie werden durch Schwankungen im Herstellungsprozess verursacht und steigen mit der Verkleinerung der Bauelemente an. Die Variationssimulationen dieser Arbeit basieren auf der Impedanz-Feld-Methode und berücksichtigen die zufälligen Dotierungsfluktuationen, die fixen Oxidladungen, die Metall-Gate-Granularität und die geometrischen Schwankungen der einzelnen Schichten.

Da mit der 22 nm-Technologie jedoch nur die aktuellen Probleme der Halbleiterindustrie betrachtet werden, soll abschließend ein Ausblick auf die 14 nm-Technologie der verschiedenen Strukturvarianten gegeben werden. Damit lässt sich ableiten, welche Konzepte für Multi-Gate-Transistoren auch die mittel- bis langfristigen Ziele der Halbleiterindustrie erfüllen können.

2 Grundlagen und Entwicklung der CMOS-Technologie

Mehrere Jahrzehnte konnten die planaren Transistoren in ihren horizontalen und vertikalen Dimensionen kontinuierlich skaliert werden. Des Öfteren waren die Halbleiterhersteller dabei an physikalische Grenzen der Lithographie sowie des Schichtauf- bzw. Schichtabtrags gestoßen und mussten neue Wege finden, die Verkleinerung der Transistoren voranzutreiben und gleichzeitig ihre Funktionalität zu erhalten. Mit der Entwicklung der 45 nm- und 32 nm- bzw. 28 nm-Technologien ist jedoch klar geworden, dass das Konzept des planaren Transistors für die weitere Skalierung zukünftiger Technologieknoten aufgrund seiner physikalischer Grenzen durch einen Multi-Gate-Ansatz mit vollständig verarmbaren Kanal ersetzt werden muss.

Zunächst wird auf den Aufbau, die Funktionsweise und die Skalierung planarer MOSFETs eingegangen. Dabei stehen die Kurzkanaleffekte und ihre Auswirkungen auf die elektrischen Eigenschaften der Transistoren im Vordergrund. Zusätzlich wird der physikalische Zusammenhang zwischen Gitterdeformierung und elektrischer Leitfähigkeit im Siliziumkristall erläutert und eine Zusammenfassung über die verschiedenen Techniken zur mechanischen Verspannung gegeben. Anschließend folgt eine Übersicht zu den wesentlichsten Multi-Gate-Strukturen sowie den mit ihnen verbundenen Überlagerungs- und Quanteneffekten, bevor die Darstellung des gegenwärtigen wissenschaftlichen und technischen Standes das Kapitel abschließt.

2.1 Planare Transistoren

2.1.1 Theoretische Grundlagen von MOSFETs

A. Aufbau und Typen von MOSFETs

Ein planarer MOSFET ist ein aktives Bauelement und kann auch als unipolarer Transistor angesehen werden, da zu dem Stromfluss zwischen Source- und Drainkontakt nur Ladungsträger einer Polarität beitragen [6]. Sein prinzipieller Aufbau (Abbildung 2.1) ist im Wesentlichen gegeben durch ein einkristallines ebenflächiges Substrat, einem Isolator, den Kontakten Source, Drain und Gate sowie dem Bulkkontakt auf der Unterseite des Substrats. Der planare Gatekontakt befindet sich oberhalb des Isolators und ist somit vom Silizium elektrisch getrennt. Der Bereich unterhalb des Isolators wird als Kanal bezeichnet und stellt eine vom Gatekontakt steuerbare Verbindung zwischen den seitlichen Source-Drain-Gebieten (SD-Gebieten) dar.

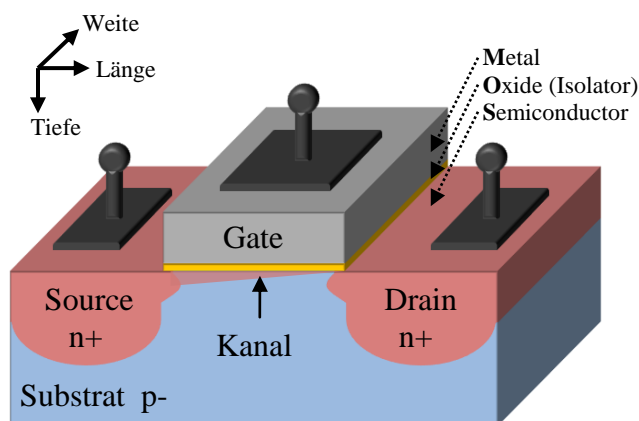


Abbildung 2.1: Schematische Darstellung eines planaren n-MOSFETs.

Funktionell gesehen ist der MOSFET eine spannungsgesteuerte (Gate-Source-Spannung) Stromquelle (Drainstrom), fungiert in hochintegrierten Schaltungen aufgrund seines großen Verhältnisses zwischen Sperr- und Durchlasswiderstand jedoch als digitaler Schalter. Ob ein Transistor elektronen- (n-MOSFET) oder löcherleitend (p-MOSFET) ist, richtet sich nach seinem Dotierungsprofil (Abbildung 2.2), welches in der Regel durch die Implantation von Ionen der dritten und fünften Hauptgruppe realisiert wird. Das Kanalgebiet des n-MOSFETs ist regulär mit einer schwachen (p-)-Dotierung versehen, die aus einer Wannenheimplantation von Borionen resultiert. Die Source- und Drain-Gebiete des n-MOSFETs hingegen benötigen eine hochkonzentrierte (n+)-Dotierung, welche mithilfe von Arsen- oder Phosphorionen erzeugt werden kann. Der Gatekontakt dient dabei als Maskierung und schützt den darunterliegenden Kanal vor der Implantation. Eine Vertauschung der Implantationsspezies führt zu einem löcherleitenden p-MOSFET. Mit diesem Aufbau des Dotierungsprofils sind die Transis-

toren im Grundzustand ($U_{GS} = 0\text{ V}$) gesperrt und werden als Anreicherungstypen bezeichnet, da erst die Gate-Source-Spannung den Kanal mit Elektronen bzw. Löchern füllt und den Transistor leitend macht. Es ist jedoch auch möglich, den Transistor mit einem Kanal zu prozessieren, der bereits im Grundzustand leitend ist und erst mithilfe der Gate-Source-Spannung gesperrt wird. Diese Art von Transistor stellt den Verarmungstyp dar, der jedoch wegen des zusätzlichen Implantationsschrittes und der schlechteren Schalteigenschaften gegenüber dem Anreicherungstyp in den Hintergrund der industriellen Fertigung gerückt ist [7] [8].

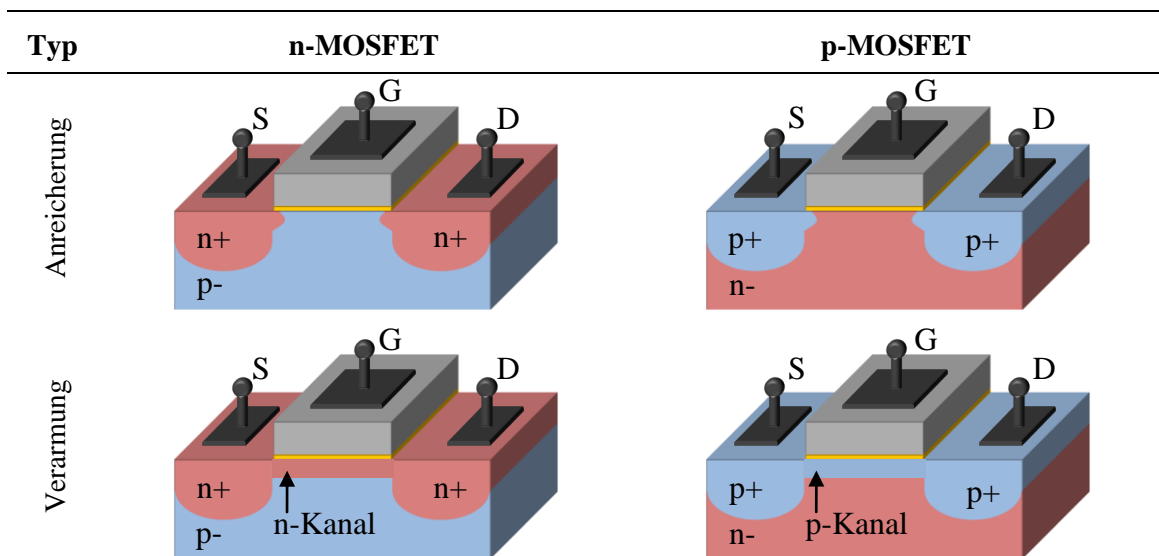


Abbildung 2.2: Schematische Darstellung der vier grundlegenden MOSFET-Typen.

Um später die Vorgänge in planaren MOSFETs und Multi-Gate-Transistoren besser beschreiben zu können, sollen im Folgenden einige grundlegende Eigenschaften von Halbleitern und pn-Übergängen erklärt werden.

Als Ausgangsmaterial für MOSFETs dient in der Regel Silizium, wie es für die Fertigung von Halbleiterbauelementen bereits seit mehreren Jahrzehnten genutzt wird [9]. Es besitzt in seiner kristallinen Form eine Diamantstruktur (kubisch flächenzentriertes Gitter) mit einer Dichte von $5 \cdot 10^{22}$ Atomen pro cm^3 und zählt zu den indirekten Halbleitern, bei denen das Valenzbandmaximum und das Leitungsbandminimum nicht im gleichen Punkt der Brillouin-Zone (Wigner-Seitz-Zelle des reziproken Gitters, Abbildung 2.3) liegen [10]. Dies führt dazu, dass die Elektronen beim Übergang vom Leitungs- in das Valenzband einen großen Impulsübertrag verursachen und die freiwerdende Energie vom Kristallgitter in Form von Gitterschwingungen (Phononen) aufgenommen wird [11]. Somit kann im Vergleich zu direkten Halbleitern kein Photon emittiert werden.

B. Bandstruktur in MOSFETs

Im hochreinen, undotierten Silizium betragen die Energien des Leitungsbandminimums und die des Valenzbandmaximums bezogen auf das Vakuumniveau $W_c = 4.08$ eV bzw. $W_v = 5.21$ eV. Daraus resultieren eine Bandlücke von $W_g = 1.13$ eV und eine Fermienergie des intrinsischen Siliziums von $W_i = 4.65$ eV. In einigen Fällen, wie in Abbildung 2.3, dient jedoch das Valenzbandmaximum als Bezugsniveau. Die Bandstruktur ist zudem ausschlaggebend für die Transporteigenschaften der Ladungsträger. Sie bestimmt, wie groß die Zustandsdichte eines Bandes an einem bestimmten Punkt der Brillouin-Zone ist und mit welcher Grundbeweglichkeit sich die Elektronen und Löcher durch das Kristallgitter des Siliziums bewegen.

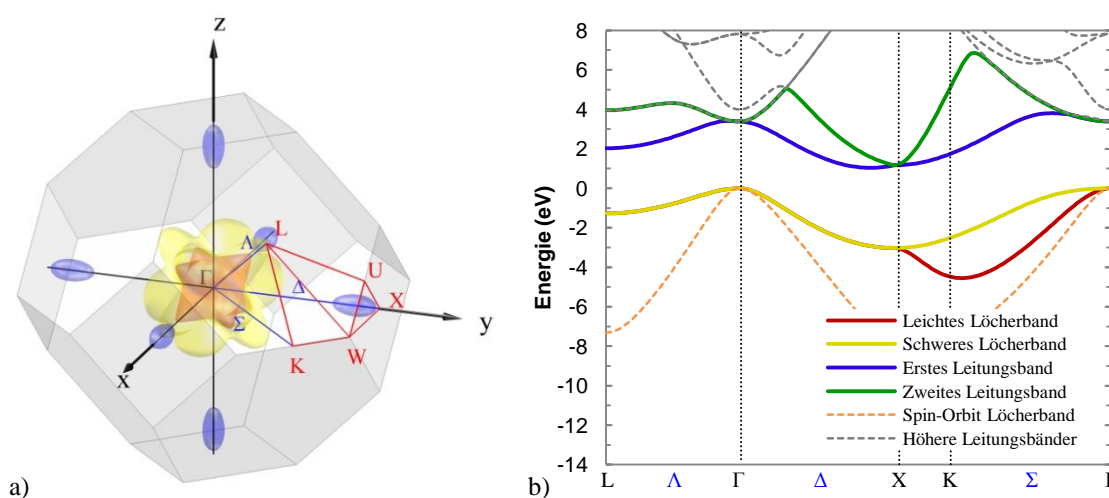


Abbildung 2.3: Siliziumkristall. a) Brillouin-Zone mit Isoenergieflächen des ersten Leitungsbandes (blau) sowie des schweren- (gelb) und leichten Löcherbandes (rot), b) zugehöriger Bandverlauf im Siliziumkristall. Die markierten Punkte und Wege der Brillouinzone entsprechen den Achsenmarkierungen und den Abschnitten des Bandverlaufs.

Mit der Dotierung des Substrats verändert sich die Lage des Fermienergielevels hin zu

$$W_{F,p} = k_B \cdot T_L \cdot \ln\left(\frac{N_D}{n_i}\right) + W_i \quad (2.1)$$

für p-dotiertes und

$$W_{F,n} = -k_B \cdot T_L \cdot \ln\left(\frac{N_A}{n_i}\right) + W_i \quad (2.2)$$

für n-dotiertes Silizium, wobei N_D und N_A die Donator- bzw. Akzeptordichte und n_i die Eigenleitungsdichte bei gegebener Gittertemperatur ($T_L = 300$ K) beschreiben [6]. Die Kanaldotierung und somit die Fermienergie an der Grenzfläche zum Gatekontakt spielen eine entscheidende Rolle bei der Wahl des Gatematerials und seiner Austrittsarbeit. Zum Beispiel liegt bei planaren Transistoren mit Wannens- und Haloimplantationen das

Ferminiveau im Kanalbereich in der Nähe des Valenzbandmaximums bzw. des Leitungsbandminimums, womit für n- und p-MOSFETs Gatematerialien mit unterschiedlicher Austrittsarbeit prozessiert werden müssen. Ein vollständig verarmungsfähiger Transistor hingegen besitzt in der Regel einen nahezu undotierten Kanal, wodurch für n- und p-MOSFETs das gleiche Gatematerial mit einer Austrittsarbeit $W_A = 4.65$ eV im mittleren Bereich der Bandlücke verwendet werden kann [12].

Auch am Verlauf der Bandkanten von Source nach Drain dicht unterhalb des Isolators kann man die gegensätzliche Dotierung der Source- und Draingebiete zum Kanal erkennen. Abbildung 2.4 stellt zwei Zustände der Beschaltung eines planaren MOSFETs anhand des Valenz- und Leitungsbandes dar. Der erste Zustand beschreibt einen gesperrten n-MOSFET mit einer Potenzialbarriere im Bereich des Gatekontakts, welche von den Elektronen nicht überwunden werden kann. Durch das Anlegen der Gate-Source-Spannung lädt sich der Gatekontakt positiv auf und es gelangen auf der anderen Seite des Isolators freibewegliche Elektronen als Gegenladungen in das Kanalgebiet. Die Hauptladungsträger (Majoritäten) sind nun die Elektronen und somit verschiebt sich das Ferminiveau des Kanals in Richtung des Leitungsbandes. Man spricht hierbei von der Inversion des Kanals. Die Potenzialbarriere sinkt ab und die Elektronen können von Source nach Drain fließen [13].

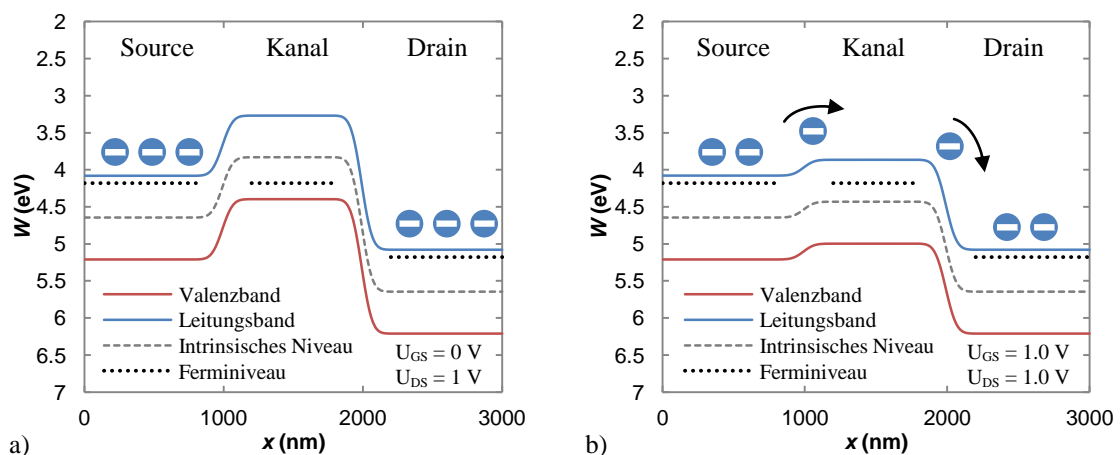


Abbildung 2.4: Energieniveaus von Source, Kanal und Drain eines n-MOSFETs im a) gesperrten Zustand und b) für $U_{GS} = U_{DS} = 1$ V.

Bei Gatelängen, die kürzer als $1 \mu\text{m}$ sind, kann es durch den drainseitigen pn-Übergang zu einer zusätzlichen Absenkung der Potenzialbarriere kommen. Dieser Effekt wird als draininduzierte Barrierenverringern (Drain Induced Barrier Lowering, DIBL) bezeichnet und wird später im Abschnitt zur Skalierung und dem Kurzkanalverhalten planarer Transistoren ausführlicher behandelt [14].

Die Kanaldotierung N_{Ch} bestimmt jedoch nicht nur die Lage des Ferminiveaus und die Höhe der Potenzialbarriere, sie ist auch ausschlaggebend für das Oberflächenpotenzial

$$\varphi_s = U_T \cdot \ln\left(\frac{N_{Ch}}{n_i}\right) \quad (2.3)$$

im Kanal (Temperaturspannung im Silizium U_T) und für die Weite

$$w_d = \sqrt{\frac{4 \cdot \varepsilon_0 \varepsilon_{Si} \cdot \varphi_s}{e \cdot N_{Ch}}} \quad (2.4)$$

der Verarmungszone, die von der Kanaloberfläche in Richtung der Transistortiefe hineinreicht [6]. Ein planarer MOSFET mit einer Dotierungskonzentration von $N_{Ch} = 5 \cdot 10^{18} \text{ cm}^{-3}$ im Kanalgebiet besitzt somit eine relativ kurze Verarmungszone von $w_d \approx 10 \text{ nm}$. Dies ist notwendig, um eine gute Steuerfähigkeit des planaren Gates zu erreichen und die Leckströme in den tieferen Gebieten unterhalb des Kanals möglichst gering zu halten. Die Weite der Verarmungszone spielt ebenfalls für Multi-Gate-Strukturen eine entscheidende Rolle und wird bei deren Betrachtung noch einmal aufgegriffen.

C. Kenngrößen von MOSFETs

Für die Charakterisierung eines Transistors hinsichtlich seiner elektrischen Eigenschaften und seiner Leistungsfähigkeit bedient man sich einer überschaubaren Anzahl an Kenngrößen, die zum Großteil aus der Transfercharakteristik gewonnen werden [6], [15], [16]. In Abbildung 2.5a sind die beiden erforderlichen Transferkennlinien im linearen und logarithmischen Maßstab dargestellt. Sie beschreiben das Verhalten des Drainstromes in Abhängigkeit von der Gate-Source-Spannung und werden unterschieden in die lineare und die Sättigungskennlinie. Der Verlauf unterhalb der Schwellspannung kann näherungsweise mit der Gleichung

$$I_{D,SS} = I_{D,th} \cdot 10^{\left(\frac{U_{GS}-U_{th}}{SS}\right)} \quad (2.5)$$

beschrieben werden, wobei SS (Subthreshold Swing) die Unterschwellsteigung und $I_{D,th}$ den zur Schwellspannung passenden Drainstrom darstellen. Die Bezeichnung „lineare“ Transferkennlinie stammt von dem zugehörigen linearen Bereich der Ausgangskennlinien aus Abbildung 2.5b, wo sich der Transistor näherungsweise wie ein ohmscher Widerstand verhält und der Drainstrom mit

$$I_{D,lin} = \mu \cdot C'_{inv} \cdot \frac{W_{Gate}}{L_{Gate}} \cdot \left(U_{GS} - U_{th} - \frac{U_{DS}}{2}\right) \cdot U_{DS} \quad (2.6)$$

angegeben werden kann [17]. Hierbei ist μ die Ladungsträgerbeweglichkeit und C'_{inv} die flächennormierte Inversionskapazität. In der Regel wird die lineare Transferkennlinie bei einer Drain-Source-Spannung von $U_{DS} = 0.05 \text{ V}$ aufgezeichnet.

Anschließend an den linearen Bereich folgt der Sättigungsbereich (Abschnürbereich), in dem die Ausgangskennlinien abflachen, da die Ladungsträger aufgrund von zunehmen-

den Stoßvorgängen eine mittlere Sättigungsgeschwindigkeit erreichen und der Kanal drainseitig abgeschnürt wird [13]. Für Langkanaltransistoren, bei denen der DIBL-Effekt eine untergeordnete Rolle spielt, lässt sich der Drainstrom in diesem Bereich über die Gleichung

$$I_{D,sat} = \mu \cdot C'_{inv} \cdot \frac{W_{Gate}}{L_{Gate}} \cdot \frac{(U_{GS} - U_{th})^2}{2} \quad (2.7)$$

ausdrücken [16].

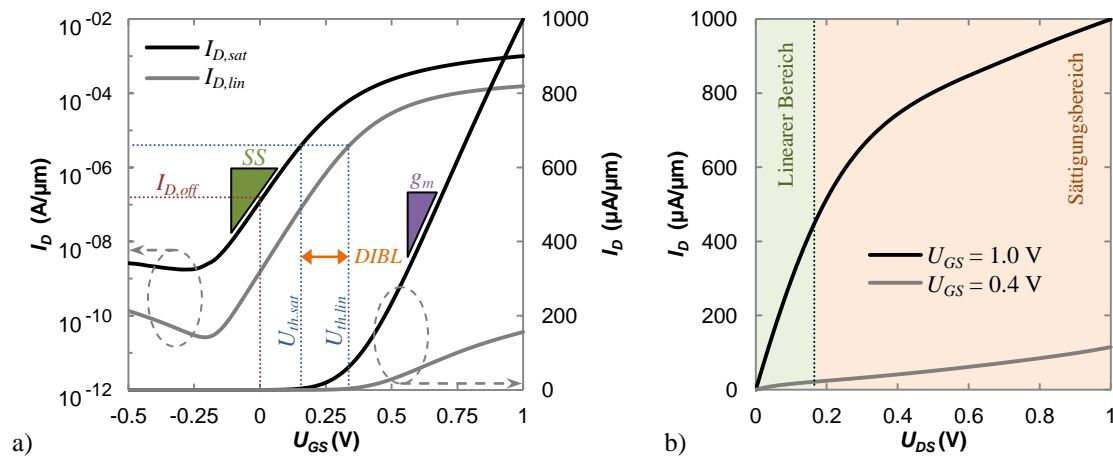


Abbildung 2.5: Kennlinien eines n-MOSFETs der 45nm-SOI-Technologie mit einer Betriebsspannung von $U_{DD} = 1$ V. a) Transferkennlinien in logarithmischer und linearer Form und b) Ausgangskennlinien.

Eine weitere Kenngröße eines MOSFETs ist die Schwellspannung

$$U_{th} = U_{GS}|_{I_D=I_{D,th}} \quad (2.8)$$

Sie bezeichnet die Gate-Source-Spannung, ab der das Kanalgebiet in die schwache Inversion übergeht und der Transistor aus dem sperrenden in den leitenden Zustand wechselt. Zur Bestimmung der Schwellspannungen existieren zwei verschiedene Ansätze, die Tangenten- und die Konstantstrommethode. Die in dieser Arbeit bevorzugte Variante ist die Konstantstrommethode, bei der die Gate-Source-Spannung für einem Drainstrom von

$$I_{D,th} = I_{ref} \cdot \frac{W_{Gate}}{L_{Gate}} \quad (2.9)$$

abgelesen wird. W_{Gate} und L_{Gate} beschreiben dabei die effektive Gateweite und die physikalische Gatelänge. Der Referenzstrom I_{ref} ist jedoch ein festgelegter Wert und hängt von Typ und Generation des Transistors ab [7].

Mithilfe der Differenz der Schwellspannungen lässt sich anschließend der Einfluss des bereits erwähnten *DIBLs* mit

$$DIBL = \frac{U_{th,lin} - U_{th,sat}}{U_{DD}} \left[\frac{\text{mV}}{\text{V}} \right]. \quad (2.10)$$

quantitativ erfassen. Hierbei ist zu beachten, dass der Wert auf die Betriebsspannung normiert wird, welche im Sättigungsbetrieb näherungsweise der Drain-Source-Spannung entspricht [14].

Trägt man die Schwellspannungen der Transistoren über ihre Gatelänge auf (Abbildung 2.6a), kann ein Abfall der Werte hin zu kürzer werdenden Gates beobachtet werden. Dieses Verhalten ist im Allgemeinen als „Roll-off“ bekannt und ein Kriterium zur Bewertung des Kurzkanalverhaltens.

Eine weitere wichtige Kenngröße ist der Sperrstrom $I_{D,off}$. Er beschreibt den Drainstrom im ausgeschalteten Zustand des Transistors ($U_{GS} = 0 \text{ V}$) und bei maximaler Drain-Source-Spannung ($U_{DS} = U_{DD}$).

Der Maximalwert der linearen Transferkennlinie ($U_{GS} = U_{DD}$, $U_{DS} = 0.05 \text{ V}$) wird gleich der Kennlinienbezeichnung als $I_{D,lin}$ geführt. Über ihn kann der Gesamtwiderstand des Transistors zu

$$R_{ges} = R_{Kanal} + R_{SD} = \frac{U_{DS}}{I_{D,lin}} \quad (2.11)$$

ermittelt werden. Ähnlich dem linearen Wert trägt auch der Maximalwert der Sättigungskennlinie ($U_{GS} = U_{DS} = U_{DD}$) die Bezeichnung $I_{D,sat}$.

Alle drei Ströme sind in der Regel auf die Transistorweite normiert und lassen sich als Verhältnis von $I_{D,off}$ zu $I_{D,lin}$ bzw. $I_{D,sat}$ darstellen. Jeder gemessene Transistor aus Abbildung 2.6b ist somit durch zwei einzelne Punkte vertreten.

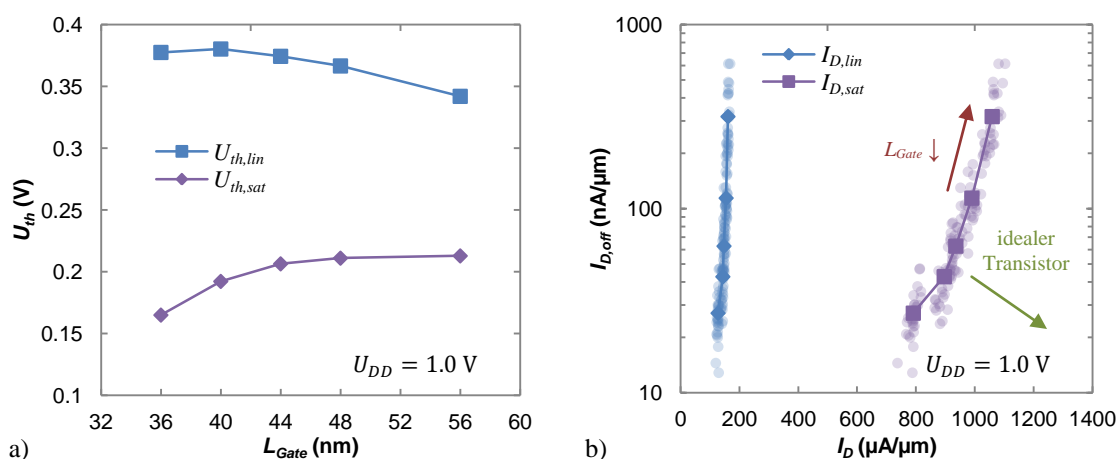


Abbildung 2.6: Kennwerte eines n-MOSFETs der 45nm-SOI-Technologie. a) Schwellspannung in Abhängigkeit von der Gatelänge (Roll-off) und b) Sperrstrom in Abhängigkeit vom Drainstrom (Universalkurve).

Die Streuung und die Gatelängenvariation einer Vielzahl von gemessenen Strukturen bilden daraus die Universalkennlinie, mit der eine Aussage zur statischen Leistungsfähigkeit einer Transistortechnologie getroffen werden kann. Der ideale Transistor besitzt demnach einen möglichst großen Drainstrom bei einem möglichst kleinen Sperrstrom. Um das Schaltverhalten eines MOSFETs zu beurteilen, ist die Betrachtung der Transferkennlinien bezüglich ihres Anstiegs notwendig. Dabei muss unterschieden werden, ob sich der Transistor im Unterschwellbereich ($U_{GS} \leq U_{th}$) oder im eingeschalteten Zustand ($U_{GS} > U_{th}$) befindet. Unterhalb der Schwellspannung bestimmt die Unterschwellsteigung

$$SS = \left(\frac{d \log[I_D(U_{GS})]}{dU_{GS}} \right)^{-1} \quad (2.12)$$

den Anstieg der Kennlinie. Sie wird mit der Einheit mV/dec angegeben und kann näherungsweise mit der analytischen Gleichung

$$SS = \ln(10) \cdot \frac{k_B T_L}{e} \cdot \left(1 + \frac{C_d}{C_{ox}} \right) \quad (2.13)$$

beschrieben werden, wobei C_d die Kapazität der Verarmungszone unterhalb des Gates und C_{ox} die Kapazität des Gateoxids repräsentieren. Somit ist die Steuerfähigkeit des Gates bei Raumtemperatur sowie $C_d/C_{ox} \rightarrow 0$ auf einen idealen Minimalwert von $SS \approx 60$ mV/dec begrenzt. Ein steilerer Anstieg ist nur mithilfe des Band-zu-Band-Tunnelmechanismus möglich, wie er bei Tunnel- oder SCHOTTKY-Transistoren ausgenutzt wird [18], [19], [20]. Im eingeschalteten Zustand bestimmt die Steilheit

$$g_m = \frac{dI_D}{dU_{GS}} \quad (2.14)$$

den weiteren Verlauf der Transferkennlinie. Auch hier ist ein möglichst großer Wert angestrebt, um ein schnelles Schalten des Transistors zu erreichen.

Die stationären elektrischen Eigenschaften allein reichen jedoch nicht aus, um Aussagen über das dynamische Verhalten des Transistors zu treffen, da hierfür die Kapazitäten des MOSFETs (Abbildung 2.7) mit einbezogen werden müssen. Eine genaue Messung bzw. Simulation der einzelnen Anteile ist nur bedingt möglich, sodass sich die Gate-Source-Kapazität (Abbildung 2.8) zur Charakterisierung des kapazitiven Verhaltens eines MOSFETs durchgesetzt hat. Zur Erfassung dieser Kennlinie wird die Gate-Source-Spannung über einen Spannungsbereich, der sowohl den akkumulierten als auch den invertierten Kanalzustand umfasst, mit einer hochfrequenten Kleinsignalspannung überlagert und die zugehörige Kapazität zwischen dem Gatekontakt und den zusammengeschlossenen Source-Drain-Kontakten gemessen.

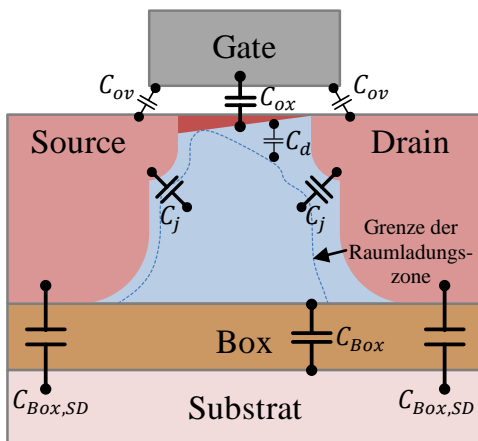


Abbildung 2.7: Kapazitäten an einem SOI-MOSFET unter Vernachlässigung parasitärer Kontaktkapazitäten. Sie unterteilen sich neben der Gateoxid- (C_{ox}) und der Verarmungskapazität (C_d) in die Überlappungskapazität (C_{ov}), die Sperrschichtkapazität (C_j) und bei SOI-Transistoren in einen Kapazitätsanteil zwischen den Source-Drain-Gebieten ($C_{Box,SD}$) bzw. dem Wannengebiet (C_{Box}) und dem Substrat.

Die Werte sind zur besseren Vergleichbarkeit mit anderen Strukturen auf die effektive Gateweite normiert. Aus ihr lassen sich die Werte für die Überlappungskapazität C_{ov} bei $U_{GS} = 0.0\text{ V}$ und die Inversionskapazität C_{inv} bei $U_{GS} = U_{DD}$ bestimmen. Die Überlappungskapazität wird auch als MILLERkapazität C_{mill} bezeichnet und zur Auswertung des Dotierungsprofils verwendet. Die Inversionskapazität setzt sich aus der Gateoxidkapazität C_{ox} und dem kapazitiven Anteil der Verarmungszone unterhalb des Gates C_d zusammen und wird zur Bestimmung der äquivalenten Gateoxiddicke (Equivalent Oxide Thickness, EOT) unter der Annahme der einfachen Geometrie eines Plattenkondensators genutzt.

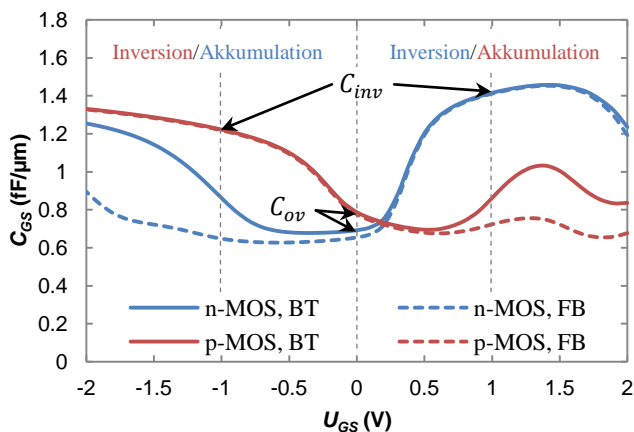


Abbildung 2.8: Gate-Source-Kapazität C_{GS} von n- und p-MOSFETs mit (Body Tight, BT) und ohne (Floating Body, BF) Kontaktierung des Bodygebiets (32 nm-SOI-Technologie).

Im Akkumulationsbereich kommt es bei den Bulk-Transistoren und den SOI-Transistoren mit angeschlossenem Bodygebiet (Body Tight, BT) ebenfalls zu einem Anstieg der Gate-Source-Kapazität, da zu den Ladungen der Dotanden weitere freibewegliche Ladungsträger gleicher Polarität über den Bodykontakt in den Kanal gelangen. Bleibt der Bodyanschluss des SOI-Transistors offen (Floating Body, BF), gilt das Wannengebiet als isoliert und es können keine freibeweglichen Ladungen durch die

gesperrten pn-Übergänge von Source und Drain in den Kanal eindringen, sodass die Gate-Source-Kapazität auf dem Niveau der Überlappungskapazität verbleibt.

Das dynamische Verhalten der MOSFETs und der aus ihnen aufgebauten Schaltungen ergibt sich aus der Interaktion der Drainströme und der Kapazitäten, die von ihnen auf bzw. umgeladen werden müssen, und wird im Wesentlichen mithilfe von Frequenzen oder Schaltzeiten charakterisiert.

Unabhängig von der Schaltung kann für jeden Transistor eine Transitfrequenz über die Kleinsignalparameter bestimmt werden. Hierbei handelt es sich um die Grenzfrequenz, bei der die Kleinsignalstromverstärkung den Wert

$$|h_{21}| = \left| \frac{y_{21}}{y_{11}} \right| = 1 \quad (2.15)$$

annimmt. Das bedeutet, dass der Transistor effektiv die gleiche Menge an Strom treiben kann, wie er selbst zum Umladen seiner eigenen Gatekapazität verbraucht. Näherungsweise kann die Transitfrequenz über

$$f_t = \frac{g_m}{2\pi \cdot C_{ISS}} \quad (2.16)$$

beschrieben werden, wobei die Eingangskapazität (Source-Schaltung) C_{ISS} und die Steilheit g_m vom Arbeitspunkt des Transistors abhängen [6].

Auch die Maximalschwingfrequenz kann über die Kleinsignalparameter bestimmt werden. Sie beschreibt die Frequenz, bei der die allgemeine Verstärkung eines Vierpols ohne Rückkopplung (Maximal Unilateral Gain, MUG)

$$MUG = \frac{|y_{DG} - y_{GD}|^2}{4 \cdot [Re\{y_{GG}\} \cdot Re\{y_{DD}\} - Re\{y_{GD}\} \cdot Re\{y_{DG}\}]} = 1 \quad (2.17)$$

ist und ein idealer Oszillator, bestehend aus derartigen MOSFETs, somit gerade noch in der Lage wäre zu schwingen [21].

Über den einzelnen Transistor hinaus ist der CMOS-Inverter ein wichtiger Anhaltspunkt zur Charakterisierung des dynamischen Verhaltens. In Abbildung 2.9 ist zur Veranschaulichung der einzelnen Schaltzeiten das Knotenpotenzial von Ein- und Ausgang der mittleren Stufe eines simulierten dreistufigen Inverters aus Tri-Gate-Transistoren dargestellt. Die Anstiegszeit t_r gibt die Dauer wieder, die das Ausgangssignal benötigt, um von 10 % auf 90 % der Betriebsspannung zu steigen. Dem gegenüber steht die Abfallzeit t_f für den umgekehrten Schaltvorgang. In diesem Fall ist t_r kürzer als t_f , was bedeutet, dass der p-MOSFET beim Einschalten (von $U_{Ein} = U_{DD}$ auf $U_{Ein} = 0$ am Eingang) durch eine niedrigere Schwellspannung eher Strom treiben kann als der n-MOSFET und die Gates der nächsten Inverterstufe schneller umgeladen werden. Das Gleiche gilt auch für die Verzögerungszeiten Δt_{hi} und Δt_{low} , mit denen der Abstand zwischen Ein- und Ausgang bei jeweils halber Betriebsspannung angegeben wird. Unabhängig vom Überschwingen am Eingang, kann am Ausgang eine Spannungsüber-

höhung als Folge des dynamischen Schaltvorgangs beobachtet werden, wenn die Flanke des Eingangssignals sich zu ändern beginnt. Somit kommt am Ausgang eine minimale und maximale Spannung von U_{min} bzw. U_{max} zustande.

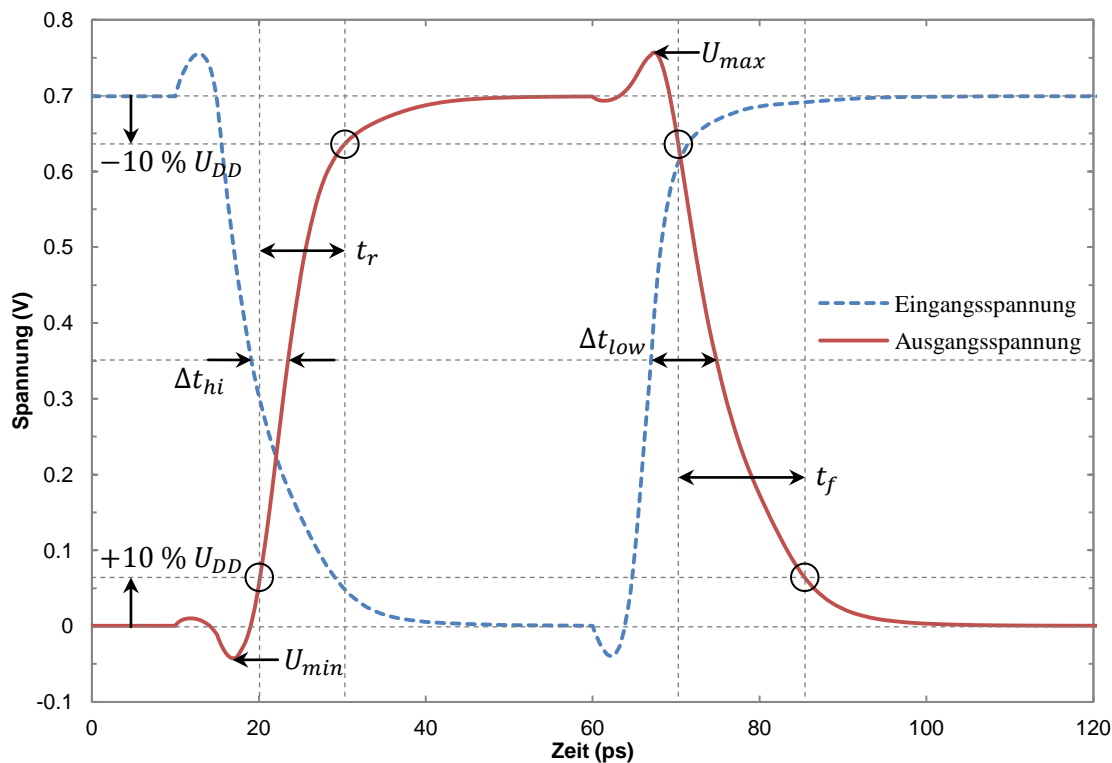


Abbildung 2.9: Schaltverhalten eines simulierten 3-stufigen Inverters (Ein- und Ausgangsspannung der mittleren Inverterstufe), aufgebaut aus Tri-Gate-Transistoren.

Der Ringoszillator (RO) ist die komplexeste Form, die hinsichtlich des dynamischen Schaltverhaltens in dieser Arbeit betrachtet werden soll. Er ist stets aus einer ungeraden Anzahl n von Inverterstufen aufgebaut, die alle in Reihe geschaltet sind und deren Ausgang auf den Eingang rückgekoppelt ist. Seine Eigenschwingfrequenz

$$f_{RO} = \frac{1}{2n \cdot t_D} \quad (2.18)$$

dient unter anderem zur Bestimmung der mittleren Verzögerungszeit t_D einer Inverterstufe. Um eine vorzeitige Rückkopplung zu vermeiden, ist für die Transistoren dieser Arbeit eine Mindestanzahl von fünf Stufen notwendig gewesen, da das Ausgangssignal seine volle Aussteuerung andernfalls nicht erreichte.

2.1.2 Skalierung und Kurzkanalverhalten planarer Transistoren

Seit Beginn der industriellen Fertigung von integrierten Schaltkreisen auf CMOS-Basis ist es das Ziel, die Produktionsverfahren und die Leistungsfähigkeit der Transistoren stetig zu optimieren. Mit der erfolgreichen Weiterentwicklung der Lithografie konnten auch die Transistoren immer kleiner und in höherer Anzahl auf einem Chip gefertigt werden, wodurch der Begriff der Technologieskalierung geprägt wurde. 1975 formulierte GORDON MOORE dazu das später nach seiner Person benannte „Gesetz“, welches besagt, dass sich die Anzahl der Bauelemente und somit die Komplexität von Schaltkreisen aller zwei Jahre verdoppelt [4]. Hierbei handelt es sich jedoch weniger um ein physikalisches Gesetz als um eine empirisch beobachtete Faustregel, die auch heute noch Bestand hat, da die Schaltungsherstellung an eine Vielzahl von Industriezweigen gebunden ist, die sich mit deren Hilfe und über Meilensteine in ihrer Entwicklung koordinieren [22]. Die Verkleinerung der Bauelementeabmessungen wird durch den Skalierungsfaktor κ beschrieben, der aufgrund der Komplexität der Transistoren auch andere Kennwerte und Parameter beeinflusst (Tabelle 2.1). Der Übergang von einer Technologiegeneration zur nächsten ist in der Regel mit einem Skalierungsschritt von $\kappa = 0.7$ verbunden [23], [24].

Tabelle 2.1: Überblick zu den Skalierungsregeln der planaren MOSFETs [24].

Transistorparameter	Formelzeichen	Skalierungsfaktor
Strukturabmessungen	L, W, t_{ox}	κ
Kanaldotierung	N_{ch}	$1/\kappa$
Betriebsspannung	U_{DD}	κ
Drainstrom	I_D	κ
Gatekapazität	C_{GS}	κ
Verzögerung pro Schaltkreis	$U \cdot C \cdot I^{-1}$	κ
Verlustleistung pro Schaltkreis	$U \cdot I$	κ^2
Leistungsdichte	$U \cdot I \cdot A^{-1}$	1

Die Weite W des Transistors bereitet bei dieser Vorgehensweise keine Probleme. Auch der mittlere Abstand der Gatekontakte zweier Transistoren (Pitch) folgt hierbei exakt dem Verlauf der Skalierungsregel, was ebenfalls entscheidend dafür ist, dass die Fläche des Transistors stetig halbiert werden konnte (Abbildung 2.10a). Doch nicht nur der Flächenbedarf profitierte von der Skalierung, auch der Gesamtwiderstand wurde dabei verringert und begünstigt somit die Leistungsfähigkeit des Transistors [25]. Der enorme Drainstromgewinn der letzten Transistorgenerationen gemäß Abbildung 2.10b ist jedoch nicht allein Resultat der Technologieskalierung, sondern konnte vor allem durch den Einfluss von modernen Konzepten zur mechanischen Verspannung erreicht werden [15], [26], [27].

Für einige Parameter der klassischen Skalierung mussten jedoch im Laufe der Entwicklung Kompromisse eingegangen werden, um die Funktionalität des Transistors zu erhalten. So konnte das Gateoxid ab der 90 nm-Technologie nicht weiter reduziert werden, da das Siliziumoxid bei einer Schichtdicke von unter 1.15 nm zu hohe Gateleckströme zugelassen hätte und die Wahrscheinlichkeit für einen Durchbruch aufgrund der gesteigerten elektrischen Feldstärke zu groß geworden wäre.

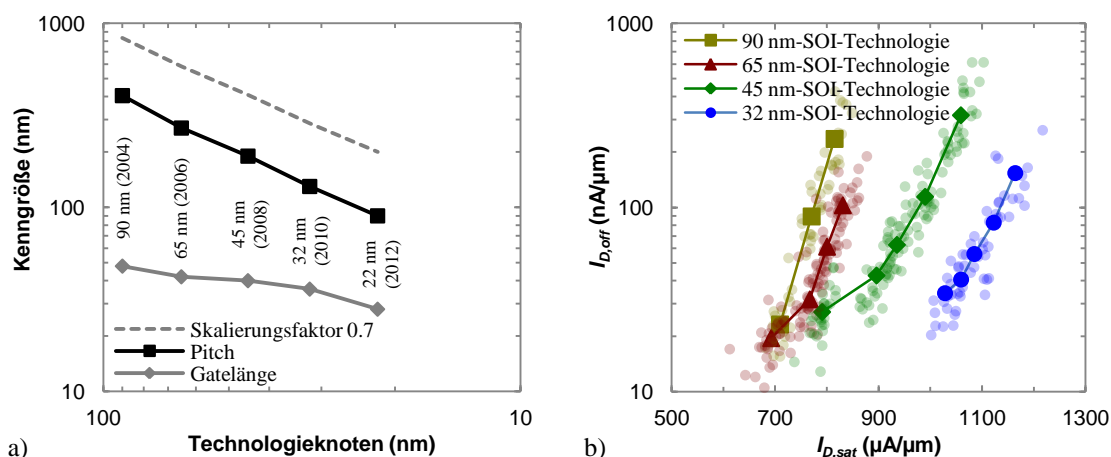


Abbildung 2.10: Skalierung von MOSFETs. a) Pitch und Gatelänge verschiedener Generationen von Transistortechnologien (SOI). Die gestrichelte Linie zeigt zum Vergleich die Steigung bei einem Skalierungsfaktor von 0.7 an. b) Universalkennlinien (Sättigung) der zugehörigen n-MOSFETs.

Ohne der vertikalen Skalierung, konnte auch die physikalische Gatelänge nicht mehr im Verhältnis von $\kappa = 0.7$ reduziert werden, da der Einfluss der Kurzkanaleffekte immer mehr zugenommen und die Schaltfähigkeit der MOSFETs mit sehr kurzem Kanal erheblich beeinträchtigt hätte. Einer dieser Effekte ist der bereits erwähnte *DIBL*, wie er anhand des Bandverlaufs in Abbildung 2.11 dargestellt ist [14]. Die beiden pn-Übergänge von Source und Drain liegen bei Kurzkanaltransistoren so dicht beieinander, dass die Potenzialbarriere des Kanals durch den drainseitigen pn-Übergang absinkt und der Transistor bereits bei einer geringeren Schwellspannung in den leitenden Zustand schaltet. Mit einer Erhöhung der Drain-Source-Spannung wird dieser Effekt zusätzlich verstärkt, da sich die drainseitige Raumladungszone weiter ausdehnt und die effektive Gatelänge somit verkürzt [17].

Um einem steilen Abfall der Schwellspannung mit kürzer werdender Gatelänge vorzubeugen, wurde für planare MOSFETs eine Haloimplantation zur Verstärkung der Gegenladung im Kanalgebiet eingeführt [28]. Diese Implantation erfolgt unter einem Winkel von ca. 30° und gelangt somit teilweise unter das Gate, wo sie die entgegengesetzt dotierten Erweiterungsgebiete von Source und Drain umschließt.

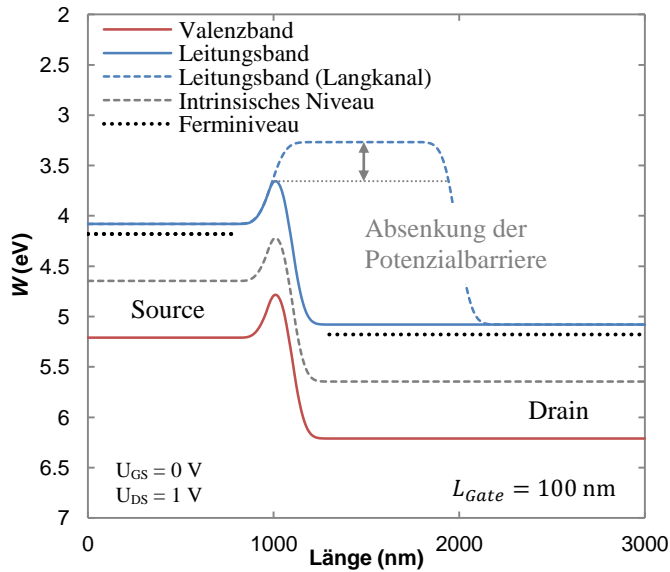


Abbildung 2.11: Absenkung der Potenzialbarriere eines planaren n-MOSFETs bei starker Verringerung der physikalischen Gatelänge (DIBL).

Zum einen werden somit die Potenzialbarriere sowie die Schwellspannung erhöht und zum anderen erhalten die pn-Übergänge von Source und Drain kanalseitig ein steileres Dotierungsprofil, wodurch sich ihre Raumladungszonen auch bei höheren Drain-Source-Spannungen nicht so weit unterhalb des Gates ausdehnen können. Gut sichtbar wird der Einfluss der Haloimplantation in Abbildung 2.12 beim Vergleich der Schwellspannungsverläufe zwischen der 90 nm- und der 32 nm-SOI-Technologie [29]. Während die Schwellspannung der 90 nm-Technologie noch maßgebend durch die Wanneneimplantation bestimmt war, somit über einen großen Gatelängenbereich gehalten wird und anschließend sehr steil abfällt, zeigen die Langkanaltransistoren der 32 nm-Technologie zuerst sehr niedrige Schwellspannungen, die mit kürzer werdendem Gate ansteigen.

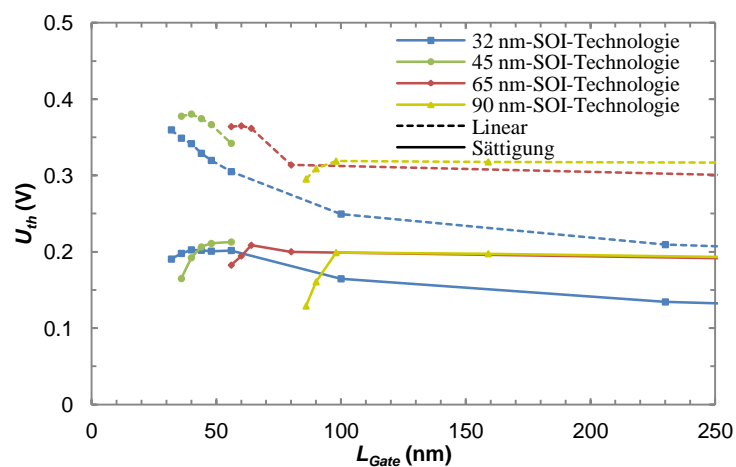


Abbildung 2.12: Abhängigkeit der linearen und der Sättigungsschwellspannung von der physikalischen Gatelänge für n-MOSFETs verschiedener Generationen.

Dieser Anstieg resultiert aus der zunehmenden Überlagerung der beiden Halogegebiete (source- bzw. drainseitig) und wird als „Roll-on“ bezeichnet. Das Maximum der Schwellspannung liegt im Bereich der jeweiligen nominellen Gatelänge. Anschließend nimmt auch hier der Kurzkanaleffekt überhand und die Schwellspannungen fallen wieder ab. Je flacher der Verlauf des Roll-offs ist, umso besser ist das Kurzkanalverhalten der Transistortechnologie.

Weitere Möglichkeiten zur Bewertung des Kurzkanalverhaltens von Transistoren finden sich in den Darstellungen der Unterschwellsteigung SS und des $DIBLs$ in Abhängigkeit von der physikalischen Gatelänge (Abbildung 2.13). Beide Werte nehmen mit kürzer werdender Gatelänge zu und verdeutlichen damit, dass die Steuerfähigkeit des Gates zunehmend verloren geht. Die Unterschwellsteigung gibt dabei Aufschluss über den Einfluss der Kurzkanaleffekte im Bereich der Verarmung bis hin zur schwachen Inversion, in dem der Drainstrom exponentiell von der Gate-Source-Spannung abhängt. Der $DIBL$ hingegen bewertet das Kurzkanalverhalten ab dem Einsetzen der Inversion und verdeutlicht, wie die lineare und die Sättigungsschwellspannung mit kürzer werdender Gatelänge auseinander driften.

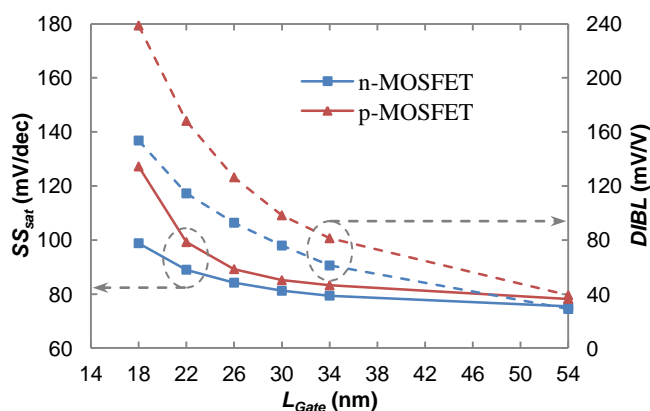


Abbildung 2.13: Unterschwellsteigung SS_{sat} (Sättigung) und $DIBL$ in Abhängigkeit von der physikalischen Gatelänge der simulierten n- und p-MOSFETs einer 22 nm-Technologie.

Ebenso zeigen sich die Kurzkanaleffekte auch in den Ausgangskennlinien der Transistoren (Abbildung 2.14a). Während der Drainstrom eines Langkanaltransistors im Sättigungsbereich sehr flach verläuft, steigt die Kennlinie eines Kurzkanaltransistors mit zunehmender Drain-Source-Spannung näherungsweise weiter linear an. Dieses Verhalten resultiert aus der Verkürzung der effektiven Gatelänge L_{eff} durch die drainseitige Abschnürung des Kanals ab dem Übergang in die Sättigung ($U_{DS} > U_{D,sat}$, Abbildung 2.14b). Mit der verkürzten effektiven Gatelänge sinkt auch die Schwellspannung, wodurch sich der Drainstrom erhöht, da der steuerbare Teil des Kanals stärker in die Inversion getrieben wird. Je steiler der Anstieg des Drainstroms im Sättigungsbereich ausfällt, umso stärker ist die Wirkung der Kurzkanaleffekte auf den Transistor.

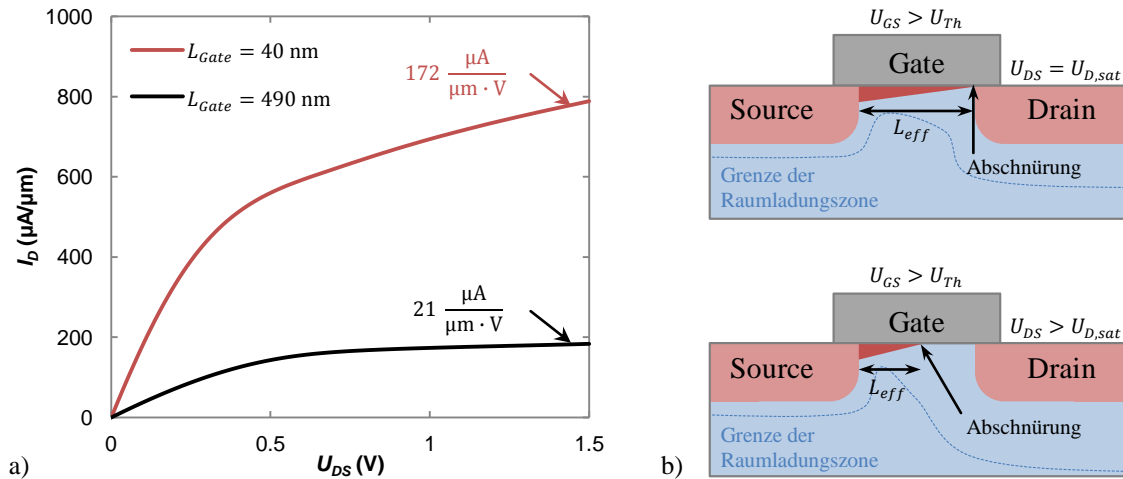


Abbildung 2.14: a) Ausgangskennlinie eines Kurz- und eines Langkanaltransistors in 45 nm-SOI-Technologie. b) Verkürzung der effektiven Gatelänge L_{eff} durch Ausdehnung der Raumladungszone und die drainseitige Abschnürung des Kanals.

Hinzu kommt, dass bei sehr kurzen longitudinalen Abmessungen die Ladungsträger beim Transport durch das Bauelement immer seltener durch elastische Stöße an Gitterdefekten und durch unelastische Stöße mit dem Gitter abgebremst werden. Unterschreitet die Kanallänge die mittlere freie Weglänge setzt der ballistische Transport ein, bei dem die Ladungsträger kontinuierlich beschleunigt werden, bis sie das Draingebiet erreichen [30], [31]. Die mittlere freie Weglänge des Siliziums ist dabei von mehreren Faktoren wie der Gittertemperatur, den Dotanden auf den Gitterplätzen, den Zwischengitteratomen und den ein- bzw. zweidimensionalen Gitterfehlern abhängig. Sie kann somit zwischen 10-100 nm betragen, weshalb sich auch kein eindeutiger Grenzwert für den Übergang zum ballistischen Transport definieren lässt.

Seit die Gatelänge von planaren Bulk-Transistoren Ende der 1980er Jahre die 1 μm -Marke unterschritten hat und erste Beobachtungen zu Kurzkanaleffekten (Short Channel Effect, SCE) veröffentlicht wurden, existieren auch Modelle zur quantitativen Beschreibung dieses Verhaltens [32]. Eine Herleitung aus [33] berechnet das Kurzkanalverhalten eines speziellen Bulk-MOSFETs mithilfe der beiden Gleichungen:

$$SCE = 0.64 \frac{\varepsilon_{Si}}{\varepsilon_{ox}} \left[1 + \frac{x_j^2}{L_{eff}^2} \right] \frac{t_{ox}}{L_{eff}} \frac{w_d}{L_{eff}} \varphi_{D,i} = 0.64 \frac{\varepsilon_{Si}}{\varepsilon_{ox}} EI \cdot \varphi_{D,i} \quad (2.19)$$

und

$$DIBL = 0.80 \frac{\varepsilon_{Si}}{\varepsilon_{ox}} \left[1 + \frac{x_j^2}{L_{eff}^2} \right] \frac{t_{ox}}{L_{eff}} \frac{w_d}{L_{eff}} U_{DS} = 0.80 \frac{\varepsilon_{Si}}{\varepsilon_{ox}} EI \cdot U_{DS} \quad (2.20)$$

Hierbei sind ε_{Si} und ε_{ox} die relativen Permittivitäten von Silizium und Siliziumoxid, x_j die Tiefe der Source-Drain-Gebiete, L_{eff} die effektive Gatelänge, t_{ox} die Dicke des

Gateoxids, $\varphi_{D,i}$ das intrinsische Potenzial des Draingebiets und w_d die Weite der Verarmungszone des Gates gemäß Gleichung (2.4). Die strukturellen Abhängigkeiten lassen sich zu dem Faktor EI zusammenfassen, der in [34] als elektrostatische Integrität bezeichnet wird (Abbildung 2.15). Er umschreibt die Vollständigkeit der Gatesteuerung auf das Kanalgebiet bzw. dessen Unversehrtheit durch die Einwirkung der drainseitigen Raumladungszone.

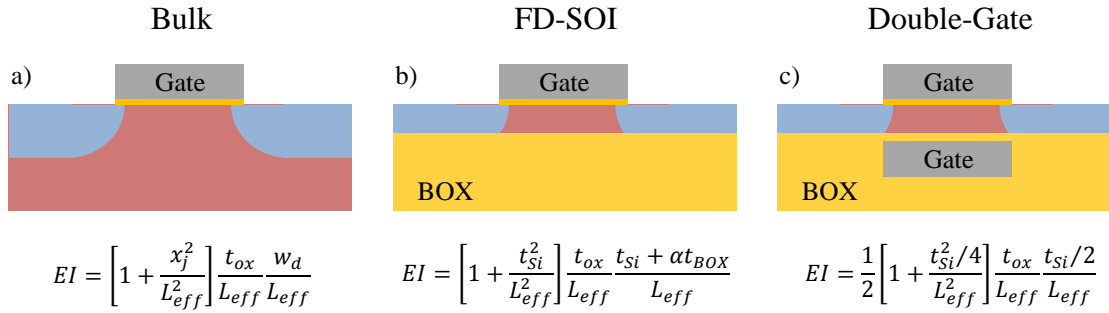


Abbildung 2.15: Elektrostatische Integrität EI von a) Bulk-, b) vollständig verarmbaren SOI- und c) Doppel-Gate-MOSFETs [34].

Dies ermöglicht eine Abschätzung der Schwellspannung

$$U_{th} = U_{th,lang} - SCE - DIBL \quad (2.21)$$

in Abhängigkeit von einzelnen Skalierungsparametern. Somit ist nachvollziehbar, dass die elektrostatische Integrität kleiner und die Kurzkanaleffekte geringer werden, wenn die Dicke des Gateoxids oder die Eindringtiefe der Source-Drain-Gebiete verringert wird. Ebenso erhält man eine bessere Steuerfähigkeit des Transistors, wenn die Kanaldotierung erhöht und somit die Weite der Verarmungszone verkleinert wird.

Diese Beziehungen gelten jedoch nur für vereinfachte Geometrien, bei denen auch eine konstante Kanaldotierung vorausgesetzt werden kann. Für die derzeitigen Kurzkanaltransistoren können sie nicht mehr angewandt werden. Um die Analyse des Kurzkanalverhaltens auf alle Multi-Gate-Strukturen auszudehnen, ist eine andere Vorgehensweise notwendig, wie sie in [34] für mehrere Beispiele hergeleitet wurde. Sie verfolgt die Lösung der Poisson-Gleichung

$$-\Delta\varphi = \frac{qN_a}{\epsilon_{Si}} \quad (2.22)$$

im Kanalbereich des Transistors [35]. Es wird davon ausgegangen, dass für die ortsfesten Ladungen die Summe der Komponenten des elektrischen Feldes in einem Punkt stets konstant ist. Somit reduzieren sich der Einfluss der lateralen Feldstärke der Drain-Source-Spannung sowie die Kurzkanaleffekte, wenn die Feldstärke bzw. die Wirkung des Gates verstärkt wird [36].

Durch Einsetzen der mit der Geometrie gegebenen Randbedingungen gelangt man für jede Struktur zu einer Doppellösung der Differenzialgleichung, die der Form

$$\varphi(x) = \varphi_{D,i} \cdot e^{\pm \frac{x}{\lambda}} \quad (2.23)$$

entspricht und den qualitativen Verlauf des Oberflächenpotenzials im Kanal beschreibt. $\varphi_{D,i}$ steht für das Bezugspotenzial auf der Drainseite, welches an der Grenze zwischen Kanal und Kontaktgebiet existiert. Für die Darstellung des Potenzialverlaufs von $\varphi(x)$ ist jedoch nur das negative Vorzeichen relevant, da die Einwirkung des drainseitigen Kontaktpotenzials auf das Kanalgebiet mit einem exponentiellen Abfall verbunden ist (Abbildung 2.16). Der Parameter λ wird als natürliche Länge bezeichnet und beschreibt den Punkt, an dem sich das Kontaktpotenzial bei seiner Ausdehnung in Kanalrichtung auf $1/e = 36\%$ reduziert hat.

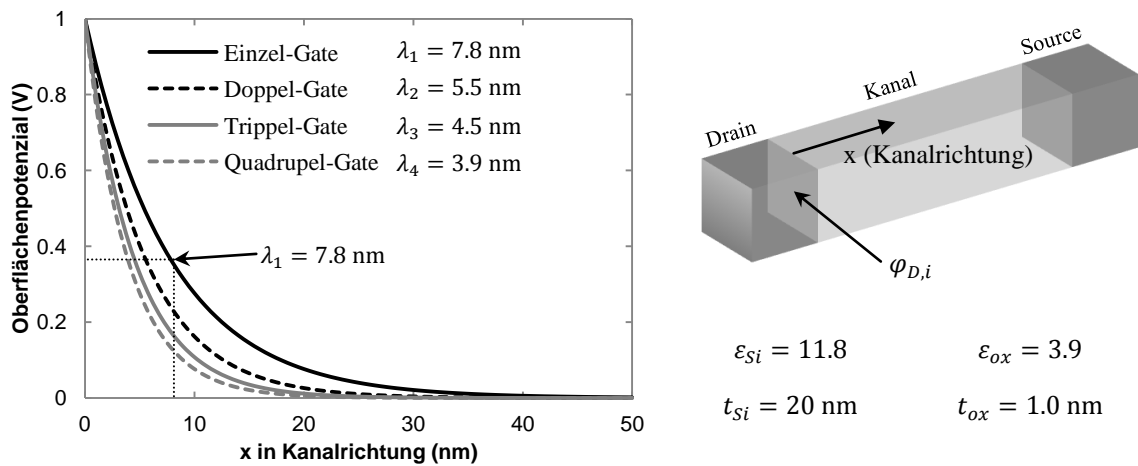
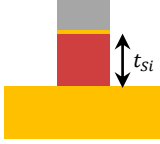
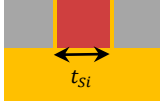
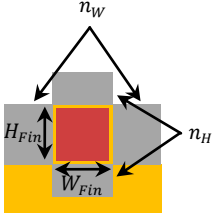


Abbildung 2.16: Berechnete Lösungsfunktion $\varphi(x)$ für den Potenzialverlauf verschiedener Transistorgeometrien entlang der Oberfläche in Kanalrichtung. Ausgangspunkt ist das Draingebiet mit dem intrinsischen Potenzial $\varphi_{D,i} = 1$ V ($t_{Si} = 20$ nm, $t_{ox} = 1.0$ nm).

Zur Vermeidung von Kurzkanaleffekten wird angegeben, dass die effektive Gatelänge 5 ... 10-mal größer als die natürliche Länge sein sollte, wodurch der Wert in der Mitte des Kanals auf unter 10 % des Bezugspotenzials $\varphi_{D,i}$ abgefallen ist und für die andere Hälfte des Kanals die komplette Steuerfähigkeit des Gates zur Verfügung steht [34], [37]. Je nach Geometrie des Kanals und Anzahl der Gateflächen ergeben sich unterschiedliche Werte für die natürliche Länge (Tabelle 2.2). Dabei muss berücksichtigt werden, dass die gegebenen Gleichungen aus [34] nur für Einzel- bzw. Doppel-Gate-Strukturen, bei denen die Kanaldicke t_{Si} senkrecht zu den wirkenden Gates verläuft, oder Multi-Gate-Strukturen gelten, für die $t_{Si, MG} = H_{Fin} = W_{Fin}$ gegeben ist und die effektive Anzahl der Gateflächen n_{MG} möglichst einen ganzzahligen Wert annimmt.

Tabelle 2.2: Natürliche Länge λ von verschiedenen Gatestrukturen [34].

Anzahl der Gates	Darstellung	Natürliche Länge
Einzel-Gate		$\lambda_1 = \sqrt{\frac{\epsilon_{Si}}{\epsilon_{ox}} t_{Si} t_{ox}}$ (2.24)
Doppel-Gate		$\lambda_2 = \sqrt{\frac{\epsilon_{Si}}{2\epsilon_{ox}} t_{Si} t_{ox}}$ (2.25)
Allgemein für Multi-Gates		$\lambda_{MG} \cong \sqrt{\frac{t_{Si, MG} \cdot \epsilon_{Si} \cdot t_{ox}}{n_{MG} \epsilon_{ox}}}$ (2.26)

Für die Berechnung von Tri-Gate-Strukturen, deren Höhe und Weite sich voneinander unterscheiden, gestaltet sich die Lösung der Poisson-Gleichung erheblich komplexer. In [38] wird hierfür die Gleichung:

$$\left(\frac{1}{\lambda_{Tri}}\right)^2 = \frac{\pi^2}{W_{Fin} \left(W_{Fin} + \frac{4\epsilon_{Si}}{\epsilon_{ox}} t_{ox}\right)} + \frac{\pi^2}{H_{Fin} \left(H_{Fin} + \frac{2\epsilon_{Si}}{\epsilon_{ox}} t_{ox}\right)} \quad (2.27)$$

angegeben, welche sich für die Simulationen der Tri-Gate-Transistoren als geeignet erwies.

Mit der natürlichen Länge und der Gatelänge lässt sich nach [39] anschließend der Skalierungsparameter

$$\alpha_n = \frac{L_{Gate}}{2\lambda_n} \quad (2.28)$$

berechnen. Er dient dem Vergleich von verschiedenen Multi-Gate-Strukturen bezüglich ihrer Kurzkanaleffekte und der Steuerfähigkeiten ihrer Gates. Simulationen von idealen Transistoren ergaben dabei Variationen der linearen Unterschwellsteigungen in Abhängigkeit von α_n , die sich unabhängig von der Anzahl der effektiven Gateflächen auf einer Linie bewegen (Abbildung 2.17). Somit kann später mit der Abweichung von diesen Berechnungen eine Aussage getroffen werden, welche der simulierten Transistorstrukturen die beste Steuerung des jeweiligen Kanalgebiets ermöglicht. Die Kurve weist zudem ab $\alpha_n \leq 2.5$ einen Knick auf, bei dem die Gatelänge den fünffachen Wert der natürlichen Länge unterschreitet und die Kurzkanaleffekte deutlich zunehmen.

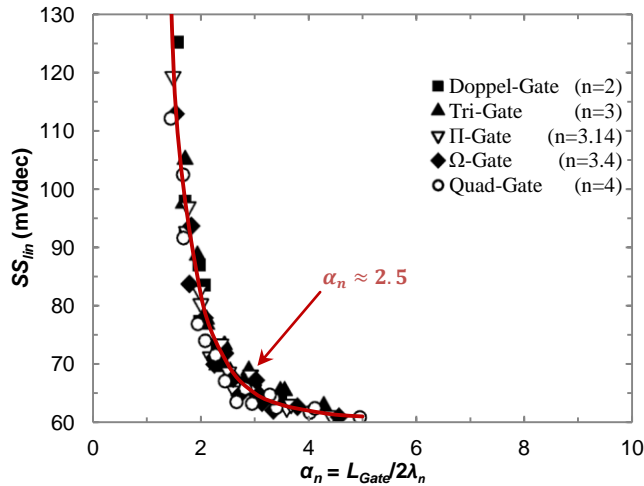


Abbildung 2.17: Lineare Unterschwellsteigung in Abhängigkeit von dem Skalierungsparameter α_n für verschiedene Gatestrukturen. Der ideale Transistor bewegt sich unabhängig von seiner Gatestruktur entlang der roten Linie [39].

2.1.3 Mechanische Verspannung von Silizium

Seit mehreren Transistorgenerationen nutzt man neben der Skalierung auch die Möglichkeit der mechanischen Verspannung des Kanalgebiets zur Leistungssteigerung der Transistoren [15], [27], [40]. Da das Silizium bis zum Bruch ein elastisches Verhalten aufweist, ist die Gitterdeformation ε über das HOOKEsche Gesetz linear mit der Verspannung σ verknüpft [41], [42]. Dadurch kommt es unter dem Einfluss von mechanischer Verspannung zur Verschiebung bzw. Verbiegung der Valenz- und Leitungsbänder (Abbildung 2.3). Die mechanische Verspannung kann als Zug- (positives Vorzeichen) oder Druckverspannung (negatives Vorzeichen) auftreten, wobei zwischen einer uniaxialen und einer biaxialen Wirkungsweise unterschieden werden muss [43]. Das Ziel ist es, das Siliziumgitter und somit seine Bandstruktur im Kanalgebiet dahingehend zu manipulieren, dass die durchschnittliche Beweglichkeit der Ladungsträger erhöht und die elektrische Leitfähigkeit

$$\sigma_{el} = e(n_n \mu_n + n_p \mu_p) \quad (2.29)$$

des Transistors verbessert wird. Dabei muss jedoch klar zwischen n- und p-MOSFET differenziert werden, da Elektronen und Löcher verschieden auf die Änderungen im Siliziumkristall reagieren.

Grundlegend gilt jedoch für beide Arten von Ladungsträgern, dass ihre Beweglichkeit

$$\mu = \frac{e \cdot \tau_s}{m^*} \quad (2.30)$$

indirekt proportional zur entsprechenden effektiven Masse m^* ist [44]. τ repräsentiert die mittlere Stoßzeit und kann nicht direkt durch mechanische Verspannung beeinflusst werden. Eine gute Ladungsträgerbeweglichkeit setzt somit auch immer eine geringe effektive Transportmasse voraus.

Für die Elektronen resultiert die effektive Transportmasse aus den einzelnen Anteilen der sechs LeitungsbandtÄler, wobei die LadungstrÄgerverteilung unter ihnen berüchsich-

tigt werden muss. Im unverspannten Fall besitzen alle sechs Täler (Orbitale) das gleiche energetische Niveau, womit die Elektronen auch gleichmäßig aufgeteilt sind. Man spricht diesbezüglich von einer sechsfachen Entartung (Δ_6) des Leitungsbandes [16]. Jedes Tal bietet den Elektronen dabei eine longitudinale Masse entlang der Achsenorientierung mit $m_l^* = 0.91 \cdot m_0$ und zwei transversalen Massen mit $m_t^* = 0.20 \cdot m_0$, die alle drei orthogonal zueinander sind. Die effektive Masse

$$m^* = \hbar^2 \left(\frac{dE(k)^2}{dk^2} \right)^{-1} \tag{2.31}$$

richtet sich indirekt nach der Krümmung des jeweiligen Bandes [41]. Eine starke Krümmung bedingt somit eine geringe effektive Masse entlang der Richtung des Vektors \mathbf{k} . Im unverspannten Fall mittelt sich die effektive Masse stets zu $m^* = 0.27 \cdot m_0$ und ist somit isotrop. Mit einer mechanischen Verspannung kommt es dagegen zu einer energetischen Anhebung bzw. Absenkung der sechs Leitungsbandestäler. Im Beispiel aus Abbildung 2.18 (biaxial zugverspannt in der xy-Ebene) senkt sich die Leitungsbandenergie der beiden Orbitale außerhalb der Verspannungsebene (out-of-plane-Täler, Δ_2) um 43 meV, wohingegen die vier Orbitale innerhalb (in-plane-Täler, Δ_4) auf ihrem Niveau verbleiben.

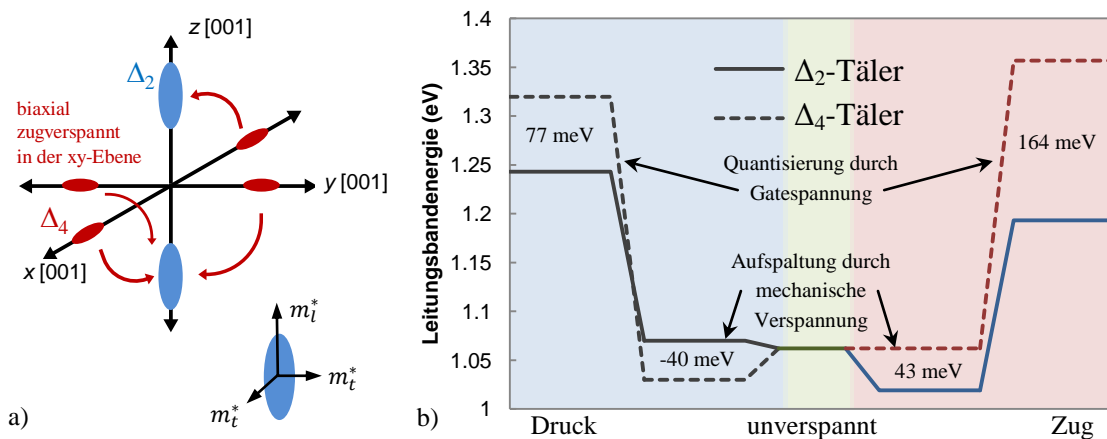


Abbildung 2.18: Einfluss der mechanischen Verspannung auf die Bandstruktur im Silizium. a) Darstellung der sechs Leitungsbandtäler der Brillouin-Zone bei biaxialer Zugverspannung in der xy-Ebene, b) Leitungsbandenergie der Δ_2 - und Δ_4 -Täler unter biaxialem Druck (blau unterlegt) bzw. Zug (rot unterlegt) sowie dem Einfluss einer Quantisierungsfeldstärke von 1 MVcm^{-1} in z-Richtung.

Da die Elektronen stets bestrebt sind, den energetisch niedrigsten Zustand einzunehmen, verteilt sich die Mehrzahl der Elektronen folglich in Richtung der Δ_2 -Tälern, die im Bezug zur xy-Ebene eine starke Krümmung und somit eine geringe effektive Masse für die Elektronen bereitstellen.

Neben der Aufspaltung der Bänder durch die mechanische Verspannung folgt ferner die Quantisierung durch die elektrische Feldstärke des Gates [45]. Diese bewirkt ebenso eine Verschiebung der Leitungsbandminima um

$$\Delta W_{Quant} = r \cdot \left(\frac{\hbar^2}{2m_q^*} \right)^{1/3} \cdot (e \cdot |E_q|)^{2/3}, \quad (2.32)$$

die sich additiv oder subtraktiv zur verspannungsabhängigen Aufspaltung überlagern kann (Abbildung 2.18). Hierbei ist m_q^* die effektive Quantisierungsmasse, \hbar das reduzierte PLANCKsche Wirkungsquantum und r die erste Wurzel der Airy-Funktion mit einem Wert von $r = 2.338$ [27].

Im unverspannten Zustand besitzen die Valenzbänder im Zentrum der Brillouin-Zone (Γ -Punkt) alle das gleich energetische Niveau, jedoch eine unterschiedliche Krümmung. Da die Bandkrümmung gemäß Gleichung (2.31) indirekt mit der effektiven Masse verbunden ist, spricht man hierbei vom leichten (light hole, lh), schweren (heavy hole, hh) und dem Spin-Orbit- (so) Löcherband. Aufgrund der unterschiedlichen Zustandsdichten der Bänder sammeln sich ca. 65 % im schweren, 30 % im leichten und die übrigen 5 % im Spin-Orbit-Löcherband [46]. Die effektiven Massen des leichten und des schweren Bandes sind jedoch stark anisotrop und ergeben sich zu:

$$\begin{array}{lll} \langle 100 \rangle\text{-Richtung:} & m_{lh}^* = 0.20 \cdot m_0 & m_{hh}^* = 0.27 \cdot m_0 \\ \langle 110 \rangle\text{-Richtung:} & m_{lh}^* = 0.15 \cdot m_0 & m_{hh}^* = 0.55 \cdot m_0 \end{array}$$

Mit der Einbringung einer mechanischen Verspannung kommt es auch unter den Valenzbändern zu einer Anhebung bzw. Absenkung. Analog zu den Elektronen streben die Löcher nach dem energetisch höchsten Zustand, womit sich ihre Verteilung unter den Subbändern ändert. Neben der Verschiebung tritt auch eine starke Verbiegung der Bänder im Bereich der Bandmaxima auf (Abbildung 2.19), wodurch sich auch die effektive Masse entlang des Vektors \mathbf{k} verändert.

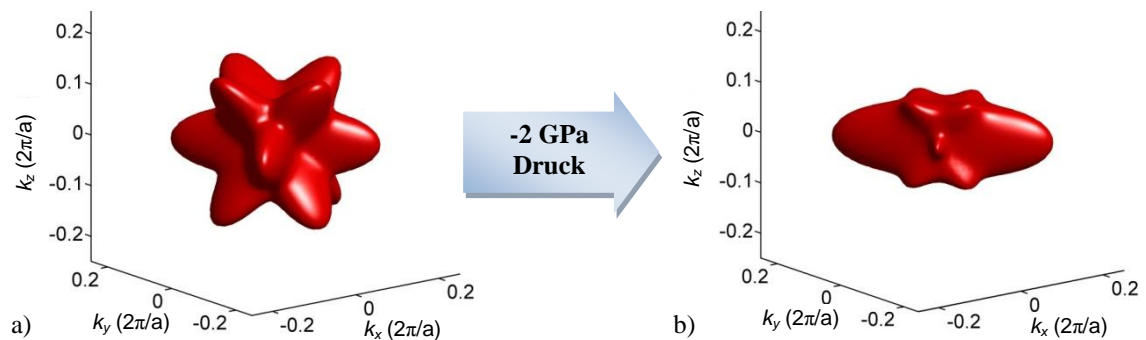


Abbildung 2.19: Darstellung des schweren Löcherbandes durch eine Isoenergiefläche mit $W_{v,iso} = 5.31$ eV für a) unverspanntes Silizium und b) druckverspanntes Silizium (Kanalbereich), wie es durch das Prozessieren von eingebettetem Silizium-Germanium (Source-Drain-Gebieten) mit 35 % Germaniumanteil entsteht.

2.1.4 Techniken zur mechanischen Verspannung

A. Verspannte Deckschichten

Die mechanische Verspannung der Transistoren kann global für große Waferbereiche oder lokal für einzelne Transistoren auftreten. Die verspannten Deckschichten (Overlayer) gehören zu den lokalen Verspannungstechniken, da sie direkt über dem Transistor abgeschieden werden und im Wesentlichen nur auf ihn einwirken (Abbildung 2.20) [47], [48]. Sie bestehen in der Regel aus Siliziumnitrid und dienen in erster Linie als Ätzstoppschicht der Source-Drain-Kontakte, können jedoch auch mit einer intrinsischen mechanischen Verspannung gefertigt werden [49], [50]. Da der n-MOSFET nur auf eine zugverspannte und der p-MOSFET nur auf eine druckverspannte Deckschicht positiv mit einer Leistungssteigerung reagieren, müssen für beide Transistortypen unterschiedliche Schichten prozessiert werden [51]. Dieses Verfahren wird als Dual Stress Liner (DSL) bezeichnet und seit der 130 nm-Technologie industriell genutzt. Die Verspannung koppelt dabei hauptsächlich über das Gate und die Source-Drain-Gebiete ein, wodurch eine uniaxiale Verspannung als Hauptkomponente lateral zur Transportrichtung in den Kanal eingepägt wird [52], [53]. Die intrinsische Zugverspannung der n-MOSFET-Deckschicht (Tensile Overlayer, TOL) kann dabei bis zu $\sigma_{TOL} = 1.7$ GPa betragen, die druckverspannte Deckschicht (Compressive Overlayer, COL) des p-MOSFETs hingegen sogar bis zu $\sigma_{COL} = -3.5$ GPa [54].

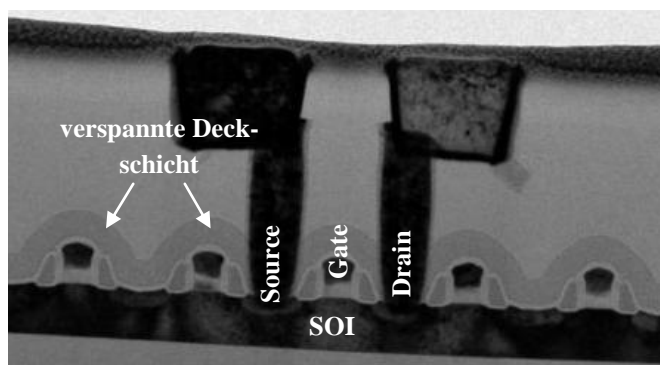


Abbildung 2.20: Transmissionselektronenmikroskopie (TEM) von MOSFETs der 45 nm-SOI-Technologie mit verspanntem Overlayer.

B. Verspannungsspeichernde Prozesse

Verspannungsspeichernde Prozesse (Stress Memorization Techniques, SMT) sind eine weitere Möglichkeit zur lokalen mechanischen Verspannung von n-MOSFETs [55]. Hierbei nutzt man die Rekristallisation von amorphen Source-Drain-Gebieten sowie temperaturabhängige Materialveränderungen während eines Ausheilungsschrittes mit einer Temperatur von über 600 °C, um im Transistor eine Verspannung einzuprägen [56]. Als Quelle der mechanischen Verspannung dienen dabei verspanntes Spacer-Material oder temporäre Deckschichten ähnlich dem TOL, welche jedoch meistens in

nachfolgenden Prozessschritten wieder entfernt werden. Der p-MOSFET kann nicht von den verspannungsspeichernden Prozessen profitieren, da das so erzeugte Spannungsfeld eine Degradation der Löcherbeweglichkeit verursacht.

C. Nutzung vorgespannter Substrate

Das verspannte Substrat ist eine Technik, die bereits seit den Anfängen der mechanischen Verspannung von Silizium bekannt ist. Hierbei bedient man sich der Tatsache, dass Silizium mit $a_{Si} = 0.357$ nm und Germanium mit $a_{Ge} = 0.357$ nm unterschiedliche Gitterkonstanten besitzen [16]. Für einen Mischkristall ergibt sich je nach Germaniumanteil eine mittlere Gitterkonstante von

$$a_{SiGe} = a_{Si}(1 - x_{Ge}) + a_{Ge} \cdot x_{Ge} = x_{Ge} \cdot 0.023 \text{ nm} + 0.357 \text{ nm} . \quad (2.33)$$

Das Substrat besteht zunächst aus einer Basisschicht mit reinem Silizium, auf der eine Silizium-Germanium-Schicht mit einer hinreichenden Dicke zur vollständigen Relaxation aufgebracht ist (Abbildung 2.21). Durch eine weitere Siliziumabscheidung kommt es zu einem pseudomorphen Aufwachsen, wobei der Siliziumfilm die Gitterkonstante des darunterliegenden Silizium-Germaniums annimmt und eine biaxiale Zugverspannung mit ca. 750 MPa pro 10 % Germaniumanteil ausbildet [15]. Da sich dieser Aufbau über den gesamten Wafer erstreckt, wirkt diese Verspannungstechnik global auf alle Transistoren, wobei in erster Linie nur der n-MOSFET eine Leistungssteigerung erfährt. Eine Weiterentwicklung kombiniert das verspannte Substrat mit der SOI-Technik (strained Silicon On Insulator, sSOI) [57]. Die Substrate sind jedoch in der Herstellung sehr aufwendig und kostenintensiv und konnten sich in der industriellen Fertigung somit nicht etablieren.

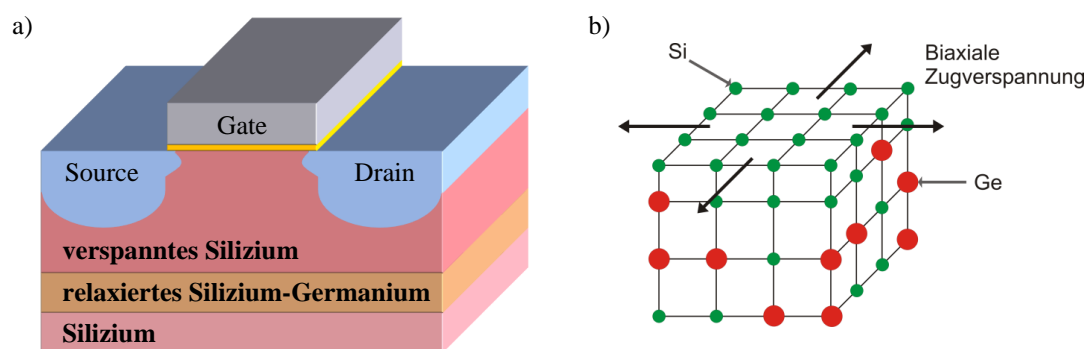


Abbildung 2.21: a) Darstellung eines Transistors mit einem Substratstapel aus Silizium, relaxiertem Silizium-Germanium und dem darauf befindlichen verspannten Silizium. Die Schichtdicke des Silizium-Germaniums benötigt zur Relaxation jedoch deutlich mehr Platz im Vergleich zur schematischen Darstellung. b) Modell einer Grenzfläche zwischen relaxiertem Silizium-Germanium und zugverspanntem Silizium.

D. Eingebettetes Silizium-Germanium

Das eingebettete Silizium-Germanium (eSiGe) ist mit dem verspannten Substrat gut vergleichbar. Hier werden die Source-Drain-Gebiete in einer bestimmten Form herausgeätzt, um sie anschließend wieder epitaktisch mit Silizium-Germanium aufzufüllen. Durch die größere Gitterkonstante des eSiGe kommt es zu einer Deformation des lokalen Kristallgitters von bis zu 1.2 % (Abbildung 2.22) und somit zu einer starken Druckverspannung in den Source-Drain-Gebieten, die sich bis in das Kanalgebiet ausbreitet [58], [59]. Diese Form der Verspannungstechnik eignet sich jedoch nur für den p-MOSFET, der mit einer Leistungssteigerung von bis zu 50 % reagiert [61]. Um einen vergleichbaren Effekt beim n-MOSFET hervorzurufen, ist eine Siliziumverbindung erforderlich, die eine kleinere Gitterkonstante besitzt. Silizium-Kohlenstoff (SiC) erfüllt zwar diese Voraussetzung, weist jedoch einige Schwierigkeiten bei der Integration auf, da es z. B. bei der Ausheilung des Gitters und der Dotandenaktivierung wieder zu einer Verspannungsrelaxation kommen kann [62].

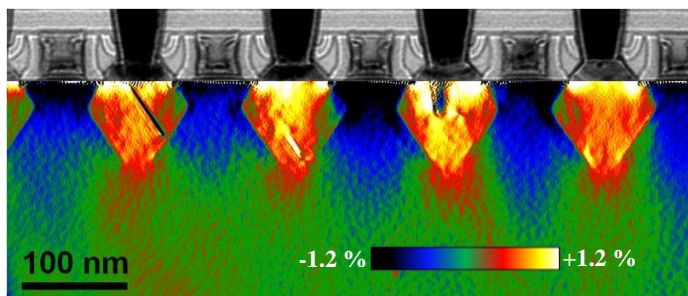


Abbildung 2.22: Gitterdeformation infolge des eingebetteten Silizium-Germaniums in den Source-Drain-Gebieten eines p-MOSFETs, gemessen mithilfe der geometrischen Phasenanalyse einer hochauflösenden Transmissionselektronenmikroskopie [60].

E. STI-Verspannung

Das Isolationsmaterial (Shallow Trench Isolation, STI), welches die Aktivgebiete der einzelnen Transistoren voneinander trennt, bietet ebenfalls die Möglichkeit zur lokalen mechanischen Verspannung. Durch die Ausdehnung des Oxids entsteht eine Druckverspannung in der Größenordnung von -400 MPa, die bei kleinen Strukturen bis in den Kanal vordringt [63].

F. Silizierung der Source-Drain-Gebiete

Die Silizierung der Source-Drain-Gebiete erzeugt aufgrund der Phasenumwandlung eine Verspannung, die von einer Vielzahl von Prozessbedingungen abhängt. Die Wahl des Metalls und die Geometrie des Transistors sind dabei entscheidend dafür, wie hoch die intrinsische Verspannung ist und wie viel in den Kanal induziert werden kann [64]. Die aktuellen Transistoren werden mit Nickelsilizid prozessiert, welches eine Zugverspannung von ca. 500 MPa aufweisen kann [65].

G. Intrinsische Verspannung

Mit einem geeigneten Material und der entsprechenden Prozessierung können im Metallgate (Metal Gate Stress, MGS) und den Source-Drain-Kontakten ebenfalls hohe Verspannungen verursacht werden, die sich schließlich bis in das Kanalgebiet des Transistors ausbreiten. Titannitrid erzeugt z. B. eine intrinsische Druckverspannung, von der hauptsächlich der n-MOSFET profitieren kann [66].

H. Einsatz von Verspannungstechniken

Nicht alle genannten Verspannungstechniken eignen sich für die Volumenproduktion. Einige besitzen durch die fortschreitende Skalierung eine begrenzte Einsatzmöglichkeit. Abbildung 2.23 fasst die verwendeten Verspannungstechniken für die aktuellen planaren Transistortechnologien schematisch zusammen.

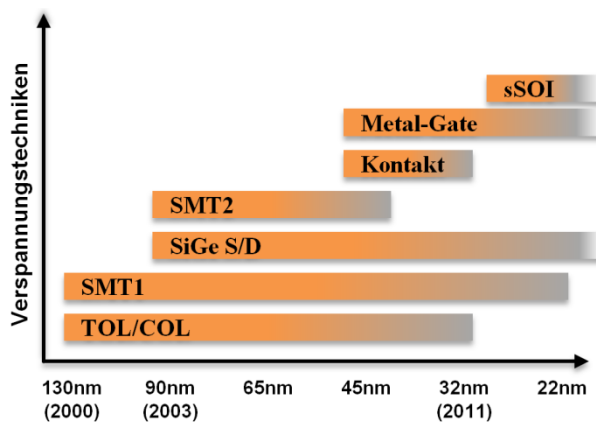


Abbildung 2.23: Angewandte Techniken zur mechanischen Verspannung von planaren Transistoren bezüglich der aktuellen Technologieknoten.




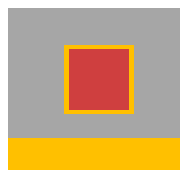




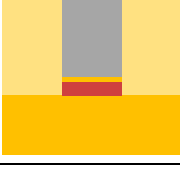

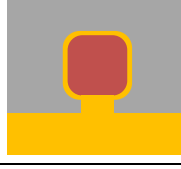
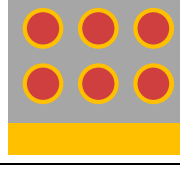

2.2 Multi-Gate-Transistoren

2.2.1 Multi-Gate-Strukturen

Schon nach wenigen Zyklen der regelmäßigen Skalierung wurde Anfang der 1980er Jahre klar, dass sich die Kurzkanaleffekte für die Weiterentwicklung von planaren MOSFETs zu einem primären Problem entwickeln könnten. Man begann daher mit der Erforschung von alternativen Transistorgeometrien [35], [67], [68]. Die wichtigsten der bisher vorgestellten Multi-Gate-Strukturen sind in Tabelle 2.3 aufgeführt. Sie stellen gleichzeitig einen evolutionären Überblick über die Entwicklung vom planaren Bulk-MOSFET hin zum gestapelten Nanowire-Transistor dar.

Im ersten Schritt wurde die Basis des planaren Bulk-MOSFETs durch ein neues Substrat (SOI) ersetzt, um parasitäre Kapazitäten zu reduzieren und die Kanaltiefe zu limitieren. So entstand der teilweise verarmte SOI-MOSFET, wie er seit Anfang der 2000er Jahre von IBM, AMD und FREESCALE gefertigt wird [69].

Tabelle 2.3: Überblick zu den gängigen Einzel- und Multi-Gate-Strukturen von MOSFETs (Querschnitt entlang der Transistorweite).

Einzel-Gate	Doppel-Gate	Tripel-Gate	Rundum-Gate
<p>Bulk</p> 	<p>Doppel-Gate (Top/Bottom)</p> 	<p>Tri-Gate</p> 	<p>Quad-Gate</p> 
<p>SOI</p> 	<p>Doppel-Gate (seitlich)</p> 	<p>Ω-Gate</p> 	<p>Nanowire</p> 
<p>FD-SOI</p> 	<p>FinFET</p> 	<p>Π-Gate</p> 	<p>Gestapelte Nanowire</p> 
<p>  Silizium Oxide Gate </p>			

Anschließend folgte der vollständig verarmte SOI-MOSFET (Fully Depleted, FD-SOI), welcher sich durch einen dünnen und niedrig- bis undotierten Kanal auszeichnet [34]. Die nächste Stufe der Entwicklung stellte der Doppel-Gate-MOSFET (DG-MOS) dar, wie er erstmals 1984 in [70] veröffentlicht wurde. Weitere Varianten, bei denen sich die beiden Gatekontakte ober- und unterhalb sowie seitlich des Kanals befinden, und der FinFET, der aufgrund des vernachlässigbaren Anteil des Top-Gates mit zu den Doppel-Gate-Transistoren gezählt werden kann, folgten [71].

Im Bestreben, die Steuerfähigkeit des Gatekontakts weiter zu verbessern, wurden Transistorgeometrien interessant, die das Kanalgebiet bestmöglich umfassen. Ohne aufwendiges Unterätzen des Silizium lassen sich jedoch bei horizontalen Strukturen maximal Tripel-Gate-Transistoren realisieren. Das Tri-Gate ist dabei die einfachste Form, da es mithilfe eines anisotropen Ätzschrittes hergestellt werden kann [72]. Ein rundes Kanalgebiet bzw. ein teilweise isotropes Überätzen ergeben Strukturen wie das Ω - oder Π -Gate [73].

Die Struktur mit den besten elektrostatischen Eigenschaften ist ein Transistor mit umschließendem Rundum-Gate (Surround-Gate). Die ersten Experimente wurden dabei an vertikalen Strukturen durchgeführt, bis später das Kanalgebiet von horizontalen Transistoren durch gezieltes Unterätzen vollständig freigelegt und somit Platz für ein Rundum-Gate geschaffen werden konnte [74].

Der Höhepunkt der derzeitigen Entwicklung sind mehrere übereinander gestapelte Nanowire mit umschließenden und untereinander verbundenen Gates. Dies ermöglicht eine wesentlich bessere Ausnutzung der gegebenen Fläche (Aktivfläche), die jedem Transistor zur Verfügung steht [75].

Eine vertikale Integration des Rundum-Gates, wie sie Abbildung 2.24 schematisch darstellt, ist jedoch nicht nur einfacher zu realisieren, sie bietet zudem Vorteile für eine weitere Skalierung [76]. Der Platzbedarf des Transistors ist in diesem Fall nicht mehr von der Gatelänge abhängig und richtet sich hauptsächlich nach dem Durchmesser und dem Abstand der Kanäle bzw. nach der Breite der Kontaktgebiete. Somit kann beim Übergang zur nächsten Technologie die Gatelänge unverändert bleiben. Durch die Verkleinerung des Kanalquerschnitts wird zudem die Steuerfähigkeit des Gates weiter verbessert [34]. Die Kontaktierung von Drain, Source und Gate muss jedoch seitlich herausgeführt werden. Die Kontaktierung von Drain, Source und Gate muss jedoch seitlich herausgeführt werden und bei der Berechnung des Platzbedarfs mitberücksichtigt werden.

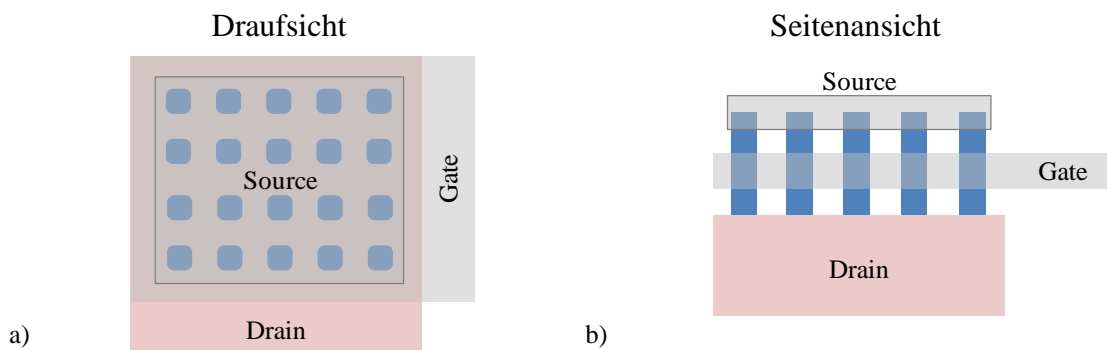


Abbildung 2.24: a) Draufsicht bzw. b) Seitenansicht eines Nanowire-MOSFETs. Drain und Gate werden seitlich herausgeführt. Das Source-Gebiet kann direkt von oben kontaktiert werden.

Um die Leistungsfähigkeit von Multi-Gate-Strukturen untereinander geeignet vergleichen zu können, ist die Normierung der Ströme von entscheidender Bedeutung, wobei zwei Varianten zur Auswahl stehen:

- Eine Normierung auf die effektive Gateweite eignet sich am besten für eine Aussage zur Effizienz der Gatekopplung und zur Leitfähigkeit des Kanalgebiets. Die Peripherie, die das Kanalgebiet und den Gatekontakt umschließt, wie z. B. die Source-Drain-Kontakte oder der Isolator zwischen zwei Finnen, bleibt dabei unberücksichtigt.
- Für einen Vergleich von verschiedenen Multi-Gate-Transistoren hinsichtlich ihrer gesamten Leistungsfähigkeit bezogen auf ihre Aktivfläche sollte auch eine Normierung auf die Strukturweite bzw. -fläche verwendet werden.

Wie verschieden die Ergebnisse dieser beiden Normierungsvarianten sein können, zeigt Abbildung 2.25, in der die Universalkennlinien zweier Tri-Gate-Transistoren mit unterschiedlicher Höhe und Weite, jedoch mit gleichem Abstand ($Pitch_{Fin} = 50 \text{ nm}$) der Finnen, dargestellt sind. In diesem Beispiel wurde die Finhöhe von 20 nm auf 25 nm und die Finweite von 25 nm auf 35 nm erhöht, wodurch die effektive Gateweite um 20 nm zugenommen hat, die Strukturweite hingegen konstant geblieben ist. Somit steht bei gleicher Aktivfläche für den Stromfluss mehr Kanalvolumen zur Verfügung, was den Transistor besser erscheinen lässt. Die auf die effektive Gateweite bezogene Leistungsfähigkeit hat jedoch abgenommen, da in diesem Fall der Zugewinn an Drainstrom durch diese Strukturänderung deutlich kleiner ausfällt als die Zunahme an effektiver Gateweite. Generell gilt für die meisten Multi-Gate-Strukturen, dass die effektive Gateweite mindestens der Strukturweite des Transistors entsprechen sollte, damit auch eine effiziente Ausnutzung der Aktivfläche gegeben ist. Andernfalls wäre ein planarer Transistor im Vergleich zum Tri-Gate-Transistor mit Strukturweitennormierung sehr wahrscheinlich leistungsfähiger und aufgrund der einfacheren Integration zu bevorzugen.

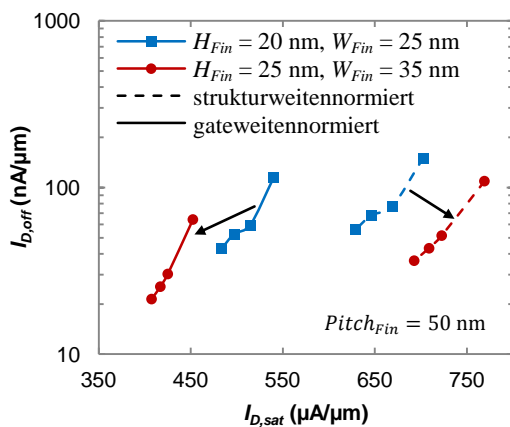


Abbildung 2.25: Universalkennlinien von zwei n-Tri-Gate-Transistoren mit unterschiedlicher Höhe H_{Fin} und Weite W_{Fin} der Finne, jedoch gleichem Abstand $Pitch_{Fin}$ zwischen den Finnen. Die Normierung der Ströme bezieht sich auf die effektive Gateweite bzw. auf die Strukturweite.

2.2.2 Überlagerungseffekte

Die Überlagerungseffekte umfassen alle auftretenden Besonderheiten von Multi-Gate-Strukturen, die durch die Überlagerung der elektrischen Felder von orthogonal zueinander stehenden Gateanteilen des Transistors auftreten. Wie stark die Effekte wirken, hängt dabei von der Geometrie und der Kanaldotierung des Transistors ab [77]. Wie Gleichung (2.4) zeigt, richtet sich die Eindringtiefe der Verarmungszone und somit der elektrischen Feldstärke des Gates nach der Stärke der Kanaldotierung. Im Beispiel von Abbildung 2.26 ist der Querschnitt eines Tri-Gate-Transistors mit konstant verteilter Kanaldotierung, jedoch unterschiedlicher Konzentration dargestellt. Für hohe Werte von $N_{Ch} = 10^{19} \text{ cm}^{-3}$ kommt es zu einer Überlagerung in den Ecken der Finne, dem so genannten Eckeneffekt [78], [79].

Bei einer niedrigeren Kanaldotierung von $N_{ch} = 10^{18} \text{ cm}^{-3}$ hingegen überschneiden sich die Verarmungszonen der beiden gegenüberliegenden Seiten-Gates und die Finne kann als vollständig verarmbar angesehen werden [34], [80].

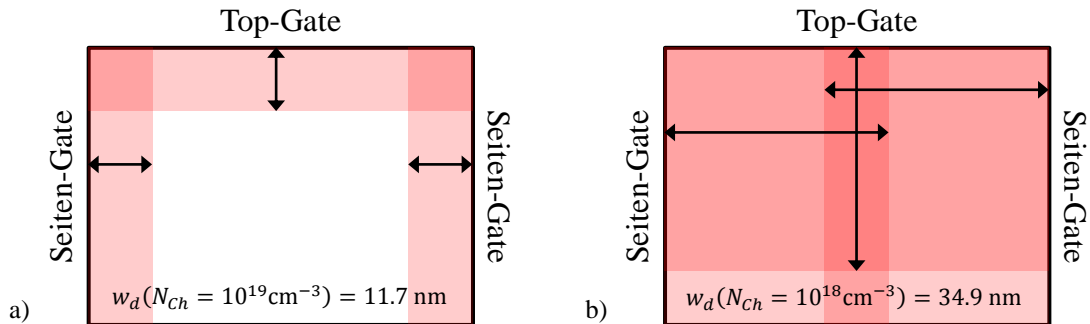


Abbildung 2.26: Ausbreitung der Verarmungsgebiete der einzelnen Gateflächen im Kanalgebiet für Dotierungskonzentrationen von a) $N_{ch} = 10^{19} \text{ cm}^{-3}$ und b) $N_{ch} = 10^{18} \text{ cm}^{-3}$.

Die Überlagerung in den Ecken der Finne bewirkt ein erhöhtes Potenzialgefälle, was bei einer Inversion des Kanalgebiets auch ein verstärktes Absenken der Bandenergien zur Folge hat (Abbildung 2.27a). Somit geraten die Bereiche der Ecken schon bei geringeren Gate-Source-Spannungen in einen verarmten bzw. invertierten Zustand und führen gegenüber den planaren Abschnitten des Gates eine deutlich erhöhte Ladungsträgerdichte, wie sie in Abbildung 2.27b dargestellt ist [81], [82].

Doch nicht allein die Kanaldotierung bestimmt, wie stark sich der Effekt ausprägt, sondern auch die Geometrie der Finne. Bei spitzen Kanten zeigt die Überlagerung eine starke Wirkung. Ein isotropes Ätzen oder aggressive Reinigungsschritte sorgen dafür, dass die Ecken abgerundet werden und sich ein Radius ausbildet, der einen kontinuierlichen Übergang von Top- und Seiten-Gate ermöglicht [78].

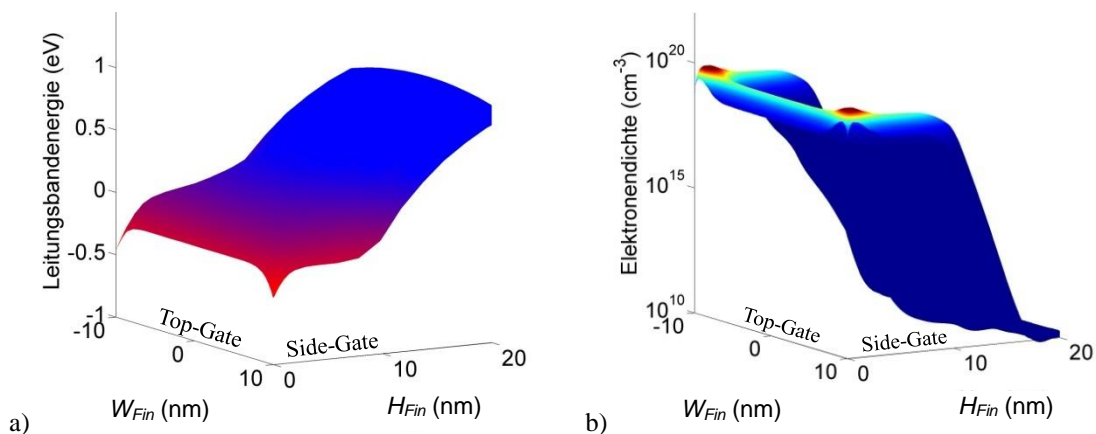


Abbildung 2.27: a) Leitungsbandenergie bezogen auf das Leitungsbandminimum des intrinsischen Siliziums und b) Elektronendichte für einen Tri-Gate-Transistor mit $H_{Fin} = 10 \text{ nm}$ und $W_{Fin} = 20 \text{ nm}$, sowie bei $U_{GS} = 0.7 \text{ V}$.

Ebenso sind die Höhe und die Weite der Finne ausschlaggebend dafür, in welchen Bereichen es zu einer Überlagerung kommt. Bei sehr kleinen Strukturen mit einem geringen Abstand der gegenüberliegenden Gateflächen von ca. 5 nm bis 10 nm kommt es in der Regel unabhängig von der Dotierung zu einer vollständigen Verarmung des Kanalgebiets bzw. einer Volumeninversion. Darüber hinaus dominieren bei einer Weite von etwa $W_{Fin} = 20$ nm die Ecken das elektrische Verhalten, was durch die eine Spitze der Steilheit in Abbildung 2.28a verdeutlicht wird. Bei größeren Strukturen ($W_{Fin} = 80$ nm, Abbildung 2.28b) mit entsprechend hoher Dotierung wird das vorzeitige Invertieren der Ecken gegenüber der planaren Bereiche anhand zweier Spitzen der Steilheit erkennbar.

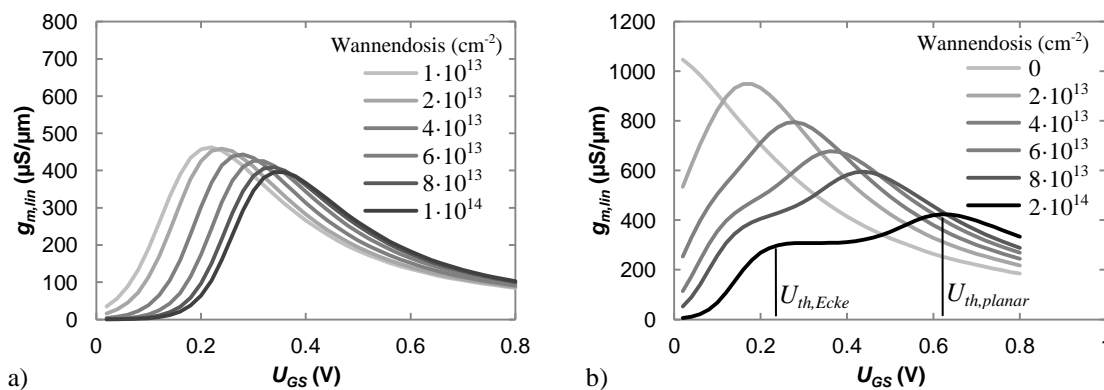


Abbildung 2.28: Lineare Steilheit von n-Tri-Gate-Transistoren für a) Standardstrukturen ($L_{Gate} = 26$ nm, $H_{Fin} = 10$ nm, $W_{Fin} = 20$ nm) und b) große Strukturen ($L_{Gate} = 54$ nm, $H_{Fin} = 20$ nm, $W_{Fin} = 80$ nm) in Abhängigkeit von der Gate-Source-Spannung für verschiedene Wannendosen.

Der Eckeneffekt als einer der Überlagerungseffekte kann das elektrische Verhalten des Transistors auch positiv beeinflussen und verbessert z. B. die Steuerfähigkeit des Gates. Er wird jedoch aufgrund der stark ungleichmäßigen Verteilung der Ladungsträger hauptsächlich als parasitäre Erscheinung von Multi-Gate-Strukturen angesehen. Um diesem Effekt vorzubeugen, sollten runde Gatestrukturen wie beim Nanowire-Transistor verwendet oder der Kanal ohne Dotierung prozessiert werden [83]. Eine weitere Möglichkeit, den Eckeneffekt zu unterdrücken, ist eine zusätzliche Eckenimplantation (siehe Abschnitt 5.2.3).

2.2.3 Quanteneffekte

Die klassischen Volumenmodelle der Ladungsträgerverteilung sowie der Valenz- und Leitungsbänder im Silizium basieren auf der Grundlage eines dreidimensional uneingeschränkten Elektronengases, wie es für kristallines Silizium ohne räumliche Begrenzung angenommen werden kann [7], [11], [16]. Eine Berücksichtigung der geometrischen

Randbedingungen von Multi-Gate-Strukturen und der daraus resultierenden Quantisierung der Ladungsträger kann von ihnen nicht nachgebildet werden. Durch die bislang hinreichend großen Dimensionen der Gatestrukturen erwies sich der Einfluss der Quantisierungseffekte vernachlässigbar klein. Für die Simulation von planaren Transistoren bis zur 22 nm-Technologie war lediglich eine Korrektur der Ladungsträgerdichte aufgrund nicht erlaubter Zustände in unmittelbarer Nähe der Grenzfläche zum Gateoxid notwendig, um das Verhalten des Transistors bezüglich dieser Quanteneffekte hinreichend genau nachzubilden [84], [85]. Mit der Einführung von Multi-Gate-Strukturen, die Abmaße von unter 20 nm aufweisen, müssen diese Aspekte bei der Simulation jedoch mit berücksichtigt werden.

Durch die Einwirkung der elektrischen Feldstärke des Gates von mehreren Seiten auf das geringe Kanalvolumen kann das Elektronengas nicht mehr als uneingeschränkt dreidimensional angenommen werden, da es zur Formierung von Subbändern kommt [34], [86]. Dabei unterscheidet man je nach Gate-Struktur zwischen ein- und zweidimensionalen Elektronengasen. Eine Doppel-Gate-Struktur erzeugt ein zweidimensionales Elektronengas, da die Wirkung des Gates nur in Richtung einer Dimension verläuft und die Elektronen sich gegenüber den beiden orthogonalen Richtungen uneingeschränkt verhalten. Im Tri- oder Rundum-Gate treten eindimensionale Elektronengase auf, da die Elektronen nur in Transportrichtung uneingeschränkt sind.

Die Formierung von Subbändern sorgt dafür, dass sich im Kanal die Zustandsdichten von Valenz- und Leitungsband anders verhalten als für ein uneingeschränktes Elektronengas. In Abbildung 2.29 ist die Zustandsdichte im Zusammenhang mit der Leitungsbandenergie bei starker Inversion für einen Tri-Gate-Transistor mit unterschiedlicher Finhöhe dargestellt.

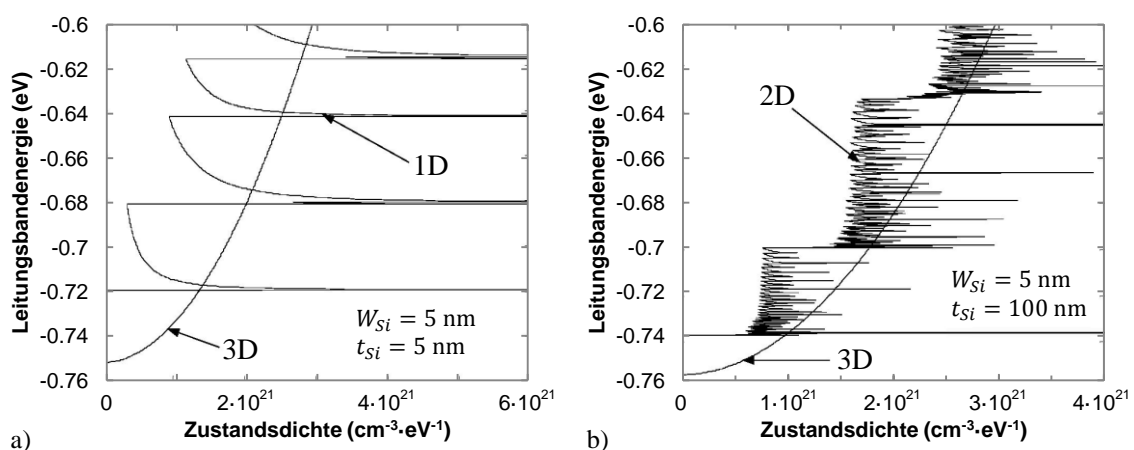


Abbildung 2.29: Zustandsdichte des Leitungsbandes eines Tri-Gate-Transistors bei starker Inversion für a) ein eindimensionales und b) ein zweidimensionales Elektronengas im Vergleich zur Zustandsdichte des dreidimensionalen Elektronengases [34].

Bei einer Finhöhe von $t_{Si} = 5$ nm liegt der eindimensionale Fall vor und es bilden sich einzelne Spitzen auf den Niveaus der Subbänder. Mit der Erhöhung der Finne auf $t_{Si} = 100$ nm wird das Elektronengas zweidimensional und es kommt zu einer treppenförmigen Abstufung der Zustandsdichte, bei der eine Vielzahl an Spitzen die Subbänder mit gleicher Zustandsdichte repräsentieren. Diese Einschränkung der Zustandsdichte sorgt dafür, dass bei Inversion in unmittelbarer Nähe zum Gateoxid aufgrund der wenigen erlaubten Zustände eine verringerte Ladungsträgerdichte auftritt. Größere Strukturen bilden trotz dieser Tatsache immer noch eine Oberflächeninversion aus, wohingegen bei kleineren Multi-Gate-Transistoren die Interaktion der gegenüberliegenden Gateflächen zu einer Volumeninversion führt, wie Abbildung 2.30 zeigt [87], [88].

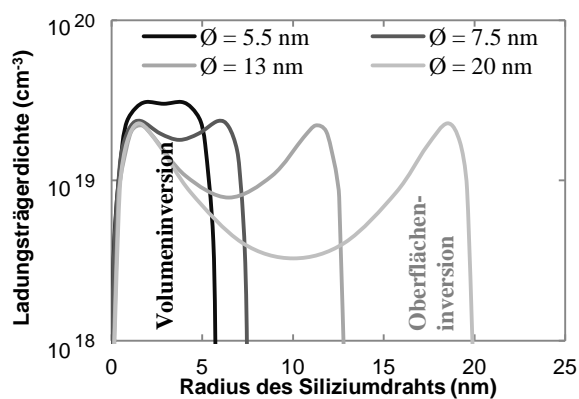


Abbildung 2.30: Radiale Verteilung der Ladungsträgerdichte in einem simulierten Nanowire-Transistor für verschiedene Durchmesser des Siliziumdrahts.

Die Quantisierung hat aber nicht nur Auswirkung auf die Zustands- und Ladungsträgerdichte, sondern auch auf die Ladungsträgerbeweglichkeit und die Schwellspannung des Transistors [89]. Wie Abbildung 2.31 zeigt, kann für Strukturen mit Abmaßen von mehr als 15 nm eine Ladungsträgerbeweglichkeit angenommen werden, die bei Inversion an die Grenzflächenrauheit planarer Transistoren angepasst ist. Die Volumeninversion verringert jedoch den Einfluss der Streuung durch die Grenzflächen, sodass die Ladungsträgerbeweglichkeit kleinerer Strukturen im Bereich der maximalen Ladungsträgerdichte deutlich höher ausfällt. Unterhalb einer Kanalweite von 5 nm nähert sich das Inversionsmaximum zunehmend der Grenzfläche und die Oberflächenrauheit wird wieder dominant [90].

Die Beweglichkeit wird noch von einem weiteren Mechanismus reduziert, der Streuung der Ladungsträger zwischen den Subbändern (Inter Subband Scattering, ISS) [91]. Dieser Effekt steigt mit wachsender Gate-Source-Spannung im Unterschwellbereich an, da zunehmend mehr Bänder besetzt werden. Zweidimensionale Elektronengase weisen dabei eine erhöhte Streuung auf, da mehrere Subbänder auf dem gleichen Energieniveau liegen können (Abbildung 2.29). Ein Tri-Gate-Transistor mit einem eindimensionalen Elektronengas verursacht aufgrund der wenigen Subbänder auch nur eine geringe Streuung unter ihnen, was eine bessere Beweglichkeit der Ladungsträger zur Folge hat.

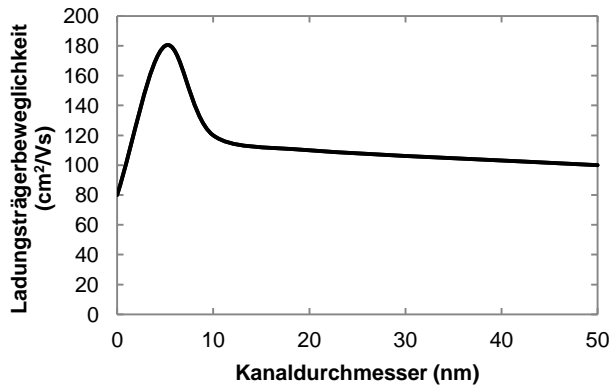


Abbildung 2.31: Ladungsträgerbeweglichkeit im Bereich der maximalen Ladungsträgerdichte eines Nanowire-Transistors in Abhängigkeit vom Kanaldurchmesser [34].

Die Schwellspannung ändert sich ebenfalls mit dem Übergang zu einem zwei- oder ein-dimensionalen Elektronengas. Die Aufspaltung in Subbänder erhöht das Leitungsbandminimum (Abbildung 2.29) und somit den Abstand zwischen Valenz- und Leitungsband bzw. die Schwellspannung des Transistors [34]. Für Siliziumstrukturen mit einem Durchmesser von weniger als 5 nm (Nanodrähte), was nur wenigen Atomlagen entspricht, geht die Bandstruktur aus Abbildung 2.3 verloren. Hierbei spricht man vom „Volumenlimit“ [92]. Die eingrenzenden Randbedingungen des Gateoxids sorgen dafür, dass sich die Welleneigenschaften der Elektronen im Silizium ändern. Dabei kommt es zu einer „Faltung“ der Bandstruktur und es entstehen Subbänder, die einen direkten Bandübergang aufweisen [93]. Experimente belegen, dass der direkte im Vergleich zum indirekten Übergang dominant werden kann und sehr dünne Siliziumschichten Licht emittieren können [94]. Mit kleiner werdendem Durchmesser erhöht sich somit die resultierende Bandlücke zunehmend und auch die effektiven Massen aller Orbitale nehmen zu [30].

2.3 Stand der Technik

Bereits kurze Zeit nachdem MOORE sein Gesetz zur Skalierung formulierte, kamen Fragen auf, wie lange diese Entwicklung planarer MOSFETs auf Siliziumbasis weitergehen kann und wo die physikalischen und technologischen Begrenzungen liegen [95]. Abgesehen von den Schwierigkeiten der Integration immer kleinerer Strukturen, welche über die Jahre stets gelöst werden konnten, wurde vor allem die Prozessierung eines planaren Transistors mit einer Gatelänge von unter 30 nm als wesentlichste Herausforderung der kontinuierlichen Skalierung angesehen [96]. Der Grund dafür liegt bei den stark ansteigenden Leckströmen und den stetig schlechter werdenden Schalteigenschaften der planaren Ultrakurzkanaltransistoren. Die Verwendung neuer Materialien und neuartiger Strukturen bieten somit die einzigen Lösungsmöglichkeiten, um noch kürzere Gate-längen zu realisieren.

In den 1980er Jahren wurde selbstjustierendes Cobaltsilizid (CoSi_2) bzw. Titansilizid (TiSi_2) eingeführt, um den parasitären Widerstand der Kontaktgebiete zu verbessern. Sie erlaubten zudem eine hohe Prozesstemperatur nachfolgender Schritte von bis zu 800°C . Mit der Verwendung des eingebetteten Silizium-Germaniums der 90 nm-Technologie musste wegen der besseren Kompatibilität zu Germanium auf Nickelsilizid (Ni_2Si) umgestellt werden [97].

Die mechanische Verspannung und der SOI-Transistor waren in den frühen 2000er Jahren der Schlüssel, um die Leistungsfähigkeit der Transistoren ohne Reduktion der Gatelänge und Erhöhung der Leckströme zu steigern [15], [69], [98]. Die Skalierung der Aktivfläche erfolgte über die Transistorweite und die Länge der Kontaktgebiete. Doch die Einsatzmöglichkeiten der Verspannungselemente nahmen mit fortschreitender Skalierung ab und auch das SOI-Prinzip kann aufgrund des Floating-Body-Effekts [99] und der fehlenden Steuerung über die Substratspannung voraussichtlich nur bis zur 32 nm-Technologie angewandt werden [100].

Ab der 90 nm-Technologie kamen neuartige Ansätze zur Dotierung (Plasma-Dotierung, Co-Implantation) und Kurzzeitausheilung der Transistoren hinzu, um die zunehmenden Kurzkanaleffekte durch steilere Dotierungsprofile zu begrenzen [101].

Um die Gatekapazität und somit die Steuerfähigkeit zu erhöhen, wurde mit der 45 nm- bzw. der 32 nm-Technologie ein Dielektrikum mit erhöhter Permittivität, das so genannte „High-k-Gateoxid“, eingeführt [102]. Somit konnte die effektive Gateoxid-dicke weiter reduziert und die Leckströme ebenfalls verringert werden [103]. In der Regel wird hierfür Hafniumoxid (HfO_2) verwendet, dessen Prozessierung jedoch mit einigen Problemen verbunden ist. Einerseits muss noch eine Siliziumoxidschicht von ca. 6 \AA zwischen dem Silizium und dem High-k-Oxid abgeschieden werden, da ein direkter Kontakt keine ausreichend gute Haftung bietet und eine starke Oberflächenrauheit hervorruft. Andererseits ist das High-k-Oxid anfälliger gegenüber fixen Ladungen und Traps (Ladungsfallen), welche die Beweglichkeit der Ladungsträger im Kanal reduzieren [104], [105].

Um den fixen Oxidladungen vorzubeugen und die Wirkung der mechanischen Verspannung des eingebetteten Silizium-Germaniums in den Source-Drain-Gebieten zu steigern, wurde mit dem High-k-Oxid der Replacement-Gate-Prozess eingeführt, bei dem ein Dummy-Gate aus polykristallinem Silizium gegen Ende des Herstellungsprozesses durch das Gatemetall ersetzt wird [106]. Parallel zu dieser Entwicklung wurde auch ein Gate-First-Prozess verfolgt, der jedoch mit Problemen der Verkapselung verbunden ist und durch eine veränderte Austrittsarbeit des Gatemetalls der p-MOSFETs den Einsatz von Silizium-Germanium im Kanal voraussetzt [107].

Die gesamte Entwicklung der planaren Transistoren bis hin zur aktuellen 28 nm-Technologie hat gezeigt, dass es trotz mechanischer Verspannung, steilen Dotierungsprofilen und der Integration von neuen Materialien nicht möglich ist, aufgrund von Kurzkanal-

effekten eine Gatelänge von deutlich weniger als 30 nm zu realisieren. Aktuell setzt die Industrie mehr auf Derivate der 28 nm-Technologie, die abgestimmt auf ihre Anwendungsmöglichkeiten von „Ultra-Low-Power“ mit möglichst geringem Energieverbrauch bis hin zu „Super-High-Performance“ mit maximaler Rechenleistung reichen. Mit großer Wahrscheinlichkeit werden die planaren Transistoren bis zum Jahr 2020 weiterhin die dominanten Bauelemente in der Halbleiterindustrie bleiben [108].

Darüber hinaus fokussiert sich die Entwicklung auf ultradünne Kanalstrukturen, deren Dicke maximal zwei Dritteln der Kanallänge entspricht [109], [110]. Die Einführung von FD-SOI- bzw. Multi-Gate-Transistoren in die Volumenproduktion ist mit hohen Entwicklungs- und Fertigungskosten verbunden, die die Industrie bis zur 22 nm-Technologie gescheut hat [111].

Auch sind Multi-Gate-Strukturen trotz ihrer deutlich besseren Steuerfähigkeit des Gates mit einigen Nachteilen verbunden. Viele experimentelle Strukturen benötigen für ihre Herstellung mehr Platz für Spacer und Source-Drain-Gebiete um die Finne herum als die Aktivfläche aktueller planarer Transistoren bietet, was dem Gesetz der Skalierung widerspricht [112]. Erhöhte Source-Drain-Widerstände sind die Folge der erschwerten Kontaktierung und des geringen Kanalquerschnitts, die auch mit aktuellen Verfahren, wie den aufgewachsenen Source-Drain-Gebieten, nicht kompensiert werden können [96]. Ebenso erhöhen sich die parasitären Anschlusskapazitäten im Vergleich zum stromdurchflossenen Kanalvolumen.

Aufgrund einer kaum kontrollierbaren starken Dotierungsfluktuation muss der Kanal eines Multi-Gate-Transistors undotiert bleiben [113]. Dies macht jedoch den Einsatz eines Gatemetalls mit einer Austrittsarbeit im Bereich der Mitte der Bandlücke erforderlich und nimmt gleichzeitig die Möglichkeit, die Schwellspannungen für verschiedene Subklassen der MOSFETs einzustellen bzw. über den Substratkontakt zu kontrollieren.

Weitere Ansätze für zukünftige Technologien sind Kohlenstoff-Nanoröhrchen und III-V-Halbleiter [114], [115]. Die Kohlenstoff-Nanoröhrchen wurden als aktive Bauelemente sowie als Verbindungsmaterial in den Metallebenen erforscht. Bei einer vergleichbaren Gatelänge derzeitig industriell gefertigter planarer Transistoren besitzen sie jedoch einen sehr hohen Leckstrom, verursacht durch Tunnelmechanismen des Gateoxids und einem direkten Source-Drain-Tunneln [116]. III-V-Halbleiter bieten zwar eine hohe Beweglichkeit der Ladungsträger, unterliegen aber einer sehr geringen Zustandsdichte. Höhere Kosten der Integration und fortbestehende Kurzkanalprobleme machen eine Implementierung von III-V-Halbleitern als Kanalmaterial auf lange Sicht unattraktiv. Auch bewegt sich die erwartete Leistungssteigerung dieser beiden Ansätze bei Weitem nicht in der Größenordnung, wie sie für einen derartig radikalen Umbruch zu einer neuen Technologie erforderlich wäre, um den Aufwand und die Kosten zu decken [96].

3 Grundlagen der Simulation

Auf dem Weg der Entwicklung und Verbesserung einer neuen Technologie bis hin zur endgültigen Produktionsreife müssen zahlreiche Experimente zu verschiedenen Prozessschritten unternommen werden. Ein sehr wichtiges Hilfsmittel, um frühzeitig Ergebnisse abschätzen zu können und zusätzliche Daten zu gewinnen, bietet eine vollständige Simulation der MOSFETs, bestehend aus Prozess- und Bauelementesimulation. Für die Simulationsergebnisse dieser Arbeit wurde das Programm *Sentaurus Workbench Advanced* (Version D-2010.03) von SYNOPSIS verwendet [117], [118]. Es ermöglicht eine komplexe Simulation des Herstellungsprozesses von 1D-, 2D- und 3D-Strukturen sowie eine stabile Bauelementesimulation, die eine umfangreiche Modellierung der wesentlichen Vorgänge im Hinblick auf den Ladungsträgertransport bietet. Um zuverlässige Daten zu erhalten, ist es jedoch erforderlich, dass sowohl die Simulation des Herstellungsprozesses bezüglich der Struktur und des Dotierungsprofils als auch die Simulation der elektrischen Eigenschaften des Bauelementes anhand entsprechender experimentell gewonnener Daten kalibriert werden.

Beginnend mit den theoretischen Grundlagen der Prozess- und Bauelementesimulation, wird zu Beginn ein Überblick über die angewandten Modellen gegeben. Der Schwerpunkt liegt hierbei auf der Bauelementesimulation, wie sie auch zur Kalibrierung der Modellparameter bezüglich der experimentellen Daten der 32 nm-SOI-Technologie genutzt wurde. Anschließend folgt eine Darstellung der Vorgehensweise zur Anpassung der Technologie- und Modellparameter mit ihren ermittelten Werten, bevor die Zusammenfassung der wesentlichen Kennlinien der kalibrierten n- und p-MOSFETs das Kapitel abschließt.

3.1 Prozesssimulation

3.1.1 Abscheiden und Abtragen von Schichten

Um die MOSFETs mit ihren verschiedenen Multi-Gate-Strukturen möglichst exakt simulieren zu können, benötigt man einen Prozesssimulator, der alle chemischen und physikalischen Abscheide- bzw. Ätzverfahren für dreidimensionale Geometrien nachbildet [118].

Zu den chemischen Abscheideverfahren zählt in erster Linie die Silizium-Gasphasenepitaxie (Chemical Vapour Deposition, CVD) [21]. Sie ermöglicht die Homo- und die Heteroepitaxie sowie das Aufbringen von Siliziumverbindungen. Besondere Formen der chemischen Gasphasenabscheidung sind die Atom- und die Moleküllagenabscheidung (Atomic/Molecular Layer Deposition, ALD/MLD), die in ihrer Reaktion selbstbegrenzend sind und so sehr dünne Schichten von einzelnen Atom- oder Moleküllagen ermöglichen [119]. Zu den physikalischen Verfahren der Schichtabscheidung gehören das Aufdampfen, die Molekularstrahlepitaxie und die Kathodenzerstäubung.

Die Verfahren zum Schichtabtrag werden in Nass- und Trockenätzen unterschieden. Die nasschemischen Verfahren resultieren in einem isotropen Abtrag und werden unter anderem zum Entfernen von Lacken und zur Reinigung des Wafers verwendet. Das Trockenätzen ermöglicht mit der Anregung von Gasen durch hochfrequente Wechselfelder einen anisotropen Schichtabtrag und wird nochmals in drei Varianten unterschieden: das rein physikalische, das rein chemische und das chemisch-physikalische Trockenätzen [16]. Je nach Prozessbedingungen kann sich die Ätzrate für einzelne Kristallrichtungen unterscheiden, wodurch auch ein direktionales Abtragen der Schichten realisiert werden kann.

Die Programme zur Prozesssimulation führen jedoch keine chemisch-physikalischen Berechnungen zum Schichtauf- bzw. Schichtabtrag durch. Sie verwenden lokale Depositions- und Ätzraten in Modellen für geometrische Operationen, um die real existierenden Strukturen möglichst authentisch nachzubilden (Abbildung 3.1). Dabei wird auch das Abschatten und Überätzen an den Kanten beim selektiven Auf- und Abtrag berücksichtigt. Anschließend wird das Simulationsgitter entsprechend der neugebildeten Grenzflächen aktualisiert. Dreidimensionale Simulationen erfordern dabei aufgrund ihrer hohen Komplexität einen vertretbaren Kompromiss zwischen Gitterauflösung und Rechenaufwand.

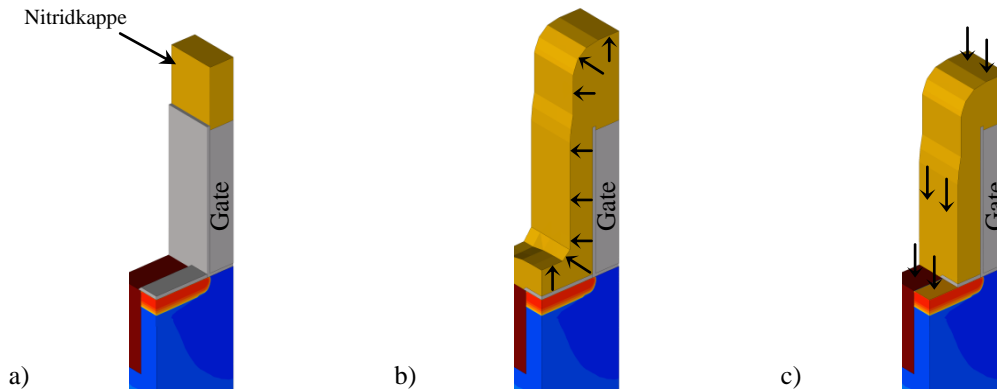


Abbildung 3.1: Darstellung einer dreidimensionalen Prozesssimulation des Siliziumnitrid-Spacers:

- a) Ausgangsstruktur mit Nitridkappe auf dem Gate
- b) Isotropes Abscheiden von Siliziumnitrid
- c) Anisotropes Ätzen des Siliziumnitrids

3.1.2 Implantation

Die Ionenimplantation ist derzeit die gängigste Methode zur Erzeugung des charakteristischen Dotierungsprofils eines MOSFETs. Dabei werden die Dotanden durch ein hochenergetisches Feld beschleunigt und treffen anschließend auf die Waferoberfläche. Ihre Eindringtiefe hängt bei diesem Vorgang von ihrer kinetischen Energie, der Kristallorientierung der Waferoberfläche und dem Implantationswinkel ab [120]. Durch Stoßprozesse mit den Gitteratomen des Siliziums verlieren die Ionen an Energie, bis sie letztendlich zum Stillstand kommen. Aufgrund der Kristallstruktur wird ein Teil der implantierten Ionen oberflächennah abgebremst, während der andere Teil entlang der Kristallorientierung tiefer eindringen kann. Dieser Vorgang wird als „Channeling-Effekt“ bezeichnet und ist in Abbildung 3.2 bei der Implantation ohne Streuoxid gut erkennbar. Bei Verwendung eines Streuoxids an der Waferoberfläche kommt es zu einer gleichmäßigeren Verteilung der Dotanden im Substrat und zu einer Verschiebung des Maximums in Richtung der Oberfläche, da die Ionen durch den Stoßvorgang mit dem Streuoxid zusätzlich abgebremst werden.

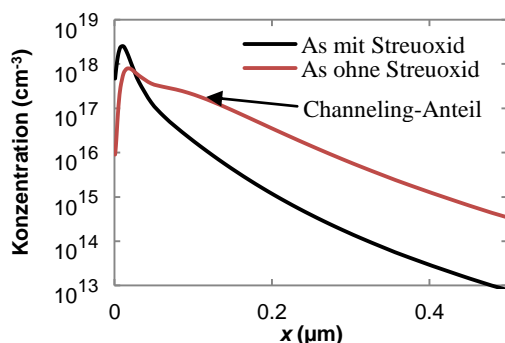


Abbildung 3.2: Simulierte Konzentration von Arsen im Siliziumsubstrat nach einer Implantation von Arsenionen mit und ohne Streuoxid an der Waferoberfläche.

Der Channeling-Effekt kann auch durch eine Preamorphisierung des Substrates unterdrückt werden. Hierbei sorgt eine vorausgehende Implantation von inerten Elementen, wie Edelgasen oder Kohlenstoff, für die Zerstörung der Kristallstruktur und somit für die Amorphisierung des Substrats.

Für die Simulation der Ionenimplantation stehen zwei Varianten zur Auswahl, die Monte-Carlo-Simulation oder die analytischen Modelle zur Berechnung des Implantationsprofils. Bei der Monte-Carlo-Simulation wird die zu implantierende Dosis in eine definierte Anzahl an Partikeln aufgeteilt, die von einem zufällig bestimmten Segment einer Fläche, welche parallel und oberhalb der Waferoberfläche erzeugt wird, aus starten und sich mit vorgegebener Energie und Winkel in Richtung der Waferoberfläche bewegen. Mit dem Eindringen in das Substrat unterliegen die Partikel auch zufälligen Stoßprozessen, die ihre Energie und Richtung beeinflussen, bis sie zum Stillstand kommen. Je höher die Anzahl der Partikel ist, um so genauer ist die resultierende Verteilung der Dotanden. Eine große Partikelanzahl bedeutet jedoch einen wesentlich höheren Rechenaufwand, besonders für dreidimensionale Strukturen. Eine schnellere Variante bietet die Berechnung des Implantationsprofils über analytische Modelle. Sie basieren auf einer GAUSS- bzw. PEARSON-Verteilung, deren Parameter aus Experimenten und Monte-Carlo-Simulationen gewonnen wurden und in Implantationstabellen hinterlegt sind. Zudem unterscheiden die analytischen Modelle zwischen amorphisierten und nichtamorphisierten Gebieten. Da der Channeling-Effekt sich aus beiden Anteilen zusammensetzt, werden zur Berechnung dieser Profile auch beide Modelle verwendet. Die bei Implantation auftretenden Punktdefekte wie Gitterfehlstellen (Vakanzen) und Zwischengitteratome werden rein analytisch nach dem HOBLER-Modell berechnet [121].

3.1.3 Thermische Ausheilung mit Diffusion

Thermische Ausheilungsschritte sind für die Herstellung von Transistoren ein wichtiger Bestandteil. Sie sorgen dafür, dass die zerstörte Gitterstruktur wiederhergestellt und die implantierten Dotanden eingebaut und somit aktiviert werden. Die maximale Anzahl an aktiven Dotanden richtet sich nach ihrer Festkörperlöslichkeit und ist von Stoff zu Stoff unterschiedlich. Diejenigen, die nicht in das Siliziumgitter eingebaut sind, tragen später auch nicht zum Ladungstransport im Transistor bei.

Neben der Aktivierung kommt es bei thermischen Schritten stets zur Diffusion der Dotanden von einem Ort höherer Konzentration zu einem Ort niedrigerer Konzentration, bis nach einer unendlich langen Zeitspanne ein vollständiger Ausgleich erreicht wäre. Nach dem ersten FICKschen Gesetz bewegen sich die Teilchen dabei mit der Teilchenstromdichte

$$J_T = -D_{Dot} \cdot \nabla N , \quad (3.1)$$

die sich aus dem Gradient der Teilchendichte N und der Diffusionskonstanten D_{Dot} zusammensetzt [122]. Die Diffusionskonstante dient in diesem Zusammenhang als Proportionalitätskonstante und berechnet sich über eine Arrhenius-Funktion zu

$$D_{Dot} = D_{Dot,0} \cdot \exp\left(\frac{-E_a}{k_B \cdot T_L}\right). \quad (3.2)$$

Sie ist somit abhängig von der Aktivierungsenergie E_a und der Gittertemperatur T_L . Das zweite FICKSchen Gesetz

$$\frac{\partial N}{\partial t} = \nabla \cdot (D_{Dot} \cdot \nabla N) \quad (3.3)$$

umfasst die Änderung der Teilchendichte nach der Zeit (Abbildung 3.3). Da sich die Aktivierungsenergie und die Diffusionskonstante für die verschiedenen Elemente stark unterscheiden, ist auch ihr Diffusionsverhalten während der thermischen Ausheilungsschritten unterschiedlich stark. Dieser Punkt ist für die Kurzkanaltransistoren der letzten Generationen immer weiter in den Vordergrund gerückt, da die Dotierungsprofile stärker und steiler geworden sind und sich somit die Geschwindigkeit, mit der die Diffusion der Dotanden abläuft, erhöht hat. Um ein Verwaschen der pn-Übergänge zu vermeiden, musste die Temperatur und die Ausheilungszeit bzw. das thermische Budget reduziert werden. Kurzzeitausheilungen (Rapid Thermal Annealing, RTA) und Laserausheilungen (Laser Spike Annealing, LSA) sind seitdem die favorisierten Verfahren der industriellen Fertigung. Eine Blitzlampenausheilung (Flash Annealing) kommt aufgrund einer erhöhten Waferbruchrate nur selten zum Einsatz.

Für die Simulationen wurde das Charged-Pair-Diffusionsmodell als Standard im Simulator integriert. Es beruht auf den FICKSchen Gesetzen und berücksichtigt als treibende Kraft das von der Dotierung erzeugte elektrische Feld [118].

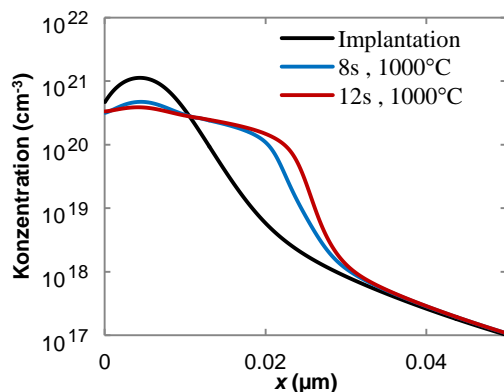


Abbildung 3.3: Konzentration von Arsen-Dotanden nach der Implantation und einem thermischen Ausheilungsschritt von 8 s bzw. 12 s und 1000 °C.

3.2 Bauelementesimulation

3.2.1 Grundgleichungen und Ladungsträgertransport

Die Berechnungen der Bauelementesimulationen dieser Arbeit beruhen auf makroskopischen Modellen, deren Modellparameter durch statistische Mittelwerte gegeben sind, da atomistische Simulationen in der Größenordnung noch nicht realisierbar sind. Zu den Grundgleichungen, die vom Simulator für jedes Halbleiterbauelement gelöst werden, zählen die POISSON-Gleichung, die Kontinuitätsgleichungen und die Transportgleichungen. Sie bilden das zentrale Gerüst der Bauelementesimulation und beinhalten Transportkoeffizienten, wie die Ladungsträgerbeweglichkeit oder die Energierelaxationszeit.

Die POISSON-Gleichung der Form

$$\nabla \cdot (\varepsilon_0 \varepsilon_r \cdot \nabla \varphi) = -e(p - n + N_D - N_A) \quad (3.4)$$

stellt den Zusammenhang zwischen dem Verlauf des elektrischen Potentials und den festen sowie den freibeweglichen Ladungsträgern her. N_A und N_D repräsentieren die Akzeptor- bzw. Donatordichte und sind über das Dotierungsprofil des Halbleiters gegeben. Die freibeweglichen Ladungsträger resultieren aus der BOLTZMANN-Statistik [16] für Elektronen

$$n = n_i \cdot \exp\left(\frac{-e(\Phi_n - \varphi)}{k_B \cdot T_L}\right) \quad (3.5)$$

und Löcher

$$p = n_i \cdot \exp\left(\frac{e(\Phi_p - \varphi)}{k_B \cdot T_L}\right), \quad (3.6)$$

wobei $\Phi_{n,p}$ für das Quasi-Fermi-Potenzial der jeweiligen Ladungsträger steht. Zudem werden die Kontinuitätsgleichungen der Elektronen

$$\nabla \cdot J_n = e \left(R - G + \frac{\partial n}{\partial t} \right) \quad (3.7)$$

und Löcher

$$\nabla \cdot J_p = -e \left(R - G + \frac{\partial p}{\partial t} \right) \quad (3.8)$$

gelöst. Sie beinhalten die Rekombinations- R und die Generationsrate G sowie die Änderung der Ladungsträgerdichte nach der Zeit.

Neben diesen Grundgleichungen bildet das Modell zum Transport der Ladungsträger einen weiteren zentralen Punkt der Bauelementesimulation. Als Grundlage hierfür dient die BOLTZMANNsche Transportgleichung, die aus der Thermodynamik hergeleitet ist und in verschiedenen Graden der Vereinfachung existiert [44].

Die gebräuchlichste und einfachste Form ist das Drift-Diffusions-Modell mit den Transportgleichungen

$$J_n = -e(n\mu_n \cdot \nabla\varphi - D_n \cdot \nabla n) \quad (3.9)$$

für Elektronen und

$$J_p = -e(p\mu_p \cdot \nabla\varphi + D_p \cdot \nabla p) \quad (3.10)$$

für Löcher, wobei sich die Stromdichten J_n und J_p aus zwei Anteilen zusammensetzen. Der Driftanteil richtet sich nach dem Gradienten des Potentials und somit dem elektrischen Feld, während der Diffusionsanteil durch den Gradienten der Ladungsträgerdichte bestimmt ist und einen Diffusionskoeffizienten

$$D_{n,p} = \mu_{n,p} \frac{k_B \cdot T_L}{e} \quad (3.11)$$

besitzt, ähnlich der Dotandendiffusion aus der Prozesssimulation. Der Diffusionskoeffizient resultiert aus der EINSTEIN-Relation und ist proportional zur Gittertemperatur T_L . Das Drift-Diffusions-Modell deckt jedoch nur räumlich schwach veränderliche elektrische Felder ab, da ein Gleichgewicht zwischen Ladungsträgertemperatur und der Temperatur des Kristallgitters angenommen wird, und berücksichtigt somit keine Quanteneffekte. Transistoren mit Kanallängen unter 100 nm bilden aufgrund der kleinen Struktur zum Teil starke elektrische Felder aus, bei denen das Drift-Diffusions-Modell an diese Grenze stößt [21]. Das vollständige hydrodynamische Transportmodell umfasst daher neben dem Ansatz der Drift-Diffusion eine Energie- sowie Impulserhaltungsgleichung und beinhaltet damit die Aufheizung der Ladungsträger [123].

Im Simulator ist jedoch nur eine vereinfachte Form des hydrodynamischen Transportmodells implementiert, bei dem die Gleichungen (3.9) und (3.10) des Drift-Diffusions-Modells um den Gradienten der Ladungsträgertemperaturen $T_{n,p}$ zu

$$J_n = -e \left(n\mu_n \cdot \nabla\varphi - D_n \cdot \nabla n - \mu_n n \frac{k_B}{e} \nabla T_n \right) \quad (3.12)$$

und

$$J_p = -e \left(p\mu_p \cdot \nabla\varphi + D_p \cdot \nabla p + \mu_p p \frac{k_B}{e} \nabla T_p \right) \quad (3.13)$$

erweitert worden sind. Die Ladungsträgertemperaturen der Elektronen und Löcher dienen der Energieerhaltung im Halbleiter und sind Bestandteile der Energiebilanzgleichungen

$$\nabla \cdot \mathbf{S}_n = \mathbf{J}_n \cdot \mathbf{E} - \frac{3}{2} k_B \cdot \frac{\partial(nT_n)}{\partial t} - \frac{3k_B}{2\tau_n^e} n(T_n - T_L) + \frac{3}{2} k_B T_n (R - G) \quad (3.14)$$

und

$$\nabla \cdot \mathbf{S}_p = \mathbf{J}_p \cdot \mathbf{E} - \frac{3}{2} k_B \cdot \frac{\partial(pT_p)}{\partial t} - \frac{3k_B}{2\tau_p^e} p(T_p - T_L) + \frac{3}{2} k_B T_p (R - G). \quad (3.15)$$

Sie beziehen alle Energieflüsse im Halbleiter mit ein, die durch Ladungsströmung, JOULEsche Wärme und Generation sowie Rekombination hervorgerufen werden [21]. Die verschiedenen Stoßprozesse der freibeweglichen Ladungsträger sind in der Energie-relaxationszeit $\tau_{n,p}^e$ zusammengefasst. Dem gegenüber stehen die Divergenzen der Energiestromdichten

$$\mathbf{S}_n = -\kappa_n^T \cdot \nabla T_n - \frac{5k_B T_n}{2e} \mathbf{J}_n \quad (3.16)$$

und

$$\mathbf{S}_p = -\kappa_p^T \cdot \nabla T_p + \frac{5k_B T_p}{2e} \mathbf{J}_p, \quad (3.17)$$

die ebenfalls die Ladungsträgertemperaturen $T_{n,p}$ beinhalten und sie mit den thermischen Leitfähigkeiten κ_n^T und κ_p^T verknüpfen.

Die genauesten Berechnungen eines Bauelements erzielen Monte-Carlo-Simulationen einzelner Ladungen auf der Grundlage der vollständigen BOLTZMANNschen Transportgleichung [124]. Diese Methode ist jedoch sehr zeitaufwendig und rechenintensiv und somit ungeeignet für die Simulationen komplexer Bauelemente.

3.2.2 Bandlückenverengung

Im Simulator sind die Bandkanten von Valenz- und Leitungsband für undotiertes Silizium definiert. Mit der Dotierung kommt es jedoch zu einer leichten Verschiebung der Bandkanten und somit zu einer Verringerung der Bandlücke zwischen ihnen [125]. Die Differenz der undotierten und der dotierungsabhängigen Bandlücke ergibt sich dabei zu

$$\Delta E_G = E_{ref} \left[\ln \left(\frac{N_{tot}}{N_{ref}} \right) + \sqrt{\left(\ln \left(\frac{N_{tot}}{N_{ref}} \right) \right)^2 + 0.5} \right]. \quad (3.18)$$

Hierbei sind N_{tot} die absolute Störstellendichte und E_{ref} und N_{ref} experimentell ermittelte Referenzwerte, die von verschiedenen Autoren mit unterschiedlichen Werten belegt wurden. Für die Simulationen dieser Arbeit wurde der Parametersatz nach [126] ausgewählt ($E_{ref} = 9 \cdot 10^{-3}$ eV, $N_{ref} = 1 \cdot 10^{17}$ cm⁻³).

3.2.3 Generation und Rekombination

Im Wesentlichen strömen die freibeweglichen Ladungsträger zwischen Source und Drain durch den Kanal und bilden so den Stromfluss durch den eingeschalteten Transistor. Dabei unterliegen sie jedoch einem ständigen Prozess der Generation und Rekombination mit anderen Ladungen.

A. SHOCKLEY-READ-HALL-Rekombination

Die SHOCKLEY-READ-HALL-Rekombination eines Elektronen-Loch-Paares erfolgt auf einem indirekten Weg zwischen Valenz- und Leitungsband über Rekombinationszentren im Bereich der Bandlücke [127]. Die Rekombinationsrate

$$R_{SRH} = \frac{pn - n_i^2}{\tau_n(p + n_i) + \tau_p(n + n_i)} \quad (3.19)$$

hängt von der Anzahl der Ladungsträger und ihrer Lebensdauer ab, die ebenfalls an das gegebene Dotierungsprofil gebunden ist. Die Dotierung steigert dabei die Anzahl an Defekten im Bereich der Bandlücke, die als Rekombinationszentren fungieren und so die Lebensdauer

$$\tau_{n,p}(N_{tot}) = \frac{\tau_{n,p}^0}{1 + \frac{N_{tot}}{N_{ref(n,p)}}} \quad (3.20)$$

der Ladungsträger reduzieren. $\tau_{n,p}^0$ repräsentiert die Grundlebensdauer, wie sie für undotiertes Silizium angenommen werden kann.

B. Trap-Assisted-Tunneling-Modell

Ebenso besteht eine erhöhte Anzahl an Defekten in der Nähe von Grenz- bzw. Oberflächen, über welche die freibeweglichen Ladungen rekombinieren können. Das SHOCKLEY-READ-HALL-Modell kann hierfür zusätzlich mit dem Trap-Assisted-Tunneling-Modell kombiniert werden. Dabei wird die Ladungsträgerlebensdauer aus Gleichung (3.20) durch hohe elektrische Feldstärken verringert [128]. Mit dem Tunneln der Ladungsträger in die Trapniveaus der verbotenen Zone verändert sich ihre Lebensdauer zu

$$\tau_{n,p} = \frac{\tau_{n,p}(N_{tot})}{1 + \Gamma_{n,p}(E)}, \quad (3.21)$$

wobei die feldstärkengebundene Komponente

$$\Gamma_{n,p}(E) = 2\sqrt{3}\pi \frac{E}{E_{n,p}^{TAT}} \exp\left(\frac{E}{E_{n,p}^{TAT}}\right)^2 \quad (3.22)$$

durch eine Referenzfeldstärke $E_{n,p}^{TAT}$ bestimmt ist.

C. AUGER-Rekombination

Die AUGER-Rekombination ist ein weiterer Rekombinationsmechanismus stark dotierter indirekter Halbleiter. Er beruht darauf, dass eines von zwei Elektronen oder Löchern bei einem Zusammenstoß in die Rekombinationszentren gelangt und sich dort mit einer entgegengesetzten Ladung vereint. Der Impulsüberschuss wird an das dritte Teilchen abgegeben, welches sich anschließend auf einem energetisch höheren Niveau befindet [6]. Die Rekombinationsrate lässt sich über die Beziehung

$$R_{AUG} = (C_n^{AUG} n + C_p^{AUG} p) \cdot (pn - n_i^2) \quad (3.23)$$

beschreiben, wobei C_n^{AUG} und C_p^{AUG} die Augerkoeffizienten darstellen [117].

D. Band-zu-Band-Tunneln

Die Generation von Ladungsträgern in MOSFETs ist im Wesentlichen durch zwei Mechanismen beschreibbar, dem Band-zu-Band-Tunneln und der Stoßionisation. Das direkte Band-zu-Band-Tunneln der Elektronen vom Valenz- in das Leitungsband wird durch eine sehr starke Verbiegung der Bänder an den pn-Übergängen ermöglicht. Hierbei handelt es sich um ein quantenmechanisches Phänomen, da die Potenzialbarriere für die Ladungsträger nach den Gesetzen der klassischen Mechanik nicht zu überwinden wäre. Die Wahrscheinlichkeit, dass es zum Tunneln eines Elektrons kommt, steigt mit der elektrischen Feldstärke und der Verringerung des Bandabstandes im Bereich des pn-Übergangs. Die Generationsrate bestimmt sich somit zu

$$G_{B2B} = A_{B2B} \cdot |\mathbf{E}|^P \cdot \exp\left(-B_{B2B} \frac{E_G^{3/2}}{|\mathbf{E}|}\right). \quad (3.24)$$

Die Modellparameter A_{B2B} , B_{B2B} und P wichten den Einfluss der elektrischen Feldstärke und sind zur Kalibrierung bezüglich experimenteller Werte gedacht [21].

E. Stoßionisation

Die Stoßionisation oder auch Avalanche-Generation (Impact Ionization) ist ein Generationsmechanismus, der umgekehrt zur AUGER-Rekombination abläuft. Eine Ladung mit einer hinreichend hohen kinetischen Energie, größer der Bandlücke, ist hierbei in der Lage, ein Bindungselektron des Kristallgitters herauszuschlagen und somit ein weiteres Elektronen-Loch-Paar zu erzeugen. Dazu ist jedoch eine hohe elektrische Feldstärke erforderlich, da sonst der Beschleunigungsweg der Ladungen zu kurz ist bzw. der abbremsende Stoß mit dem Gitter eine höhere Wahrscheinlichkeit besitzt und statistisch gesehen überwiegt. Die Generationsrate

$$G_{ii} = \alpha_n^{ii} \cdot n v_n + \alpha_p^{ii} \cdot p v_p \quad (3.25)$$

steigt mit der Anzahl der Ladungsträger und ihrer Geschwindigkeit $v_{n,p}$ sehr schnell an, bis es zu einem lawinenartigen Durchbruch (Avalanche-Effekt) kommt [129].

Die Ionisationsrate $\alpha_{n,p}^{ii}$ ist standardmäßig von der elektrischen Feldstärke abhängig, kann aber bei Verwendung des hydrodynamischen Transportmodells auch von der Ladungsträgertemperatur bestimmt werden. Bei Ultrakurzkanal-Transistoren ist der Bereich, in dem die Stoßionisation wirkt, durch die kurze Kanallänge stark begrenzt, sodass der Effekt trotz hoher Feldstärken gegenüber den Rekombinationsmechanismen an Einfluss verliert.

3.2.4 Ladungsträgerbeweglichkeit

A. Grundbeweglichkeiten

Die Grundbeweglichkeiten der Elektronen und Löcher in reinem Silizium betragen $\mu_n^0 = 1400 \text{ cm}^2/\text{Vs}$ bzw. $\mu_p^0 = 450 \text{ cm}^2/\text{Vs}$. Sie werden im Siliziumkristall und speziell im Kanalgebiet des MOSFETs durch verschiedene Streumechanismen wie Gitterstöße, fixe Ladungen und Grenzflächenrauheit sowie der Hochfeldsättigung verringert. Dabei kann jedem Mechanismus ein separates Modell zugewiesen werden, da sie unabhängig voneinander agieren. Jedes Modell berechnet somit eine reduzierte Beweglichkeit, die nach der MATTHIESEN-Regel

$$\frac{1}{\mu_{res}} = \sum_i \frac{1}{\mu_i} \quad (3.26)$$

zu einer resultierende Beweglichkeit zusammengefasst werden [117].

B. Dotierungsabhängigkeit

Das in den Simulationen angewandte Philips-Unified-Beweglichkeitsmodell nach KLAASSEN [130] umfasst zwei Abhängigkeiten der Ladungsträgerbeweglichkeit. Zum einen wird die Dotierungsabhängigkeit erfasst [131]. Hierbei wird über die Gleichung

$$\mu_{n,p}^{Dop} = \mu_{n,p}^{min} + \frac{\mu_{n,p}^{max} - \mu_{n,p}^{min}}{1 + (N_{tot}/N_{n,p}^{ref})^{\alpha_{n,p}^{Dop}}} \quad (3.27)$$

der Übergang zwischen einer maximalen und einer minimalen Beweglichkeit nachgebildet. Die maximale Ladungsträgerbeweglichkeit $\mu_{n,p}^{max}$ unterliegt der Annahme eines hochreinen Siliziumkristalls, der den Elektronen und Löchern die anfangs beschriebene Grundbeweglichkeit zur Verfügung stellt. Die minimale Ladungsträgerbeweglichkeit $\mu_{n,p}^{min}$ resultiert aus Beweglichkeitsmessungen der Ladungsträger in Silizium mit einer Dotierungskonzentration, die an die jeweilige Festkörperlöslichkeit der Dotanden heranreicht. $N_{n,p}^{ref}$ und $\alpha_{n,p}^{Dop}$ sind Modellparameter zur Kalibrierung des Übergangsbereichs.

C. Temperaturabhängigkeit

Der zweite Sachverhalt, den das Philips-Unified-Beweglichkeitsmodell beinhaltet, ist die Temperaturabhängigkeit durch das Aufheizen des Kristalls. Dies verursacht Gitterschwingungen, akustische Phononen, die die Wahrscheinlichkeit des Abbremsens der Ladungsträger durch Stöße mit Atomrümpfen erhöht. Daraus resultiert eine Beweglichkeit von

$$\mu_{n,p}^{LT} = \mu_{n,p}^{max} \left(\frac{T_L}{300 \text{ K}} \right)^{-\alpha_{n,p}^{LT}}. \quad (3.28)$$

$\mu_{n,p}^{max}$ repräsentiert die Grundbeweglichkeit der jeweiligen Ladungsträger bei einer Temperatur von 300 K. $\alpha_{n,p}^{LT}$ ist ein Modellparameter, der zur Anpassung an experimentelle Daten genutzt wird. Er kann von Struktur zu Struktur variieren, da die Modellierung der Wärme und die Ausbildung von heißen Punkten im Bauteil nur bedingt vom Simulator nachgestellt werden kann und der Einfachheit halber oft als konstant im gesamten Transistor angenommen wird.

D. Hochfeldsättigung

Die Hochfeldsättigung der Ladungsträgerdriftgeschwindigkeit ist ein weiterer begrenzender Mechanismus der Beweglichkeit von Elektronen und Löchern. Für hohe Feldstärken verhält sich die Driftgeschwindigkeit nicht mehr proportional zum parallel verlaufenden elektrischen Feld. Die Ladungsträger unterliegen einer begrenzenden Sättigungsgeschwindigkeit v_{sat} , sodass die resultierende Beweglichkeit

$$\mu \approx \frac{|v_{sat}|}{|E_{hoch}|} \quad (3.29)$$

mit zunehmender Feldstärke abnimmt. Im Simulator wird die Beweglichkeitskomponente der Hochfeldsättigung im Allgemeinen über

$$\mu_{n,p}^{HS} = \frac{\mu_{n,p}^{LF}}{\left[1 + \left(\frac{\mu_{n,p}^{LF} \cdot E_{||}}{v_{n,p}^{sat}} \right)^{\beta_{n,p}^{HS}} \right]^{1/\beta_{n,p}^{HS}}} \quad (3.30)$$

berechnet, wobei $\mu_{n,p}^{LF}$ die Ladungsträgerbeweglichkeit bei geringer Feldstärke und $\beta_{n,p}^{HS}$ ein Modellparameter zur Anpassung ist [21]. Unter Verwendung des hydrodynamischen Transportmodells kann das parallele elektrische Feld $E_{||}$ durch ein effektives elektrisches Feld

$$E_{n,p}^{eff} = \frac{3k_B(T_{n,p} - T_L)}{2e \cdot v_{n,p}^{sat} \cdot \tau_{n,p}^0} \quad (3.31)$$

ersetzt werden, welches sich aus der Ladungsträgertemperatur $T_{n,p}$ und der Energielaxationszeit $\tau_{n,p}^0$ der Elektronen bzw. Löcher berechnet [132].

E. Ladungsträgerbeweglichkeit an Grenz- und Oberflächen (LOMBARDI-Modell)

Die Degradation der Ladungsträgerbeweglichkeit an Grenz- und Oberflächen ist im Vergleich zu den anderen Beweglichkeitsmodellen deutlich komplexer. Es beruht auf der Streuung der Ladungsträger, verursacht durch die Rauheit (Surface Roughness, SR) und erhöhte Gitterschwingungen (Acoustic Phonon, AP) an Grenz- und Oberflächen, wie sie zwischen Kanalgebiet und Gateoxid vorkommen [117], [133]. Dabei spielen das Gateoxidmaterial und die Oberflächenorientierung eine entscheidende Rolle [134]. Der Beweglichkeitsanteil der grenznahen Gitterschwingung

$$\mu_{AP} = \frac{B}{E_{\perp}} + \frac{C((N_A + N_D)/N_0)^{\lambda}}{E_{\perp}^{1/3}(T/300 \text{ K})^k} \quad (3.32)$$

ist maßgeblich durch die orthogonale Feldstärke E_{\perp} und die Anzahl der Dotanden an der Grenzfläche bestimmt und kommt somit erst bei starker Inversion zum Tragen. Die Parameter B , C , N_0 sowie die Exponenten λ und k dienen der Kalibrierung des Modells. Der Anteil der Oberflächenrauheit lässt sich über

$$\mu_{SR} = \left(\frac{(E_{\perp}/E_{ref})^{A^*}}{\delta} + \frac{E_{\perp}^3}{\eta} \right)^{-1} \quad (3.33)$$

beschreiben und unterliegt ebenfalls der orthogonalen Feldstärke. Der Exponent

$$A^* = A + \frac{(\alpha_n n + \alpha_p p) N_{ref}^{\vartheta}}{(N_A + N_D + N_1)^{\vartheta}} \quad (3.34)$$

ist zudem von den Ladungsträgerdichten und der Dotierung abhängig. Die Parameter A , $\alpha_{n,p}$, δ und η sowie N_1 und N_{ref} dienen der Anpassung des Modells an experimentelle Daten. Ein vollständiger Parametersatz für beide Mechanismen ist in [133] und [135] veröffentlicht.

Die erweiterte Form des Lombardi-Beweglichkeitsmodells umfasst zusätzlich den Einfluss der fernen Ladungsstreuung (Remote Coulomb Scattering, RCS) und der fernen Gitterschwingungsstreuung (Remote Phonon Scattering, RPS). Diese beiden Modelle kommen speziell bei High-k-Gateoxiden zum Tragen, da sie im Vergleich zum herkömmlichen Siliziumoxid eine rauere Grenzfläche aufweisen und verstärkt Traps ausbilden bzw. prozessbedingt feste Ladungen aufnehmen [136], [137]. Im Simulator ist der Beweglichkeitsanteil für ferne Ladungsstreuung mit

$$\mu_{RCS} = \frac{\mu_{RCS}^0}{f(E_{\perp})} \left(\frac{N_{A,D}}{3 \cdot 10^{16} \text{ cm}^{-3}} \right)^{\gamma_1} \left(\frac{T}{300 \text{ K}} \right)^{\gamma_2} \left(\frac{N_S}{10^{11} \text{ cm}^{-2}} \right)^{\gamma_3 + \gamma_4 \cdot \ln\left(\frac{N_{A,D}}{3 \cdot 10^{16} \text{ cm}^{-3}}\right)} \quad (3.35)$$

gegeben, wobei N_S die Inversionsladungsträgerdichte darstellt und mit

$$f(E_{\perp}) = 1 - \exp\left(\frac{-E_{\perp}}{N_{depl}}\right) \quad (3.36)$$

eine Abhängigkeit von der orthogonalen Feldstärke E_{\perp} zur Grenzfläche sowie von der Dichte der festen Ladungen des Verarmungsgebiets N_{depl} implementiert ist. Das Modell zur fernen Gitterschwingungsstreuung

$$\mu_{RPS} = \mu_{RPS}^0 \left(\frac{E_{\perp}}{10^6 \text{ V/cm}}\right)^{\gamma_5} \left(\frac{T}{300 \text{ K}}\right)^{\gamma_6} \quad (3.37)$$

gestaltet sich im Vergleich etwas einfacher und ist hauptsächlich durch das elektrische Feld senkrecht zur Grenzfläche bestimmt. Beide Modelle können über die Grundbeweglichkeiten μ_{RCS}^0 und μ_{RPS}^0 bzw. die Exponenten γ_1 bis γ_6 angepasst werden. Mit der Definition einer der Exponenten zu Null kann der Einfluss der jeweiligen Größe ausgeschlossen werden.

Alle Komponenten des erweiterten LOMBARDI-Beweglichkeitsmodells werden über die MATTHIESEN-Regel der Form

$$\frac{1}{\mu_{Lomb}} = \alpha_{Lomb} \left(\frac{1}{\mu_0} + \frac{D_{Lomb}}{\mu_{AP}} + \frac{D_{Lomb}}{\mu_{SR}}\right) + \alpha_{RCS} \frac{D_{RCS}}{\mu_{RCS}} + \alpha_{RPS} \frac{D_{RPS}}{\mu_{RPS}} \quad (3.38)$$

verknüpft. Die einzelnen Anteile werden dabei nochmal mit einem Wichtungsfaktor α multipliziert und besitzen einen Dämpfungsfaktor

$$D_i = \exp(-dist/l_i^{crit}), \quad (3.39)$$

der den Einfluss auf die Beweglichkeit des jeweiligen Modells in Relation zum Abstand bezüglich der Grenzfläche ($dist$) setzt und den Referenzwert l_{RCS}^{crit} bzw. l_{RPS}^{crit} beinhaltet.

3.2.5 Effekte der mechanischen Verspannung

Wie bereits in Abschnitt 2.1.3 ausgeführt wurde, äußert sich die mechanische Verspannung σ des Siliziumkristalls durch elastische und plastische Verformung ε des Materials und verursacht eine Verschiebung bzw. Verbiegung von den Valenz- und Leitungsbandern. Die Verspannungen und Deformationen

$$\sigma_j = \begin{bmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \\ \sigma_{12} \\ \sigma_{13} \\ \sigma_{23} \end{bmatrix} \text{ bzw. } \varepsilon_i = \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{22} \\ \varepsilon_{33} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{23} \end{bmatrix} \quad (3.40)$$

werden dabei in die drei uniaxialen Komponenten der Hauptkristallrichtungen und in drei Scherkomponenten unterschieden.

In der Regel führen plastische Deformationen zu Gitterfehlern bzw. resultieren bei hinreichender Stärke im Bruch des Wafers. Der elastische Anteil der Deformation kann über das HOOKEsche Gesetz

$$\varepsilon_i = \sum_{j=1}^6 S_{ij} \sigma_j \quad (3.41)$$

direkt aus der eingprägten Verspannung berechnet werden. Die Elastizitätskonstanten des Tensors

$$S_{ij} = \begin{bmatrix} S_{11} & S_{12} & S_{12} & 0 & 0 & 0 \\ S_{12} & S_{11} & S_{12} & 0 & 0 & 0 \\ S_{12} & S_{12} & S_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & S_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & S_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & S_{44} \end{bmatrix} \quad (3.42)$$

bestimmen sich nach [138] zu $S_{11} = 0.768 \cdot 10^{-11} \text{ Pa}^{-1}$, $S_{12} = -0.214 \cdot 10^{-11} \text{ Pa}^{-1}$ und $S_{44} = 1.26 \cdot 10^{-11} \text{ Pa}^{-1}$.

A. Deformationspotenzialmethode

Die deformationsbedingte Verschiebung der einzelnen Leitungsbandkanten der sechs Orbitale berechnet der Simulator über die Deformationspotenzialmethode zu

$$\Delta E_{C,i} = \Xi_d(\varepsilon_{11} + \varepsilon_{22} + \varepsilon_{33}) + \Xi_u \varepsilon_{ii} . \quad (3.43)$$

Hierbei ist $\Xi_d = 8.6 \text{ eV}$ die hydrostatische und $\Xi_u = -9.5 \text{ eV}$ die uniaxiale Deformationsenergie [15].

Die Verschiebung des leichten und schweren Löcherbandes berechnet sich mit

$$\Delta E_{V,i} = -a(\varepsilon_{11} + \varepsilon_{22} + \varepsilon_{33}) \pm \delta E \quad (3.44)$$

und

$$\delta E = \sqrt{\frac{b^2}{2}((\varepsilon_{11} - \varepsilon_{22})^2 + (\varepsilon_{22} - \varepsilon_{33})^2 + (\varepsilon_{11} - \varepsilon_{33})^2) + d^2(\varepsilon_{12}^2 + \varepsilon_{13}^2 + \varepsilon_{23}^2)} \quad (3.45)$$

aufgrund ihrer energetischen Entartung etwas komplexer. Die Deformationsenergien wurden nach [139] mit $a = 2.1 \text{ eV}$, $b = -0.5 \text{ eV}$ und $d = -4.0 \text{ eV}$ angenommen.

Da der Simulator jedoch nur einen Wert für alle Subbänder berücksichtigen kann, werden die Verschiebungen der einzelnen Bandkanten gewichtet nach ihrer Besetzung zu einem Mittelwert

$$\overline{\Delta E}_C = -k_B T_L \cdot \ln \left[\frac{1}{3} \sum_{i=1}^3 \exp \left(\frac{-\Delta E_{C,i}}{k_B T_L} \right) \right] \quad (3.46)$$

bzw.

$$\overline{\Delta E}_V = k_B T_L \cdot \ln \left[\frac{1}{2} \sum_{i=1}^2 \exp \left(\frac{\Delta E_{V,i}}{k_B T_L} \right) \right] \quad (3.47)$$

zusammengefasst.

B. Elektronenbeweglichkeit unter mechanischer Verspannung

Die Modellierung der Elektronenbeweglichkeit unter mechanischer Verspannung basiert auf der Umbesetzung der sechs Leitungsbandtaler und der daraus veranderten Leitfahigkeitsmasse [117]. Die deformationsabhangige Elektronenbeweglichkeit

$$\mu_n^i = \mu_n^0 \left[1 + \frac{1 - m_i^*/m_t^*}{1 + 2(m_i^*/m_t^*)} \left(\exp \left(\frac{\overline{\Delta E}_C - \Delta E_{C,i}}{k_B T_L} \right) - 1 \right) \right], \quad i = x, y, z \quad (3.48)$$

der Hauptkristallrichtungen beinhaltet das erwahnte Deformationspotenzial und die longitudinalen und transversalen Anteile der effektiven Massen, welche als unabhangig von der mechanischen Verspannung angenommen werden konnen. Zusatzlich kann die veranderte Ladungstragerstreuung und eine veranderte effektive Masse durch eine Scherverspannung berucksichtigt werden, wobei dieser Einfluss jedoch vernachlassigbar gering ist [140].

C. Locherbeweglichkeit unter mechanischer Verspannung

Die verspannungsabhangige Locherbeweglichkeit gestaltet sich aufgrund der ungleichmaigen Form (Abbildung 2.19) der beiden Locherbander und der daraus resultierenden Anisotropie der Locherbeweglichkeit sehr komplex. Zunachst kann jedoch die Vereinfachung getroffen werden, dass sich der Hauptteil der Locher (ca. 75 %) im schweren Locherband aufhalt und somit das leichte Locher- und das Spin-Orbit-Band bei der Modellierung vernachlassigt werden konnen [117]. Die Form des schweren Locherbandes ist durch sechs sich berschneidende Ellipsen beschreibbar, von denen je zwei mit ihren Hauptachsen in einer Ebene liegen (Abbildung 3.4).

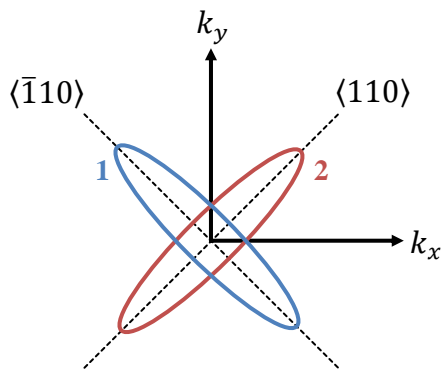


Abbildung 3.4: Schematische Darstellung der beiden gedachten Ellipsen mit ihren Hauptachsen in der xy -Ebene.

Die Grundbeweglichkeit

$$\mu_p^0 = e\tau_s \left[\frac{0.5}{m_{p,t}^0} + \frac{0.5}{m_{p,l}^0} \right] \quad (3.49)$$

der Löcher in beiden Ellipsen dieser Ebene richtet sich im unverspannten Zustand zu gleichen Teilen nach ihrer longitudinalen $m_{p,l}^0$ und ihrer transversalen Masse $m_{p,t}^0$ sowie einer Streuungsrate τ_s . Die simulierten Werte der Verspannungskomponenten in einer Ebene

$$\sigma_{ij} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} b_p + a_p & s_p \\ s_p & b_p - a_p \end{bmatrix} \quad (3.50)$$

unterteilen sich in einen biaxialen Anteil b_p , einen uniaxial asymmetrischen Anteil a_p und eine Scherkomponente s_p . Mit ihrer Hilfe werden der Energieunterschied

$$\Delta E_{hh} = d_1 s_p \quad (3.51)$$

und die veränderten transversalen Massen

$$\frac{1}{m_{p,t}^i} = \frac{1}{m_{p,t}^0} \left(1 \pm s_{t1} s_p + s_{t2} s_p^2 + b_{t1} b_p + b_{t2} b_p^2 \right) \quad (3.52)$$

der beiden Ellipsen bestimmt, die sich durch die Verschiebung bzw. Verbiegung des schweren Löcherbandes aufgrund der mechanischen Verspannung ergeben. Die Faktoren d_1 , s_{t1} , s_{t2} , b_{t1} und b_{t2} dienen zur Anpassung des Modells. Die longitudinale Masse gilt als verspannungsunabhängig und bleibt somit unverändert. Damit berechnen sich die veränderten Beweglichkeiten für die Ellipsen zu

$$\begin{bmatrix} \Delta\mu_p^{(110)} \\ \Delta\mu_p^{(\bar{1}10)} \\ \Delta\mu_p^{(001)} \end{bmatrix} = \mu_p^0 \begin{bmatrix} \left(\frac{f_1}{m_{p,t}^1} + \frac{f_2}{m_{p,l}^0} \right) - 1 & 0 & 0 \\ \frac{0.5}{m_{p,t}^0} + \frac{0.5}{m_{p,l}^0} & \left(\frac{f_1}{m_{p,l}^0} + \frac{f_2}{m_{p,t}^2} \right) - 1 & 0 \\ 0 & \frac{0.5}{m_{p,t}^0} + \frac{0.5}{m_{p,l}^0} & \frac{m_{p,t}^0}{m_{p,t}^{(001)}} - 1 \end{bmatrix}, \quad (3.53)$$

wobei

$$f_1 = \frac{1}{1 + \exp(-\Delta E_{hh}/k_B T_L)} \quad (3.54)$$

und

$$f_2 = 1 - f_1 \quad (3.55)$$

die Besetzungswahrscheinlichkeiten der Ladungsträger unter den beiden Ellipsen beschreiben. Die Beweglichkeitskomponente $\Delta\mu_p^{(001)}$ steht orthogonal zur Ebene und wird nur bei dreidimensionalen Simulationen berücksichtigt. Sie ist für beide Ellipsen gleich und berechnet sich aus der transversalen Masse in $\langle 001 \rangle$ -Richtung:

$$\frac{1}{m_{pt}^{(001)}} = \frac{1}{m_{p,t}^0} (1 + b_{tt}b_p). \quad (3.56)$$

3.2.6 Ladungsträgerquantisierung

Von den in Abschnitt 2.2.3 angesprochenen Quanteneffekten kann der Simulator nur die Korrektur der Ladungsträgerdichte im Bereich der Grenz- und Oberflächen berücksichtigen. Potenzialmulden in der Nähe von Grenzflächen sorgen dafür, dass sich die Ladungsträger verstärkt in diesem Bereich sammeln. Innerhalb der Potenzialmulde sind jedoch nur diskrete Zustände (Subbänder) der Ladungsträger erlaubt, wodurch es zu einer veränderten Ladungsträgerdichte an Grenzflächen kommt. Zur Modellierung dieses Verhaltens wird das Quantenkorrekturpotenzial von Elektronen

$$\lambda_n^{QP} = \frac{\gamma_n \cdot \hbar^2}{6 \cdot e \cdot m_{q,n}^*} \cdot \frac{\Delta\sqrt{n}}{\sqrt{n}} \quad (3.57)$$

und Löchern

$$\lambda_p^{QP} = \frac{-\gamma_p \cdot \hbar^2}{6 \cdot e \cdot m_{q,p}^*} \cdot \frac{\Delta\sqrt{p}}{\sqrt{p}}, \quad (3.58)$$

hergeleitet aus der SCHRÖDINGER-Gleichung [141], in das Gleichungssystem integriert. Es ist abhängig von der effektiven Quantisierungsmasse $m_{q,n}^*$ bzw. $m_{q,p}^*$ der Ladungsträger und kann über die Wirkungsfaktoren γ_n und γ_p in seiner Intensität an experimentelle Daten angepasst werden. Die berechneten Quantenkorrekturpotenziale werden dann in die Transportgleichung des Drift-Diffusions- bzw. des hydrodynamischen Transportmodells und ebenso in die Energiestromdichtegleichung sowie die Energiebilanzgleichung integriert [21].

3.3 Kalibrierung der Modellparameter

3.3.1 Prozessparameter

Bevor mit der Simulation neuer Transistorstrukturen begonnen werden kann, ist es notwendig, die Modellparameter der Prozess- und Bauelementesimulation anhand experimentell gewonnener Daten zu kalibrieren. Aufgrund umfangreich vorliegender Messergebnisse und einer starken Ähnlichkeit zu FD-SOI-Transistoren bot sich hierfür

die 32 nm-SOI-Technologie von GLOBALFOUNDRIES an [29]. Weiterhin ist das Kanalgebiet, ähnlich der meisten Multi-Gate-Strukturen, nur geringfügig durch das Substrat- oder ein Bulkpotenzial beeinflusst.

Der erste Schritt der Prozesssimulation ist die geometrische Nachbildung von n- und p-MOSFETs mithilfe der Prozesse des Schichtauf- bzw. Schichtabtrags nach Vorlage von Transmissionselektronenmikroskopien (TEM, Abbildung 3.5). Die genauen Abmessungen, die für die Simulation verwendet wurden, sind in Tabelle 3.1 aufgelistet. Zu den Besonderheiten der 32 nm-SOI-Technologie gehören ein High-k-Oxid, eingebettetes Silizium-Germanium in den Source-Drain-Gebieten sowie im Kanal des p-MOSFET, das Entfernen des Spacermaterials vor der Abscheidung der verspannten Deckschichten und der Gate-First-Ansatz.

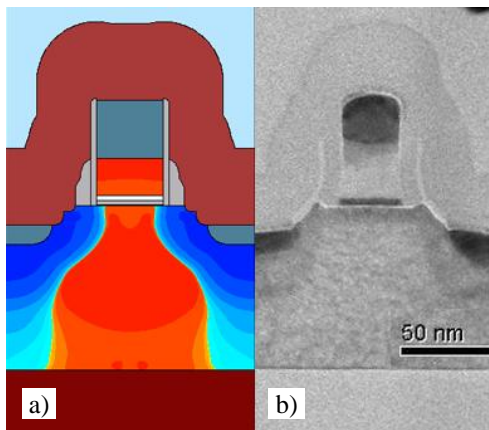


Abbildung 3.5: Querschnitt eines p-MOSFETs der 32 nm-SOI-Technologie mit verspannter Deckschicht. a) Simulation, b) TEM-Bild

Die Kalibrierung der Implantations- und Diffusionsmodelle von Dotanden ist eng an die Geometrie des jeweiligen Transistors gebunden. Da das kalibrierte Simulationsprojekt jedoch als Ausgangsbasis für nachfolgende Multi-Gate-Strukturen mit verschiedenen Geometrien dienen sollte, sind an dieser Stelle die Standardparameter beibehalten worden. Die Anpassung der Simulationen an die experimentellen Daten musste somit zum Großteil über die Modellparameter der Bauelementesimulation realisiert werden.

Tabelle 3.1: Geometrieparameter für MOSFETs der 32 nm-SOI-Technologie.

Kenngröße	Formelzeichen	Wert n-/p-MOSFET (nm)
BOX-Dicke	t_{BOX}	145
SOI-Filmdicke	t_{SOI}	80
Pitch	L_{Pitch}	130
Gatehöhe	h_{Gate}	60
Nominelle Gatelänge	L_{Gate}	40
Physikalische Gateoxiddicke (SiO ₂)	t_{ox}	1.0 / 1.3
Spacerbreite (Erweiterungsgebiete)	$B_{Sp,Erw}$	11 / 13
Spacerbreite (tiefe SD-Gebiete)	$B_{Sp,SD}$	6
Schichtdicke des TOL / COL	d_{OL}	34 / 55

Da die TEM-Bilder lediglich den Querschnitt eines Transistors an einer bestimmten Stelle zeigen und die geometrischen Abmessungen über die Länge bzw. Weite schwanken können, erwies es sich als erforderlich, die Gateoxiddicke und die Spacerbreite anhand der Gate-Source-Kapazität bzw. des Schwellspannungsverhaltens anzupassen (Abbildung 3.6). Im Vergleich zu den normierten Kapazitätsmessungen aus dem Experiment besteht ein konstanter Unterschied (Offset) von ca. $180 \text{ aF}/\mu\text{m}$. Dieser Unterschied ist systematisch bedingt, da in der Simulation parasitäre Kapazitäten, wie z. B. die Kontaktkapazität, nicht berücksichtigt werden. Für den n-MOSFET wurde somit eine äquivalente Gateoxiddicke von $EOT_n = 1.05 \text{ nm}$ bestimmt und für die nachfolgenden Bauelementesimulationen übernommen. Analog dazu ergab sich für den p-MOSFET eine äquivalente Gateoxiddicke von $EOT_p = 1.30 \text{ nm}$.

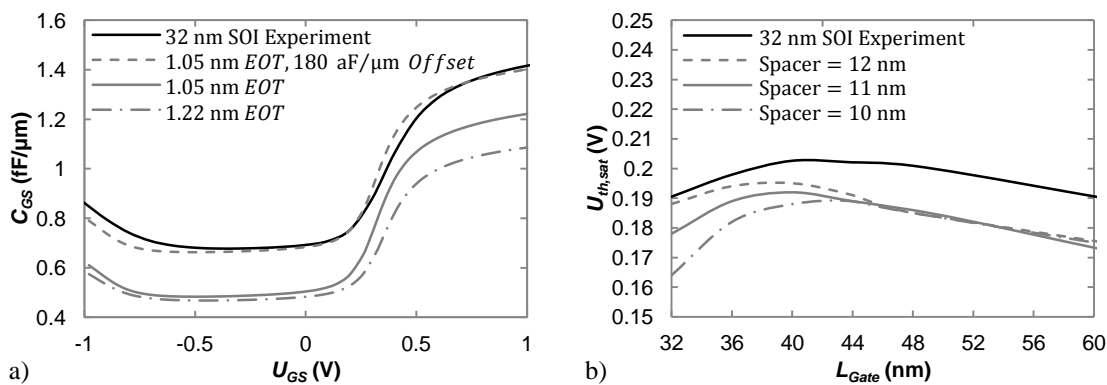


Abbildung 3.6: Vergleich der simulierten Werte eines n-MOSFETs der 32 nm-SOI-Technologie mit den experimentell ermittelten Daten für a) die Gate-Source-Kapazität mit unterschiedlichen äquivalenten Oxiddicken und b) das Verhalten der simulierten Schwellspannungen, bezogen auf die Gatelänge für verschiedene Breiten des Spacers (Erweiterungsgebiete).

Der Spacer bestimmt mit seiner Breite, wie nah die Implantationen der Erweiterungsgebiete am Kanalgebiet liegen. Je kleiner er wird, desto größer wird die Überlappung von Gate und Source-Drain-Erweiterung. Um die Beschaffenheit des Spacers auf die elektrischen Daten abzustimmen, wurde der Kurzkanaleffekt betrachtet, da er für dünne Spacer deutlich ausgeprägter erscheint als für dickere. Der Vergleich mit den experimentell gewonnenen Daten zeigt, dass der Spacer des n-MOSFETs eine minimale Breite von 11 nm nicht unterschreiten darf, da die Schwellspannungen für die Transistoren mit 32 nm und 36 nm Gatelänge sonst zu weit absinken. Auf der Seite des p-MOSFET ergaben die Simulationen mit einer absoluten Spacerbreite von 13 nm eine gute Übereinstimmung mit dessen Schwellspannungsverhalten.

3.3.2 Modellparameter

Anschließend an die Prozesssimulation erfolgte die Anpassung einiger Modellparameter an die experimentellen Daten von n- und p-MOSFETs (Abbildung 3.7). Zusätzlich zu den Grundgleichungen der Bauelementesimulation für den Transport der Ladungsträger und ihrer Dichte sind je nach Transistortyp die in Tabelle 3.2 aufgeführten Modellannahmen getroffen worden. Um Rechenzeit zu sparen und Konvergenzprobleme zu vermeiden, wurden letztere meist nur für die Majoritätsladungsträger verwendet.

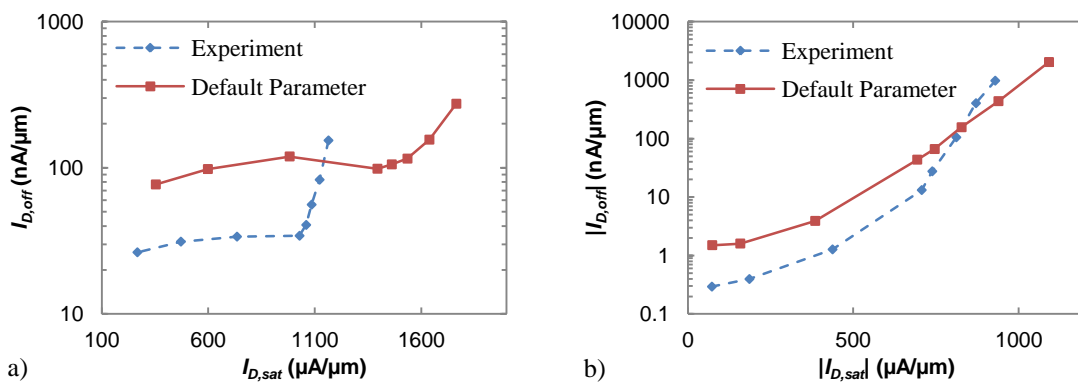


Abbildung 3.7: Simulierte Universalkennlinie von a) n- und b) p-MOSFETs mit den Default-Modellparametern (Standardwerte) im Vergleich zu den experimentellen Daten.

A. Austrittsarbeit

Die Austrittsarbeit des Gatemetalls besitzt einen wesentlichen Einfluss auf das Einsetzen der Kanalinversion sowie der daraus resultierenden Schwellspannung und ist abhängig von der Qualität des Schichtübergangs sowie der Körnung des Metalls. Somit unterliegt die Austrittsarbeit einer statistischen Schwankung und variiert für den n-MOSFET im Bereich von 4.0 ... 4.2 eV und für den p-MOSFET im Bereich von 4.6 ... 4.8 eV. Anhand der Schwellspannung kann jedoch keine exakte Anpassung der Austrittsarbeit für einen simulierten Transistor vorgenommen werden, da sie über die Konstantstrommethode ermittelt wird und bei leichter Inversion auch von der Beweglichkeit der Ladungsträger abhängt. Eine geeignete Lösung bietet die Simulation der Gate-Source-Kapazität. In diesem Fall spielen die Transport- und Beweglichkeitsmodelle eine untergeordnete Rolle und können mit ihren voreingestellten Parametern verwendet werden. Für die n- und p-MOSFETs wurden somit Austrittsarbeiten von $W_{A,n} = 4.1$ eV bzw. $W_{A,p} = 5.1$ eV ermittelt.

Tabelle 3.2: Modellannahmen für die Bauelementesimulation von n- und p-MOSFETs [117].

Modellkategorie	n-MOSFET	p-MOSFET
Ladungsträgertransport	Hydrodynamisches Transportmodell	
	<ul style="list-style-type: none"> mit Diskretisierung der Elektronentemperatur 	<ul style="list-style-type: none"> mit Diskretisierung der Löchertemperatur
Bandlückenverengung	Old-SLOTBOOM	
Generation und Rekombination	SHOCKLEY–READ–HALL-Rekombination <ul style="list-style-type: none"> Dotierungsabhängige Ladungsträgerlebensdauer Trap-Tunneling unterstützt <ul style="list-style-type: none"> Lokale Zustandsdichtenmodifizierung Feldabhängige Ladungsträgerlebensdauer 	
	Band-zu-Band-Tunneln (SCHENK-Modell mit lokaler Zustandsdichtenkorrektur)	
Ladungsträgerbeweglichkeit	Philips-Einheits-Beweglichkeitsmodell	
	Feldstärkenabhängige Beweglichkeitsmodelle <ul style="list-style-type: none"> Hochfeldsättigung (ladungstemperaturgesteuert) Ladungsträgerstreuung an Grenzflächen (LOMBARDI-Modell mit High-k-Korrektur) 	
Effekte der mechanischen Verspannung	Deformationspotenzial (kp-Methode) mit Mittelwertbildung der Bandkanten	
	Verspannungsabhängige Elektronenbeweglichkeit mit Dotierungsabhängigkeit	Verspannungsabhängige Löcherbeweglichkeit
Ladungsträgerquantisierung	Quantenpotenzialmodell ausschließlich für	
	<ul style="list-style-type: none"> Elektronen 	<ul style="list-style-type: none"> Löcher

B. Energierelaxationszeiten

Die Energierelaxationszeiten $\tau_{n,p}^e$ fassen die verschiedenen Stoßprozesse zusammen und sind Bestandteil der Energiebilanzgleichungen (3.14) und (3.15). Sie sind nur geringfügig von der Ladungsträgertemperatur abhängig und können somit als konstant angenommen werden [123]. Je größer die Energierelaxationszeit definiert wird, umso geringer werden die Verluste durch Stoßprozesse berücksichtigt. Somit kann bei gleicher elektrischer Feldstärke ein größerer Strom durch den Transistor fließen. Für den n-MOSFET musste die Energierelaxationszeit auf $\tau_n^e = 0.2$ ps reduziert werden, um eine akzeptable Übereinstimmung der Bauelementesimulation mit den elektrischen Daten aus dem Experiment zu erzielen. Im Gegensatz dazu war für den p-MOSFET eine leichte Erhöhung des Parameters auf $\tau_p^e = 0.4$ ps notwendig.

C. Band-zu-Band-Tunnel-Modell

Das Band-zu-Band-Tunnel-Modell (Band to Band, B2B) sorgt für eine zusätzliche Generationsrate, verursacht durch Bandverbiegungen wie sie an den pn-Übergängen und dem Oxid-Halbleiter-Übergängen der MOSFETs auftreten. Deutlich wird dieser Effekt vor allem durch erhöhte Sperrströme bei Transistoren mit großen elektrischen Feldstärken. Die Sättigungsströme sind in diesem Fall jedoch nicht betroffen und weisen nur eine geringe Abhängigkeit auf.

Das Modell besitzt zwei Parameter A_{B2B} und B_{B2B} , die zur Anpassung genutzt werden können. Beide sind Vorfaktoren, wobei A_{B2B} die gesamte Generationsrate des Band-zu-Band-Tunnelns und B_{B2B} die kritische elektrische Feldstärke beeinflusst. Um die erhöhten Sperrströme des n-MOSFETs zu verringern, ist der Wert von $B_{B2B} = 2.147 \cdot 10^{-7} \frac{\text{V}}{\text{cm}} \text{eV}^{-3/2}$ auf $B_{B2B} = 2.5 \cdot 10^{-7} \frac{\text{V}}{\text{cm}} \text{eV}^{-3/2}$ vergrößert worden. Für die Simulation des p-MOSFET ist das Band-zu-Band-Modell nicht verwendet worden, da es die Distanz zwischen linearer und Sättigungsschwellspannung (*DIBL*) deutlich erhöht hat und die Simulationen mit dem Modell stark von den experimentellen Daten abgewichen sind.

D. Trap-Assisted-Tunneling-Modell

Das Trap-Assisted-Tunneling-Modell (TAT) beschreibt im Gegensatz zum Band-zu-Band-Tunnel-Modell einen Rekombinationseffekt mit einer Abhängigkeit von der elektrischen Feldstärke. Im Simulator wurde für dieses Modell mithilfe des Parameters E_{min}^{TAT} eine Anpassung (Glättung) für niedrige Feldstärken integriert. Die Voreinstellung von SYNOPSIS für diesen Wert beträgt $E_{min}^{TAT} = 0 \text{ V/cm}$. Je größer dieser Grenzwert wird, um so geringer ist die Rekombinationsrate im Bereich der pn-Übergänge, bis eine schwache Inversion des Kanals einsetzt. Für den p-MOSFET ist die Mindestfeldstärke auf $E_{min}^{TAT} = 5 \cdot 10^5 \text{ V/cm}$ erhöht worden. Der n-MOSFET bedurfte keiner weiteren Anpassung.

E. LOMBARDI-Modell

Das einfache LOMBARDI-Modell betrachtet mit Gitterschwingungen und Oberflächenrauheit zwei Komponenten, die zur Korrektur der Ladungsträgerbeweglichkeit beitragen. Der Beweglichkeitsanteil μ_{AP} , der auf Gitterschwingungen beruht, ist abhängig von der Dotierungskonzentration und somit gut geeignet, um zwischen Kurzkanal- (Überlagerung der Halogebiete) und Langkanaltransistoren zu unterscheiden. Eine Variation der Parameter führte jedoch nicht zu den gewünschten Simulationsergebnissen. Mithilfe der High-k-Erweiterung des Modells konnte der Verlauf der Sperrströme über den Sättigungsströmen in guter Näherung zu den experimentellen Daten simuliert werden. Die dotierungsabhängige Dämpfung der Beweglichkeit, kontrolliert über den Exponenten γ_1 , musste entsprechend dominant gewählt werden. Zum Ausgleich wurde

für den n-MOSFET die Grundbeweglichkeit μ_{RCS}^0 etwas erhöht und ab dem Einsetzen der Inversion eine leichte Steigerung der Ladungsträgerbeweglichkeit über den Parameter γ_3 eingestellt. Um die durch die SOI-Technologie begrenzte Wannentiefe der Transistoren zu berücksichtigen, wurde die kritische Eindringtiefe auf $l_{RCS}^{crit} = 1$ nm gesetzt, da das Modell gegenüber der Geometrie sonst zu dominant gewesen wäre. Für den p-MOSFET wurde die Grundbeweglichkeit μ_{RCS}^0 etwas reduziert, um den Einfluss des High-k-Anteils zu verstärken. Die anderen Parameter entsprechen denen des n-MOSFETs.

F. Zusammenfassung der geänderten Modellparameter

Zusammenfassend sind die Modellparameter, die zur Kalibrierung geändert wurden, in Tabelle 3.3 aufgelistet. Alle anderen Parameter wurden mit den vorgegebenen Standardwerten übernommen [117]. Zum Vergleich sind die Simulationsergebnisse mit den experimentellen Daten anhand der wesentlichen Kennlinien in Anhang 1 und Anhang 2 dargestellt.

Tabelle 3.3: Auflistung der zur Kalibrierung veränderten Modellparameter der Bauelementesimulation.

Modell/Parameter	Symbol	Elektronen (n-MOSFET)	Löcher (p-MOSFET)
Austrittsarbeit	W_A	4.1 eV	4.61 eV
Energierelaxationszeit	τ_e	0.2 ps	0.4 ps
B2B-Tunneln	B_{B2B}	$2.5 \cdot 10^{-7} \frac{\text{V}}{\text{cm}} \text{eV}^{-3/2}$	nicht aktiv
Trap-Assisted-Tunneling	E_{min}^{TAT}	0 V/cm	$5 \cdot 10^5$ V/cm
LOMBARDI- High-k-Modell	α_{RCS}	0.5	1.0
	l_{RCS}^{crit}	10^{-7} cm	10^{-7} cm
	μ_{RCS}^0	350 cm ² /Vs	100 cm ² /Vs
	γ_1	-0.70	-0.50
	γ_3	0.20	0.20

4 Planare Transistoren auf Basis einer 22 nm-Technologie

Grundsätzlich basieren alle untersuchten Technologien dieser Arbeit auf den Designregeln der ITRS für den 22 nm-Technologieknoten. Die planaren Transistoren stellen dabei den Ausgangspunkt für die nachfolgenden Tri-Gate-Transistoren bzw. die Grundlage für einen angestrebten hybriden Prozessansatz dar, bei dem sowohl planare als auch Tri-Gate-Transistoren auf einem Chip integriert sind. Ziel war es, die planaren Bauelemente mithilfe möglichst weniger Prozessschritte in ihrer Weite zu strukturieren, um die gewünschte Tri-Gate-Struktur zu erzeugen. Dafür war jedoch ein bereits hochoptimierter planarer Transistor erforderlich, um den zusätzlichen Aufwand gering zu halten. Die Vorteile einer hybriden Integration bestehen darin, dass Langkanaltransistoren, analoge Bauelemente und Designblöcke, welche von den vorangegangenen Technologien übernommen wurden, auf einfachem Wege weiterhin mit planaren Transistoren umgesetzt werden können, wohingegen die Tri-Gate-Transistoren für Hochleistungsanwendungen eingesetzt werden.

Beginnend mit der Beschreibung der Prozesssimulation von n- und p-MOSFETs wird auf die Besonderheiten der planaren 22 nm-Transistoren eingegangen. Dazu gehören die epitaktisch erzeugten Source-Drain-Gebiete und der Replacement-Gate-Prozess. Anschließend folgt eine Auswertung der stationären Kenngrößen und des Kurzkanalverhaltens beider Transistortypen, bevor die Analyse des dynamischen Schaltverhaltens das Kapitel abschließt.

4.1 Transistoraufbau

4.1.1 Replacement-Gate-Prozess

Folgt man den Regeln der Skalierung aus der ITRS [22], gelangt man für die 22 nm-Technologien zu Dimensionen, bei denen die nominelle Gatelänge 26 nm und der Pitch 86 nm beträgt. Mit diesen Abmaßen rücken der Kurzkanaleffekt und der zunehmende Verlust der Steuerfähigkeit des Gates deutlich in den Vordergrund. Für den hier beschriebenen Ansatz einer 22 nm-Bulk-Technologie wurde somit das Hauptaugenmerk auf möglichst steile pn-Übergänge der Source-Drain-Gebiete und auf den Replacement-Gate-Prozess gelegt.

Der Prozess ist in Tabelle 4.1 dargestellt. Er startet nach der Strukturierung der Aktivgebiete (STI) mit der Implantation der Wannengebiete. Nach einer Wannenausheilung kommt es zur Oxidation der Siliziumoberfläche, wodurch sich eine 14 Å starke Siliziumoxidschicht bildet. Auf ihrer Oberfläche wird eine 70 nm dicke Schicht polykristallines Silizium mit einer Siliziumnitridschicht abgeschieden, bevor anschließend unter Verwendung einer Maske die Gatestruktur geätzt wird. Hierbei handelt es sich jedoch nur um das vorläufige Dummy-Gate, welches gegen Ende des Herstellungsprozesses durch das Replacement-Gate ersetzt wird.

Eine Reoxidation der verbleibenden Source-Drain-Fläche und des polykristallinen Siliziums bilden das Streuoxid bzw. den Platzhalter (Spacer0) seitlich des Dummy-Gates für die nachfolgenden Implantationen. Beide Transistortypen erhalten eine Haloimplantation unter einem Winkel von 25° zur Gatekante als Gegendotierung im Kanalgebiet. Die Implantationen der Erweiterungsgebiete von Source und Drain werden jedoch nur für die n-MOSFETs umgesetzt, da die Arsenionen in den nachfolgenden Prozessschritten mit erhöhter Temperatur einer deutlich geringeren Diffusion unterliegen. Zusätzlich wird vor der Erweiterungsimplantation eine Preamorphisierung mit Xenon durchgeführt, um den Channeling-Effekt zu vermeiden. Erweiterungsgebiete aus Borionen auf der Seite des p-MOSFETs würden zu weit in das Kanalgebiet diffundieren, sodass der Transistor seine Schalteigenschaften verlieren würde.

Mit der Prozessierung des Platzhalters (Spacer1) aus Siliziumnitrid mit einer Breite von ca. 15 nm werden die Epitaxiemodule der in-situ-dotierten Source-Drain-Gebiete von n- und p-MOSFETs vorbereitet, deren Details aufgrund ihrer Komplexität in einem separaten Abschnitt (4.1.2) erläutert werden.

Das thermische Budget der nachfolgenden Gitterausheilung bzw. Dotandenaktivierung musste gegenüber der 32 nm-Technologie reduziert werden, um die Diffusion der Dotanden gering zu halten [142]. Die Kurzzeitausheilung (RTA) wurde dafür bei einer Dauer von 5 s auf 920°C gesenkt. Damit es dennoch zur Aktivierung einer hinreichenden Anzahl an Dotanden kommt, wurde eine Laserausheilung mit 1225 °C vorgesehen.

Vor dem Austausch des Dummy-Gates erfolgt die Silizierung der Source-Drain-Gebiete und deren Auffüllung mit dem Kontaktmetall sowie ein chemisch-mechanischer Polierschritt (CMP), bei welchem die schützende Nitridkappe über dem Dummy-Gate entfernt wird. Somit kann das polykristalline Silizium herausgelöst und der Graben mit einem Basisoxid (SiO_2 , 6 Å), einem High-k-Material (HfO_2 , 12 Å) sowie dem Gate-Metall wieder aufgefüllt werden. Der entscheidende Vorteil liegt darin, dass dem High-k-Oxid infolge von thermischen Ausheilschritten mit höheren Temperaturen keine Beschädigungen mehr zugefügt werden [106]. Die resultierenden Prozesssimulationen von n- und p-MOSFET mit den zugehörigen Dotierungsprofilen sind in Abbildung 4.1 dargestellt.

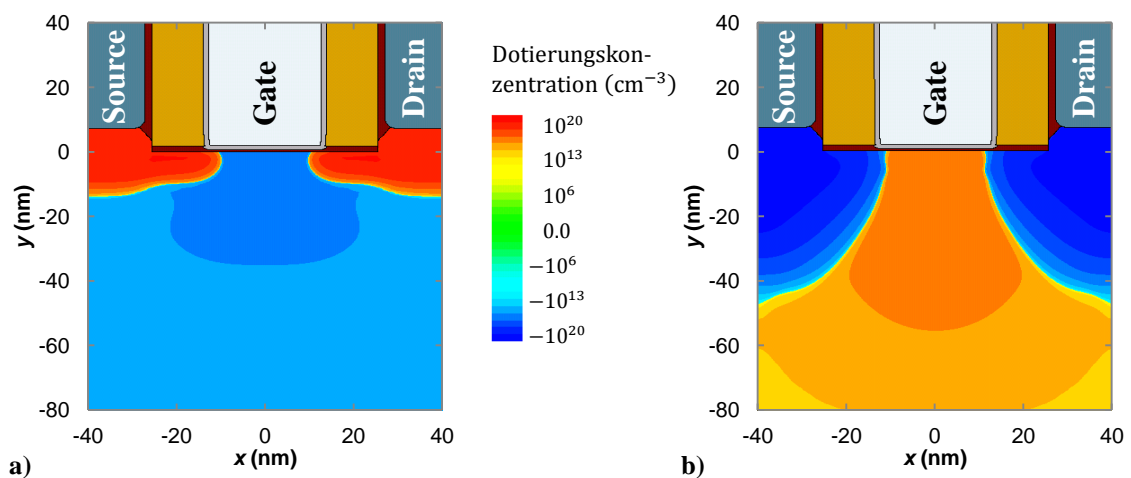
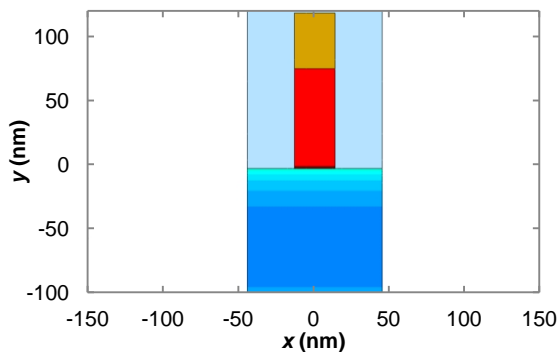


Abbildung 4.1: Berechnete zweidimensionale Dotierungskonzentration in a) n- und b) p-MOSFETs. n-dotierte Gebiete sind positiv, p-dotierte Gebiete negativ dargestellt.

Da es sich um planare Transistoren ohne eine Strukturierung der Weite handelt, ist der Großteil der Prozess- und Bauelementesimulation für die planare 22 nm-Bulk-Technologie nur zweidimensional realisiert worden, um die Rechenzeit gering zu halten und eine hohe Genauigkeit der Gitterauflösung zu erreichen. Um auch die Rechenzeit für die 3D-Prozesssimulationen in einem akzeptablen Rahmen zu halten, mussten entsprechende Vereinfachungen der Strukturen getroffen werden. So sind z. B. einige isotrope Ätzschritte durch anisotrope ersetzt worden, um teilweise runde Materialecken bzw. -kanten, welche eine unnötig hohe Anzahl an Gitterpunkten verursachen, zu vermeiden. Weiterhin sind alle Oxidationsschritte auf einfache Diffusionsschritte mit anschließendem Abscheiden einer Oxidschicht reduziert worden. Es zeigte sich jedoch, dass die Ergebnisse für die 3D-Simulation trotz der getroffenen Vereinfachungen hinreichend gut mit den 2D-Simulationen übereinstimmten.

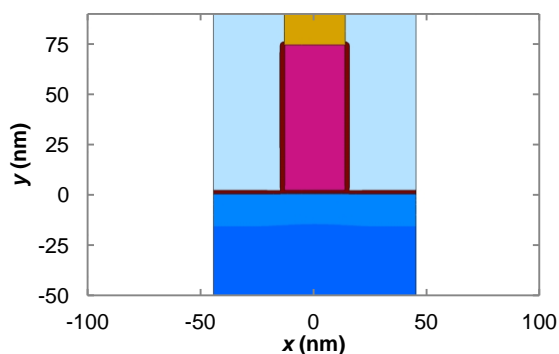
Tabelle 4.1: Ablauf des planaren 22 nm-CMOS-Prozesses

2D Prozesssimulation - Wannenheimplantation ... Epitaxiemodul (p-MOSFET)



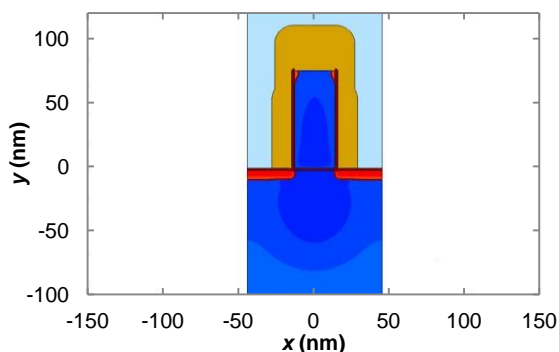
Wannenheimplantation und Dummy-Gate

- Wannenheimplantation
 - BF_2 , $5 \cdot 10^{12} \text{ cm}^{-2}$, 70 keV (n-MOSFET)
 - As, $5 \cdot 10^{12} \text{ cm}^{-2}$, 100 keV (p-MOSFET)
- Wannenausheilung (1000°C, 5 s)
- Dummy-Gate-Strukturierung
 - 14 Å Oxid
 - 70 nm Poly-Si
 - 40 nm Nitridkappe



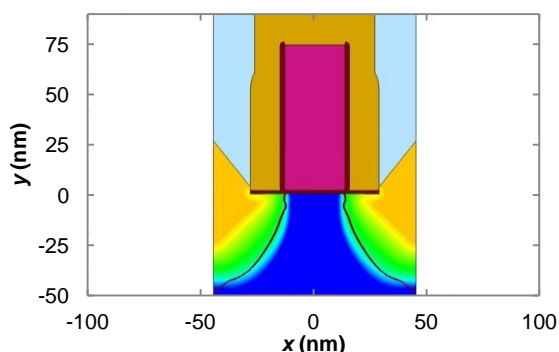
Halo-/Erweiterungsimplantation

- Reoxidation (Spacer 0)
 - 800°C, 28 min
 - 20 Å Oxid
- n-MOSFET
 - Xe-Preamorphisierung
 - BF_2 , $5.7 \cdot 10^{13} \text{ cm}^{-2}$, 40 keV, 25°
 - As, $1 \cdot 10^{14} \text{ cm}^{-2}$, 1.5 keV
- p-MOSFET
 - As, $3.7 \cdot 10^{13} \text{ cm}^{-2}$, 65 keV, 25°



Spacer 1

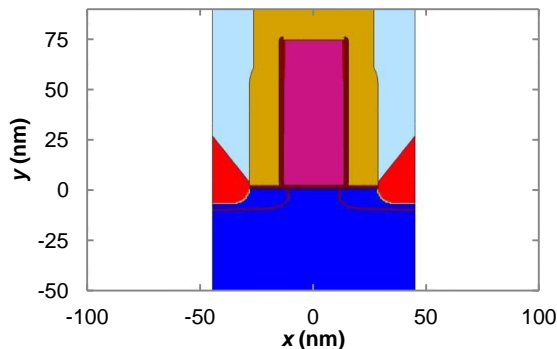
- Isotrope Abscheidung von 20 nm Siliziumnitrid
- Anisotropes/isotropes Ätzen (Spacerformung mit 15 nm Spacerbreite)



Epitaxie-Modul (p-MOSFET)

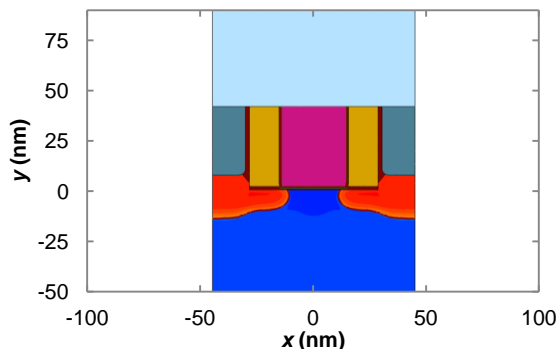
- n-MOSFET abgedeckt durch Maske
- Directionales Ätzen eines sigmaförmigen Grabens bis unter die Gatekante
- Oberflächenausheilung; 850°C; 1 min
- Aufwachsen der Source-Drain-Gebiete
 - 13 nm, 20 % Ge, undotiert
 - Auffüllen, 30 % Ge, Bor, 10^{20} cm^{-3} (in-situ, oberflächenlimitiert)

2D Prozesssimulation - Epitaxie-Modul (n-MOSFET) ... Replacement-Gate



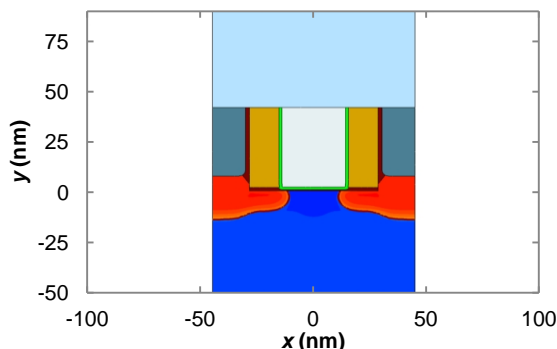
Epitaxie-Modul (n-MOSFET)

- p-MOSFET abgedeckt durch Maske
- Isotropes Ätzen eines Grabens von 70 Å
- Oberflächenausheilung; 850°C; 1 min
- Aufwachsen der Source-Drain-Gebiete
 - Auffüllen, P, $2 \cdot 10^{20} \text{ cm}^{-3}$ (in-situ, oberflächenlimitiert)



Source-Drain-Kontakte und CMP

- Ausheilung und Aktivierung
 - Kurzeitenausheilung (920°C, 5 s)
 - Laserausheilung (1225°C)
- Silizierung
- Auffüllen mit Metall
- Chemisch-Mechanisches Polieren



Replacement-Gate

- Entfernen des Dummy-Gates
- 6 Å Basisoxid (ALD)
- 12 Å HfO₂ (High-k-Material)
- Auffüllen mit Titanitrid
- CMP auf endgültige Gatehöhe (40 nm)

4.1.2 In-situ-dotierte Source-Drain-Gebiete

Die Prozessierung der Source-Drain-Gebiete wird bei diesem Technologieansatz für beide Transistortypen über ein epitaktisches Aufwachsen von Silizium in Kombination mit Germanium bzw. Dotanden realisiert. Diese Vorgehensweise ist bereits seit einigen Jahren bekannt und auch umsetzbar [143], verlangt jedoch einen gewissen Mehraufwand und verursacht somit erhöhte Kosten gegenüber der herkömmlichen Implantation. Im Gegenzug erlaubt dieses Verfahren die Realisierung steiler pn-Übergänge, wie sie für diesen Prozess erforderlich sind.

Der Begriff „in-situ“ bedeutet, dass die Dotanden während der epitaktischen Abscheidung des Siliziums der Source-Drain-Gebiete direkt in das aufwachsende Gitter

integriert werden. Dies geschieht sowohl in der Form von Zwischengitteratomen als auch direkt auf den Gitterplätzen des Siliziumkristalls. Dabei können p- bzw. n-Gebiete mit der höchstmöglichen Konzentration an Dotanden (Grenze der Festkörperlöslichkeit) erzeugt werden, wodurch ihre Leitfähigkeit erhöht und der Kontaktwiderstand zum Silizid verringert wird.

Im Falle des n-MOSFETs ist die tiefe Source-Drain-Implantation durch einen epitaktischen Schritt ersetzt worden, bei dem zuerst isotrop ein ca. 7 nm tiefer Graben in die Source-Drain-Gebiete geätzt und anschließend mit Silizium und einer Phosphorkonzentration von $2 \cdot 10^{20} \text{ cm}^{-3}$ wieder aufgefüllt wird. Dies geschieht über ein facettierte Wachstum, welches in $\langle 100 \rangle$ -Richtung begünstigt ist, sodass sich eine pyramidenförmige Spitze bildet und der Prozess selbstlimitierend stoppt.

Für den p-MOSFET ist die epitaktische Dotierung mit dem eingebetteten Silizium-Germanium (embedded Silicon-Germanium, eSiGe) verbunden. Nach dem directionsalen Ätzen der 55 nm tiefen sigmaförmigen Gräben in die Source-Drain-Gebiete, deren seitliche Ecken bis knapp unter das Gate reichen, wird zunächst eine undotierte, ca. 13 nm dicke Silizium-Germanium-Schicht mit einem Germaniumgehalt von 20 % abgeschieden. Ein höherer Germaniumanteil ist zwar in der Simulation umsetzbar, jedoch in der Realität mit einem erschwerten Aufwachsen und zweidimensionalen Gitterfehlern verbunden, sodass der Effekt der mechanischen Verspannung deutlich geringer ausfallen und die elektrischen Eigenschaften des Transistors degradieren würden. Im Anschluss an die dünne undotierte Schicht folgt ein Auffüllen der Gräben mit 30 % Germaniumgehalt und einer Borkonzentration von $2 \cdot 10^{20} \text{ cm}^{-3}$. Auch hier wird das facettierte Wachstum in $\langle 100 \rangle$ -Richtung genutzt, um den Prozess selbstlimitierend zu stoppen, wobei die letzten Atomlagen mit einer abfallenden Konzentration an Germanium abgeschieden werden, um die Oberfläche beständiger gegenüber Ätz- und Reinigungsschritten zu halten. Eine Implantation von Erweiterungsgebieten wird bei dieser Vorgehensweise nicht benötigt.

Die dabei erzeugte mechanische Verspannung des p-MOSFETs im Kanalgebiet kann zusätzlich durch den Replacement-Gate-Prozess weiter gesteigert werden. Mit dem Herauslösen des Dummy-Gates wird quasi der Puffer zwischen den Source-Drain-Gebieten entfernt, wodurch diese stärker relaxieren können und sich die Druckverspannung im Kanalbereich auf bis zu 2 GPa erhöht (Abbildung 4.2).

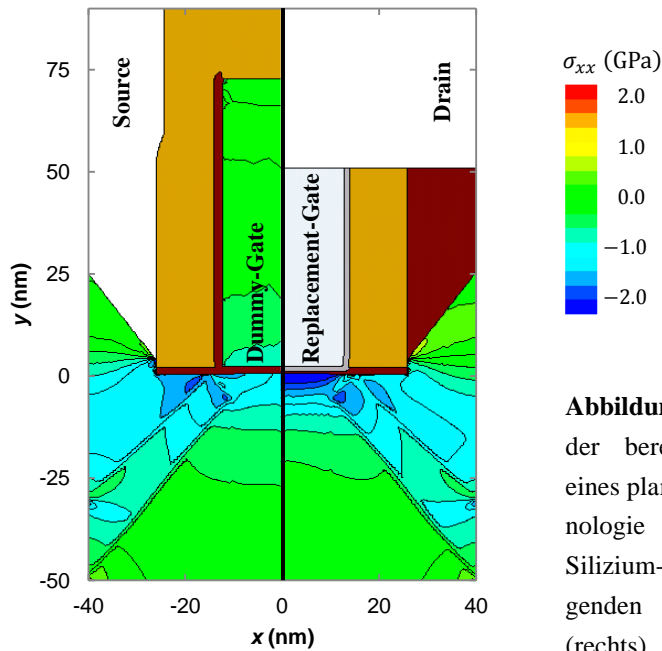


Abbildung 4.2: Zweidimensionale Darstellung der berechneten mechanischen Verspannung eines planaren p-MOSFETs in 22 nm-Bulk-Technologie nach der Source-Drain-Epitaxie mit Silizium-Germanium (links) und der nachfolgenden Prozessierung des Replacement-Gates (rechts).

4.1.3 Haloimplantation

Die Simulationen mit den angedachten Prozessparametern des n-MOSFETs bei variierender Gatelänge von 22 ... 54 nm zeigten einen leichten Knick der Schwellspannungen im Linear- und Sättigungsbereich zwischen der 30 nm- und der 34 nm-Gatelänge (Abbildung 4.3). Dieser Effekt resultiert aus einer unvermeidlichen Abschattung der Source-Drain-Gebiete durch die benachbarten Gates. Für die Halogebiete, die unter einem Winkel von 25° zur Gatekante implantiert werden, bedeutet dies eine Reduktion der Konzentration im Kanal, was einen stärkeren Kurzkanaleffekt zur Folge hat. Für alle Gatelängen von mehr als 30 nm wird aus diesem Grund der Pitch um die Gatelängenänderung erweitert. Mit einem durch die Designregeln gegebenen maximalen Pitch von 110 nm ist demnach nur eine maximale Gatelänge von $L_{Gate} = 54$ nm für den Standard-MOSFET möglich. Eine Anpassung des Implantationswinkels auf 10° bewirkt zwar, dass die Abschattungseffekte keinen Einfluss mehr auf das elektrische Verhalten des Transistors zeigen, reduziert jedoch auch die Schwellspannungen aller Transistoren. Bei noch kleineren Winkeln bis hin zu 0° gelangt nicht mehr genug Gegendotierung zu den Erweiterungsimplantationen in den Kanal, sodass eine Zunahme der Kurzkanaleffekte beobachtet werden kann.

Bei den Schwellspannungen der p-MOSFETs trat dieser Effekt nicht auf, da die Lage der n-p-Übergänge und das Kurzkanalverhalten hauptsächlich durch die epitaktisch aufgewachsenen Source-Drain-Gebiete bestimmt werden. Eine Verringerung des Winkels der Haloimplantation vermindert lediglich bei allen Gatelängen die Schwellspannungen.

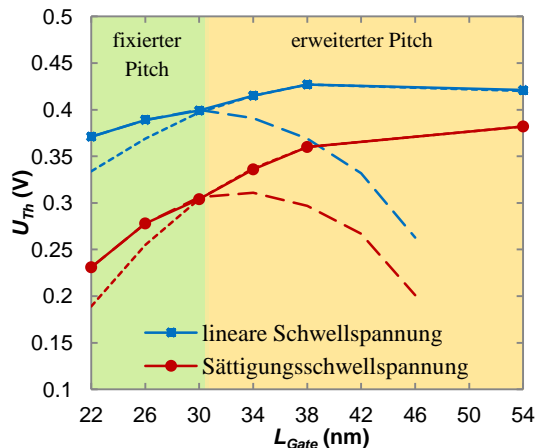


Abbildung 4.3: Berechnetes Schwellspannungsverhalten eines planaren n-MOSFETs in 22 nm-Bulk-Technologie. Der Winkel der Haloimplantation betrug 25°. Die gestrichelten Fortführungen deuten den Verlauf für einen komplett fixierten bzw. erweiterten Pitch an.

4.1.4 Elemente der mechanischen Verspannung

Infolge der Skalierung und dem Replacement-Gate-Ansatz können nicht mehr alle Techniken zu mechanischen Verspannung der planaren Transistoren angewandt werden (Abbildung 4.4). So entfallen die verspannten Deckschichten, die in den vorangegangenen Technologieknoten für eine deutliche Leistungssteigerung gesorgt hatten [43]. Der Grund dafür ist die Verringerung des Pitches um ca. 24 nm im Vergleich zur 32 nm-Technologie. Die Kontaktflächen der Source-Drain-Gebiete, über welche der Großteil der Verspannung einkoppelt, bietet nur noch verhältnismäßig wenig Angriffsfläche. Auch müssen die verspannten Schichten selbst dünner gewählt werden, um ein hinreichend gutes Füllverhalten beim Abscheiden gewährleisten zu können.

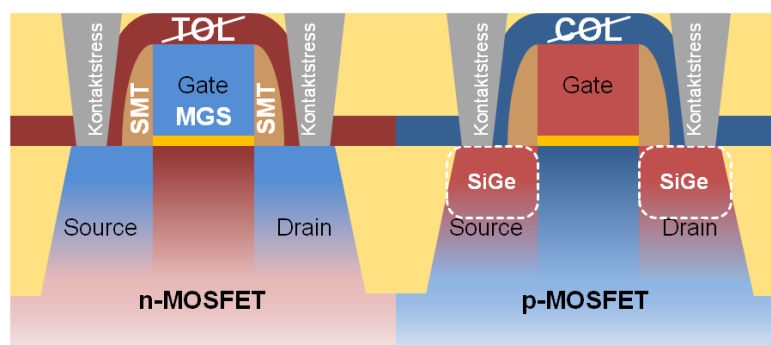


Abbildung 4.4: Übersicht zu den Techniken der mechanischen Verspannung der planaren Transistoren einer 22 nm-Technologie. Die verspannten Deckschichten (TOL/COL) lassen sich nicht mehr anwenden. Möglich bleiben: SiGe in den Source-Drain-Gebieten des p-MOSFETs, verspannungsspeichernde Prozesse (SMT), verspanntes Gatemetall (MGS) und verspannte Kontakte.

Wie in Abbildung 4.5 deutlich wird, geht die eingebrachte Kanalverspannung des TOLs durch den Replacement-Gate-Prozess und die Source-Drain-Kontakte wieder verloren. Eine Verschiebung der Universalkennlinie und somit eine Veränderung der Leistungsfähigkeit konnte nicht beobachtet werden. Die gleichen Aussagen können über das

Verhalten des p-MOSFETs mit und ohne COL getroffen werden. Zudem ist beim p-MOSFET das eSiGe wesentlich dominanter gegenüber einer druckverspannten Nitridschicht.

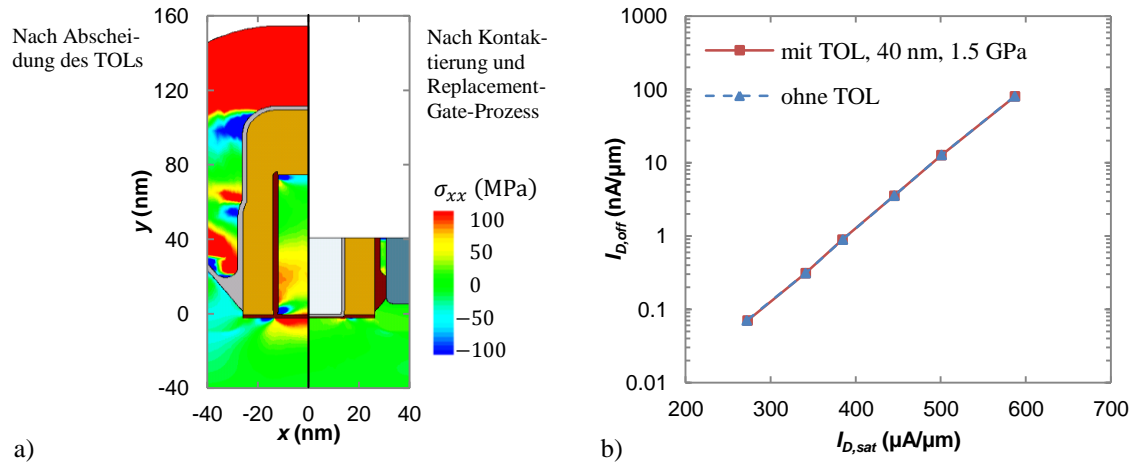


Abbildung 4.5: a) Berechnetes Verspannungsprofil nach Abscheidung des TOL sowie nach der Prozessierung der Kontaktgebiete und des Replacement-Gates, b) Universalkennlinie eines n-MOSFETs mit und ohne Integration eines TOL von 40 nm Dicke und einer intrinsischen Verspannung von 1.5 GPa.

Als sinnvolle Verspannungselemente für den n-MOSFET eignen sich somit nur noch der Einsatz von SMT und die Prozessierung von zugverspannten Source-Drain-Kontakten. Die intrinsische Verspannung der Source-Drain-Gebiete nach der Silizierung kann bestenfalls mit maximal 500 MPa Zugverspannung angenommen werden [64]. Die Simulationen hierzu ergaben jedoch nur eine geringe Steigerung der Leistungsfähigkeit von 3 ... 5 % gegenüber dem unverspannten n-MOSFET. Hinzu kommt eine ungünstige Lage der silizidierten Gebiete bezüglich des Kanals, wodurch sich die Wirkung der mechanischen Verspannung nicht direkt bis in das Kanalgebiet ausbreiten kann (Abbildung 4.6).

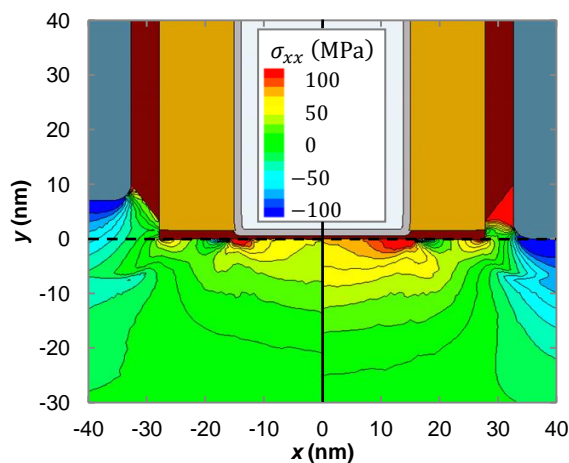


Abbildung 4.6: Berechnetes Verspannungsprofil eines n-MOSFETs mit einer mechanischen Verspannung der Kontaktgebiete auf Höhe des Kanals (links) und 7 nm oberhalb des Kanals (rechts).

Ebenso wie die Source-Drain-Kontakte kann auch das Gatemetall mit einer mechanischen Verspannung prozessiert werden [66], [144]. Es nimmt dabei eine intrinsische Druckverspannung von bis zu 500 MPa an, was zu einer Verbesserung des n-MOSFETs genutzt werden könnte. Der p-MOSFET würde jedoch in seiner Leistungsfähigkeit gemindert werden, wodurch sich dieses Verfahren nur für CMOS-Prozesse eignet, bei denen zwei getrennte Metallgates gefertigt werden.

Eine weitere Variante der mechanischen Verspannung des n-MOSFETs wäre das epitaktische Aufwachsen von Silizium-Kohlenstoff (SiC) in den Source-Drain-Gebieten [145], ähnlich dem Silizium-Germanium bei p-MOSFETs. Dies konnte an Teststrukturen zwar nachgewiesen werden [146], bietet jedoch noch keine hinreichende Prozessstabilität und scheidet somit ebenfalls aus.

Die Silizium-Germanium-Epitaxie der Source-Drain-Gebiete ist für diesen Technologieknoten das wichtigste Verspannungselement der p-MOSFETs [147]. Mithilfe eines direktionalen Ätzschrittes gelangt das eSiGe bei dem anschließenden Abscheiden bis weit unter den Nitridspacer und ermöglicht so eine hohe Druckverspannung des Kanalgebiets von ca. 1.5 GPa. Hinzu kommt der Replacement-Gate-Prozess, bei dem das Entfernen des Dummy-Gates (mechanischer Puffer) eine Relaxation der Source-Drain-Gebiete um 30 % und eine Steigerung der Kanalverspannung auf knapp 2 GPa ermöglicht. Die berechneten Werte ergaben dabei eine Steigerung des Drainstromes um 50 ... 100 % im Vergleich zu einem p-MOSFET ohne Germaniumanteil in den Source-Drain-Gebieten. Diese Werte müssen jedoch kritisch betrachtet werden, da die Modelle zur Simulation der verspannungsabhängigen Löcherbeweglichkeit ursprünglich für geringere Verspannungen entwickelt und im Nachhinein mit experimentellen Daten an Wert mit bis zu 2 GPa angepasst wurden. Darüber hinaus sind die Modelle für größere Verspannungen, wie sie an Kanten im Transistor auftreten können, nicht verifiziert.

4.2 Charakterisierung des elektrischen Verhaltens

4.2.1 Stationäres Verhalten

Im Vergleich zur 32 nm-SOI-Technologie, an der die Modelle der Bauelementesimulation kalibriert wurden, zeigt das berechnete Verhalten der 22 nm-Bulk-Transistoren deutliche Unterschiede. Dabei muss stets berücksichtigt werden, dass zwischen den beiden Technologieknoten ein Skalierungsschritt und diverse Änderungen im Herstellungsprozess, wie die Einführung des Replacement-Gate-Ansatzes und die Reduzierung der Verspannungselemente, stehen.

Die Schwellspannungen beider Transistortypen weisen ein unterschiedlich stark ausgeprägtes Kurzkanalverhalten auf (Abbildung 4.7a). Während der n-MOSFET mit dem Knick aufgrund der Änderung des Pitches nur einen leichten Abfall der Schwellspan-

nungen zeigt, fällt die des p-MOSFETs mit kürzer werdender Gatelänge wesentlich steiler ab (Abbildung 4.3). Die Ursache für das stärker ausgeprägte Kurzkanalverhalten sind auch hier die epitaktisch erzeugten Source-Drain-Gebiete des p-MOSFETs, welche stets den gleichen Abstand zur Gatekante besitzen. Somit ändert sich die Lage der pn-Übergänge auch bei einer Variation der Halodosis bzw. des Halowinkels nur geringfügig. Bei Betrachtung der Universalkennlinien fällt auf, dass sich die Leistungsfähigkeit des p-MOSFETs an die des n-MOSFETs angenähert hat (Abbildung 4.7b). In den vorangegangenen Technologien hatte der n-MOSFET aufgrund der besseren Ladungsträgerbeweglichkeit der Elektronen stets einen höheren Drainstrom. Da jedoch der n-MOSFET der 22 nm-Technologie unverspannt simuliert wurde und der p-MOSFET mit dem eingebetteten Silizium-Germanium in den Source-Drain-Gebieten eine sehr starke mechanischen Verspannung aufweist, besitzen beide Transistortypen nun fast die gleiche Leistungsfähigkeit.

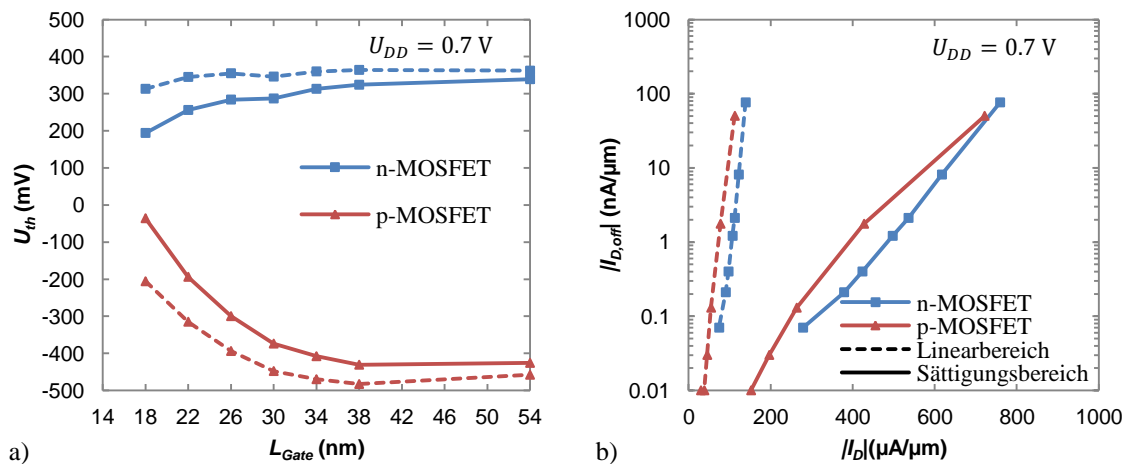


Abbildung 4.7: Berechnetes Verhalten planarer 22 nm-Transistoren. a) Schwellspannungen in Abhängigkeit von der Gatelänge und b) Universalkennlinie von n- und p-MOSFETs im Linear- und Sättigungsbereich.

Auch wenn die berechneten Werte wegen der Annahmen zur verspannungsabhängigen Ladungsträgerbeweglichkeit des p-MOSFETs kritisch betrachtet werden müssen, bestätigen die experimentellen Daten, dass der p-MOSFET mit den Skalierungsschritten von der 45 nm- zur 22 nm-Technologie durch die mechanische Verspannung eine deutlich höhere Leistungssteigerung als der n-MOSFET erfahren hat [148]. Auch der nichtlineare Anstieg der Steilheit des p-MOSFETs bestätigt, dass der Einfluss des eingebetteten Silizium-Germaniums mit der Skalierung der Gatelänge steigt, wohingegen der n-MOSFET nur eine lineare Abhängigkeit aufgrund der Kanallängenverkürzung zeigt (Abbildung 4.8). Alle weiteren stationären Kenndaten der nominellen Gatelänge von n- und p-MOSFETs sind in Tabelle 4.2 zusammengefasst.

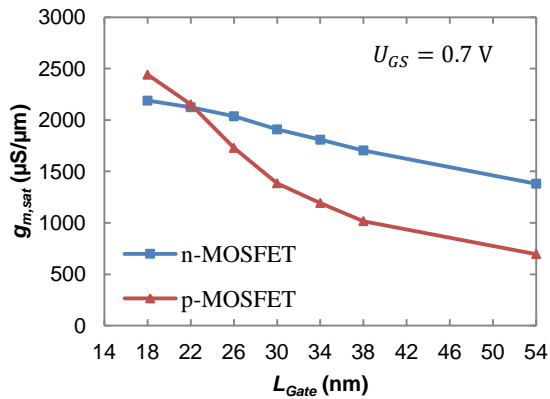


Abbildung 4.8: Berechnete Steilheit planarer 22 nm-Transistoren in Abhängigkeit von der Gatelänge (Sättigungsbereich).

Tabelle 4.2: Berechnete Kenndaten planarer 22 nm-Transistoren im linearen (lin.) und im Sättigungsbereich (sat.) mit $L_{Gate} = 26$ nm und $U_{DD} = 0.7$ V.

Typ	U_{th} (mV)		$DIBL$ (mV/V)	SS (mV/dec)		g_m ($\mu\text{S}/\mu\text{m}$)		I_D ($\mu\text{A}/\mu\text{m}$)		$I_{D,off}$ (nA/ μm)	R_{on} ($\Omega \cdot \mu\text{m}$)
	lin.	sat.		lin.	sat.	lin.	sat.	lin.	sat.		
n-MOS	355	284	101	85.6	84.1	408	2036	112	536	2.11	506
p-MOS	-394	-300	134	87.7	87.1	300	1728	78	428	1.76	1006

4.2.2 Gatesteuerung und Kurzkanaleffekte

Für den $DIBL$ und die Unterschwellsteigung konnten im Bereich der mittleren Gate-längen von 30 ... 54 nm Werte simuliert werden, die auf eine hinreichend gute Gatesteuerung dieses Technologieansatzes schließen lassen, wobei die p-MOSFETs ein stärker ausgeprägtes Kurzkanalverhalten aufweisen (Abbildung 4.9). Positiv ist hier der flache Verlauf der Unterschwellsteigung von ca. 80 mV/dec bis zur nominellen Gate-länge, was auf die steilen pn-Übergänge zurückzuführen ist.

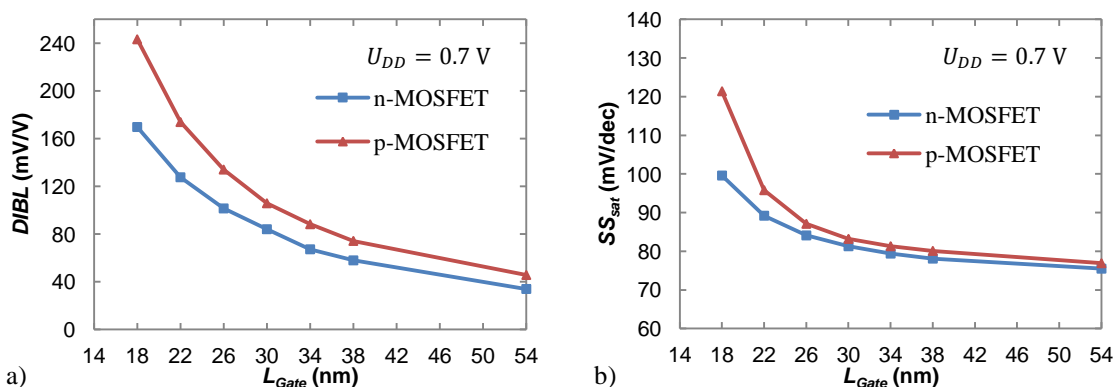


Abbildung 4.9: Berechnetes Kurzkanalverhalten planarer 22 nm-Transistoren. a) $DIBL$ und b) Unterschwellsteigung (Sättigung) in Abhängigkeit von der Gatelänge.

Im Bereich der sehr kurzen Gatelängen steigen beide Werte jedoch rapide an. Vor allem die Schwellspannungen des Linear- und Sättigungsbereichs driften stark auseinander und vergrößern den *DIBL* beider Transistortypen auf über 100 mV/V.

Im Vergleich mit der berechneten Ideallinie zur linearen Unterschwellsteigung aus [39] schneiden die planaren MOSFETs jedoch deutlich schlechter ab (Abbildung 4.10). Dies liegt zum einen daran, dass die Kanäle nur durch einen planaren Gatekontakt gesteuert werden. Aber auch an der Tatsache, dass die Transistoren auf einem Bulk-Substrat produziert wurden, welches trotz der flachen Erweiterungsgebiete von Source und Drain eine große Ausdehnung des Kanalgebiets in der Tiefe bedeutet. Die Werte für die natürlichen Längen wurden über die Gleichung einer Einzel-Gate-Struktur ermittelt. Sie unterscheiden sich, da die Erweiterungsgebiete von den p-MOSFETs aufgrund der epitaktischen Realisierung etwas tiefer liegen als die implantierten Erweiterungsgebiete der n-MOSFETs.

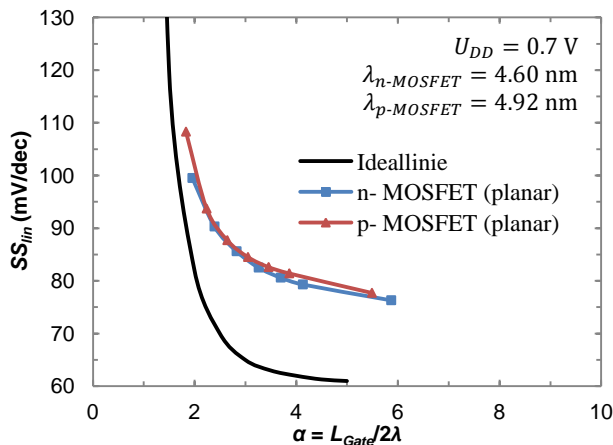


Abbildung 4.10: Berechnete lineare Unterschwellsteigung in Abhängigkeit von dem Skalierungsparameter α für die planaren n- bzw. p-MOSFETs der 22 nm-Technologie.

Ersichtlich wird dies auch durch den deutlichen Anstieg der Ausgangskennlinien im Sättigungsbereich beider Transistoren mit einer Gatelänge von 26 nm (Abbildung 4.11). Ein sättigendes Verhalten ist nur noch bedingt erkennbar, da die Verkürzung der effektiven Gatelänge durch das Abschnüren des Kanals in beiden Fällen einen starken Einfluss auf die Schwellspannung hat und der Drainstrom im Bereich der Betriebsspannung mit ca. $0.29 \mu\text{A}/\mu\text{mV}$ weiter ansteigt.

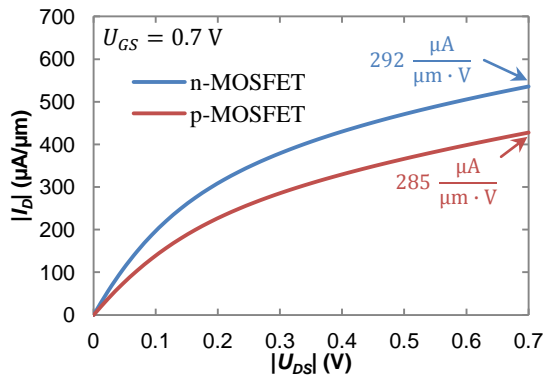


Abbildung 4.11: Berechnete Ausgangskennlinie eines planaren n- bzw. p-MOSFET der 22 nm-Technologie mit einer nominellen Gatelänge $L_{Gate} = 26$ nm.

4.2.3 Dynamisches Verhalten

Das dynamische Verhalten der planaren Transistoren resultiert aus dem Zusammenspiel der Kapazitäten und den Drainströmen von n- und p-MOSFETs. Idealerweise sollten beide Transistortypen dafür einen gleichgroßen Sättigungsstrom haben und vergleichbare Kapazitäten (Abbildung 4.12) aufweisen, damit die Schaltzeiten einer Inverterstufe möglichst kurz sind und sich symmetrisch bezüglich des Ein- und des Ausschaltvorgangs verhalten.

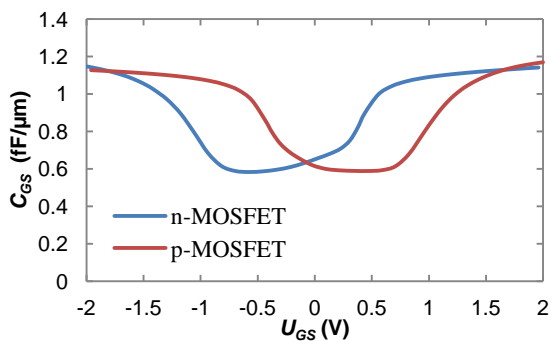


Abbildung 4.12: Gate-Source-Kapazität C_{GS} von planaren n- und p-MOSFETs der 22 nm-Bulk-Technologie.

Da jedoch n- und p-MOSFETs in der Regel bei identischen Strukturabmessungen nicht den gleichen Drainstrom besitzen, wurde für die planaren Transistoren in den vorangegangenen Technologieknoten eine Anpassung über die Transistorweite vorgenommen, sodass die absoluten Werte übereinstimmten. Dieses Vorgehen ist bei Multi-Gate-Strukturen zwar auch möglich, muss aber über eine ganzzahlige Anzahl an Finnen erfolgen. Bei einem Stromverhältnis von n- und p-MOSFETs mit 1.25 : 1 wären somit $4 \times \text{n-MOSFET} + 5 \times \text{p-MOSFET} = 11$ Multi-Gate-Strukturen für die dreidimensionale Simulation einer ausgeglichenen Inverterstufe erforderlich. Der Aufwand für einen Ringoszillator mit fünf Inverterstufen würde die zur Verfügung stehende Rechentechnik überfordern, weshalb für alle transienten Simulationen der unangepasste Fall einer Inverterstufe, bestehend aus einem n- und einem p-MOSFET mit identischen Abmessungen und den stationären Werten aus Tabelle 4.2, simuliert wurde.

Die Unterschiede beider Transistortypen bezüglich der Überlappungs- und der Inversionskapazitäten C_{ov} bzw. C_{inv} fallen jedoch sehr gering aus, ebenso wie die Werte ihrer Transit- und Maximalschwingfrequenzen (Tabelle 4.3). Dies wird auch durch die Schaltzeiten des dreistufigen Inverters deutlich, dessen Anstiegszeit t_r , getrieben durch den p-MOSFET, nur um 1.5 ps länger ist als die Abfallzeit t_f . Die Ströme des fünfstufigen Ringoszillators sind neben der Ringoszillatorfrequenz f_{RO} und der mittlere Verzögerungszeit einer Inverterstufe t_D als flächennormierte Werte aufgeführt, wobei der Stromverbrauch im schwingenden Zustand durch den Effektivwert $I_{Ro,eff}$ und der Sperrstrom (ruhender Zustand ohne Rückkopplung) durch $I_{Ro,off}$ ausgedrückt wird. Die benötigte Fläche aller simulierten Transistoren beläuft sich dabei auf $0.043 \mu\text{m}^2$.

Tabelle 4.3: Berechnete dynamische Kenndaten planarer MOSFETs in 22 nm-Transistoren, $L_{Gate} = 26 \text{ nm}$, $U_{DD} = 0.7 \text{ V}$.

Einzeltransistor	Typ	C_{ov} (fF/ μm)	C_{inv} (fF/ μm)	f_{max} (GHz)	f_t (GHz)	
	n-MOSFET	0.657	1.043	1146	380	
p-MOSFET	0.618	1.023	1131	365		
Dreistufiger Inverter	t_r (ps)	t_f (ps)	Δt_{hi} (ps)	Δt_{low} (ps)	U_{max} (V)	U_{min} (V)
	8.23	6.73	5.34	5.18	0.779	-0.086
Fünfstufiger Ringoszillator	f_{RO} (GHz)	$I_{Ro,eff}$ ($\mu\text{A}/\mu\text{m}^2$)	$I_{Ro,off}$ (nA/ μm^2)	t_D (ps)	Gesamte Aktivfläche (μm^2)	
	19.59	607.37	11.66	5.11	0.043	

5 Tri-Gate-Transistoren

Die Tri-Gate-Transistoren, die in diesem Abschnitt vorgestellt werden, sind Bestandteil der hybriden 22 nm-Technologie und basieren auf dem beschriebenen Herstellungsprozess der planaren 22 nm-MOSFETs. Dieser wurde zur Formierung der Finne um wenige Prozessschritte erweitert, sodass der Hauptteil der Technologieparameter für beide Strukturvarianten identisch ist und zusätzliche Kosten weitgehend vermieden werden können. Gegenüber den planaren MOSFETs bieten die Tri-Gate-Transistoren eine bessere Steuerfähigkeit des Kanalgebiets, da der dreiseitige Gatekontakt, bestehend aus dem Top-Gate und den beiden Seiten-Gates, mehr Gatefläche pro Kanalvolumen besitzt. Dies sichert die angestrebten Schalteigenschaften der 22 nm-Technologie und ermöglicht die Skalierung des Aktivgebiets hin zu den Vorgaben des 14 nm-Technologieknotts.

Zu Beginn des Kapitels werden die Anforderungen hochintegrierter Schaltkreise an Multi-Gate-Transistoren erläutert, bevor der Ansatz für eine hybride CMOS-Technologie beschrieben wird. Dabei wird speziell auf die Strukturierung der Finne und die damit verbundenen Besonderheiten des Herstellungsprozesses eingegangen. Anschließend folgt eine Auswertung der elektrischen Eigenschaften der Tri-Gate-Transistoren, wobei der Schwerpunkt auf dem Einfluss des Eckeneffektes und der Finnegeometrie liegt. Zur Optimierung der Tri-Gate-Transistoren wurden abschließend die Epitaxie der Source-Drain-Gebiete, eine Möglichkeit zur mechanischen Verspannung und der Einfluss der Substratorientierung näher untersucht. Die Darstellungen der jeweiligen Ergebnisse schließen das Kapitel ab.

5.1 Prozessintegration und Transistoraufbau

5.1.1 Anforderungen an hochintegrierte Schaltkreise

Die Anforderungen an die konventionellen CMOS-Technologien hochintegrierter Schaltkreise sind mit ihrer fortschreitenden Miniaturisierung stetig gewachsen. Die Zunahme der Komplexität, die hohe Packungsdichte, steigende Taktraten und ein ökonomischer Leistungsverbrauch machen eine strukturierte Vorgehensweise bei der Entwicklung elektronischer Systeme unverzichtbar. Sie umfasst unter anderem die Wiederverwendung von Entwurfsdaten der Vorgängergeneration und strebt eine kurze und somit wirtschaftliche Entwicklungszeit (Time to Market) an. Vor diesem Hintergrund ist die Einführung einer Multi-Gate-Technologie im Vergleich zur Skalierung der planaren Transistoren ein sehr großer Schritt für die Halbleiterindustrie, da sich nicht nur die Entwicklung von n- und p-MOSFETs umfangreicher gestaltet, sondern sich die Schaltungsdesigner bedingt durch die Neudimensionierung vieler Grundbausteine auch mit einem erhöhten Aufwand konfrontiert sehen.

Als kurzfristiges Ziel und speziell für die 22 nm-Technologie ist eine schrittweise Substitution planarer Transistoren durch Tri-Gate- bzw. FinFET-Strukturen der Weg, den die Industrie gehen wird, bis in naher Zukunft auch andere Multi-Gate-Strukturen oder Nanotechnologien den Einzug in die industrielle Fertigung halten können. Die neuen Multi-Gate-Transistoren müssen dabei derart geschaffen sein, dass sie die Aktivfläche optimal ausnutzen und durch ihre Integration nicht der Skalierung der Packungsdichte widersprechen. Dabei prognostizieren Untersuchungen von dotierten Multi-Gate-Strukturen eine Steigerung der Fehlerrate durch statistische Schwankungen im Dotierungsprofil oder der geometrischen Struktur [113], [149]. Auch die größere Belastung der Bauelemente durch erhöhte Stromdichten und Feldstärken steigert die Wahrscheinlichkeit von Ausfällen und senkt die Zuverlässigkeit der integrierten Schaltungen. Somit besteht die Herausforderung darin, aus den vorgestellten planaren 22 nm-Transistoren mit einem möglichst geringen Aufwand funktionsfähige und zuverlässige Tri-Gate-Transistoren herzustellen, um den ersten Schritt in Richtung einer hybriden 22 nm-Technologie zu gehen, die sowohl planare als auch Tri-Gate-Strukturen vereint.

5.1.2 Hybride CMOS-Technologie

Der Ansatz für eine hybride CMOS-Technologie verfolgt das Ziel, dass beide Transistorstrukturen den Großteil aller Prozessschritte aus Tabelle 4.1 gemeinsam nutzen, um den Mehraufwand für den Tri-Gate-Transistor so gering wie möglich zu halten. Dies bedeutet, dass sowohl die planaren als auch die Tri-Gate-Transistoren bezüglich des Implantationsprofils, der Source-Drain-Epitaxie, der Austrittsarbeit des Metalls und der Gateoxidstärke gleichermaßen gestaltet sein werden.

Realisiert werden kann dieser Ansatz, indem im Prozess frühzeitig eine Strukturierung der Tri-Gate-Finnen vorgenommen wird (Abbildung 5.1) [150], [151]. Dafür wird eine zusätzliche Maske benötigt, welche die späteren Finnen abdeckt und gleichzeitig die planaren Transistoren vor dem nachfolgenden anisotropen Ätzschritt zur Erzeugung der Gräben schützt. Anschließend werden die Gräben mit einem Isolationsmaterial wieder aufgefüllt und die Oberfläche chemisch-mechanisch poliert, bevor die Strukturierung des Gates erfolgt. Das Dummy-Gate umfasst die Finne dabei noch nicht und liegt lediglich als Platzhalter oben auf.

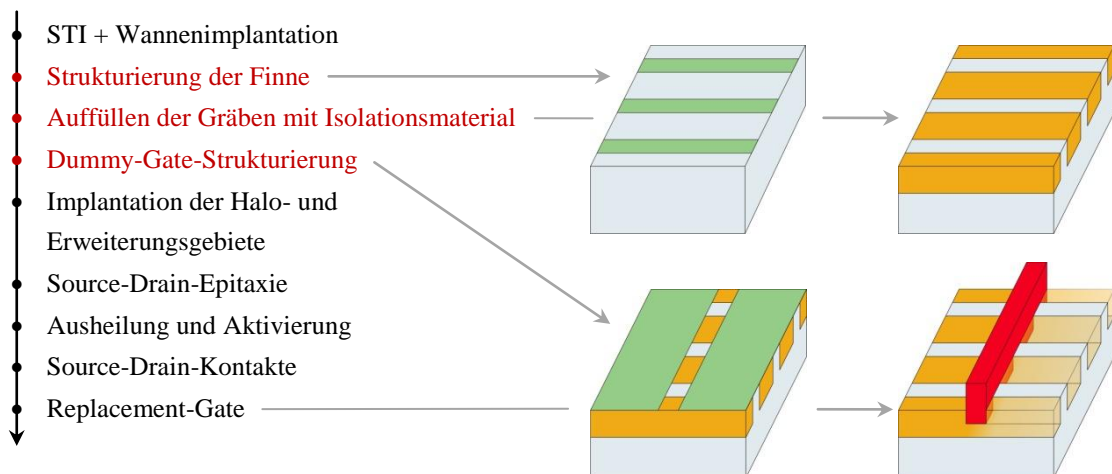


Abbildung 5.1: Prozessablaufplan zur Strukturierung der Tri-Gate-Finnen innerhalb der hybriden 22 nm-Technologie.

Im später folgenden Replacement-Gate-Prozess wird das Isolationsmaterial um die angestrebte Finnhöhe abgesenkt, wodurch sich aus den Seitenflächen der beschriebenen Gräben die zusätzlichen Seiten-Gates ergeben. Eine erste Dimensionierung des Tri-Gate-Transistors sah vor, dass die Finnen eine Weite von 20 nm und eine Höhe von 20 nm haben und sich diese Struktur aller 50 nm wiederholt ($Pitch_{Fin}$).

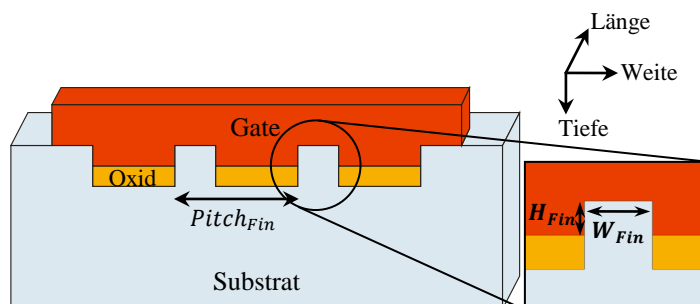


Abbildung 5.2: Querschnitt des Tri-Gate-Transistors mit H_{Fin} als Höhe, W_{Fin} als Weite und $Pitch_{Fin}$ als Abstand der Finnen.

5.1.3 Strukturierung der Finne

Der im vorhergehenden Abschnitt beschriebene Ansatz zur Strukturierung des Tri-Gate-Transistors entspricht der Variante 1 aus Abbildung 5.3 mit einem Graben für das Isolationsmaterial, der sich über die gesamte Länge des Transistors erstreckt. Das bedeutet jedoch auch, dass die Source-Drain-Gebiete automatisch mit verkleinert werden, wodurch ein erhöhter Kontaktwiderstand zu erwarten ist. Aus diesem Grund wurde nach weiteren Möglichkeiten zur Strukturierung der Finnen gesucht, die ebenso einfach in die hybride 22 nm-Technologie zu integrieren sind und eine verbesserte Leitfähigkeit der Source-Drain-Gebiete ermöglichen.

Variante 2 beruht auf der Vertauschung von Dummy-Gate- und Finnenstrukturierung. Dabei wird zunächst eine negative Hartmaske für das Dummy-Gate auf das Substrat aufgebracht, sodass ein Graben entlang der Transistorweite den Bereich für den späteren Gatekontakt markiert. Anschließend wird über diesen Graben die Maske zur Strukturierung der Finne gelegt, die Isolationsgräben geätzt und mit Oxid wieder aufgefüllt. Dadurch werden die Source-Drain-Gebiete vor der Finnenstrukturierung geschützt und bleiben vollständig erhalten.

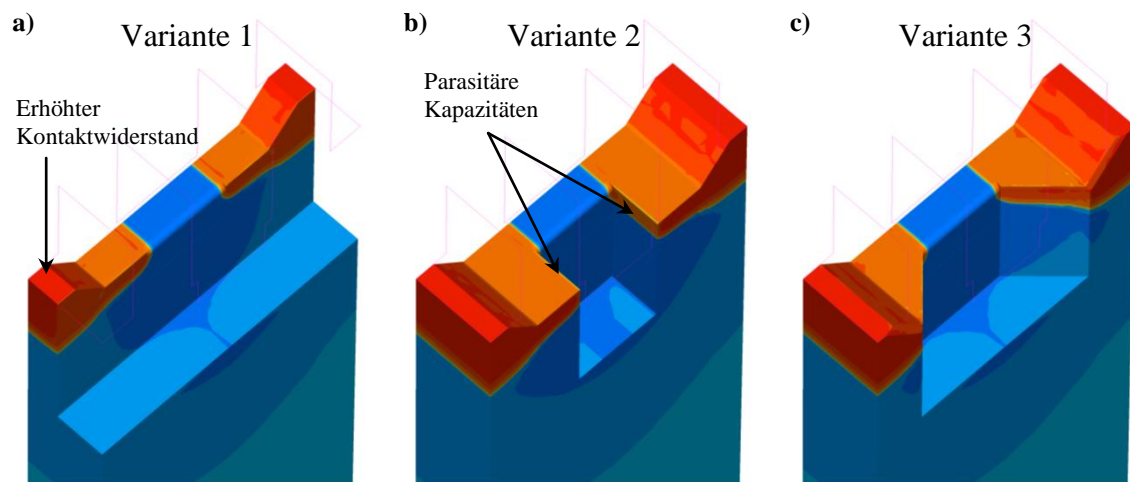


Abbildung 5.3: Varianten zur Strukturierung der Finne. a) Variante 1 mit durchgängigem Isolationsmaterial über der gesamten Transistorlänge, b) Variante 2 mit Isolationsmaterial im Bereich des Gates und durchgängigen Source-Drain-Gebieten und c) Variante 3 mit zulaufenden Source-Drain-Gebieten.

Der Nachteil dieser Variante sind die zusätzlichen parasitären Gatekapazitäten, die zwischen den abgesenkten Seiten-Gates und den Source-Drain-Gebieten entstehen. Diese sind für das dynamische Verhalten des Transistors nachteilig, da sich an den Flächen Ladungen sammeln, die nicht zum Stromfluss beitragen, jedoch bei jedem Schaltvorgang mit umgeladen werden. Als Kompromiss beider Lösungen dient die Variante 3, bei der sich die Source-Drain-Gebiete zur Finne hin verjüngen. Die Realisierung dieses Ansatzes ähnelt der ersten Variante, weist jedoch als Besonderheit auf, dass die Maske

zur Finnenstrukturierung den zulaufenden Source-Drain-Gebieten entsprechend angepasst wird. Die Umsetzung könnte aufgrund von auftretender Kantenrauigkeit zwar mit einer erhöhten Ungenauigkeit der resultierenden Finweite verbunden sein, wird aber als möglich eingeschätzt [21].

Ein Vergleich der Gate-Source-Kapazitäten bestätigte die Annahme, dass die parasitären Kapazitätsanteile der Variante 2 gegenüber der Variante 1 eine deutliche Erhöhung bewirken (Abbildung 5.4). Auffällig war dabei der markierte Anteil vor dem Einsetzen der Inversion des Kanalgebiets. Andererseits konnte für Variante 2 auch eine Reduzierung des Durchgangswiderstandes beobachtet werden. Als geeigneter Mittelweg erwies sich die Variante 3 mit den zulaufenden Source-Drain-Gebieten. Die leicht erhöhte Gate-Source-Kapazität gegenüber Variante 1 und der vergleichbare Widerstand zu Variante 2 sollten das beste Schaltverhalten für die Tri-Gate-Transistoren erzielen.

Für den p-MOSFET gelten die gleichen Zusammenhänge, allerdings variiert durch die Veränderung der Form der Source-Drain-Gebiete auch der Verspannungseintrag des eingebetteten Silizium-Germaniums. Mit dem verringerten Volumen der Source-Drain-Gebiete von Variante 1 ergibt sich auch eine schwächere Verspannung des Kanalgebiets, was in einem geringeren Drainstrom und einer erhöhten Schwellspannung resultiert. Welche der Varianten für die Eigenschaften der Inverter und Ringoszillatoren vorteilhafter ist, wird im Abschnitt 5.2.6 näher untersucht.

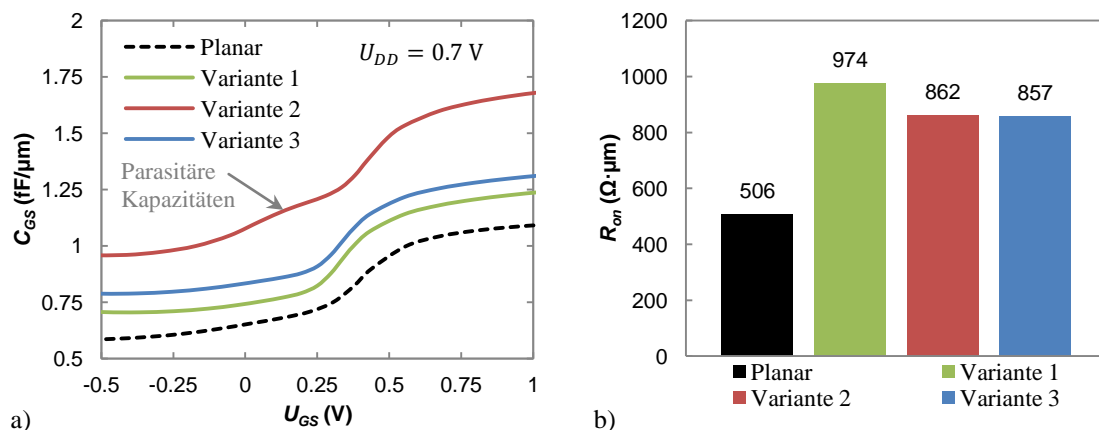


Abbildung 5.4: Vergleich der drei Varianten zur Strukturierung der Tri-Gate-Finne. a) Gate-Source-Kapazität und b) Durchgangswiderstand bei konstanter Inversion der n-MOSFETs. ($L_{Gate} = 26$ nm)

5.1.4 Geometrieabhängiges Dotierungsprofil

Bei den Betrachtungen der Tri-Gate-Transistoren muss das geometrieabhängige Dotierungsprofil berücksichtigt werden (Abbildung 5.5). Während der planare Transistor in der Weite uneingeschränkt ist und sein Dotierungsprofil in diese Richtung vereinfacht als konstant angesehen werden kann, bewirkt die Strukturierung im Bereich der Finne, dass sich sowohl die Implantation als auch die Diffusion der Dotanden nach der

Geometrie der Finne richten. Eine erhöhte Streuung der implantierten Ionen an den Grenzflächen der Seiten-Gates und das verringerte Kanalvolumen sorgen dafür, dass bereits bei der Implantation weniger Dotanden das Kanalgebiet erreichen und durch Zusammenstöße mit den Atomrüpfen in der Finne abgebremst werden. Hinzu kommt die Segregation der Ionen an den Seiten-Gates und das zusätzliche Ausgasen der Borionen, welche die Dotierungskonzentration sinken lassen [11].

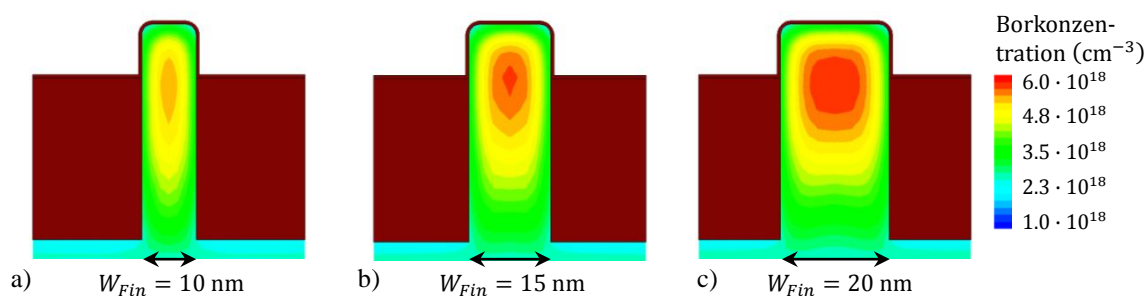


Abbildung 5.5: Konzentration aktivierter Borionen in der Finne von n-Tri-Gate-Transistoren mit unterschiedlicher Finweite, jedoch identischen Implantationsparametern.

Dieser Effekt tritt vor allem an den Seiten der Finne auf und besitzt eine Eindringtiefe von ca. 10 nm, was bedeutet, dass erst eine Finne mit einer Weite $W_{Fin} \geq 20$ nm im Zentrum eine zum planaren Transistor vergleichbare Kanaldotierung aufweisen kann. Je schmaler das Kanalgebiet wird, um so geringer ist die endgültige Konzentration an Dotanden im Kanal, auch wenn alle anderen Technologieparameter unverändert bleiben. Die fehlende Gegendotierung bewirkt, dass die Schwellspannung des Tri-Gate-Transistors rapide abfällt, da die Austrittsarbeit des Gatemetalls im hybriden Prozessansatz nicht an die verringerte Kanaldotierung angepasst werden kann. Aus diesem Grund wurde für die meisten Simulationen der Tri-Gate-Transistoren eine Weite von $W_{Fin} = 20$ nm verwendet.

5.2 Charakterisierung des elektrischen Verhaltens

5.2.1 Stationäres Verhalten

Vergleichend zu den planaren MOSFETs der hybriden 22 nm-Technologie soll das stationäre Verhalten der Tri-Gate-Transistoren betrachtet werden (Abbildung 5.6). Schon bei der Analyse des Schwellspannungsverlaufs über der Gatelänge fällt auf, dass die Tri-Gate-Transistoren eine deutlich bessere Steuerfähigkeit des Gates und einen flacheren Roll-off-Verlauf aufweisen. Somit bietet der Tri-Gate-Transistor mit seinen elektrischen Eigenschaften vorläufig die Möglichkeit, die physikalische Gatelänge weiter zu skalieren. Auch im Hinblick auf die Leistungsfähigkeit konnte gegenüber den

planaren MOSFETs eine Steigerung um 9 % (n-MOSFET) bzw. 35 % (p-MOSFET) erzielt werden. Die Normierung der Ströme erfolgte dabei auf die effektive Gateweite mit $W_{Gate} = 34$ nm, um die Leistungsfähigkeit des Kanals an sich zu betrachten (Tabelle 5.1). Mit der Normierung auf die Strukturweite bei $Pitch_{Fin} = 50$ nm würden die Tri-Gate-Transistoren jedoch schlechter dargestellt als die planaren MOSFETs. Um dies zu vermeiden, sollte der Abstand der Finnen maximal so groß sein wie die effektive Gateweite selbst.

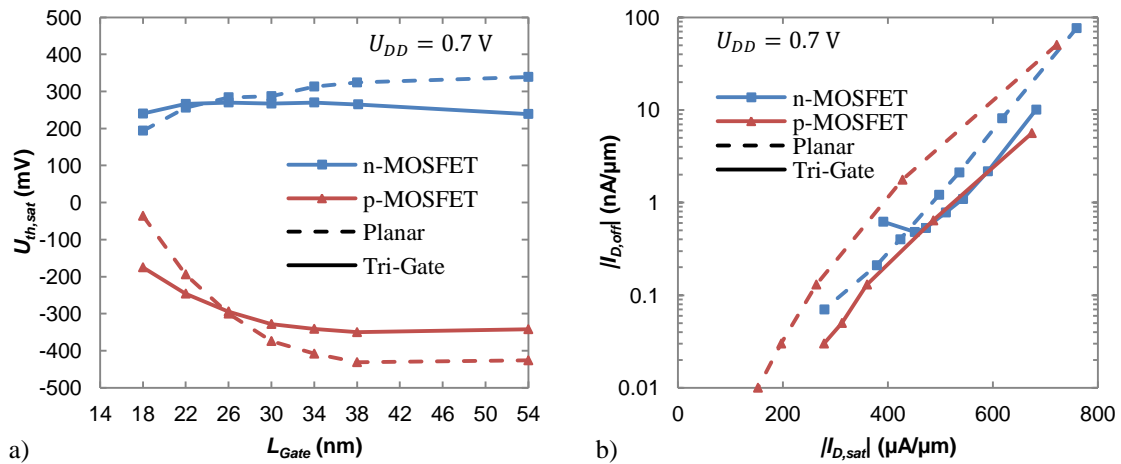


Abbildung 5.6: Berechnetes Verhalten der Tri-Gate-Transistoren in 22 nm-Technologie im Vergleich zu planaren MOSFETs. a) Schwellspannungsverhalten in Abhängigkeit von der Gatelänge und b) Universalkennlinie von n- und p-MOSFETs im Sättigungsbereich mit Normierung auf die effektive Gateweite. ($H_{Fin} = 7$ nm, $W_{Fin} = 20$ nm, $Pitch_{Fin} = 50$ nm)

Die deutlichere Leistungssteigerung des p-Tri-Gates ist auf die Verspannung des eingebetteten Silizium-Germaniums zurückzuführen, da die Strukturierung der Finne einen ähnlichen Effekt wie der Replacement-Gate-Prozess hervorruft und sich entsprechend eine stärkere Kanalverspannung ergibt [152]. Somit unterscheiden sich beide Transistortypen in ihrer Leistungsfähigkeit kaum noch und können bei den integrierten Schaltkreisen mit den gleichen Aktivflächen prozessiert werden.

Tabelle 5.1: Berechnete stationäre Kenndaten von Tri-Gate-Transistoren in 22 nm-Technologie mit Normierung auf die effektive Gateweite, $Pitch_{Fin} = 50$ nm, $L_{Gate} = 26$ nm, $W_{Fin} = 20$ nm, $H_{Fin} = 7$ nm bzw. $U_{DD} = 0.7$ V.

Typ	U_{th} (mV)		$DIBL$ (mV/V)	SS (mV/dec)		g_m (μ S/ μ m)		I_D (μ A/ μ m)		$I_{D,off}$ (nA/ μ m)	R_{on} ($\Omega \cdot \mu$ m)
	lin.	sat.		lin.	sat.	lin.	sat.	lin.	sat.		
n-MOS	312	270	60	74.8	74.3	240	1248	111	543	1.09	857
p-MOS	-350	-294	80	75.7	75.9	184	1436	88	486	0.64	1011

5.2.2 Kurzkanaleffekte und Gatesteuerung

Im Vergleich zum planaren 22 nm-Transistor zeigt sich die verbesserte Steuerfähigkeit des Tri-Gates auch anhand der Darstellung des *DIBLs* und der Unterschwellsteigung (Abbildung 5.7). Für beide Kenngrößen liegen die Werte deutlich unter denen der planaren MOSFETs. Es konnte jedoch beobachtet werden, dass der Punkt, an dem die Unterschwellsteigung und der *DIBL* ansteigen, den planaren Transistoren gleich, bei einer Gatelänge von ca. $L_{Gate} = 30$ nm liegt.

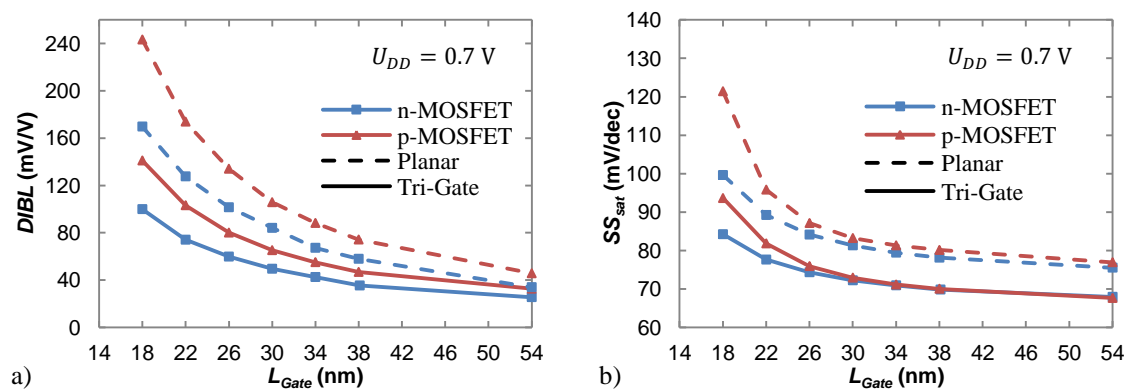


Abbildung 5.7: Berechnetes Kurzkanalverhalten der Tri-Gate-Transistoren in 22 nm-Technologie im Vergleich zu planaren MOSFETs. a) *DIBL* und b) Unterschwellsteigung (Sättigung) in Abhängigkeit von der Gatelänge. ($H_{Fin} = 7$ nm, $W_{Fin} = 20$ nm, $Pitch_{Fin} = 50$ nm)

Im Bezug auf die linearen Unterschwellsteigungen liegen die Tri-Gate-Transistoren näher an der berechneten Ideallinie als die planaren MOSFETs (Abbildung 5.8). Dies gilt vor allem für den Bereich der größeren Gatelängen, wohingegen die kürzeren Transistoren sich schnell dem Verlauf des planaren Niveaus annähern. Für diesen Vergleich wurde die natürliche Länge der n- und p-Tri-Gate-Transistoren nach Gleichung (2.27) auf $\lambda_{Tri} = 3.34$ nm bestimmt.

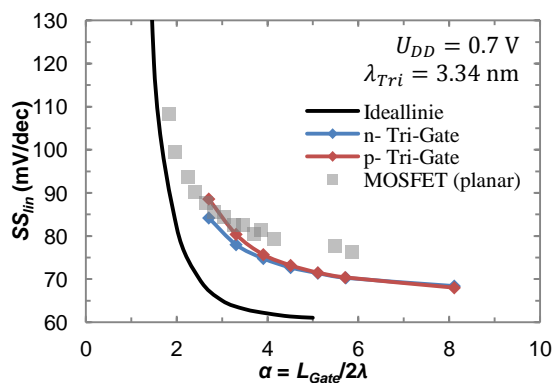


Abbildung 5.8: Berechnete lineare Unterschwellsteigung in Abhängigkeit von dem Skalierungsparameter α für die n- bzw. p-Tri-Gate-Transistoren der 22 nm-Technologie.

Die Strukturierung zum Tri-Gate konnte die Kurzkanaleffekte jedoch nicht signifikant unterdrücken, sondern verbesserte hauptsächlich die Steuerfähigkeit des Kanalgebiets durch den dreiflächigen Gatekontakt. Die Betrachtung der Ausgangskennlinien bestätigt dies, da sich auch das Sättigungsverhalten beider Transistortypen nicht deutlich verbesserte (Abbildung 5.9). Positiv hervorzuheben ist jedoch die Tatsache, dass sowohl die Werte der Unterschwellsteigung als auch des *DIBLs* für die Tri-Gate-Transistoren mit einer Gatelänge von $L_{Gate} = 18$ nm denen der planaren MOSFETs mit $L_{Gate} = 26$ nm entsprechen und somit die Skalierung hin zur 14 nm-Technologie mit den Tri-Gate-Transistoren in Hinblick auf die Steuerfähigkeit ohne einen Verlust vollzogen werden kann.

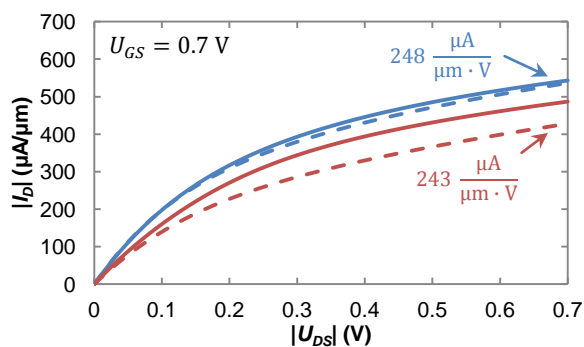


Abbildung 5.9: Simulierte Ausgangskennlinien der n- bzw. p-Tri-Gate-Transistoren der hybriden 22 nm-Technologie im Vergleich zu den planaren MOSFETs. ($Pitch_{Fin} = 50$ nm, $L_{Gate} = 26$ nm, $H_{Fin} = 7$ nm, $W_{Fin} = 20$ nm)

5.2.3 Eckeneffekt

Der Eckeneffekt spielt für den beschriebenen Tri-Gate-Transistor eine entscheidende Rolle, da er aufgrund der hohen Dotierung im Kanalgebiet und der überlagernden Wirkung der drei Gateflächen sehr stark ausgeprägt ist. Deutlich wird dies bei einem Vergleich der Transferkennlinien mit dem planaren Transistor der hybriden 22 nm-Technologie (Abbildung 5.10).

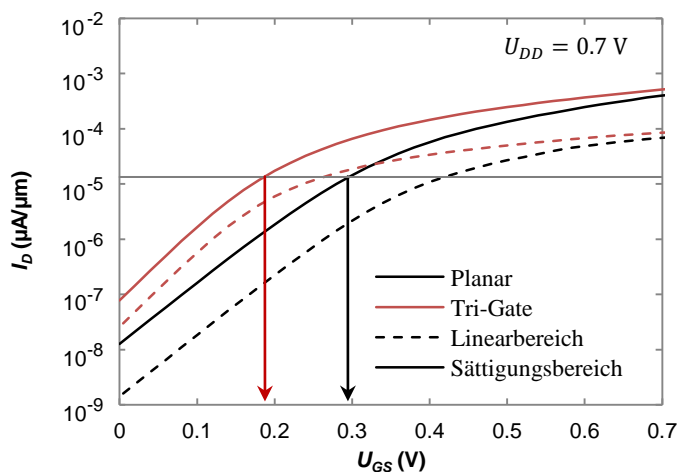


Abbildung 5.10: Transferkennlinien eines planaren und eines Tri-Gate-Transistors zur Darstellung der Schwellspannungverschiebung aufgrund des Eckeneffektes und der veränderten Unterschwellsteigung.
 $U_{Th,planar} = 284$ mV
 $U_{Th,Tri-Gate} = 180$ mV

Trotz der gleichen Implantationsparameter für die Halogegebiete besitzt der Tri-Gate-Transistor einen höheren Sperrstrom und eine niedrigere Schwellspannung. Es sind jedoch keine Leckströme, die dieses Verhalten verursachen, sondern vielmehr sorgt der Eckeneffekt dafür, dass die Kanalgebiete in den Ecken der Finne gegenüber den planaren Abschnitten vorzeitig in die Inversion getrieben werden und somit auch im gesperrten Zustand ($U_{GS} = 0.0 \text{ V}$) ein höherer Strom fließt. Hinzu kommt der steilere Anstieg der Unterschwellsteigung, wodurch der Wert für die Bestimmung der Schwellspannung mit der Konstantstrommethode eher erreicht wird.

Zur Optimierung des Tri-Gate-Transistors konnten für den Eckeneffekt analytische Modelle erarbeitet werden, die verschiedene Abhängigkeiten berücksichtigen. Hierfür wurde der Eckeneffekt (Corner Effect, CE) mit

$$CE = \frac{\text{Ladungsträgerdichte(Ecke)}}{\text{Ladungsträgerdichte(Top-Gate)}} \quad (5.1)$$

als Quotient der Ladungsträgerdichten im Bereich der Ecke und zentral unterhalb des Top-Gates definiert. Der Abstand zum Gateoxid beträgt jeweils 1 nm (Abbildung 5.11).

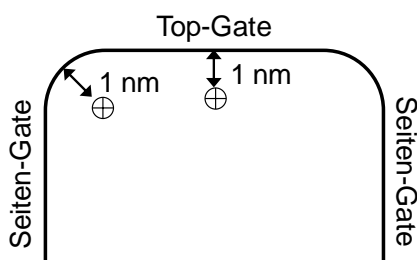


Abbildung 5.11: Positionen zur Bestimmung der Ladungsträgerdichten in der Finne eines Tri-Gate-Transistors.

Mit der Darstellung von CE in Abhängigkeit von der Gate-Source-Spannung kann man das spannungsversetzte Schalten der Tri-Gate-Bereiche ebenfalls gut verdeutlichen (Abbildung 5.12a) [153]. Für hohe Kanaldotierungen erkennt man bereits bei negativen Spannungen einen deutlichen Anstieg der freibeweglichen Ladungsträger in den Ecken, der erst bei einer positiven Gate-Source-Spannung kompensiert werden kann, die größer als die planare Schwellspannung ist. Bei starker Inversion ($U_{GS} \geq 1.0 \text{ V}$) nähern sich die Kurven dem Grenzwert $CE = 2$ an, welcher aufgrund der Überlagerung der beiden orthogonalen Felder in den Ecken von Top- und Seiten-Gate nicht unterschritten werden kann. Im negativen Bereich von U_{GS} wird zudem am Punkt $CE = 1$ der Flachbandzustand des Kanals sichtbar.

A. Dotierungskonzentration

Um ein analytisches Modell für den Eckeneffekt aufzustellen, wurde ausschließlich der Einfluss auf den Sperrstrom bei $U_{GS} = 0 \text{ V}$ untersucht, wobei das Verhältnis der Ladungsträgerdichten

$$CE(N_{Ch}) = 3 \cdot e^{\left(\frac{N_{Ch}}{1.7 \cdot 10^{18} \text{ cm}^{-3}}\right)} \quad (5.2)$$

als Funktion der Kanalkonzentration einen exponentiellen Anstieg aufweist (Abbildung 5.12b). Es ist demnach sinnvoll, die Dotierungskonzentration im Kanal so gering wie möglich zu halten. Andererseits ergibt sich damit eine erhöhte Unterschwellsteigung und somit ein schlechteres Schaltverhalten des Transistors. Um gegenüber dem planaren Transistor noch eine bessere Steuerfähigkeit des Gates zu sichern, sollte die Konzentration der Kanaldotierung bei dieser Tri-Gate-Architektur nicht unter den Wert von $N_{Ch} = 2 \cdot 10^{18} \text{ cm}^{-3}$ fallen. Mit den angenommenen Implantationsparametern des hybriden Prozesses ergibt sich daraus jedoch eine Dotierungskonzentration von $N_{Ch} = 5 \cdot 10^{18} \text{ cm}^{-3}$.

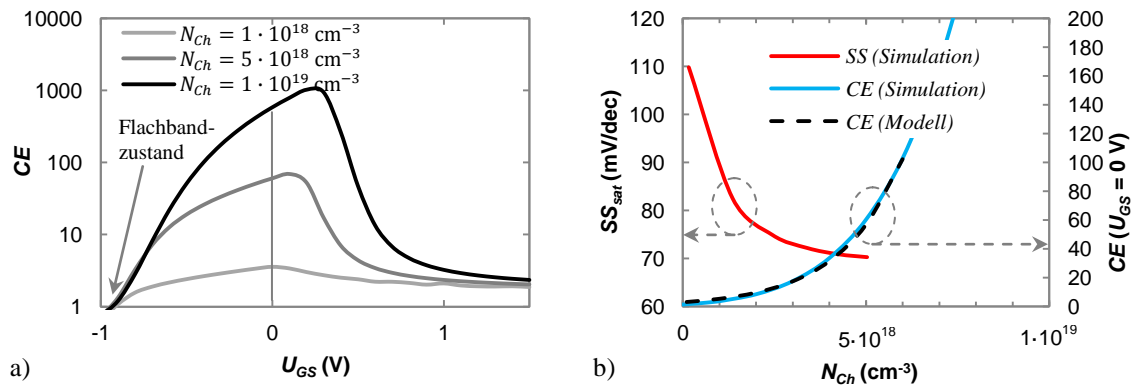


Abbildung 5.12: Ausprägung des Eckeneffektes CE in Abhängigkeit von der Dotierungskonzentration im Kanalgebiet eines Tri-Gate-Transistors als Funktion der a) Gate-Source-Spannung und b) der Dotierungskonzentration bei $U_{GS} = 0 \text{ V}$. ($L_{Gate} = 26 \text{ nm}$, $H_{Fin} = 10 \text{ nm}$, $W_{Fin} = 20 \text{ nm}$)

B. Finweite

Da der Eckeneffekt eine Folge der Tri-Gate-Geometrie ist, besteht auch eine Abhängigkeit von Höhe und Weite der Finne. Jedoch beeinflussen sich beide Größen wechselseitig, sodass die ermittelten Funktionen keine allgemeine Gültigkeit besitzen und nur den jeweilig untersuchten Spezialfall repräsentieren. Die qualitativen Verläufe der Funktionen ändern sich jedoch nicht.

Wie in [154] und [155] gezeigt wurde, nähert sich die Unterschwellsteigung mit schmaler werdender Finne einem theoretischen Optimum von rund 60 mV/dec an. In Abbildung 5.13a wird jedoch deutlich, dass eine dotierte Tri-Gate-Finne für $W_{Fin} \leq 15 \text{ nm}$ einen vollständig verarmten Zustand annimmt, wodurch sich die Schwellspannung stark absenkt. Dies resultiert aus einer Verschiebung der Transferkennlinie in negative Richtung der Gate-Source-Spannung, wodurch sich auch die Unterschwellsteigung bei $U_{GS} = 0 \text{ V}$ erhöht. Bei einer Vergrößerung der Finweite

kommt es neben der Erhöhung der Unterschwellsteigung durch die verminderte Steuerfähigkeit des Tri-Gates zu einer Zunahme des Eckeneffektes (Abbildung 5.13b). Somit ergibt sich der analytische Zusammenhang

$$CE(W_{Fin}) = K_W(N_{Ch}, H_{Fin}) \cdot \left(1 - e^{-\left(\frac{W_{Fin}}{47 \text{ nm}}\right)^5}\right), \quad (5.3)$$

wobei die Konstante K_W den maximal erreichbaren Wert beschreibt und von Kanal-dotierung sowie Finhöhe abhängt. Das Verhalten zeigt zunächst einen steilen Anstieg mit der fünften Potenz, bis die Kurve bei $W_{Fin} = 47 \text{ nm}$ einen Wendepunkt durchläuft und zum Maximum hin abflacht. Ab diesem Punkt kann davon ausgegangen werden, dass die Seiten-Gates so gut wie keinen Einfluss mehr auf die Ladungsträgerdichte in der Mitte des Top-Gates haben und der Bereich als rein planar angesehen werden kann.

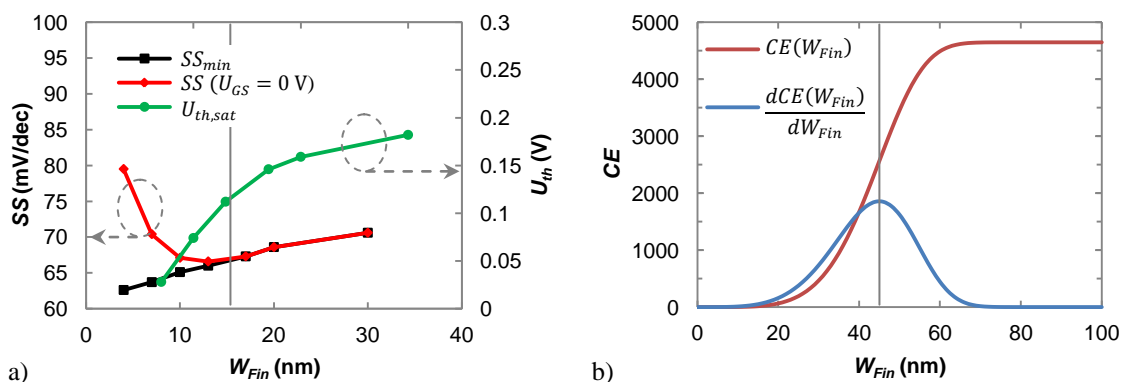


Abbildung 5.13: Einfluss der Finweite W_{Fin} auf das stationäre Verhalten von Tri-Gate-Transistoren. a) Unterschwellsteigung und Schwellspannung sowie b) Eckeneffekt in Abhängigkeit von der Finweite bei $U_{GS} = 0 \text{ V}$. ($L_{Gate} = 26 \text{ nm}$, $H_{Fin} = 10 \text{ nm}$, $N_{Ch} = 5 \cdot 10^{18} \text{ cm}^{-3}$)

C. Finhöhe

Mit der Finhöhe verändert sich, ähnlich zur Finweite, die Unterschwellsteigung und der Eckeneffekt (Abbildung 5.14). Ausgehend von einem Wert $H_{Fin} = 0 \text{ nm}$, welcher den Eigenschaften des planaren Transistors sehr nahe kommt, fällt die Unterschwellsteigung mit zunehmender Höhe der Finne ab, bis sie gegen ein Minimum läuft. Als Ausgangspunkt für die Simulationen wurde eine Finweite vom $W_{Fin} = 20 \text{ nm}$ und eine Kanalkonzentration von $N_{Ch} = 5 \cdot 10^{18} \text{ cm}^{-3}$ gewählt. Anhand des Verlaufs wird deutlich, dass die Finhöhe mindestens einen Wert von 7 nm besitzen sollte, um die Steuerfähigkeit des Tri-Gates zu erhalten. Zur Untersuchung des Eckeneffektes in Abhängigkeit von der Finhöhe wurde die Finweite auf 400 nm erhöht, um ein gegenseitiges Überlagern der Seiten-Gates vollständig ausschließen zu können. Somit ergab sich

$$CE(H_{Fin}) = K_H(N_{Ch}, W_{Fin}) \cdot \left(1 - e^{-\left(\frac{H_{Fin}}{15 \text{ nm}}\right)^2}\right). \quad (5.4)$$

Die Konstante K_H beschreibt hierbei den Wert, den der Eckeneffekt aufgrund der Dotierungskonzentration und der Finweite annimmt. Der Wendepunkt der Kurve bei $H_{Fin} = 10 \text{ nm}$ macht deutlich, dass bei der gegebenen Kanaldotierung die Überlagerung der Feldstärken von Top- und Seiten-Gate nahezu vollständig ist und eine weitere Erhöhung der Finne keine signifikante Verbesserung der Steuerfähigkeit bzw. der Schalteigenschaften mehr bewirkt.

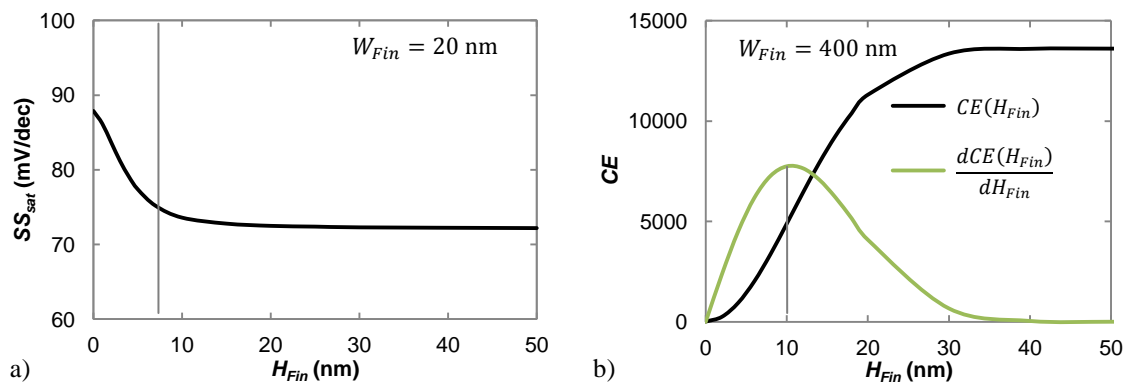


Abbildung 5.14: Einfluss der Finhöhe H_{Fin} auf das stationäre Verhalten des Tri-Gate-Transistors. a) Unterschwellsteigung sowie b) Eckeneffekt in Abhängigkeit von der Finhöhe bei $U_{GS} = 0 \text{ V}$. ($L_{Gate} = 26 \text{ nm}$, $N_{Ch} = 5 \cdot 10^{18} \text{ cm}^{-3}$)

D. Eckenradius

Zur Betrachtung der Geometrie der Finne gehört auch der Eckenradius CR , mit dem sich der Übergang zwischen Top- und Seiten-Gate beschreiben lässt. Je runder sich die Ecke des Tri-Gates gestaltet, um so geringer ist die Wirkung des Eckeneffektes. In der Realität wird stets eine Abrundung vorhanden sein, da ein Überätzen während der Finnenformierung und ein Abtrag durch Reinigungsschritte unvermeidbar sind. Die Simulationen des Eckeneffektes mit der Abhängigkeit vom Radius zeigten einen Verlauf, der einer Normalverteilung gleich kommt (Abbildung 5.15). Als analytische Lösung ergab sich

$$CE(CR) = 1 + K_{CR}(H_{Fin}, W_{Fin}, N_{Ch}) \cdot e^{-\frac{1}{2}\left(\frac{CR}{2,5 \text{ nm}}\right)^2}. \quad (5.5)$$

Der Summand „1“ als konstanter Anteil beschreibt den Extremfall eines sehr großen Radius (Nanowire), bei dem keine Unterscheidung zwischen Top- und Seiten-Gates mehr möglich ist und jeder Teil der Finne das gleiche Potenzial besitzt. Für alle weiteren Simulationen wurde ein Radius von 2 nm verwendet, welcher auch bei einer experimentelle Umsetzung der Tri-Gate-Struktur eingehalten werden sollte. Ab einem

Radios von 3 nm erkennt man, dass sich sowohl die Unterschwellsteigung als auch die Schwellspannung deutlich verändert. Die Ursache hierfür liegt in der Quantisierung der Ladungsträger an den Grenzflächen. Bei kleinen Radien wird das Konzentrationsmaximum aus der Spitze heraus getrieben, sodass die Abrundung keinen Einfluss besitzt. Vergrößert sich der Radius weiter, wird auch relevantes Kanalgebiet entfernt, was zum einen den Eckeneffekt reduziert und die Schwellspannung steigen lässt und zum anderen die Überlagerung der elektrischen Felder von Top- und Seiten-Gate bzw. die Steuerfähigkeit mindert.

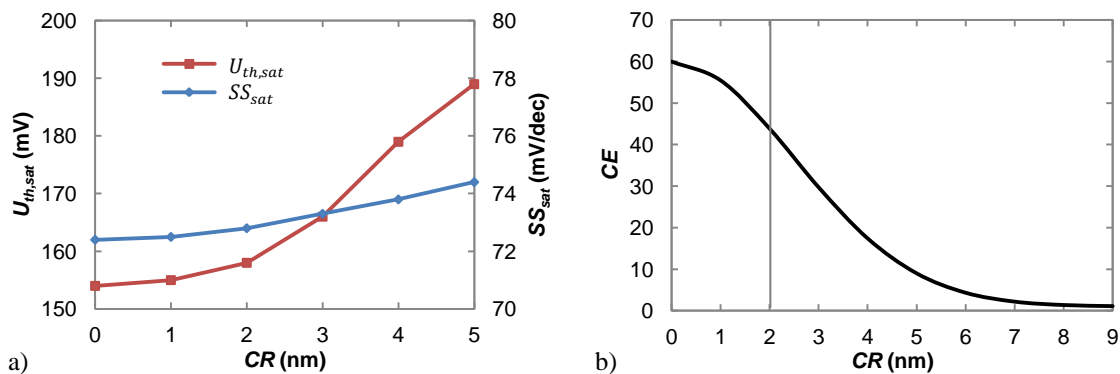


Abbildung 5.15: Einfluss des Eckenradius CR auf das stationäre Verhalten eines Tri-Gate-Transistors. a) Unterschwellsteigung SS_{sat} und Schwellspannung $U_{Th,sat}$ sowie b) Eckeneffekt CE bei $U_{GS} = 0$ V. ($H_{Fin} = 10$ nm, $W_{Fin} = 20$ nm, $N_{Ch} = 5 \cdot 10^{18}$ cm $^{-3}$)

E. Ladungsträgerverteilung in einer Finne

Mithilfe der analytischen Modelle lässt sich nun, ausgehend von einer bekannten Struktur mit simuliertem CE , die Ladungsträgerverteilung in einer Finne mit veränderter Geometrie näherungsweise prognostizieren. Somit kann zum Beispiel bei einer Reduzierung der Finne schon vor einer zeitintensiven dreidimensionalen Simulation abgeschätzt werden, ob eine vollständige Verarmung der Finne möglich ist. Im Beispiel aus Abbildung 5.16 wurde die Ausgangsstruktur mit $H_{Fin} = 10$ nm, $W_{Fin} = 20$ nm und einem simulierten Eckeneffekt von $CE_{Sim} = 22$ in der Höhe bzw. der Weite halbiert und somit CE_{Mod} berechnet. Unter Berücksichtigung eines Eckenradius von $CR = 2$ nm ergab sich folglich ein Verhältnis von $CE_{Mod} = 8$ im Vergleich zum simulierten Wert $CE_{Sim} = 10$ für den Tri-Gate-Transistor mit $H_{Fin} = 5$ nm. Auch für die 10 nm weite Finne stimmt der berechnete Wert von $CE_{Mod} = 1.4$ mit dem simulierten Wert $CE_{Sim} = 1.5$ gut überein und bestätigt die qualitative Aussagekraft der analytischen Modelle.

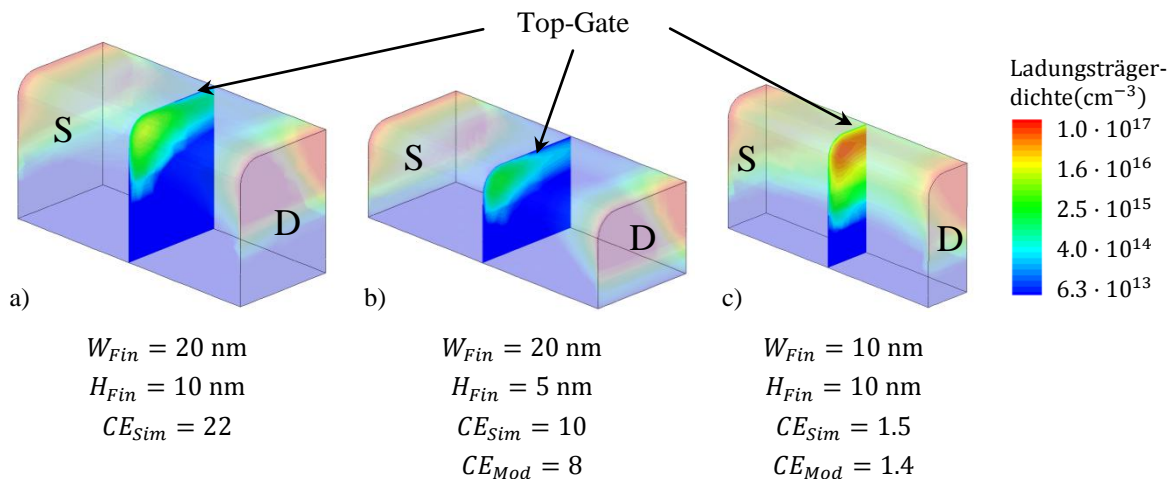


Abbildung 5.16: Simulierte Ladungsträgerverteilung in einer Finne (halbiert) mit den analytisch berechneten Werten des Eckeneffektes für unterschiedliche Geometrien eines Tri-Gate-Transistors:

a) $H_{Fin} = 10$ nm, $W_{Fin} = 20$ nm, b) $H_{Fin} = 5$ nm, $W_{Fin} = 20$ nm, c) $H_{Fin} = 10$ nm, $W_{Fin} = 10$ nm

5.2.4 Eckenimplantation

Die Eckenimplantation verfolgt das Ziel, die deutlich niedrigere Schwellspannung eines Tri-Gate-Transistors an die des planaren Transistors anzupassen. Da sich die Überlagerung von Top- und Seiten-Gates über die komplette Länge des Gates erstreckt, bedarf es zur Erhöhung der Schwellspannung auch einer Anpassung der Kanalgegendotierung über die gesamte Gatelänge [156]. Eine separate Haloimplantation für die Tri-Gate-Transistoren ist zwar umsetzbar, verhindert jedoch nicht die überhöhte Ladungsträgerkonzentration in den Ecken der Finne und somit den Effekt selbst. Die Schwellspannung von Langkanaltransistoren, bei denen sich die Halogebiete von Source und Drain nicht mehr unterhalb des Gates überlagern, würde somit im Vergleich zum planaren Langkanaltransistor dennoch deutlich niedriger ausfallen.

Die Integration der Eckenimplantation gestaltet sich verhältnismäßig einfach, da hierfür die existierende Maske der Finnenformation genutzt werden kann (Abbildung 5.17). Dieser Ansatz ermöglicht es zudem, dass die Implantation selbstjustierend zur späteren Finne ist. Nach der Aufbringung der Maske muss der p-Tri-Gate-Transistor jedoch noch mithilfe von Photolack vor der Eckenimplantation des n-Tri-Transistors geschützt werden. Die Implantation erfolgt mit der gleichen Spezies (BF_2 für den n-Tri-Gate-Transistor) wie die Haloimplantation, einem Winkel von 0° und einer vergleichsweise geringen Energie von 0.25 ... 1.0 keV, da nur oberflächennahe Bereiche, wie die Ecken der Finne selbst, dotiert werden sollen. Dabei gelangen einige Ionen durch laterale Streuung in den Randbereich unter der Hartmaske. Sie stellen später die gewünschte Erhöhung der Kanaldotierung in den Ecken der Finne dar.

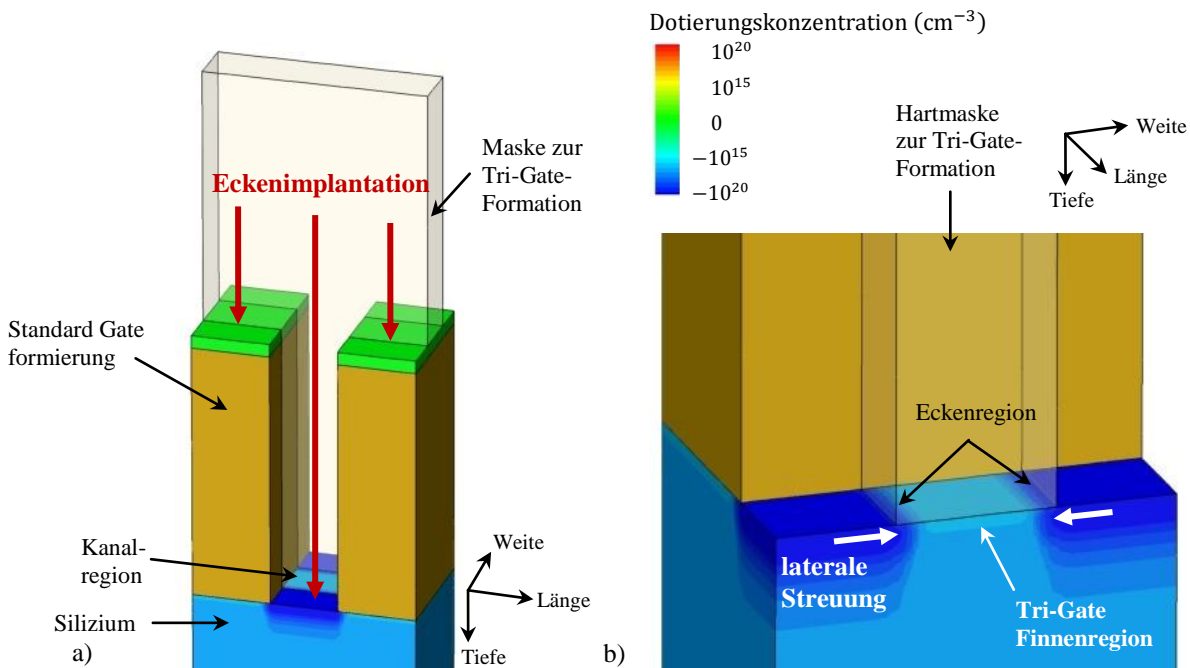


Abbildung 5.17: Realisierung der Eckenimplantation. a) Implantationen in das freiliegende Silizium mit Abschattung des Kanalgebiets durch die Maske der Finnenformation. b) Querschnitt mit Darstellung des Dotierungsprofil nach der Eckenimplantation mit lateraler Streuung der Dotanden.

Im Anschluss an die Implantation des n-Tri-Gate-Transistors wird der Photolack entfernt und der Vorgang für den p-Tri-Gate-Transistor mit der entsprechenden Implantationsspezies wiederholt. Das freiliegende Silizium, welches den Hauptteil der Implantationen abgefangen hat, wird nachfolgend mit dem anisotropen Ätzschritt zur Finnenformierung entfernt.

Um die Vorteile im elektrischen Verhalten eines Tri-Gate-Transistors mit Eckenimplantation deutlich zu machen, wurden vergleichende Simulationen zur Anpassung der Schwellspannung mithilfe der Haloimplantation durchgeführt. Hierfür wurde die ursprüngliche Dosis von $HD = 5 \cdot 10^{13} \text{ cm}^{-2}$ auf $HD = 7 \cdot 10^{13} \text{ cm}^{-2}$ bzw. $HD = 1 \cdot 10^{14} \text{ cm}^{-2}$ angehoben (Abbildung 5.18). Die Schwellspannungen der Transistoren mit den kürzeren Gatelängen sind zwar wie gewollt erhöht, die Reaktion der 54 nm-Gatelängen auf die Erhöhung der Halodosis fällt jedoch deutlich geringer aus. Der verstärkte Roll-on bzw. Roll-off der Kurve resultiert aus der ungleichmäßigen Verteilung der hohen Halodosis, was bereits in Abbildung 2.12 anhand der planaren Transistoren gezeigt wurde. Somit wird auch offenbar, dass die Haloimplantation keinen Einfluss auf den Eckeneffekt ausübt, der die Verringerung der Schwellspannung ursprünglich hervorgerufen hat. Die Universalkennlinien zeigen zudem, dass diese Variante zusätzlich mit einer Degradation der längeren Transistoren verbunden ist und sich ein unterschiedliches Verhalten zwischen den kurzen und den längeren Gatelängen einstellt.

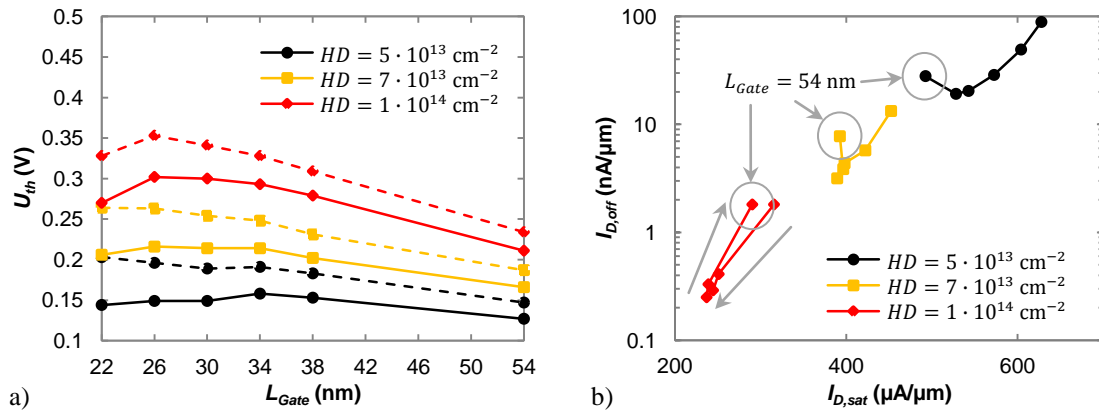


Abbildung 5.18: n-Tri-Gate-Transistoren ohne Eckenimplantation, jedoch mit unterschiedlichen Halo-dosen HD . a) Berechneter Verlauf der Schwellspannung im Linear- und Sättigungsbereich und b) berechnete Universalkennlinien.

Die Eckenimplantation bietet im Vergleich zu einer getrennten Haloimplantation hierbei nicht nur einen geringeren Aufwand in der Prozessintegration, sondern auch das bessere elektrische Verhalten der Tri-Gate-Transistoren (Abbildung 5.19). Die Schwellspannungen der einzelnen Gatelängen ließen sich somit alle um einen konstanten Wert verschieben, sodass der Verlauf für die unterschiedlichen Dosen der Eckenimplantation beibehalten werden konnte. Zudem kam es auch nicht zu einer Degradation der Universalkennlinie, sondern bedingt durch die geänderten Schwellspannungen lediglich zu einer Verschiebung entlang der Kurve. Die Dosis der Eckenimplantation, mit der letztendlich eine vergleichbare Schwellspannung zum planaren 22 nm-Transistor erreicht werden konnte, ist mit $CD = 5 \cdot 10^{14} \text{ cm}^{-2}$ sehr hoch, was jedoch notwendig ist, da nur ein geringer Teil davon durch die laterale Streuung in das Kanalgebiet gelangt.

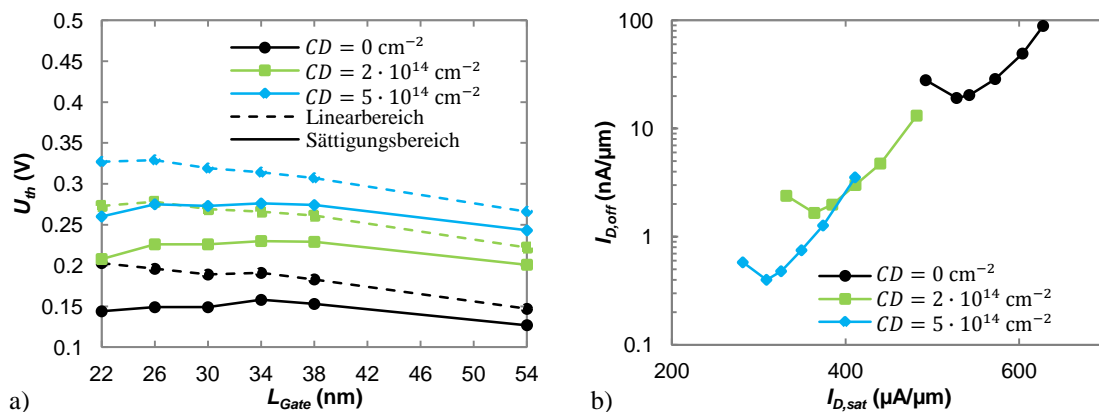


Abbildung 5.19: n-Tri-Gate-Transistoren mit Eckenimplantation verschiedener Dosen CD bei konstanter Halodosis $HD = 5 \cdot 10^{13} \text{ cm}^{-2}$. a) Berechneter Verlauf der Schwellspannung im Linear- und Sättigungsbereich und b) berechnete Universalkennlinien.

5.2.5 Finnengeometrie

Durch eine Änderung von Höhe und Weite der Finne verändern sich auch die elektrischen Eigenschaften des Tri-Gate-Transistors [157]. Für die Leistungsfähigkeit und die Unterschwellsteigung eines FinFETs ist eine hohe und schmale Finne von Vorteil [158]. Der vorgestellte Tri-Gate-Transistor zeigte jedoch hinsichtlich der gateweitennormierten Drainströme ein abweichendes Verhalten (Abbildung 5.20), da die Prozessierung der Source-Drain-Gebiete mithilfe einer Implantation nur einen oberflächennahen Kanalschluss der Seiten-Gates zulässt. Eine große Finnhöhe bedeutet somit auch eine schlechte Ausnutzung der Seiten-Gates, wodurch der Sättigungsstrom ab- und der Sperrstrom zunimmt. Betrachtet man die Abhängigkeit von der Finweite, fällt auf, dass sich der Tri-Gate-Transistor wie angenommen verhält und eine Verringerung der Finweite demnach eine Steigerung der Transistorleistung zur Folge hat.

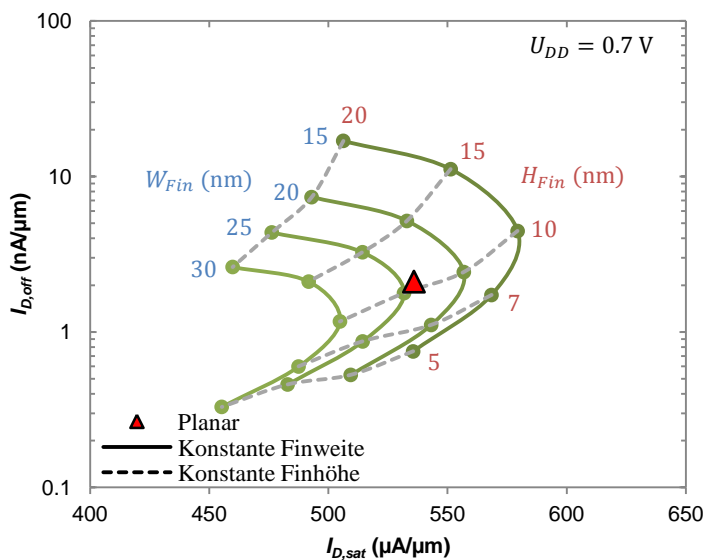


Abbildung 5.20: Verhältnis von Sättigungs- zu Sperrstrom für n-Tri-Gate-Transistoren mit unterschiedlicher Finnhöhe (5 ... 20 nm) und -weite (15 ... 30 nm). ($L_{Gate} = 26 \text{ nm}$)

Im Falle sehr schmaler Finnen (unter 10 nm) wird der gesamte Transistorkanal vollständig verarmt. Berücksichtigt man, dass die Implantationen und die Austrittsarbeit des Metalls wegen der planaren Transistoren unverändert bleiben, führt das zwangsläufig zu einer sehr niedrigen Schwellspannung, sodass das Bauelement nicht mehr eindeutig ausgeschaltet werden kann. Die minimale Finnhöhe sollte demnach einen Wert von $W_{Fin} = 15 \text{ nm}$ nicht unterschreiten.

Für die Betrachtung der Sensitivität von Strom und Unterschwellsteigung gegenüber möglichen Prozessschwankungen wurde eine Ausgangsstruktur mit einer Finnhöhe von 20 nm und einer Finweite von 7 nm gewählt. Als mögliche Prozessschwankung der Finnhöhe wurde eine Toleranz der entsprechenden Maske von ca. $\pm 4 \text{ nm}$ angenommen. Die Ungenauigkeit der Finnhöhe wird durch das anisotrope Ätzen des Isolationsoxids verursacht und kann ebenfalls eine Abweichung von bis zu $\pm 4 \text{ nm}$ aufweisen.

Der Einfluss auf den Sperrstrom ist im Vergleich zum Sättigungsstrom für Höhe und Weite der Finne hierbei wesentlich stärker (Abbildung 5.21). Am anfälligsten reagiert der Tri-Gate-Transistor auf die Schwankung der Finnhöhe, die im Vergleich zu denen der Finweite relativ gesehen deutlich höher sind. Dabei variiert nicht nur das Seiten-Gate in seiner Fläche, sondern auch der Eckeneffekt fällt unterschiedlich stark aus. Dass der Unterschwellbereich durch die geometrischen Schwankungen mehr variiert, erkennt man auch an der prozentualen Änderung der Schwellspannung. Die Unterschwellsteigung bleibt dabei jedoch nahezu konstant.

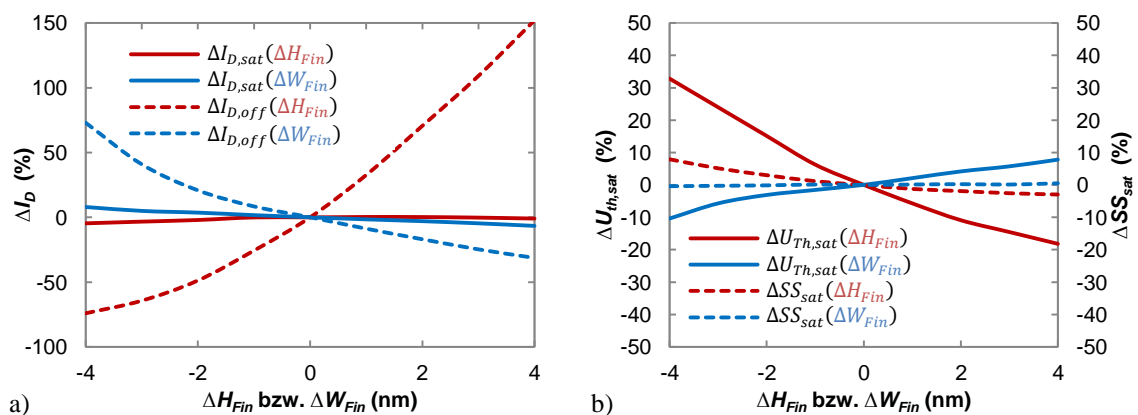


Abbildung 5.21: Sensitivität elektrischer Parameter von n-Tri-Gate-Transistoren gegenüber Prozessschwankungen von Höhe und Weite der Finne. Änderungen von a) Sperr- und Sättigungsstrom sowie b) der Schwellspannung und der Unterschwellsteigung im Sättigungsbereich in Abhängigkeit von der absoluten Schwankung der Finnhöhe und -weite. ($H_{Fin} = 7$ nm, $W_{Fin} = 20$ nm, $L_{Gate} = 26$ nm)

5.2.6 Dynamisches Verhalten

Die Gate-Source-Kapazitäten der drei beschriebenen Varianten zur Umsetzung der Finnenformation sind in Abbildung 5.4 und alle weiteren Kenndaten in Tabelle 5.2 dargestellt. Erwartungsgemäß besitzt Variante 2 mit den höchsten Werten für C_{ov} und C_{inv} auch die längsten Schaltzeiten des dreistufigen Inverters und die geringste Ringoszillatorfrequenz. Auch wenn die Variante 3 von n- und p-MOSFETs mit den zulaufenden Source-Drain-Gebieten gegenüber der Variante 1 die deutlich höheren Maximalschwing- bzw. Transitfrequenzen aufweist, so ähnelt sich das Verhalten der beiden Inverter sehr. Auch die mittleren Verzögerungszeiten, bestimmt durch die Ringoszillatorfrequenzen, unterscheiden sich nur um wenige Femtosekunden, sodass unter diesen Gesichtspunkten beide Varianten gleichwertig erscheinen.

Ähnlich der Universalkennlinie im stationären Verhalten kann mithilfe einer Darstellung der Stromaufnahme über der Ringoszillatorfrequenz eine Aussage über die dynamische Leistungsfähigkeit einer CMOS-Technologie getroffen werden (Abbildung 5.22). $I_{RO,eff}$ beschreibt dabei den effektiven Mittelwert über einer Periode der Schwingungen.

Tabelle 5.2: Berechnete Kenndaten des dynamischen Schaltverhaltens von Tri-Gate-Transistoren in 22 nm-Technologie. Die Kapazitäten der Einzeltransistoren und die Ströme des Ringoszillators sind auf die effektive Gateweite bzw. die Aktivfläche normiert. ($L_{Gate} = 26$ nm, $U_{DD} = 0.7$ V)

Einzeltransistor	Typ	C_{ov} (fF/ μ m)	C_{inv} (fF/ μ m)	f_{max} (GHz)	f_t (GHz)
Variante 1	n-MOSFET	0.741	1.188	853	259
	p-MOSFET	0.706	1.171	1113	322
Variante 2	n-MOSFET	1.147	1.629	865	229
	p-MOSFET	1.047	1.618	1002	283
Variante 3	n-MOSFET	0.835	1.259	1061	283
	p-MOSFET	0.800	1.247	1278	365

Dreistufiger Inverter	t_r (ps)	t_f (ps)	Δt_{hi} (ps)	Δt_{low} (ps)	U_{max} (V)	U_{min} (V)
Variante 1	6.50	8.30	5.20	5.74	0.782	-0.074
Variante 2	8.65	9.01	5.89	5.77	0.772	-0.071
Variante 3	6.75	8.27	5.30	5.80	0.783	-0.077

Fünfstufiger Ringoszillator	f_{RO} (GHz)	$I_{RO,eff}$ (μ A/ μ m ²)	$I_{RO,off}$ (nA/ μ m ²)	t_D (ps)	Gesamte Aktivfläche (μ m ²)
Variante 1	18.94	485.11	8.72	5.28	0.043
Variante 2	17.92	676.08	8.92	5.58	0.043
Variante 3	18.66	545.44	9.40	5.36	0.043

Der mittlere Leckstrom im nichtschwingenden Zustand wird durch $I_{RO,off}$ dargestellt. Anhand der mittleren Stromaufnahme des schwingenden Zustands in Abhängigkeit von der Ringoszillatorfrequenz wird deutlich, dass Variante 1 mit dem durchgängigen Isolationsmaterial gegenüber Variante 3 mit den zulaufenden Source-Drain-Gebieten eine etwas höhere Frequenz bei geringerem Verbrauch besitzt.

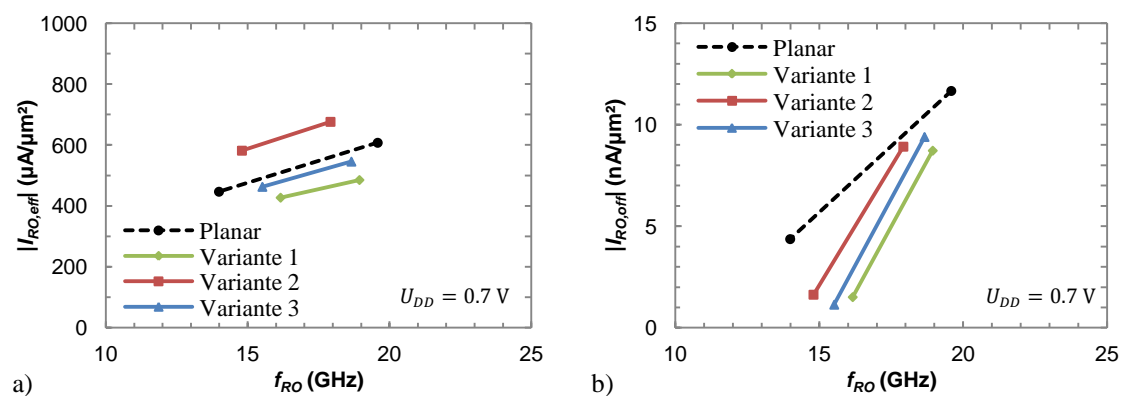


Abbildung 5.22: Stromaufnahme von fünfstufigen Ringoszillatoren der drei Varianten von Tri-Gate-Transistoren in 22 nm-Technologie im a) schwingenden und b) ausgeschalteten Zustand in Abhängigkeit von der Ringoszillatorfrequenz. Zum Vergleich sind auch die Werte für planare MOSFETs in 22 nm-Technologie dargestellt und die Stromaufnahme auf die Aktivfläche normiert. ($H_{Fin} = 7$ nm, $W_{Fin} = 20$ nm, $L_{Gate} = 26 \dots 30$ nm)

Das bedeutet, dass eine Reduzierung der parasitären Kapazitäten für das dynamische Verhalten vorteilhafter ist als die Erhöhung der Durchgangsströme zur Umladung dieser Kapazitäten. Aus diesem Grund zeigt Variante 2 mit dem Isolationsoxid nur im Bereich des Gatekontakts eine im Vergleich zu den planaren 22 nm-Transistoren schlechtere dynamische Leistungsfähigkeit.

Bei den Simulationen wurde stets davon ausgegangen, dass eine Inverterstufe jeweils mit dem Ausgang der vorherigen betrieben wird. Es muss jedoch bedacht werden, dass ein Ausgang unter Umständen auch mehrere Eingänge als Last besitzt und hierbei einen höheren Strom liefern muss. Dann ist zu erwarten, dass die Variante 3 mit den zulaufenden Source-Drain-Gebieten besser abschneidet.

5.3 Optimierung der Tri-Gate-Struktur

5.3.1 Gestaltung der epitaktischen Source-Drain-Gebiete

Die Gestaltung der epitaktisch aufgewachsenen Source-Drain-Gebiete entspricht jener der planaren Transistoren, um den Herstellungsprozess der hybriden 22 nm-Technologie für beide Transistorstrukturen nach Möglichkeit identisch zu halten und somit Aufwand und Kosten zu sparen. Für die planaren Transistoren wurde das Dotierungsprofil allerdings so angelegt, dass die pn-Übergänge nur in der Nähe der Waferoberfläche lateral bis unter das Gateoxid reichen. Damit Leckströme in tieferen Bereichen des Kanals vermieden und die Steuerfähigkeit des Transistors gesichert werden. Diese Form der Source-Drain-Gebiete verhindert jedoch die Kanalbindungen der tieferen Regionen von hohen Tri-Gate-Finnen, weshalb die optimale Finhöhe für n- und p-MOSFET höchstens 7 nm betragen sollte (Abbildung 5.23). Um die Effizienz der Tri-Gate-Transistoren bezüglich der Ausnutzung ihrer Aktivfläche zu steigern, ist es notwendig, die Finne höher zu gestalten und eine über den gesamten Bereich der Seitengates reichende Anbindung der Source-Drain-Gebiete zu realisieren. Aus diesem Grund wurde der Ätzprozess der Source-Drain-Gräben für beide Transistortypen umgestellt.

Nun wird ein anisotroper Schichtabtrag genutzt, um die angestrebte Tiefe zu erreichen. Der anschließende isotrope Ätzschritt rundet die Kanten ab und erzeugt die gewünschte U-Form. Diese Änderung bewirkt eine Leistungssteigerung der Tri-Gate-Transistoren, bezogen auf ihre Strukturweite W_{Str} , da sich die effektive Gateweite gegenüber der ursprünglichen Variante fast verdoppelt (Abbildung 5.24). Die auf die effektive Gateweite normierten Universalkennlinien haben sich jedoch kaum verschoben, was darauf hindeutet, dass die Effizienz des Gatekontakts in Verbindung mit den U-förmigen Source-Drain-Gebieten nicht gemindert wurde.

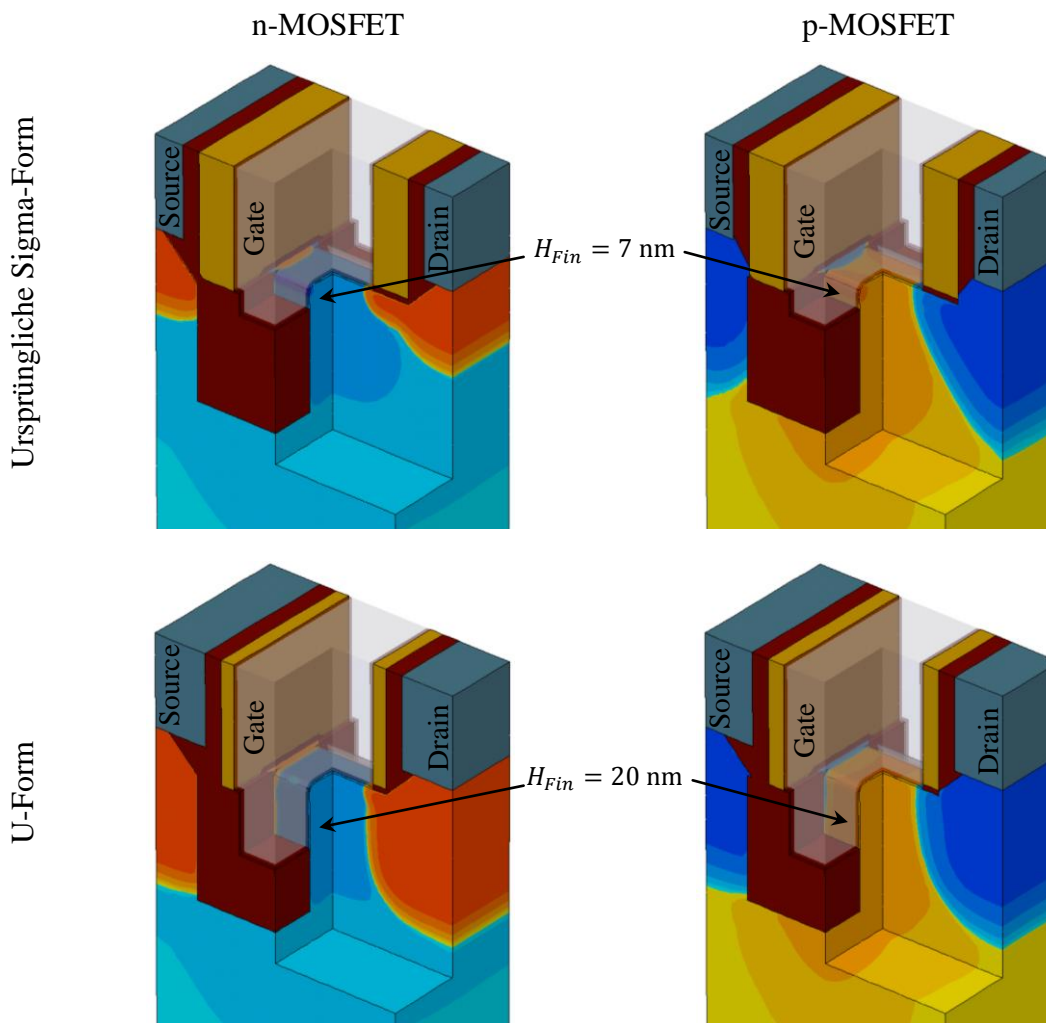


Abbildung 5.23: Verbesserte U-förmige Gestalt der epitaktisch erzeugten Source-Drain-Gebiete für n- und p-Tri-Gate-Transistoren.

Die U-förmigen Source-Drain-Gebiete verändern die elektrischen Eigenschaften der Tri-Gate-Transistoren noch in weiterer Hinsicht. So musste z. B. der Nitridspacer für n- und p-MOSFET um 7 ... 8 nm schmaler gestaltet werden, damit die Überlappung der Source-Drain-Gebiete und somit die effektive Gatelänge an die ursprünglichen Werte angepasst ist. Der zusätzliche Platz kann folglich für die Kontaktierung von Source und Drain genutzt werden, wodurch der Kontaktwiderstand verringert wird. Durch die schmalen Spacer und die tiefen Source-Drain-Gebiete erhöht sich die Gate-Source-Kapazität der Transistoren, was jedoch durch den deutlich höheren Drainstrom ausgeglichen wird. Weiterhin bewirkt die U-Form für den p-MOSFET im Vergleich zur ursprünglichen Sigma-Form eine Veränderung beim Eintrag der mechanischen Verspannung, was in dieser Arbeit jedoch nicht näher untersucht werden soll.

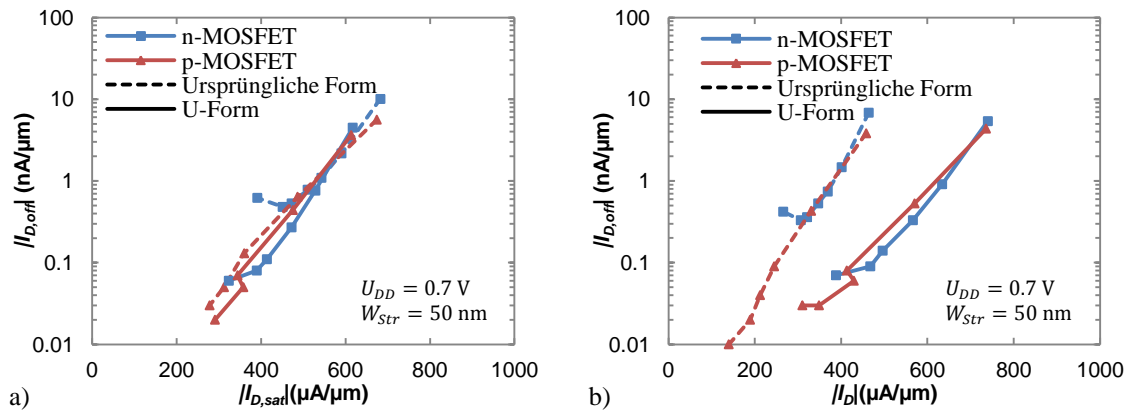


Abbildung 5.24: Universalkennlinien von n- und p-Tri-Gate-Transistoren normiert auf a) die effektive Gateweite W_{Gate} sowie b) auf die Strukturweite W_{Str} . Ursprüngliche Sigma-Form der epitaktisch erzeugten Source-Drain-Gebiete: $H_{Fin} = 7$ nm, $W_{Fin} = 20$ nm, $W_{Gate} = 34$ nm; U-Form: $H_{Fin} = 20$ nm, $W_{Fin} = 20$ nm, $W_{Gate} = 60$ nm.

5.3.2 Mechanisch verspanntes Isolationsoxid

A. Ansatz und Wirkungsweise

Die mechanische Verspannung von Transistoren ist ein sehr wichtiger Aspekt zur Erhöhung der Leistungsfähigkeit der Bauelemente. Für die Verspannung von FinFET-Strukturen kann das bereits existierende Isolationsoxid zwischen den Finnen mit einer intrinsischen Verspannung versehen werden [159]. Dabei ist es nicht zwingend erforderlich, Siliziumoxid zu verwenden. Das Material dient lediglich als Dielektrikum, um die Gatekapazität unterhalb des Seiten-Gates zu reduzieren. Es könnte daher mit einer Zug- oder Druckverspannung versehen werden (Abbildung 5.25) [160], ähnlich dem Siliziumnitrid der verspannten Deckschichten.

Die dominante Verspannung tritt dabei in Richtung der Transistorweite auf, da die Finnen auf beiden Seiten an das Oxid angrenzen. Somit kann ein direkter Gegendruck aufgebaut werden, ohne dass es zur Relaxation kommt. Von dieser Verspannungskomponente kann im Wesentlichen jedoch nur der Bereich um die pn-Übergänge profitieren, da das restliche Kanalgebiet durch die Finnenformierung den direkten Kontakt zum verspannten Isolationsmaterial verliert. Für die Komponente der mechanischen Verspannung in Kanalrichtung verhindert das Substrat mit seinem deutlich größeren Volumen gegenüber dem Isolationsoxid, dass die Finne die Verspannung des Oxids annimmt. Den geringsten Wert besitzt die vertikale Spannungskomponente (Tiefe), da sie im intrinsisch verspannten Material durch eine fehlende Begrenzung an der Oberfläche leicht relaxieren kann und somit auch keine Verspannung übertragen wird.

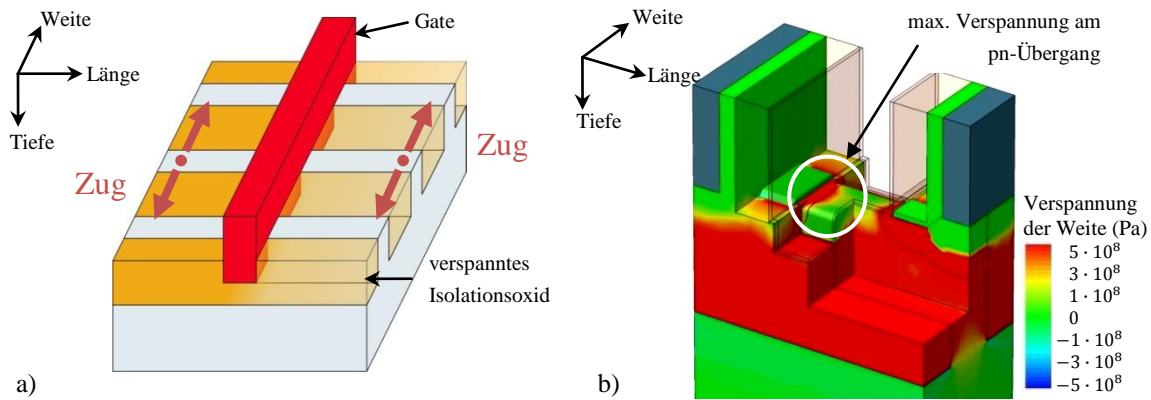


Abbildung 5.25: Verspanntes Isolationsoxid zwischen den Finnen eines Tri-Gate-Transistors. a) Prinzipielle Umsetzung und b) Profil der Verspannung in Richtung der Gateweite (1 GPa zugverspanntes Isolationsoxid).

B. Intrinsische Verspannung

Im Folgenden soll die intrinsische Verspannung der drei Hauptrichtungen quantitativ betrachtet werden. In Tabelle 5.3 sind dazu für n- und p-Tri-Gate-Transistoren die Werte der Verspannungen in die Länge, die Tiefe und die Weite des Transistors (Abbildung 5.25) an bestimmten Punkten im Kanalgebiet (zentral im Top-Gate, am pn-Übergang und im Durchschnitt) zusammengefasst. Für den n-Tri-Gate-Transistor erkennt man, dass entlang der Tiefe nur eine sehr geringe Verspannung auftritt. In Richtung des Kanals (Länge) zeigt sich dagegen eine leichte Verspannung mit einem Maximum auf Höhe der Kanalmitte. Am besten entfaltet sich die Wirkung entlang der Weite des Transistors an den pn-Übergängen. Dies wird zusätzlich durch die Amorphisierung des Siliziums infolge der Implantation der Erweiterungsgebiet begünstigt. Die Kanalmitte relaxiert jedoch stark durch den Replacement-Gate-Prozess und die Finnenformierung. Zur Kompensation könnte ein verspanntes Gatemetall eingesetzt werden [144], [161], was jedoch nicht weiter untersucht wurde.

Tabelle 5.3: Berechnete Kanalverspannung von n- und p-Tri-Gate-Transistoren mit verspanntem Isolationsoxid zentral, am pn-Übergang und als mittlere Verspannung über der gesamten Kanallänge im Top-Gate-Bereich. ($H_{Fin} = 7$ nm, $W_{Fin} = 20$ nm, $Pitch_{Fin} = 50$ nm, Werte in MPa)

	Verspannung Isolationsoxid	Länge			Tiefe			Weite		
		zentral	pn	\emptyset	zentral	pn	\emptyset	zentral	pn	\emptyset
n-MOS	0	-2	-2	-2	0	0	0	2	3	2
	1000	199	89	183	5	2	3	410	929	531
	-1000	-209	-107	-186	-3	-5	-4	-385	-881	-531
p-MOS	0	138	107	143	0	-15	-4	1	-9	0
	0 (eSiGe)	-1359	-1133	-1326	-2	35	12	-2	254	69
	1000 (eSiGe)	-966	-959	-966	8	18	11	380	1319	646
	-1000 (eSiGe)	-1756	-1333	-1677	9	45	17	-382	-809	-513

Der p-Tri-Gate-Transistor weist aufgrund der linearen Überlagerung mit der Verspannung des eingebetteten Silizium-Germaniums in den Source-Drain-Gebieten ein leicht verändertes Verhalten auf. Dabei zeigte sich die Verspannung in Richtung der Tiefe gegenüber der deutlichen Druckverspannung der Länge des Transistors ohne Wirkung. Entlang der Weite ist die Verspannung des Isolationsoxids jedoch deutlich ausgeprägter und erreicht bei intrinsischer Zugverspannung sogar Werte im GPa-Bereich.

Ein großer Vorteil des verspannten Isolationsoxids ist die gleichzeitig verbesserte Wirkung auf die Leistungsfähigkeit sowohl von n- als auch von p-Tri-Gate-Transistoren (Abbildung 5.26). Für die verspannten Deckschichten konnte man in Abhängigkeit von der intrinsischen Verspannung stets ein gegenläufiges Verhalten beobachten, da die Hauptkomponente dieser Verspannungstechnik mit der Transportrichtung der Ladungsträger gleichgerichtet war [15]. Im Falle des verspannten Isolationsoxids liegt die Hauptkomponente orthogonal zur Transportrichtung, wodurch beide Transistortypen auf ein zugverspanntes Oxid mit einer Leistungssteigerung von ca. 10 % reagieren. Dies ist ein enormer Vorteil für die Prozessintegration, da anders als bei den verspannten Deckschichten keine getrennte Prozessierung für n- und p-MOSFET erforderlich ist und somit die Anzahl an Schritten und die Herstellungskosten unverändert bleiben.

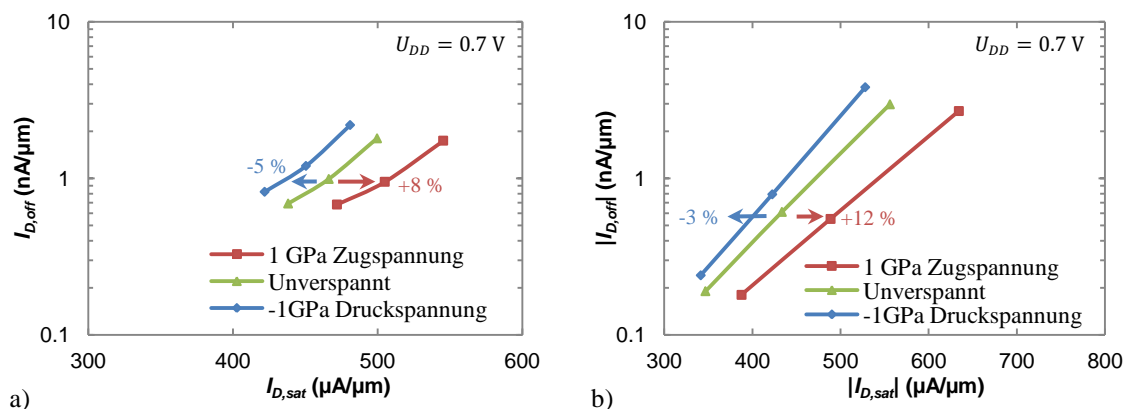


Abbildung 5.26: Universalkennlinie von a) n- und b) p-Tri-Gate-Transistoren mit verspanntem Isolationsoxid. ($H_{Fin} = 7 \text{ nm}$, $W_{Fin} = 20 \text{ nm}$, $Pitch_{Fin} = 50 \text{ nm}$)

C. Abhängigkeiten von den Technologieparametern

Da die Verspannungen hauptsächlich an den Rändern des Kanals zu den Source- und Drain-Gebieten auftreten, können auch noch die nachfolgenden Technologien mit kürzeren Gatelängen und kleineren Kontaktgebieten von dieser Verspannungsmethode profitieren. Zudem steigt die durchschnittliche Verspannung für kleinere Strukturen weiter an (Abbildung 5.27a). Die Kanalverspannung ist zudem linear abhängig von der intrinsischen Verspannung des Isolationsoxids, was bereits für die verspannten Deckschichten beobachtet werden konnte (Abbildung 5.27b) [15]. Dies ist ein Indiz dafür,

dass sich das verspannte Isolationsoxid auch mit anderen Verspannungstechniken kombinieren lässt und sich deren Wirkungen linear überlagern.

Da die Verspannung hauptsächlich entlang der Weite des Transistors auftritt, ändert sich auch der Verspannungseintrag, wenn der Abstand der Finnen und somit die Breite des Isolationsoxids variiert wird (Abbildung 5.27c). Für die Untersuchungen wurde die Weite der Finnen stets mit 20 nm angenommen. Zunächst tritt ein Anstieg der Kanalverspannung im Bereich der pn-Übergänge auf, der jedoch schnell abflacht und gegen einen maximalen Wert läuft. Ein größerer Abstand zwischen den Finnen bedeutet jedoch auch, dass auf einem gleich großen Aktivgebiet weniger Finnen Platz hätten, wodurch sich auch der Gesamtstrom des Transistors verringern würde. Dabei sollte der Abstand $Pitch_{Fin}$ der Finnen die effektive Gateweite W_{Gate} nicht überschreiten [162]:

$$Pitch_{Fin} \leq W_{Gate} = W_{Fin} + 2 \cdot H_{Fin} . \quad (5.6)$$

Um bei einem $Pitch_{Fin} \geq 50$ nm den Transistor entsprechend auszunutzen, ist gleichzeitig eine hohe Finne mit $H_{Fin} \geq 15$ nm erforderlich. Dies widerspricht jedoch dem ermittelten Optimum aus Abschnitt 5.2.5, nach welchem bei einer Finweite von $W_{Fin} = 20$ nm der Abstand der Finnen maximal 34 nm betragen sollte.

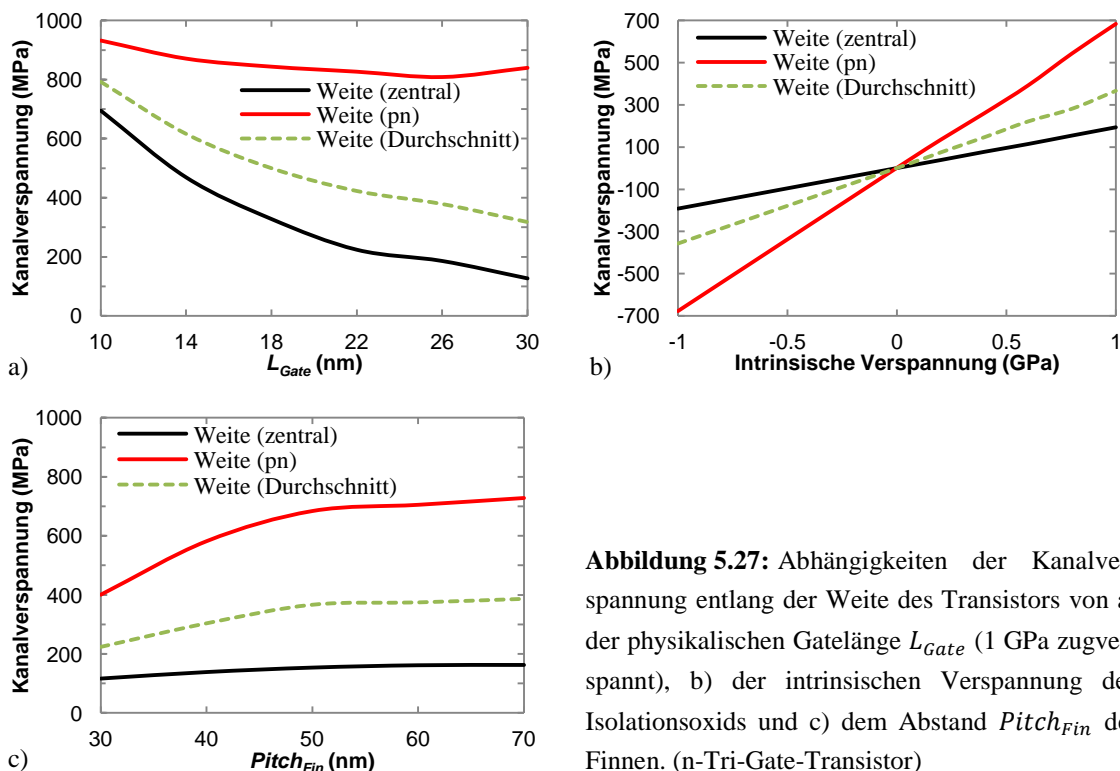


Abbildung 5.27: Abhängigkeiten der Kanalverspannung entlang der Weite des Transistors von a) der physikalischen Gatelänge L_{Gate} (1 GPa zugverspannt), b) der intrinsischen Verspannung des Isolationsoxids und c) dem Abstand $Pitch_{Fin}$ der Finnen. (n-Tri-Gate-Transistor)

Auch bei Tri-Gate-Transistoren besteht die Möglichkeit Silizium-Germanium epitaktisch einzusetzen um Deformationen im Kristallgitter zu erzeugen. Eine Variante davon ist bereits vom planaren p-MOSFET her bekannt, bei der das Kanalgebiet aus einer

dünnen Schicht Silizium-Germanium (Channel-SiGe, cSiGe) auf einem relaxierten Siliziumsubstrat besteht [15], [163]. Die Finne des Tri-Gate-Transistors müsste dafür nur um wenige Nanometer isotrop zurückgeätzt werden, um mit der aufgewachsenen Schicht wieder die gewünschte Höhe und Weite zu erreichen. Eine weitere Möglichkeit besteht darin, die komplette Finne aus Silizium-Germanium zu formen, welches sich auf einem relaxierten Si- bzw. SiGe-Substrat mit verändertem Germaniumgehalt befindet [164]. Beide Ansätze wären jedoch nur einseitig für die n- oder die p-Tri-Gate-Transistoren nutzbar und sollen hier nicht weiter betrachtet werden.

5.3.3 Substratorientierung

Die Substratorientierung beeinflusst die Verteilung der Ladungsträger in der Finne und die damit verbundene Grundbeweglichkeit sowie die mechanische Verspannung und die feldgebundene Quantisierung der Ladungsträger [165], [166]. Bei einem Tri-Gate-Transistor müssen daher die Oberflächenorientierungen der Finne in drei Bereichen, dem Top-Gate, dem Seiten-Gate und der Ecke betrachtet werden. Die Ecke selbst ist abhängig von ihrem Radius und bietet somit eine mehr oder weniger große Oberfläche. Die Kanalrichtung ist für alle drei Abschnitte gleich, jedoch unterscheiden sie sich im Bezug auf ihre Oberflächenorientierung und der damit verbundenen Quantisierungsmasse. Ebenso werden die Bereiche von der mechanischen Verspannung des Isolationsoxids unterschiedlich stark beeinflusst. Wegen der Komplexität der Einflüsse, wurde die Bandstruktur der Finne an sechs ausgewählten Punkten untersucht, die sich im Top-Gate, im Seiten-Gate und im Eckenbereich sowie in der Mitte des Kanals und am pn-Übergang befinden (Abbildung 5.28).

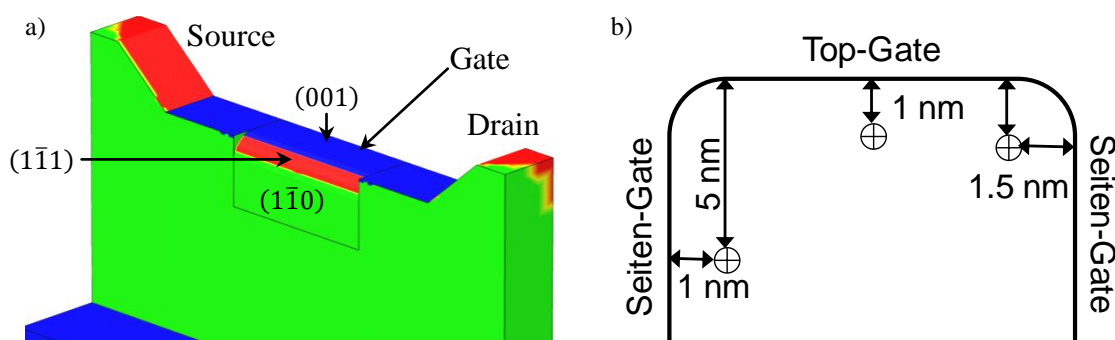


Abbildung 5.28: Simulation der Oberflächen im Tri-Gate-Transistor. a) Orientierung der Oberflächen im Bereich der abgerundeten Finne und b) der ausgewählten Punkte zur Untersuchung der Bandstruktur in der Finne, jeweils bei halber Kanallänge und am pn-Übergang.

Für die Betrachtungen wurden die vier am meisten genutzten Varianten von (Top-Gate-Orientierung) bzw. (Seiten-Gate-Orientierung) und (Kanalrichtung) ausgewählt (Abbildung 5.29) [167]. Zu ihnen gehören:

- $(001)(100)\langle 010 \rangle$
- $(001)(\bar{1}\bar{1}0)\langle 110 \rangle$
- $(011)(01\bar{1})\langle 100 \rangle$
- $(011)(100)\langle 01\bar{1} \rangle$

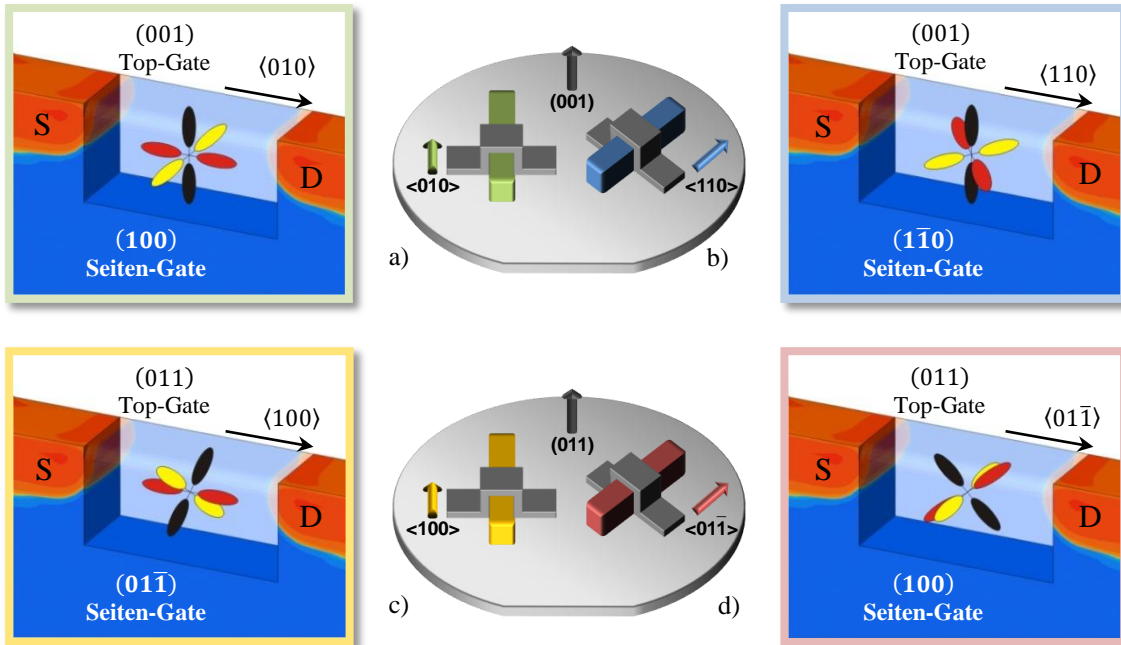


Abbildung 5.29: Oberflächenorientierung und Kanalrichtung für vier häufig verwendete Kristallorientierungen anhand der sechs Elektronenorbitale. Je zwei gleichfarbige Täler zeigen in eine der drei Hauptkristallrichtungen. Die vier Farben werden fortlaufend für die jeweilige Kristallorientierung genutzt.

A. Bandveränderung durch Quantisierung und mechanischer Verspannung

Aufgrund der Komplexität der Bandstruktur müssen die einzelnen Subbänder Leitungsbandes separat betrachtet werden. Je nach Ausrichtung der Kristallstruktur reagieren sie unterschiedlich auf den Einfluss der mechanischen Verspannung und zeigen keine einheitliche Bandverschiebung bzw. Änderung der Bandverbiegung. Unabhängig von der Kristallorientierung in der Finne gibt es stets Subbänder mit einer starken Bandkrümmung und einer geringen effektiven Masse, sowie einer hohen Ladungsträgerbeweglichkeit in Transportrichtung. Welche Grundbeweglichkeiten sich in den unterschiedlichen Bereichen der Finne einstellen, hängt jedoch von der Verteilung der Ladungsträger zwischen den Subbändern ab. Die eingebrachte mechanische Verspannung und die Wirkung der Bandverschiebung durch die Quantisierung der Ladungsträger sollte so gerichtet sein, dass die Orbitale mit einer hohen Ladungsträgerbeweglichkeit in Transportrichtung den Elektronen den niedrigsten Energiezustand bieten.

Für das Valenzband gilt analog, dass die Löcher nach dem höchsten energetischen Zustand streben, wobei das leichte und das schwere Löcherband im unverspannten Fall das gleiche Valenzbandmaximum besitzen. Unter mechanischer Verspannung bzw. dem quantisierenden Einfluss der Gate-Feldstärke kommt es auch hier zu einer Verschiebung der Bandkante, sodass man von einem oberen und einem unteren Löcherband spricht.

Da die Bauelementesimulation nur mit den Werten der Bandkanten arbeitet und nicht den Verlauf der Subbänder berechnet, wurden die Bandstruktur und die zugehörigen effektiven Massen mithilfe der empirischen Pseudopotenzialmethode (EPM) simuliert (Anhang 3). Diese Vorgehensweise setzt voraus, dass das Kristallgitter unendlich weit ausgedehnt ist und sich die eingestellten mechanischen Verspannungen der einzelnen Richtungen im gesamten Raum konstant verhalten. Dies gilt jedoch nicht für den Kanalbereich des Tri-Gate-Transistors, weshalb dieser Ansatz nicht für beliebig kleine Strukturen angewandt werden kann. Mit Werten von $W_{Fin} = 20$ nm und $H_{Fin} = 7$ nm sind die minimal zulässigen Abmessungen jedoch noch nicht unterschritten.

Mithilfe der effektiven Masse, die ein Elektron bei einer Bewegung senkrecht zur Grenzfläche aufweist, und der Gleichung (2.32) kann die Bandverschiebung infolge einer Quantisierungsfeldstärke mit $E_q = 1$ MV/cm für jedes Subband bestimmt werden. Die resultierende Aufspaltung der Subbänder wurde anschließend über

$$\Delta E_{Sub} = E_{\mu,min} - E_{\mu,max} \quad (5.7)$$

berechnet, wobei $E_{\mu,max}$ und $E_{\mu,min}$ die Bandkanten des jeweiligen Subbandes mit der besten bzw. der schlechteren Beweglichkeiten in Transportrichtung beschreiben. Ein positiver Wert bedeutet demnach für das Leitungsband eine Neuverteilung der Elektronen hin zu den Orbitalen mit der besseren effektiven Transportmasse (Anhang 4, grün hinterlegt). Im Gegensatz dazu muss ein positiver Wert für das Valenzband mit einer Neuverteilung der Löcher hin zu dem Band mit der schlechteren Transportmasse interpretiert werden (Anhang 5, rot hinterlegt).

Da bei den Elektronen die Orbitale auf alle drei Hauptrichtungen verteilt sind, gibt es bei jeder der betrachteten Kristallorientierungen stets zwei Täler mit der gleichen Orientierung zur Transportrichtung. Demnach ergibt sich für die Elektronen im unverspannten Fall auch der gleiche Bandverlauf sowie die gleiche effektive Masse in Transportrichtung. Die mittlere Bandverschiebung durch Verspannung bzw. Quantisierung dieser beiden Täler ergibt sich zu [117]

$$\frac{\Delta E_c}{k_B \cdot T_{300K}} = -\ln \left[\frac{1}{n_\Delta} \cdot \sum_{i=1}^{n_\Delta} \exp \left(\frac{-\Delta E_{c,i}}{k_B \cdot T_{300K}} \right) \right]. \quad (5.8)$$

Hierbei ist $n_{\Delta} = 2$ die Anzahl der gleichorientierten Subbänder. Anhand der Bandaufspaltungsbeiträge ΔE_{Stress} und $\Delta E_{Quantisierung}$ durch die mechanische Verspannung bzw. durch die quantisierende Feldstärke kann nun beurteilt werden, ob die Verschiebung der Subbänder beider Effekte in den einzelnen Bereichen der Finne additiv oder subtraktiv wirkt.

Für den n-Tri-Gate-Transistor zeigt sich die (001)(100)(010)-Orientierung mit dem zugverspannten Isolationsoxid als optimale Variante. Das druckverspannte Material verursacht in fast allen Fällen eine negativ wirkende Aufspaltung der Bandstruktur.

Auf der Seite des p-Tri-Gate-Transistors wirkt zunächst jede Variante als positiv, da die Dominanz des eingebetteten Silizium-Germaniums in den Source-Drain-Gebieten beachtet werden muss. Hinzu kommt, dass nach der Aufspaltung durch die quantisierende Feldstärke stets das Löcherband mit der geringeren Transport- jedoch der höheren Quantisierungsmasse die obere Valenzbandkante bildet. Da die Verbesserung der Löcherbeweglichkeit zu einem Großteil durch die verringerte Ladungsträgerstreuung bewirkt wird, kann hier die gesamte überlagerte Aufspaltung der Subbänder zur Beurteilung herangezogen werden. Die stärksten Effekte zeigten sich für die (110)-Oberflächenorientierungen bei einer $\langle 110 \rangle$ -Kanalrichtung.

B. Grundbeweglichkeit der Ladungsträger

Im Simulator wird die Abhängigkeit der Ladungsträgerbeweglichkeit von:

- der Hochfeldsättigung
- der Dotierung (PHILIPS-Model)
- der Beweglichkeitsdegradation an Grenzflächen (High-k-LOMBARDI-Model)
- der mechanischen Verspannung

berücksichtigt. Bis auf die verspannungsabhängige Beweglichkeit, bei der sich die vektorielle Stromdichte $J_{n,p}$ aus den unverspannten Größen $J_{n,p}^0$, $\mu_{n,p}^0$ und einem berechneten Beweglichkeitstensor $\boldsymbol{\mu}_{n,p}$ zu

$$J_{n,p} = \boldsymbol{\mu}_{n,p} \cdot \left(\frac{J_{n,p}^0}{\mu_{n,p}^0} \right) \quad (5.9)$$

ergibt, gehen alle anderen Beweglichkeitseinflüsse lediglich in die isotrope Grundbeweglichkeit ein. Für die Simulationen, bei denen das Bauelement mechanisch unverspannt ist, kann die Modellierung der Verspannungseffekte ausgeschaltet werden. Die Beweglichkeit der Ladungsträger wird dann vom Simulator nur noch als rein isotrope Größe behandelt und kann somit bei gleicher Oberflächenorientierung nicht mehr zwischen einer $\langle 010 \rangle$ - und einer $\langle 110 \rangle$ -Kanalrichtung unterscheiden.

Für die Elektronenbeweglichkeit und ihre effektive Transportmasse, die auch als Leitfähigkeitmasse bezeichnet wird, spielt diese Tatsache eine untergeordnete Rolle. Der

Grund dafür liegt in der sechsfachen Entartung des Leitungsbandes, wodurch die gemittelte effektive Transportmasse bei einer Gleichverteilung im unverspannten Fall mit

$$p_{\Delta_x} = p_{\Delta_y} = p_{\Delta_z} = \frac{1}{3} \quad (5.10)$$

nach der MATTHIESENSCHEN Regel zu

$$\frac{1}{\bar{m}_n^*} = \frac{p_{\Delta_x}}{m_{\Delta_{2x}}^*} + \frac{p_{\Delta_y}}{m_{\Delta_{2y}}^*} + \frac{p_{\Delta_z}}{m_{\Delta_{2z}}^*} \quad (5.11)$$

wird und für jede untersuchte Kristallorientierung einen Wert von rund $\bar{m}_n^* = 0.27 \cdot m_0$ annimmt (Tabelle 5.4).

Tabelle 5.4: Leitfähigkeitsmasse \bar{m}_n^* der Elektronen (Top-Gate) im unverspannten Kristallgitter für vier Kristallorientierungen sowie für unterschiedlich stark quantisierende Feldstärken E_q .

\bar{m}_n^*	Top-Gate			
	$(001)(1\bar{1}0)\langle 110 \rangle$	$(011)(100)\langle 01\bar{1} \rangle$	$(001)(100)\langle 010 \rangle$	$(011)(01\bar{1})\langle 100 \rangle$
E_q (V/m)				
0	0.267	0.267	0.270	0.270
$0.5 \cdot 10^8$	0.208	0.293	0.209	0.227
$1.0 \cdot 10^8$	0.202	0.304	0.202	0.215
$1.5 \cdot 10^8$	0.200	0.310	0.200	0.209

Mit der Aufspaltung der Bandstruktur infolge einer quantisierenden Feldstärke E_q aus Gleichung (5.8) kommt es zur Umverteilung der Ladungsträger unter den Subbändern. Über die veränderten Ladungsträgerdichten [9]

$$n_j = N_c \cdot \exp\left(\frac{W_F - W_C - e \cdot \Delta E_j}{k_B \cdot T_L}\right) \quad (5.12)$$

kann das prozentuale Verhältnis

$$p_{\Delta_j} = \frac{n_j}{n_g} = \frac{\exp\left(\frac{-e \cdot \Delta E_j}{k_B \cdot T_L}\right)}{\sum_i \exp\left(\frac{-e \cdot \Delta E_i}{k_B \cdot T_L}\right)} \quad (5.13)$$

unter den Subbändern neu bestimmt werden. Die Ladungsträgerdichte n_j steht dabei für beide Orbitale einer Raumrichtung (Δ_{2x} , Δ_{2y} bzw. Δ_{2z}). Die effektive Zustandsdichte

$$N_c = 2 \cdot \vartheta_n \cdot \left(\frac{m_{q,n}^* \cdot k_B T_L}{2\pi \cdot \hbar^2}\right)^{3/2} \quad (5.14)$$

mit $\vartheta_n = 2$ als Anzahl der äquivalenten Täler kann aufgrund der nahezu gleichbleibenden effektiven Quantisierungsmasse $m_{q,n}^*$ aller Orbitale als konstant angenommen werden und kürzt sich aus Gleichung (5.13) heraus. Somit ergeben sich je nach Kristall-

orientierung und Höhe der Quantisierungsfeldstärke unterschiedliche effektive Leitfähigkeitsmassen (Tabelle 5.4).

Für die Leitfähigkeitsmasse der Löcher kann dieses Vorgehen größtenteils analog angewendet werden. Jedoch ergibt sich durch die unterschiedlichen Zustandsdichten keine Gleichverteilung der Ladungsträgerdichten zwischen dem leichten und dem schweren Löcherband. Nach [46] kann davon ausgegangen werden, dass sich unter Vernachlässigung des Split-Orbit-Bandes ca. $p_{lh} = 30\%$ im leichten und $p_{hh} = 70\%$ aller freibeweglichen Ladungsträger im schweren Löcherband befinden. Somit ergeben sich die mittleren effektiven Leitfähigkeitsmassen der beiden Kanalrichtungen zu $\bar{m}_{p\langle 110 \rangle}^* = 0.306 \cdot m_0$ bzw. $\bar{m}_{p\langle 100 \rangle}^* = 0.244 \cdot m_0$ (Tabelle 5.5).

Tabelle 5.5: Leitfähigkeitsmasse \bar{m}_p^* der Löcher (Top-Gate) im unverspannten Kristallgitter für vier Kristallorientierungen sowie für unterschiedlich stark quantisierende Feldstärken E_q .

\bar{m}_p^*	Top-Gate			
	$(001)\langle 1\bar{1}0 \rangle \langle 110 \rangle$	$(011)\langle 100 \rangle \langle 01\bar{1} \rangle$	$(001)\langle 100 \rangle \langle 010 \rangle$	$(011)\langle 01\bar{1} \rangle \langle 100 \rangle$
E_q (V/m)				
0	0.306	0.306	0.244	0.244
$0.5 \cdot 10^8$	0.248	0.166	0.233	0.207
$1.0 \cdot 10^8$	0.222	0.153	0.226	0.202
$1.5 \cdot 10^8$	0.204	0.151	0.221	0.200

Damit stellt sich deutlich eine Anisotropie der Löcherbeweglichkeit für verschiedene Oberflächenorientierung und Kanalrichtungen ein, die auch durch die experimentell ermittelte Ladungsträgerbeweglichkeit in Abhängigkeit von der Inversionsladungsträgerdichte N_{inv} bestätigt wird (Abbildung 5.30) [168].

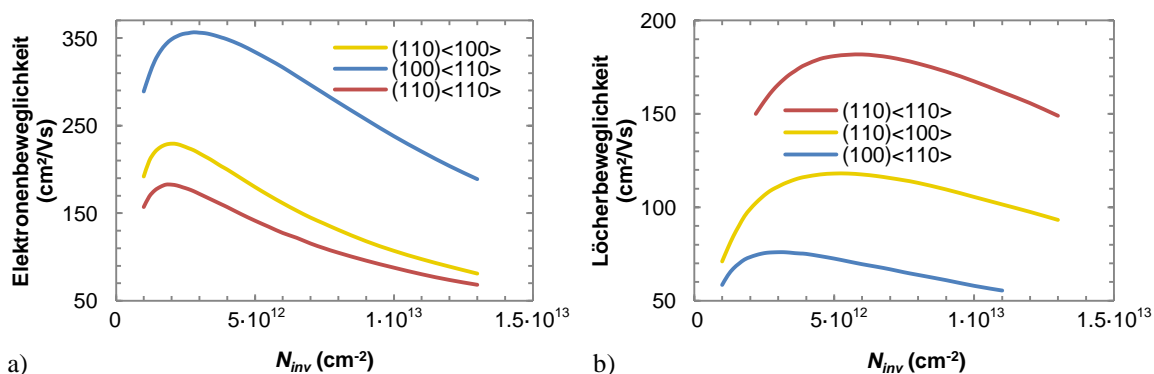


Abbildung 5.30: Ladungsträgerbeweglichkeit von Elektronen (a) und Löchern (b) in Abhängigkeit von der Inversionsladungsträgerdichte für verschiedene Kristallorientierungen [168].

Der Unterschied zwischen der $\langle 100 \rangle$ - und der $\langle 110 \rangle$ -Kanalrichtungen bei $\langle 110 \rangle$ -Oberflächenorientierung zeigt sich im Verhältnis der Löcherbeweglichkeit auch bei verstärkter Inversion mit:

$$\frac{\mu_{(110)\langle 110 \rangle}}{\mu_{(110)\langle 100 \rangle}} = \frac{3}{2}. \quad (5.15)$$

Im Gegensatz dazu liegen die entsprechenden Elektronenbeweglichkeiten eng beieinander, sodass der Unterschied mit zunehmender Inversionsladungsträgerdichte vernachlässigbar gering wird.

Für die Simulationen der Tri-Gate-Transistoren ist daher die isotrope Grundbeweglichkeit der Löcher über den Parameter des dotierungsabhängigen Beweglichkeitsmodells (Philips Unified Mobility Model) mit dem Verhältnis aus Gleichung (5.15) korrigiert worden (Abbildung 5.31). Die Differenz zwischen der $\langle 110 \rangle$ - bzw. $\langle 100 \rangle$ -Kanalrichtung mit $\langle 110 \rangle$ -Oberflächenorientierung entspricht jedoch nicht mehr dem ursprünglichen Verhältnis von 3:2 aus Gleichung (5.15), da sich die Modelle zur Degradation der Ladungsträgerbeweglichkeit nicht linear verhalten. Aufgrund eines unteren Grenzwertes für die Beweglichkeit in den Modellen kann man auch zwischen den beiden Simulationen der $\langle 100 \rangle$ -Oberflächenorientierung, die im Kanalbereich eine Löcherbeweglichkeit von unter $50 \text{ cm}^2/\text{Vs}$ aufweisen, keinen sichtbaren Unterschied mehr beobachten.

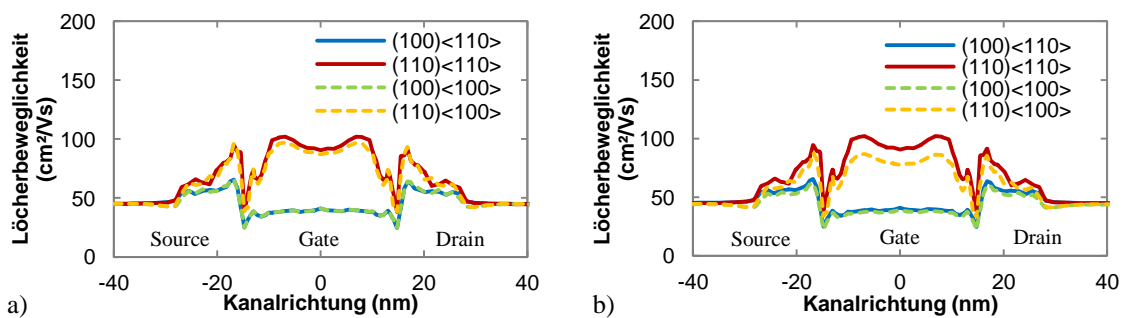


Abbildung 5.31: Berechnete Beweglichkeit der Löcher im Top-Gate eines unverspannten p-Tri-Gate-Transistors für alle vier untersuchten Kristallorientierungen. Verlauf durch den Kanal a) ohne und b) mit Korrektur. ($U_{GS} = 0.7 \text{ V}$, $U_{DS} = 0.05 \text{ V}$)

C. Leitfähigkeit der Tri-Gate-Bereiche

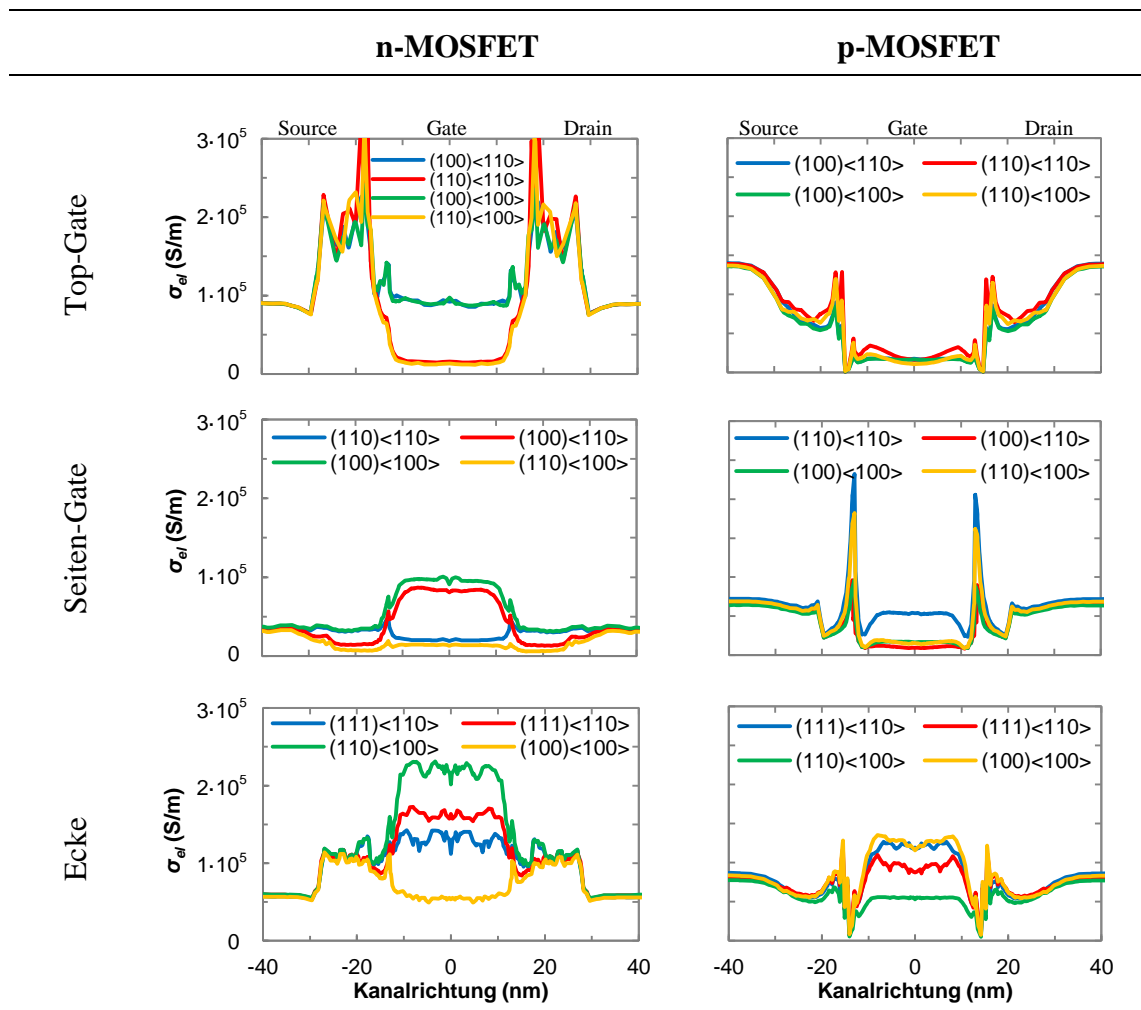
Zur Optimierung des Ladungsträgertransports in den Tri-Gate-Transistoren ist nicht nur eine hohe Beweglichkeit gewünscht, sondern auch eine große Ladungsträgerdichte in den jeweiligen Bereichen. Über die elektrische Leitfähigkeit

$$\sigma_{el} = e(n \cdot \mu_n + p \cdot \mu_p) \quad (5.16)$$

lassen sich beide Faktoren in Kombination betrachten, wobei die Minoritätsladungsträger des jeweiligen Transistortyps vernachlässigt werden können. Die Finnen wurden dafür wie bei der Analyse der Bandstruktur in drei Bereiche (Top-Gate, Seiten-Gate und Ecke) unterteilt, in denen jeweils ein Schnitt von Source nach Drain angelegt wurde (Tabelle 5.6).

Für das Top-Gate und Seiten-Gate des n-Tri-Gate-Transistors bietet die (100)-Oberfläche aufgrund der geringeren Oberflächenstreuung der Ladungsträger die bessere Leitfähigkeit. Dabei wird wiederum die Isotropie der effektiven Elektronenmasse deutlich, da lediglich zwischen der (100)- und der (110)-Oberflächenorientierung ein Unterschied auftritt. In der Ecke weicht die simulierte jedoch von der erwarteten Leitfähigkeit ab. Die schlechte Leitfähigkeit der (100)-Eckenorientierung und die deutlich bessere Leitfähigkeit der (110)-Orientierung lassen vermuten, dass die Orientierung der Ecke gegenüber den angrenzenden von Top- und Seiten-Gate bei einem Radius von 2 nm nur eine untergeordnete Rolle einnimmt.

Tabelle 5.6: Leitfähigkeit der Tri-Gate-Bereiche aller vier untersuchten Kristallorientierungen in unverspannten n- und p-Tri-Gate-Transistoren. ($U_{GS} = 0.7 \text{ V}$; $U_{DS} = 0.05 \text{ V}$)



Das Gleiche gilt für die beiden (111)-Eckenorientierungen, die aufgrund der unterschiedlichen Ausrichtungen von Top- und Seiten-Gate voneinander abweichen. Dennoch tritt in den Ecken des n-Tri-Gate-Transistors die höchste Leitfähigkeit auf, da die zugehörigen Ladungsträgerdichten aufgrund des Eckeneffektes höher sind als in den planaren Abschnitten. Die günstigste Kristallorientierung für die Finne des n-MOSFETs ist demnach die (001)(100)⟨010⟩-Orientierung.

Die simulierte Leitfähigkeit im p-Tri-Gate-Transistor verhält sich aufgrund der Anisotropie der Löcherbeweglichkeit etwas anders. Hinzu kommt, dass sich die Arsen-Halo-Gebiete zwischen den beiden Top-Gate-Orientierungen unterscheiden. Für die (110)-Substratorientierung wurde eine erhöhte Konzentration der Gegendotierung im Kanalgebiet beobachtet, womit sich auch eine geringere Ladungsträgerdichte im eingeschalteten Zustand ergibt. Erwartungsgemäß hätte die (100)-Orientierung mit der geringeren Ladungsträgerbeweglichkeit auch die schlechtere Leitfähigkeit zeigen müssen. Bedingt durch die oben genannten Gründe liegen die Leitfähigkeitswerte für beide Oberflächenorientierungen jedoch dicht beieinander. Betrachtet man alle drei Bereiche der Finne, so bietet die (011)(01 $\bar{1}$)⟨100⟩-Orientierung für den p-Tri-Gate-Transistor die beste Leitfähigkeit. Bei den übrigen Kristallorientierungen zeichneten sich kaum Unterschiede ab. Lediglich die (001)(100)⟨010⟩-Orientierung, die am günstigsten für den n-Tri-Gate-Transistor ist, besitzt in den Ecken eine sichtbar geringere Leitfähigkeit gegenüber den Löchern, was sich wegen der erhöhten Ladungsträgerdichten in diesen Bereichen umso negativer auf die Leistungsfähigkeit des gesamten Transistors auswirkt. Die zugehörigen Transistorkenndaten für alle Kristallorientierungen bestätigen die getroffenen Annahmen für n- und p-Tri-Gate-Transistoren bezüglich ihrer Leistungsfähigkeiten (Tabelle 5.7). Der normierte Durchgangswiderstand bei konstanter Inversion (Constant Overdrive, CO) $R_{on,CO}(U_{th} + 300 \text{ mV})$ gibt dabei die orientierungsgebundene Leistungsfähigkeit am besten wieder.

Tabelle 5.7: Abhängigkeit der Transistorkenndaten unverspannter n- und p-Tri-Gate-Transistoren von der Kristallorientierung. Normiert auf die effektive Gateweite.

Typ	Orientierung	$U_{th,lin}$ (V)	$U_{th,sat}$ (V)	$DIBL$ (mV/V)	STS_{sat} (mV/dec)	$I_{D,lin}$ ($\mu\text{A}/\mu\text{m}$)	$I_{D,sat}$ ($\mu\text{A}/\mu\text{m}$)	$I_{D,off}$ (nA/ μm)	$R_{on,CO}$ ($\Omega^*\mu\text{m}$)
n-MOS	(001)(1 $\bar{1}$ 0)⟨110⟩	0.310	0.272	54.7	73.6	90	466	0.99	974
	(011)(100)⟨01 $\bar{1}$ ⟩	0.310	0.275	49.7	71.7	65	352	0.69	1297
	(001)(100)⟨010⟩	0.308	0.270	53.9	73.5	97	498	1.03	931
	(011)(01 $\bar{1}$)⟨100⟩	0.327	0.293	48.4	72.6	42	244	0.43	1749
p-MOS	(001)(1 $\bar{1}$ 0)⟨110⟩	-0.383	-0.306	109.4	79.3	54	266	0.77	1481
	(011)(100)⟨01 $\bar{1}$ ⟩	-0.360	-0.301	84.0	73.1	77	341	0.39	1139
	(001)(100)⟨010⟩	-0.350	-0.274	108.0	79.3	53	286	1.86	1577
	(011)(01 $\bar{1}$)⟨100⟩	-0.343	-0.290	75.9	73.4	89	390	0.52	1038

Leider verhalten sich beide Transistortypen gegensätzlich, sodass die optimale Kristallorientierung des n-MOSFETs eine starke Degradation des p-MOSFETs bewirkt und umgekehrt. Da beide durch den CMOS-Prozess jedoch mit der gleichen Kristallorientierung gefertigt werden, kommt als Kompromiss vor allem die $(001)(\bar{1}\bar{1}0)(110)$ -Orientierung infrage, welche bereits für derzeitige Technologieknoten genutzt wird. Die Degradation im Vergleich zum Optimum ist für den n-Tri-Gate-Transistor in diesem Fall gering. Zudem lässt sich für den p-Tri-Gate-Transistor der Verlust der Leistungsfähigkeit mit einer mechanischen Verspannung in $\langle 110 \rangle$ -Kanalrichtung ausgleichen.

D. Mechanische Verspannung

Abschließend wurde die Leistungsfähigkeit der Transistoren in Abhängigkeit von der mechanischen Verspannung untersucht (Abbildung 5.32). Auf der Seite des n-Tri-Gate-Transistors kann die (100) -Substratorientierung mit einer Steigerung um 5 ... 8 % von dem mit 1 GPa zugverspannten Isolationsoxid profitieren. Insgesamt zeigte die $(001)(100)\langle 010 \rangle$ -Orientierung dank der guten Leitfähigkeit auch hier in allen Bereichen der Finne das beste Verhältnis von Sperr- zu Sättigungsstrom. Für den p-Tri-Gate-Transistor ist wiederum die $\langle 110 \rangle$ -Kanalrichtung besonders geeignet, um mittels der sehr starken Druckverspannung des eingebetteten Silizium-Germaniums eine erhebliche Steigerung der Sättigungsströme zu erzielen. Im Gegensatz zum p-MOSFET, bei dem die beiden verbleibenden Kristallorientierungen keine Änderungen zeigten, degradierten die n-MOSFETs mit der $(011)(100)\langle 01\bar{1} \rangle$ -Orientierung sogar um 6 %.

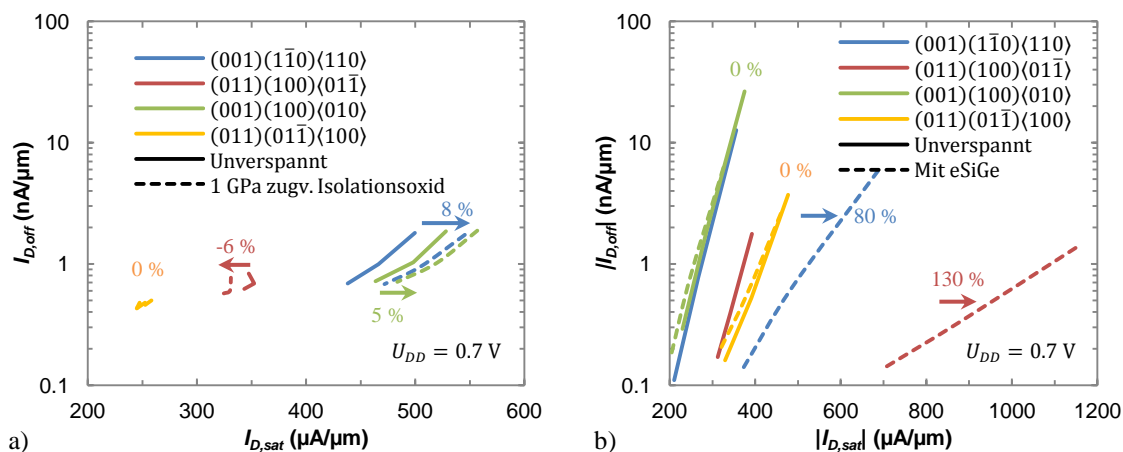


Abbildung 5.32: Orientierungsabhängigkeit der Universalkennlinien von a) n- und b) p-Tri-Gate-Transistoren. Dabei ist zusätzlich die prozentuale Änderung der Leistungsfähigkeit mit und ohne mechanische Verspannung dargestellt (n-MOSFET mit 1 GPa zugverspanntem Isolationsoxid, p-MOSFET mit eSiGe, $U_{DD} = 0.7$ V).

6 Transistoren mit vollständig verarmtem Kanal

Ein Transistor mit idealer Steuerfähigkeit des Gatekontakts benötigt einen Kanal, der vollständig verarmt bzw. mit freibeweglichen Ladungsträgern angereichert werden kann (Inversion). Dabei tragen alle Bereiche des Kanalgebiets, in denen Leckströme auftreten direkt zum Stromfluss bei. Diese Eigenschaft lässt sich sowohl bei Einzel-, Doppel- als auch Multi-Gate-Strukturen durch einen geeigneten Herstellungsprozess realisieren. Im Zusammenhang mit dieser Arbeit wurde für jede Kategorie eine Beispielstruktur nach den Richtlinien der 22 nm-Technologie ausgewählt. Der ultradünne SOI-Transistor ist dabei eine Spezialform des planaren SOI-Transistors. Der FinFET mit einer schmalen, hohen Finne stellt die Weiterentwicklung des Tri-Gate-Transistors aus der hybriden 22 nm-Technologie dar. Der Nanowire-MOSFET vereint alle theoretischen Vorzüge, die aus bisherigen Forschungsergebnissen und Veröffentlichungen der Fachpresse bekannt sind.

Das Kapitel ist in drei Abschnitte unterteilt, in denen jeweils eine der Transistorstrukturen behandelt wird. Dabei erfolgt anfangs eine Beschreibung des Herstellungsprozesses mit einzelnen Besonderheiten der entsprechenden Struktur. Anschließend werden die stationären und dynamischen Eigenschaften der Transistoren charakterisiert, wobei stets ein Vergleich zu den planaren und den Tri-Gate-Transistoren der 22 nm-Technologie gezogen wird.

6.1 Ultra-Dünne-SOI-MOSFETs

6.1.1 Prozessintegration

Der ultradünne SOI-MOSFET ist eine Weiterentwicklung eines klassischen SOI-Transistors. Als Grundlage diente ein teilweise verarmter SOI-Transistor von SYNOPSIS [169], welcher mithilfe der Prozessskizze von ST-MICROELECTRONICS und SOITEC [170] zu einem UTB-MOSFET modifiziert wurde. Das Ziel dieser Struktur ist es, das Kanalgebiet auf eine Schichtdicke zu reduzieren, die vollständig vom planaren Gatekontakt kontrolliert werden kann. Somit kann der Sperrstrom reduziert und der Kurzkanaleffekt gemindert werden.

Das Kanalgebiet besitzt in diesem Fall eine Schichtdicke von 8 nm, eine Länge von 26 nm und eine nur sehr geringe Dotierungskonzentration, womit es quasi als undotiert angesehen werden kann (Abbildung 6.1). Im Gegensatz dazu ist das Substrat unterhalb des 25 nm dicken vergrabenen Oxids mit einer tiefen Wannenheimplantation versehen, die in Verbindung mit einer Back-Gate-Spannung zur Korrektur der Schwellspannung genutzt werden kann. Anschließend erfolgt die Prozessierung des Gatekontakts mit Hafniumoxid als Dielektrikum und einem Metall, dessen Austrittsarbeit mit $W_A = 4.6$ eV auf die Mitte der Bandlücke des undotierten Kanals eingestellt ist. Die Source- und Draingebiete werden nach einem isotropen Zurückätzen epitaktisch aufgewachsen. Sie besitzen für n- und p-UTB-MOSFETs eine In-Situ-Dotierung mit einer Konzentration von $2 \cdot 10^{20} \text{ cm}^{-3}$ Phosphor- bzw. Borionen. Die Ausheilung des Siliziums erfolgt, wie auch bei den anderen Transistortechnologien, über eine Kurzzeit- bzw. Laserausheilung, bevor die Source-Drain-Gebiete silizidiert werden.

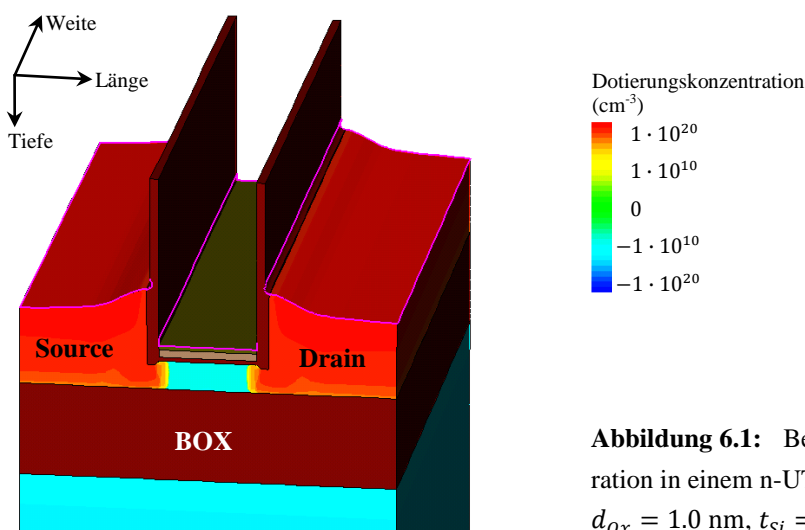


Abbildung 6.1: Berechnete Dotierungskonzentration in einem n-UTB-MOSFET, $L_{Gate} = 26$ nm, $d_{Ox} = 1.0$ nm, $t_{Si} = 8$ nm.

6.1.2 Charakterisierung des elektrischen Verhaltens

Der UTB-Transistor stellt die planare Variante eines MOSFETs mit vollständig verarmtem Kanal dar. Seine Schwellspannung ähnelt im Kurzkanalverhalten der des planaren Bulk-Transistor aus Abschnitt 4.1 (Abbildung 6.2). Im Vergleich der Universalkennlinien schneidet er im Bereich der kürzeren Gatelängen jedoch deutlich schlechter ab, da der Sperrstrom drastisch ansteigt. Mit der gegebenen Leistungsfähigkeit eignet er sich somit mehr für Anwendungen mit geringem Leistungsverbrauch (Low-Power), bei denen Transistoren mit längeren Kanälen zum Einsatz kommen.

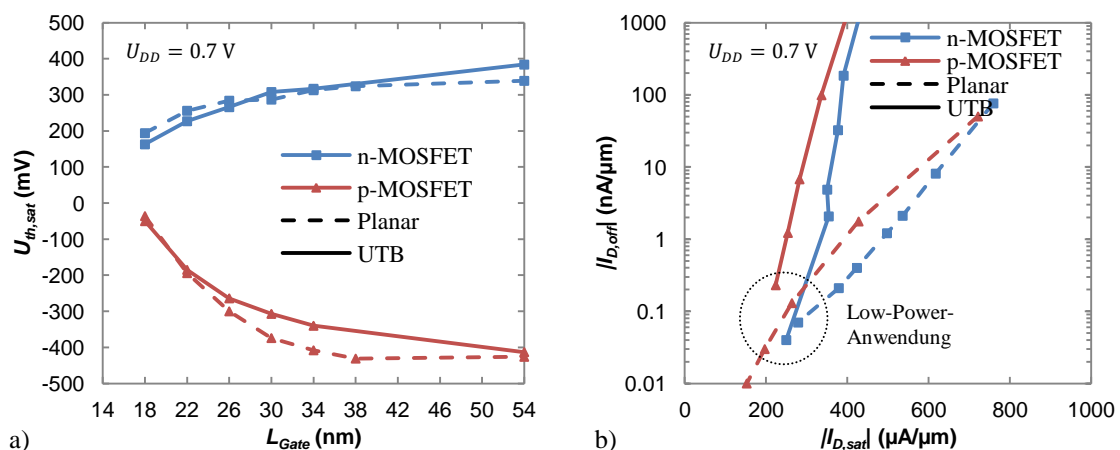


Abbildung 6.2: Berechnetes stationäres Verhalten von UTB- im Vergleich zu planaren Transistoren einer 22 nm-Technologie. a) Schwellspannung in Abhängigkeit von der physikalischen Gatelänge und b) Universalkennlinie.

Der Grund hierfür ist unter anderem der niedrigdotierte Kanal, der ohne Gegendotierung in den gateferneren Regionen eine zu starke Ausbreitung der drainseitigen Raumladungszone in den Kanal hinein zulässt. Dies verursacht auch eine hohe Unterschwelsteigung sowie einen erhöhten *DIBL* (Tabelle 6.1). Dennoch erreichen die UTB-Transistoren einen geringeren Widerstand, der hauptsächlich durch die Source-Drain-Gebiete und deren schmale Kanalbindung bestimmt wird und mit zunehmender Gatelänge nur geringfügig ansteigt.

Tabelle 6.1: Berechnete stationäre Kenndaten von UTB-Transistoren einer 22 nm-Technologie, $L_{Gate} = 26$ nm, $T_{Si} = 8$ nm, $U_{DD} = 0.7$ V, $U_{BG} = 0.0$ V.

Typ	U_{th} (mV)		<i>DIBL</i> (mV/V)	SS (mV/dec)		g_m (μS/μm)		I_D (μA/μm)		$I_{D,off}$ (nA/μm)	R_{on} (Ω·μm)
	lin.	sat.		lin.	sat.	lin.	sat.	lin.	sat.		
n-MOS	378	266	159	93.5	98.3	691	1366	188	377	32	449
p-MOS	-393	-264	184	95.4	99.6	530	1572	94	282	7	902

Um die Steuerfähigkeit des Gates zu steigern, müsste die Siliziumschicht mit der angegebenen Dicke von 8 nm noch dünner gestaltet werden, damit das Potenzial des Gates in einem Großteil des Kanals bis zur Grenzfläche zwischen Silizium und vergrabenen Oxid annähernd konstant und unabhängig von der Drain-Source-Spannung bleibt (Abbildung 6.3). Der Sättigungsstrom verhält sich dabei nahezu unverändert, während der Sperrstrom offensichtlich reduziert ist. Dabei wird jedoch auch deutlich, welchen Schwankungen die elektrischen Parameter des Bauelements unterliegen können, wenn die Siliziumschichtdicke aufgrund der Toleranzen beim Schichtauf- bzw. Schichtabtrag um 1 ... 2 nm variiert.

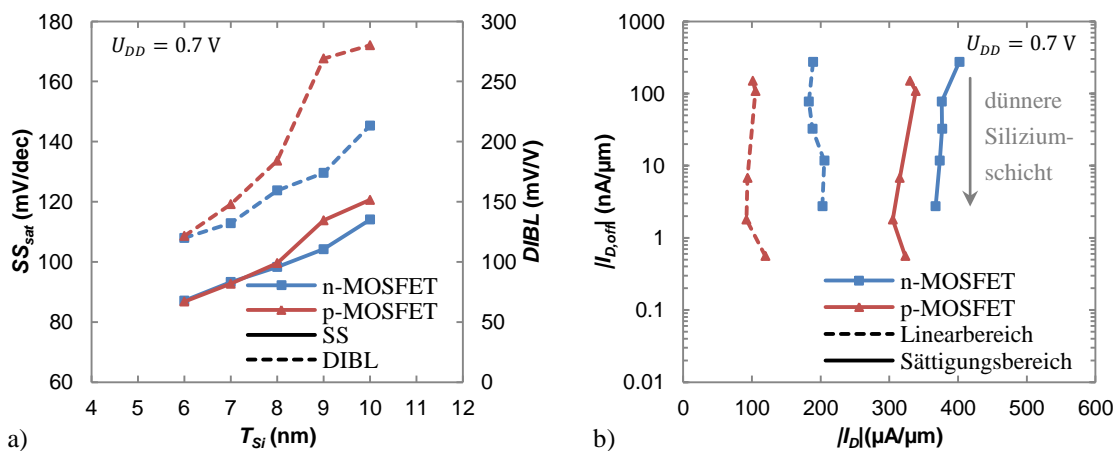


Abbildung 6.3: Berechnetes Verhalten von UTB-Transistoren in Abhängigkeit von der Siliziumschichtdicke. a) Unterschwellsteigung bei Sättigung SS_{sat} bzw. $DIBL$ und b) Verhältnis von Sperr- zu Sättigungsstrom. ($L_{Gate} = 26$ nm)

Ein wesentlicher Vorteil, den die Architektur der UTB-MOSFETs mit sich bringt, ist die Möglichkeit, die Schwellspannung über die Spannung des Substratkontakts (Backgate) zu steuern (Abbildung 6.4). Somit können Wafer-zu-Wafer-Schwankungen des Herstellungsprozesses im Nachhinein ausgeglichen oder auch Hochleistungsbaulemente mit erhöhtem Sperrstrom in einen Stand-by-Zustand geschaltet werden, in dem sie wesentlich weniger Leckstrom verursachen.

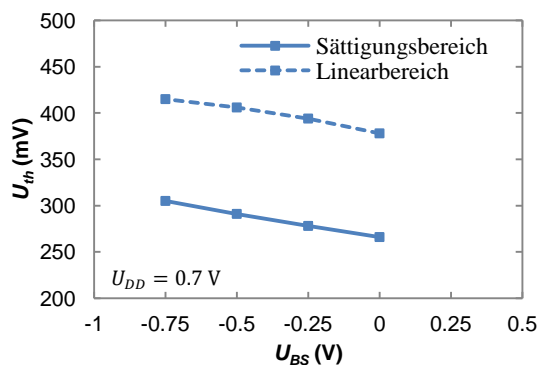


Abbildung 6.4: Berechnete Schwellspannungen von UTB-Transistoren in Abhängigkeit von der Backgate-Source-Spannung U_{BS} .

Bei den dynamischen Eigenschaften fällt auf, dass die Gate-Source-Kapazitäten der UTB-MOSFETs im Vergleich zu den planaren 22 nm-Transistoren geringere Werte aufweisen (Tabelle 6.2). Gründe dafür sind unter anderem das dünne Kanalgebiet und die Nähe des Gatekontakts zum vergrabenen Oxid, wodurch die Summe an Gegenladungen in den tieferen Regionen des Kanalgebiets erheblich reduziert wird. Somit ergeben sich auch höhere Frequenzen der Bauelemente bzw. kürzere Schaltzeiten der Inverterstufen. Der hohen Ringoszillatorfrequenz steht jedoch auch ein wesentlich größerer Sperrstrom gegenüber, sodass die UTB-Struktur gegenüber der planaren Architektur im Endeffekt nur eine geringe Steigerung der dynamischen Leistungsfähigkeit aufweist.

Tabelle 6.2: Berechnete dynamische Kenndaten der UTB-Transistoren in 22 nm-Technologie, $L_{Gate} = 26$ nm, $T_{Si} = 8$ nm, $U_{DD} = 0.7$ V, $U_{BS} = 0.0$ V.

Einzeltransistor	Typ	C_{ov} (fF/ μm)	C_{inv} (fF/ μm)	f_{max} (GHz)	f_t (GHz)	
	n-MOSFET	0.360	0.595	2270	455	
p-MOSFET	0.356	0.572	2420	435		
Dreistufiger Inverter	t_r (ps)	t_f (ps)	Δt_{hi} (ps)	Δt_{low} (ps)	U_{max} (V)	U_{min} (V)
	3.85	3.51	3.42	3.43	0.761	-0.065
Fünfstufiger Ringoszillator	f_{RO} (GHz)	$I_{RO,eff}$ ($\mu\text{A}/\mu\text{m}^2$)	$I_{RO,off}$ (nA/ μm^2)	t_D (ps)	Gesamte Aktivfläche (μm^2)	
	24.50	441.61	15.29	4.08	0.02655	

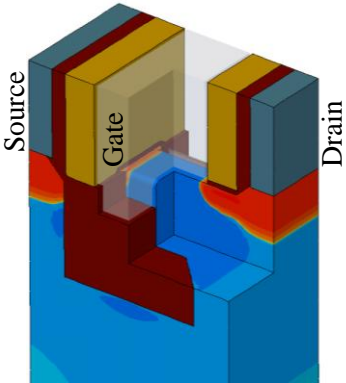
6.2 FinFETs

6.2.1 Prozessintegration

Der FinFET-Transistor ist eine spezielle Form des Trippel-Gate-Transistores, kann aber aufgrund seiner hohen schmalen Finne bzw. der Dominanz der beiden Seiten-Gates gegenüber dem Top-Gate auch als Doppel-Gate-Struktur angesehen werden. Der in dieser Arbeit simulierte FinFET stellt zudem eine Weiterentwicklung des Tri-Gate-Transistors der hybriden 22 nm-Technologie dar. Der Grundgedanke dieses Ansatzes ist die beschriebene schrittweise Implementierung neuer Transistorstrukturen in die Volumenproduktion der Halbleiterhersteller, die aus Kostengründen und aufgrund einer kurzen Entwicklungszeit harte Umbrüche bei den Übergängen von einem zum nächsten Technologieknoten vermeiden wollen. Vor diesem Hintergrund stellt der Tri-Gate-Transistor nur eine Zwischenstufe auf dem Weg zu einer Technologie mit vollständig verarmten Multi-Gate-Transistoren dar (Tabelle 6.3).

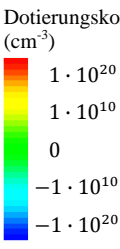
Tabelle 6.3: Gegenüberstellung der Prozessablaufpläne und der Dotierungsstruktur von Tri-Gate-Transistoren und von FinFETs.

Tri-Gate	FinFET
• STI	• STI
• Wannenimplantation	• Tiefe Wannenimplantation (Retrograde Well)
• Strukturierung der Finne (flach, weit)	• Strukturierung der Finne (hoch, schmal)
• Eckenimplantation	• Eckenimplantation
• Dummy-Gate-Strukturierung	• Dummy-Gate-Strukturierung
• Implantation der Halo- (n-/p-MOS) und Erweiterungsgebiete (nur n-MOS)	• Implantation der Halo- (n-/p-MOS) und Erweiterungsgebiete (nur n-MOS)
• Nitridspacer (15 nm)	• Nitridspacer (8 nm)
• Source-Drain-Epitaxie (Sigma-Form)	• Source-Drain-Epitaxie (U-Form)
• Ausheilung und Aktivierung	• Ausheilung und Aktivierung
• Source-Drain-Kontakte	• Source-Drain-Kontakte
• Replacement-Gate	• Replacement-Gate

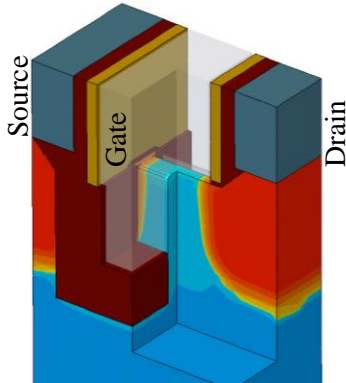


Source Gate Drain

Dotierungskonzentration (cm⁻³)



1 · 10²⁰
1 · 10¹⁰
0
-1 · 10¹⁰
-1 · 10²⁰



Source Gate Drain

Ausgehend vom beschriebenen Tri-Gate-Prozess wurde die Höhe der Finne auf 30 nm eingestellt und gleichzeitig die Weite auf 10 nm reduziert. Um eine vollständige Anbindung an den Kanal zu realisieren, ist die U-Form für die Source-Drain-Gebiete übernommen und der Spacer für die Epitaxie entsprechend auf 8 nm angepasst worden. Somit benötigt auch der n-MOSFET wie schon der p-Tri-Gate-Transistor keine weitere Implantation für die Erweiterungsgebiete. Weiterhin wurden die Ecken- und Haloimplantationen aus dem Herstellungsprozess entfernt, damit der Kanal niedrigdotiert bleibt und eine Volumeninversion ermöglicht bzw. den Ladungsträgern eine höhere Beweglichkeit bietet [171]. Dies erforderte jedoch eine Anpassung der Austrittsarbeit des Gatemetalls, welche für den n- und p-FinFET auf $W_{A,n} = 4.53$ eV bzw. $W_{A,p} = 4.72$ eV eingestellt wurde, um eine Sättigungsschwelspannung von ca. $|U_{th,sat}| = 290$ mV zu erreichen. Wie es für Bulk-FinFETs üblich ist, wurde die Wannenimplantation bezüglich Energie und Dosis modifiziert, um eine Gegendotierung unterhalb der Finne und den Source-Drain-Gebieten zu erzeugen (Retrograde Well). Diese soll die Leckstrompfade in diesen Regionen des Transistors einschränken.

Abgesehen davon, dass die Strukturierung der Finne eine erhöhte Präzision des anisotropen Schichtabtrags verlangt, hat der Herstellungsprozess gegenüber dem des Tri-Gate-Transistors nicht an Komplexität gewonnen. Es wurden lediglich existierende Prozessschritte modifiziert oder rationalisiert, wie im Beispiel der Ecken- und Haloimplantation, wodurch keine bedeutende Steigerung der Herstellungskosten mit der Substitution der Tri-Gate-Transistoren durch die FinFETs zu erwarten ist. Allerdings stehen die Anpassung der Austrittsarbeit und die Epitaxie der Source-Drain-Gebiete im Konflikt zu einer möglichen Kombination mit planaren Transistoren einer hybriden Technologie.

6.2.2 Charakterisierung des elektrischen Verhaltens

Aufgrund seiner schmalen hohen Finne, die mehr Gatefläche pro Kanalvolumen bietet, sollte der FinFET im Vergleich zum Tri-Gate-Transistor eine verbesserte Steuerfähigkeit des Gates, eine reduzierte Anfälligkeit gegenüber Kurzkanaleffekten und einen höheren Drainstrom bezüglich der Aktivfläche bieten. Doch bei Betrachtung der Abhängigkeit der Schwellspannung von der physikalischen Gatelänge (Abbildung 6.5) fällt auf, dass die Werte für den n-MOSFET schlechter ausfallen und der FinFET im Roll-off-Verhalten eher mit dem planaren Transistor vergleichbar ist. Ursache hierfür sind die veränderten pn-Übergänge im Vergleich zum Tri-Gate-Transistor. Durch die fehlende Gegendotierung der Haloimplantation und die diffusionsbedingte Ausbreitung der Arsenionen in die tieferen Regionen des Transistors besitzen sie einen flacheren Übergang zum Kanal und somit eine leicht erhöhte Überlappung zum Gatekontakt. Die effektive Kanallänge kann zwar auf einfachem Wege über die Vergrößerung der Spacerbreite erzielt werden, die verstärkte Ausbreitung der drainseitigen Raumladungszone bedingt durch den abgeflachten pn-Übergang jedoch nicht. Als Resultat kann beim n-FinFET ein größerer Anstieg der Unterschwellsteigung mit kürzer werdender Gatelänge beobachtet werden (Abbildung 6.5b).

Der p-FinFET hingegen verhält sich erwartungsgemäß und zeigt trotz der fehlenden Halodotanden eine verbesserte Steuerfähigkeit. Ausschlaggebend dafür ist die verringerte Diffusion der Borionen im eingebetteten Silizium-Germanium, die verhindert, dass die Überlappung zunimmt und die pn-Übergänge abflachen [172]. Dieser Effekt könnte auch für den n-FinFET genutzt werden, indem bei der in-situ-dotierten Epitaxie ein weiteres, diffusionsreduzierendes Element mit in die Source-Drain-Gebiete eingebaut wird. Dies wurde jedoch im Rahmen dieser Arbeit nicht näher untersucht.

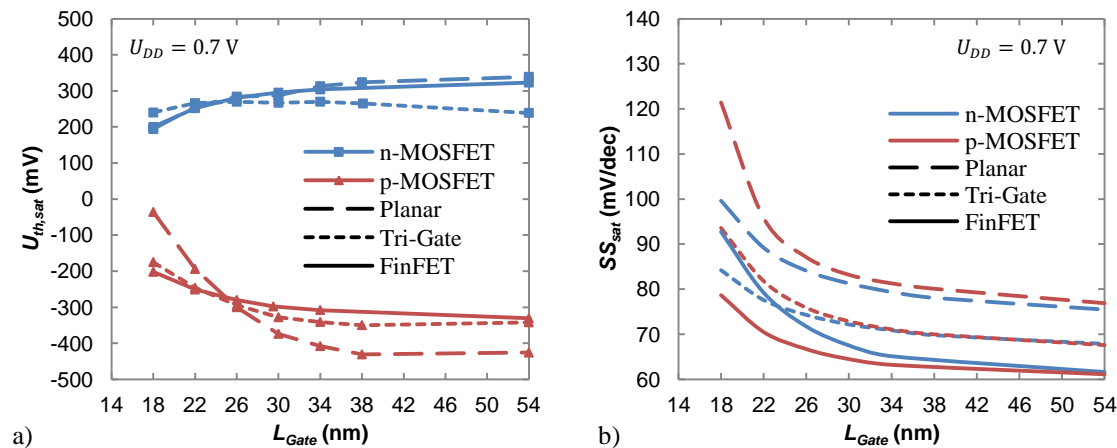


Abbildung 6.5: Berechnetes stationäres Verhalten von FinFETs im Vergleich zu planaren und Tri-Gate-Transistoren einer 22 nm-Technologie. a) Schwellspannung und b) Unterschwellsteigung in Abhängigkeit von der physikalischen Gatelänge.

Im Vergleich mit den verschiedenen Transistorstrukturen bezüglich der linearen Unterschwellsteigung liegen die FinFETs deutlich näher an der Ideallinie als die zuvor betrachteten Transistorvarianten (Abbildung 6.6). Aufgrund der Dominanz ihrer Seiten-Gates gegenüber ihren Top-Gates wurde für die FinFETs dabei die Gleichung einer Doppel-Gate-Struktur zur Berechnung der natürlichen Länge von $\lambda_2 = 3.89 \text{ nm}$ verwendet. Auch hier erkennt man das etwas schlechtere Kurzkanalverhalten der n-MOSFETs mit sehr kurzen Gatelängen, wobei die größeren Gatelängen beider Transistortypen dicht beieinander liegen und sich nahe dem Minimum von ca. 60 mV/dec bewegen.

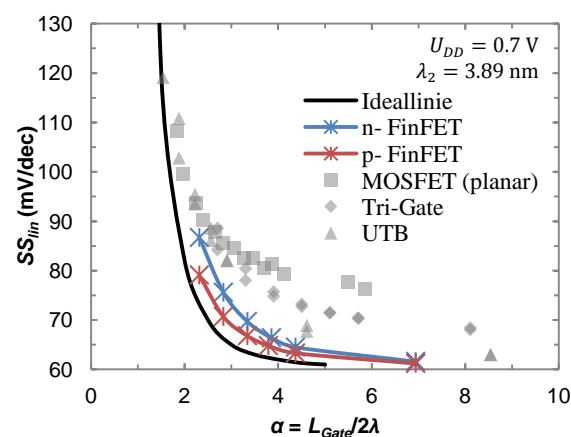


Abbildung 6.6: Berechnete lineare Unterschwellsteigung in Abhängigkeit von dem Skalierungsparameter α für die n- bzw. p-FinFETs der 22 nm-Technologie.

Die Leistungsfähigkeit der FinFETs hat sich bei der Betrachtung der Universalkennlinie mit Normierung auf die effektive Gateweite gegenüber den Tri-Gate-Transistoren um ca. 25 % gesteigert (Abbildung 6.7).

Dies bedeutet, dass sich der Kanalwiderstand der Finne erheblich reduziert haben muss,

was zum einen an der geringen Dotierung und der damit reduzierten Ladungsträgerstreuung liegt und zum anderen durch die schmale Finne begründet ist, die bei der verstärkten Überlagerung der beiden Seiten-Gates zu einer Volumeninversion mit reduzierter Ladungsträgerstreuung an den Grenzflächen führt. Beide Fakten ermöglichen eine höhere Ladungsträgerbeweglichkeit sowie Leitfähigkeit des Transistors und spiegeln sich in den stationären Eigenschaften der Bauelemente wider (Tabelle 6.4).

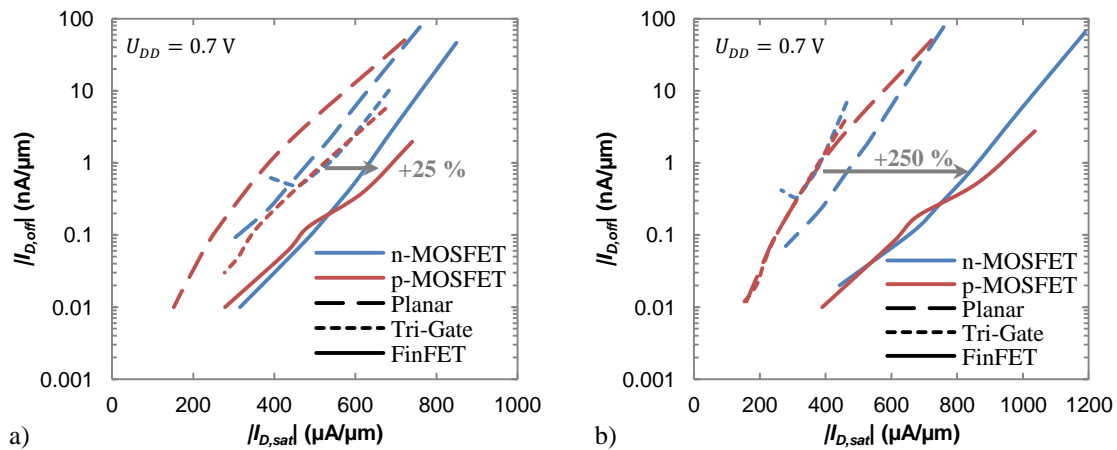


Abbildung 6.7: Berechnete Universalkennlinien von n- und p-MOSFETs der planaren, Tri-Gate- und FinFET-Transistoren einer 22 nm-Technologie, normiert auf a) die effektive Gateweite und b) die Strukturweite.

Bei Betrachtung der auf die Strukturweite ($W_{Str} = 50$ nm) normierten Universalkennlinien kommt zusätzlich der Zugewinn der effektiven Gateweite ($W_{Gate} = 70$ nm) durch die schmalen, aber hohen Finnen hinzu, sodass die FinFETs gegenüber den Tri-Gate-Transistoren den zweieinhalbfachen Drainstrom bei gleichem Flächenbedarf liefern können.

Tabelle 6.4: Berechnete stationäre Kenndaten der FinFETs einer 22 nm-Technologie, Werte normiert auf die effektive Gateweite, $L_{Gate} = 26$ nm, $H_{Fin} = 30$ nm, $W_{Fin} = 10$ nm, $U_{DD} = 0.7$ V.

Typ	U_{th} (mV)		$DIBL$ (mV/V)	SS (mV/dec)		g_m ($\mu S/\mu m$)		I_D ($\mu A/\mu m$)		$I_{D,off}$ (nA/ μm)	R_{on} ($\Omega \cdot \mu m$)
	lin.	sat.		lin.	sat.	lin.	sat.	lin.	sat.		
n-MOS	317	281	52	69.7	71.7	417	2171	135	604	0.58	760
p-MOS	-309	-280	42	66.8	66.7	443	2106	145	619	0.39	698

Auch bei den FinFETs konnte der Effekt des mechanisch verspannten Isolationsoxids beobachtet werden (Abbildung 6.8). Der n-FinFET profitiert im Vergleich zum p-FinFET, welcher nach wie vor primär durch die mechanische Verspannung des eingebetteten Silizium-Germaniums bestimmt ist, wesentlich stärker von der orthogonalen

Zugverspannung im Bereich der pn-Übergänge des Kanals. Es zeigte sich jedoch, dass auch hier beide Transistortypen positiv auf eine Zugverspannung reagieren. Dabei ist im Gegensatz zu den Tri-Gate-Transistoren, deren Ladungsträger hauptsächlich an die (100)-Orientierung des Top-Gates gebunden sind, der überwiegende Teil der Ladungsträger des FinFETs durch die (110)-Oberflächenorientierung der Seiten-Gates bestimmt.

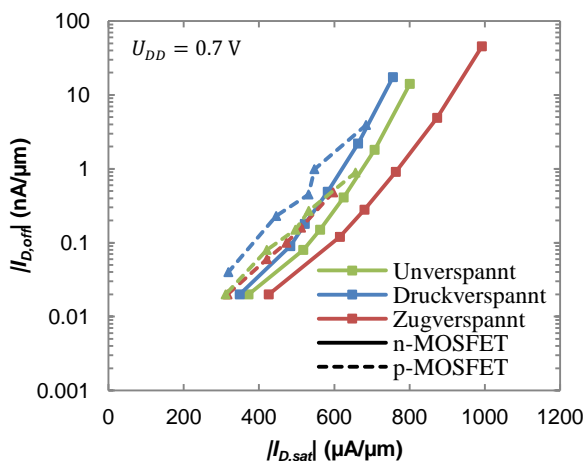


Abbildung 6.8: Berechnete Universalkennlinien von n- und p-FinFET-Transistoren einer 22 nm-Technologie mit neutralem und zug- bzw. druckverspanntem Isolationsoxid zwischen den Finnen. (Intrinsische Verspannung 1 GPa)

Bei Betrachtung der dynamischen Kenndaten der 22 nm-FinFETs hat sich gezeigt, dass die Überlappungs- und Inversionskapazitäten C_{ov} bzw. C_{inv} nach der Normierung auf die effektive Gateweite vergleichbare Werte zu den planaren Transistoren aufweisen (Tabelle 6.5). Insofern kann davon ausgegangen werden, dass sich die Effektivität der Gatekapazität und somit auch das kapazitive Verhalten des Transistors mit dem Übergang vom Tri-Gate zum FinFET verbessert hat und keine weiteren parasitären Anteile hinzugekommen sind.

Tabelle 6.5: Berechnete dynamische Kenndaten der FinFETs einer 22 nm-Technologie, $L_{Gate} = 26\text{ nm}$, $H_{Fin} = 30\text{ nm}$, $W_{Fin} = 10\text{ nm}$, $U_{DD} = 0.7\text{ V}$.

Einzeltransistor	Typ	C_{ov} (fF/ μm)	C_{inv} (fF/ μm)	f_{max} (GHz)	f_t (GHz)	
	n-MOSFET	0.615	1.023	1609	409	
	p-MOSFET	0.553	0.991	1682	436	
Dreistufiger Inverter	t_r (ps)	t_f (ps)	Δt_{hi} (ps)	Δt_{low} (ps)	U_{max} (V)	U_{min} (V)
	4.60	4.95	3.45	3.40	0.78	-0.08
Fünfstufiger Ringoszillator	f_{RO} (GHz)	$I_{RO,eff}$ ($\mu\text{A}/\mu\text{m}^2$)	$I_{RO,off}$ (nA/ μm^2)	t_D (ps)	Gesamte Aktivfläche (μm^2)	
	33.02	1350	3.684	3.03	0.043	

Als Resultat konnten in Kombination mit der erhöhten Leistungsfähigkeit die Schaltzeiten der Inverterstufen verkürzt und die Ringoszillatorfrequenz gesteigert werden. Es muss jedoch festgehalten werden, dass die Gesamtkapazität, bezogen auf die Aktivfläche des Transistors, aufgrund der vergrößerten effektiven Gateweite deutlich zugenommen hat und somit auch eine höhere effektive Stromaufnahme beim Schalten verursacht (Abbildung 6.9a). Da die Werte des FinFETs und des planaren Transistors jedoch näherungsweise auf einer Geraden liegen, bieten auch beide Transistorstrukturen eine vergleichbare Effizienz bezüglich des dynamischen Verhaltens.

Unter Berücksichtigung der Stromaufnahme im ausgeschalteten Zustand schneidet der FinFET gegenüber dem Tri-Gate und dem planaren Transistor wesentlich besser ab (Abbildung 6.9b). Die vollständige Kontrolle des Kanals und die tiefe Wannenimplantation reduzieren die parasitären Leckstrompfade dabei auf ein Minimum.

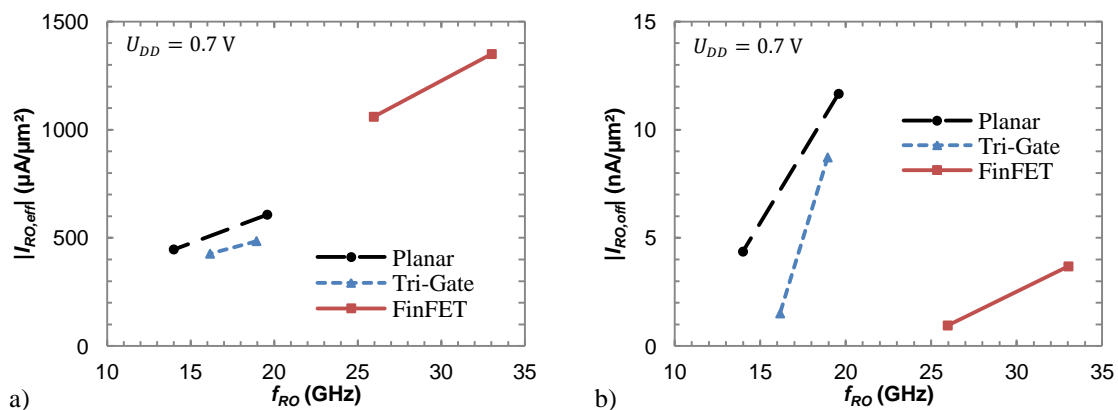


Abbildung 6.9: Berechnete Stromaufnahme im a) schwingenden und b) ausgeschalteten Zustand über der Ringoszillatorfrequenz (fünftufig) für die planaren, Tri-Gate- und FinFET-Transistoren einer 22 nm-Technologie, normiert auf die Aktivfläche, $H_{Fin} = 30\text{ nm}$, $W_{Fin} = 10\text{ nm}$, $L_{Gate} = 26 \dots 30\text{ nm}$.

6.3 Vertikale Nanowire-MOSFETs

6.3.1 Prozessintegration

Mit den vertikalen Nanowire-MOSFETs soll abschließend die Transistorstruktur mit der besten Steuerfähigkeit des Kanals vorgestellt werden [173], [174], [175]. Für die Prozessstabilität derart kleiner Bauelemente ist es von großem Vorteil, wenn die Strukturierung, Dotierung und Kontaktierung zum Großteil mithilfe von selbstjustierenden Prozessen erfolgt. Die klassischen Spacer seitlich des Gates, die zur Strukturierung und Implantation der horizontalen Transistoren dienen, lassen sich bei vertikalen Strukturen nicht mehr anwenden. Sämtliche Abstände, wie z. B. die Gatelänge, müssen mithilfe von Schichtauf- bzw. Schichtabtrag realisiert werden, wobei Toleranzen von 1 ... 4 nm auftreten können. Weiterhin sind bis auf den lithographischen Schritt zur Strukturierung

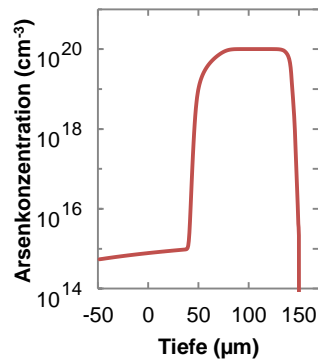


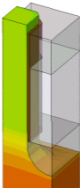
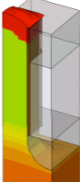
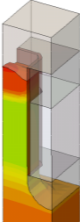
der Siliziumdrähte nur großflächige Masken zur Abdeckung der jeweiligen Aktivgebiete von n- und p-MOSFET erforderlich.

Der Herstellungsprozess der untersuchten Nanowire-MOSFETs (Tabelle 6.6) beginnt mit einem herkömmlichen SOI-Wafer, bei welchem das Silizium oberhalb des 150 nm dicken vergrabenen Oxids nach dem STI-Prozess auf eine Schicht von 5 ... 10 nm Stärke zurückgeätzt bzw. poliert wird. Es ist auch möglich, gleich mit derartigen SOI-Wafern zu beginnen, wie sie für UTB-Transistoren angeboten werden. Diese sind jedoch deutlich teurer im Einkauf und bieten eine Genauigkeit der Schichtdicke, die nicht zwangsläufig erforderlich ist. Anschließend wird getrennt für n- und p-MOSFET das Drain- bzw. Kanalgebiet ganzflächig mittels dotierter Epitaxie aufgewachsen. Auf dem Gebiet des n-MOSFETs wird zuerst eine 60 nm dicke Schicht mit $1 \cdot 10^{20} \text{ cm}^{-3}$ Arsen abgeschieden. Danach wird bei kontinuierlichem Wachstum von weiteren 33 nm Silizium die beigemischte Arsenkonzentration linear bis auf $1 \cdot 10^{15} \text{ cm}^{-3}$ zurückgefahren, von wo an undotiertes Silizium, welches zu dem späteren Kanalgebiet gehört, aufgewachsen wird. Für den p-MOSFET verläuft die dotierte Epitaxie analog, wobei für das p-Draingebiet Bor statt Arsen verwendet wird. Die hochdotierte Schicht ist hier jedoch etwas dünner gehalten, da Bor bei den nachfolgenden thermischen Schritten stärker diffundiert. Der anschließende Ausheilungsschritt von 5 s mit $1000 \text{ }^\circ\text{C}$ ist von der Wannenausheilung planarer 22 nm-Transistoren übernommen worden. Auch wenn dieser Temperaturschritt für die epitaktisch aufgewachsenen Gebiete keine rekristallisierende Bedeutung hat, ist der Einsatz von Bauelementen mit einer implantierten Wanne auf einem Chip und somit auch im thermischen Budget der Nanowire-MOSFETs zu berücksichtigen.

Die Strukturierung der Nanowire erfolgt mithilfe zweier orthogonaler Maskierungen in Verbindung mit einem anisotropen Ätzen des Siliziums. Die veränderte Dotierungskonzentration zwischen Kanal- und Draingebiet kann als Ätzstopp genutzt werden, da sich die Ätzrate ab einer Dotierungskonzentration von $1 \cdot 10^{19} \text{ cm}^{-3}$ gegenüber undotiertem Silizium deutlich reduziert [176]. Anschließend werden die Gräben mit Oxid, hergestellt aus einem hoch dichten Plasma (High Density Plasma, HDP), gefüllt und anisotrop auf die Übergangstiefe von Drain und Kanal zurück geätzt. Das HDP-Oxid bietet hierbei eine geringe Permittivität zur Isolierung von Draingebiet und Gate Metall und mindert somit die Wirkung dieser parasitären Kapazität.

Die freistehenden Nanowire-Strukturen werden dann mit einem 20 min langen und $850 \text{ }^\circ\text{C}$ heißen thermischen Schritt an der Oberfläche oxidiert. Hierbei bildet sich das ca. 1 nm dicke Gateoxid. Die Prozessierung von Siliziumoxid ist bei sehr dünnen Nanowire-MOSFETs von Vorteil, da es im Vergleich zu Hafniumoxid einfacher aufzubringen ist, die bessere Grenzfläche bildet und deutlich weniger fixierte Ladungen aufnimmt, wodurch die Grundbeweglichkeit der wenigen Ladungsträger im Nanowire-MOSFET besser erhalten bleibt [177].

Tabelle 6.6: Prozessablaufplan mit simulierter Entwicklung der 3D-Struktur (1/4-Struktur) für vertikale 22 nm-Nanowire-Transistoren.

Struktur	Prozessablauf
	<p>Schichtförmiger Aufbau von Drain- und Kanalgebiet</p> <ul style="list-style-type: none"> ➤ SOI-Wafer (Grenzfläche Silizium/Oxid bei 150 nm Tiefe) ➤ STI + Zurückätzen auf 5 ... 10 nm Siliziumdicke ➤ n-MOSFET <ul style="list-style-type: none"> • As; 60 nm; $1 \cdot 10^{20} \text{ cm}^{-3}$ • As; 33 nm; auf $1 \cdot 10^{15} \text{ cm}^{-3}$ abfallend • undotiertes Auffüllen (Kanalgebiet) ➤ p-MOSFET <ul style="list-style-type: none"> • Bor; 50 nm; $1 \cdot 10^{20} \text{ cm}^{-3}$ • Bor; 20 nm; auf $1 \cdot 10^{15} \text{ cm}^{-3}$ abfallend • undotiertes Auffüllen (Kanalgebiet) ➤ Ausheilung (5 s; 1000 °C)
	<p>Wire-Maskierung + HDP-Oxide</p> <ul style="list-style-type: none"> ➤ Lithographie <ul style="list-style-type: none"> • zweifache orthogonale Maskierung • anisotropes Ätzen des Silizium • veränderte Dotierungskonzentration als Ätzstopp ➤ HDP-Isolationsoxid <ul style="list-style-type: none"> • Auffüllen mit HDP-Oxid • anisotropes Zurückätzen
	<p>Gateoxid + Metallgate</p> <ul style="list-style-type: none"> ➤ Thermische Oxidation <ul style="list-style-type: none"> • 20 min; 850 °C • 1 nm Siliziumoxid ➤ Auffüllen der Gräben mit Metall <ul style="list-style-type: none"> • Füllhöhe entspricht der späteren Gatelänge • Austrittsarbeit bei 4.6 eV (Mitte der Bandlücke)
	<p>HDP-Oxid (Source)</p> <ul style="list-style-type: none"> ➤ HDP-Isolationsoxid <ul style="list-style-type: none"> • Auffüllen mit HDP-Oxid • anisotropes Zurückätzen • 12 nm Schichtdicke ➤ Chemisch-mechanisches Polieren
	<p>Dotierte Epitaxie (Source)</p> <ul style="list-style-type: none"> ➤ isotropes Zurückätzen des Siliziums um 3 nm ➤ n-MOSFET <ul style="list-style-type: none"> • Si-Epitaxie; 5 nm; As; $1 \cdot 10^{21} \text{ cm}^{-3}$ ➤ p-MOSFET <ul style="list-style-type: none"> • Si-Epitaxie; 5 nm; Bor; $1 \cdot 10^{20} \text{ cm}^{-3}$
	<p>RTA/LSA + Silizierung (Source)</p> <ul style="list-style-type: none"> ➤ RTA (5 s; 920 °C) ➤ LSA (2 ms; 1225 °C) ➤ Abscheidung; 15 nm; Nickel ➤ Diffusionsschritt; 0.5 s; 500 °C ➤ Entfernen des restlichen Nickels ➤ Auffüllen mit Kontaktmetall

Im nächsten Schritt werden die Gräben weiter mit dem Gatemetall aufgefüllt, wobei die Füllhöhe der physikalischen Gatelänge entspricht. Da die Kanäle von n- und p-MOSFETs zum größten Teil gering bis undotiert sind, könnte für beide das gleiche Metall mit einer Austrittsarbeit von 4.6 eV zum Vakuumniveau (Mitte der Siliziumbandlücke) verwendet werden. Die Schwellspannungen wären dann in beiden Fällen jedoch etwas höher als die Vorgaben der ITRS, weshalb für die Simulationen des p- und des n-MOSFET eine Austrittsarbeit von 4.74 eV bzw. 4.54 eV gewählt wurde. Anschließend wird zur Isolation zwischen Gatemetall und dem später folgenden Sourcekontakt erneut ein HDP-Oxid (Schichtdicke 12 nm) prozessiert.

Bevor ein weiterer epitaktischer Schritt folgt, wird das freiliegende Silizium isotrop um 3 nm zurück geätzt. Anschließend muss das Aufwachsen der Sourcegebiete von n- und p-MOSFET wieder getrennt realisiert werden, da für den n-MOSFET eine Konzentration von $1 \cdot 10^{21} \text{ cm}^{-3}$ Arsen und für den p-MOSFET $1 \cdot 10^{21} \text{ cm}^{-3}$ Bor in die 5 nm dicke Sourceschicht eingebunden wird.

Nun schließen sich die beiden üblichen Ausheilschritte an, wobei die RTA mit 5 s und 920 °C der bedeutsamere Prozess ist. Hierbei kommt es erneut zur Diffusion der Dotanden, nach der sich das finale Dotierungsprofil einstellt. Die Laserausheilung beträgt zwar 1225 °C, wirkt jedoch effektiv nur 2 ms lang, sodass sich die Dotanden kaum verlagern. Wie auch die Wannenausheilung ist die Laserausheilung für die Nanowire-MOSFETs nicht zwangsläufig erforderlich. Sie ist jedoch Bestandteil aktueller Technologien und wird in Hinblick auf die parallele Fertigung verschiedener Bauelemente in der Simulation mitberücksichtigt.

Zuletzt erfolgt die Silizierung der Sourcegebiete. Hierzu wird eine 15 nm Nickelschicht aufgebracht und ein Diffusionsschritt von 0.5 s bei 500 °C getätigt. Das überschüssige Nickel wird anschließend entfernt und die Kontaktgebiete werden mit Metall aufgefüllt.

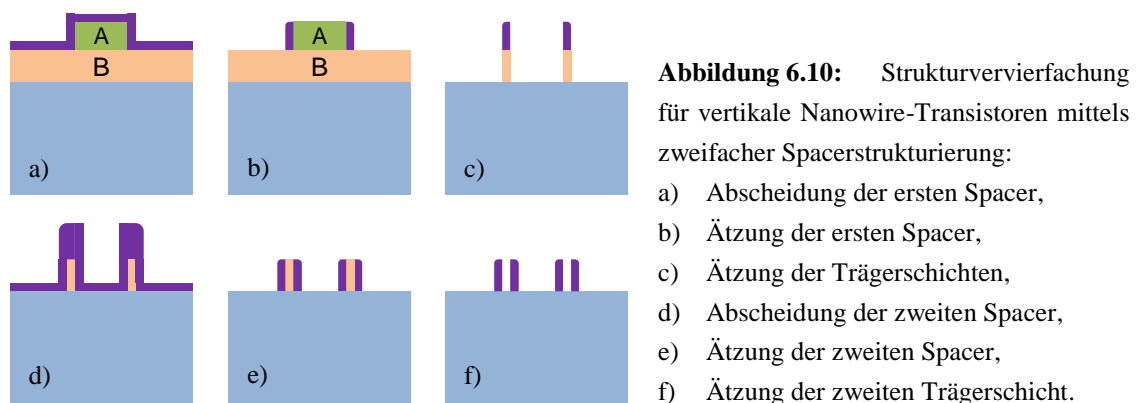
6.3.2 Strukturierung des Aktivgebiets

Für die Herstellung von vertikalen Nanowire-Strukturen existieren zwei grundlegende Varianten der Strukturierung, die in „Bottom-Up“ bzw. „Top-Down“ unterschieden werden.

Im Sinne eines Bottom-Up-Verfahrens wird auf dem Substrat ein Katalysator, wie z. B. Gold, aufgebracht, der das epitaktische Wachstum von Silizium fördert [178]. An dieser Stelle wächst dann auf dem Substrat ein nanometerfeiner Siliziumdraht gerichtet und mit dem Durchmesser des Katalysatortropfens auf. Um vertikale Siliziumdrähte zu erzeugen, bedarf es jedoch eines (111)-Substrats, da die Drähte bevorzugt in $\langle 111 \rangle$ -Richtung wachsen. Allerdings bietet die damit einhergehende $\langle 111 \rangle$ -Kanalrichtung für die späteren elektrischen Eigenschaften der Transistoren keine guten Grundbeweglichkeiten von Elektronen und Löchern. Einen weiteren Nachteil stellt die Platzierung des Katalysators auf dem Substrat dar, da es nur für sehr grobe Strukturen

mit großem Abstand zwischen den Goldtröpfchen möglich ist, ein Array vorzugeben und die Position der Nanowire-Strukturen gezielt zu steuern. Mit einem Aufsprühverfahren lassen sich zwar feinere und dichter gepackte Tröpfchen realisieren, jedoch in ihrer Lage nicht mehr beeinflussen, sodass zufällig verteilte Nanowire-Strukturen mit einem Durchmesser von mehr als 20 nm entstehen. Noch feinere Katalysatortröpfchen ließen sich durch zeitlich und thermisch gesteuertes Aufschmelzen einer dünnen Goldschicht erzeugen. Hierbei können Tröpfchen mit einem Durchmesser von unter 10 nm und in einer zufälligen Verteilung entstehen.

Im Top-Down-Verfahren nutzt man für die Formierung der Siliziumdrähte einen anisotropen Ätzschritt. Um vertikale Nanowire-Strukturen zu erzeugen, sind jedoch zwei Lithographieschritte erforderlich, bei denen orthogonal zueinander liegende Gräben in das Aktivgebiet geätzt werden. Entscheidend ist dabei die sehr genaue Steuerung von Breite und Abstand der Nanowire-Strukturen. Nach den Richtlinien der ITRS sollten die Durchmesser und Abstände der Siliziumdrähte für die 22 nm-Technologie bei ca. 17 nm liegen. Diese Abmessung ist jedoch mit einer einfachen Maske nicht mehr realisierbar und bedarf der Hilfe der Mehrfachstrukturierung (Multiple Patterning) [179]. Dies erfordert mehrere Prozessschritte und Oxid- bzw. Nitridmasken (Spacer), um von der minimalen Strukturbreite (ca. 45 nm) einer Fotolackmaske zu einer sehr schmale und vervielfältigte Struktur zu gelangen (Abbildung 6.10).



Dieser Ansatz kann dahingehend modifiziert werden, dass mehrere Spacer neben- bzw. aneinander prozessiert werden, wobei der Nitridspacer z. B. die Breite und der Oxidspacer den Abstand der Nanowire-Strukturen bestimmen (Abbildung 6.11). Die Mehrfach-Spacer-Strukturierung (Multiple Spacer Patterning) ermöglicht eine gute Steuerung der Nanowire-Geometrie, da sich die Schichtdicken von Oxid und Nitrid anhand der Prozessparameter bei der Abscheidung der Schichten bis auf 1 nm genau einstellen lassen. Im Anschluss erfolgt ein chemisch-mechanischer Polierschritt und ein Herauslösen des Oxids bevor die Gräben in das freiliegende Silizium geätzt werden.

Eine kombinierte Lösung aus Bottom-Up und Top-Down ist ebenfalls denkbar. So besteht nach [180] die Möglichkeit, Zinkoxiddrähte mithilfe eines Katalysators senkrecht auf einem beliebig orientierten Substrat aufwachsen zu lassen. Diese würden dann als Hartmaske für den anisotropen Ätzschritt des Siliziums dienen. Der Nachteil liegt jedoch auch hier bei der räumlich zufälligen Verteilung des Katalysators und dem weitläufigen Durchmesser der Zinkoxiddrähte.

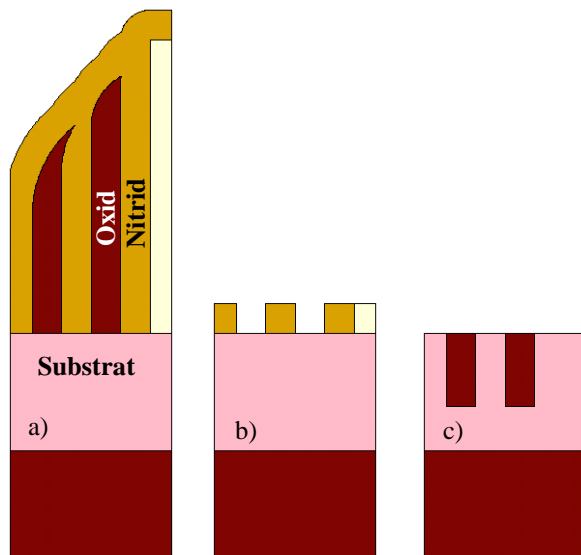


Abbildung 6.11: Mehrfach-Spacer-Strukturierung von Nanowire-Transistoren:

- a) Abscheidung der Spacer,
- b) CMP; Herauslösen des Oxids,
- c) Ätzen und Auffüllen der Gräben.

6.3.3 Charakterisierung des elektrischen Verhaltens

Zunächst sollen die stationären elektrischen Größen der beiden Transistortypen betrachtet werden (Tabelle 6.7). Der Abstand L_{Pitch} zwischen zwei Transistoren wird vorerst nicht bei der Charakterisierung der Nanowire-MOSFETs berücksichtigt, da er lediglich für die planaren bzw. Tri-Gate-Transistoren eine wesentlich begrenzende Rolle spielt. Aus diesem Grund werden alle Größen auf die effektive Gateweite normiert. Später soll ein Vergleich der Universalkennlinien mit einer Normierung auf die Aktivfläche zeigen, ob die vertikalen Nanowire-MOSFETs in Bezug auf eine effektive Platzausnutzung ebenso vorteilhaft sind wie für die Steuerfähigkeit des Gates.

Tabelle 6.7: Berechnete stationäre Kenndaten der Nanowire-Transistoren einer 22 nm-Technologie, $L_{Gate} = 26$ nm, $W_{NW} = 17$ nm, $U_{DD} = 0.7$ V.

Typ	U_{th} (mV)		$DIBL$ (mV/V)	SS (mV/dec)		g_m (μ S/ μ m)		I_D (μ A/ μ m)		$I_{D,off}$ (nA/ μ m)	R_{on} ($\Omega \cdot \mu$ m)
	lin.	sat.		lin.	sat.	lin.	sat.	lin.	sat.		
n-MOS	308	285	34.0	65.9	66.3	419	2181	136	585	0.22	463
p-MOS	-317	-286	44.6	65.1	65.0	213	1256	69	386	0.22	860

Da in den Nanowire-MOSFETs keine Verspannungskomponenten eingebaut sind, kann für n- und p-MOSFETs ein Verhältnis der Sättigungsströme von ca. 2:1 beobachtet werden, wie es aufgrund der unterschiedlichen Grundbeweglichkeiten der Ladungsträger bereits bei älteren Transistorgenerationen (ohne mechanische Verspannung) üblich war (Abbildung 6.12). Die lineare bzw. Sättigungsschwellspannung verhält sich im Bereich von $L_{Gate} > 26$ nm bei einem Kanaldurchmesser von ca. 17 nm verhältnismäßig konstant, da das Kriterium $L_{Gate}/W_{NW} \geq 1.5$ noch erfüllt ist. Erst bei kürzeren Gatelängen setzt der Kurzkanaleffekt ein und die Sättigungsschwellspannungen fallen ab.

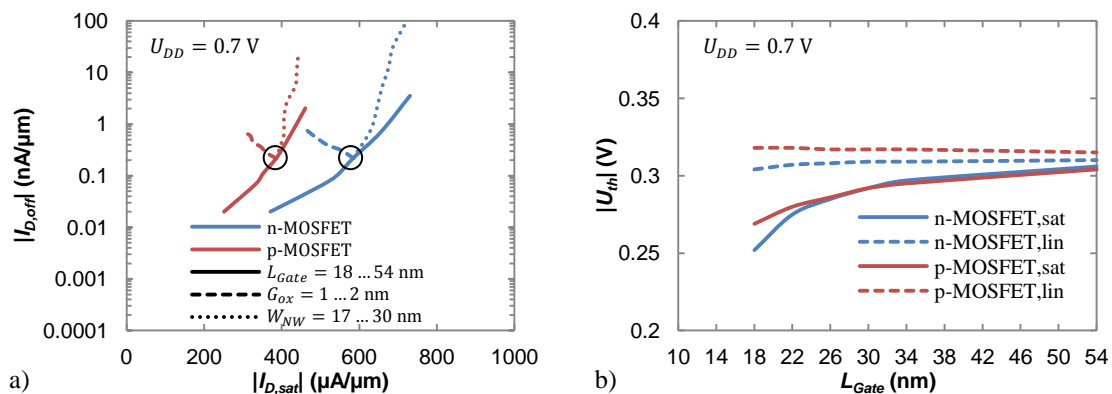


Abbildung 6.12: Stationäres Verhalten von n- und p-Nanowire-MOSFETs. a) Berechnete Universal-kennlinie für Variationen der Gatelänge, Gateoxiddicke und Kanalweite. Die Kreise markieren die nominellen Transistoren mit $L_{Gate} = 26$ nm, $G_{ox} = 1$ nm und $W_{NW} = 17$ nm. b) Schwellspannungs-verhalten.

Eine Variation der Gateoxiddicke resultiert hauptsächlich in einer Veränderung der Gatekapazität und der damit verbundenen Ladungsträgerdichte im invertierten Kanal. Ein dünneres Gateoxid erhöht jedoch nicht nur die Anzahl an verfügbaren Ladungsträgern bzw. die Leitfähigkeit des Transistors, sondern reduziert auch die Unterschwellsteigerung und somit den Sperrstrom des Transistors. Deutlich wird dies durch den erhöhten Drainstrom bei geringerem Leckstrom der Nanowire-MOSFETs.

Betrachtet man die Schwellspannung in Abhängigkeit von der Gateoxiddicke, fällt auf, dass sie für Nanowire-MOSFETs im Gegensatz zu planaren bzw. Tri-Gate-Transistoren nahezu konstant ist (Abbildung 6.13). Der Anstieg der Schwellspannung von planaren und Tri-Gate-Transistoren mit größer werdender Gateoxiddicke resultiert aus einer starken Verschiebung der Transferkennlinie im Unterschwellbereich. Durch die Austrittsarbeit des Gatemetalls der beiden Transistoren von 4.1 eV befindet sich der ausgeschaltete n-MOSFET bereits in einem verarmten Zustand. Eine Variation der Gateoxiddicke verändert demnach auch die Ladungsträgerdichte und die Leitfähigkeit des gesperrten planaren bzw. Tri-Gate-Transistors. Der Nanowire-MOSFET mit seinem

undotierten Kanal und einer Austrittsarbeit im Bereich der Bandlückenmitte zeigt dagegen nur eine Änderung der Unterschwellsteigung, jedoch keine Verschiebung der Transferkennlinie mit der Gateoxidicke.

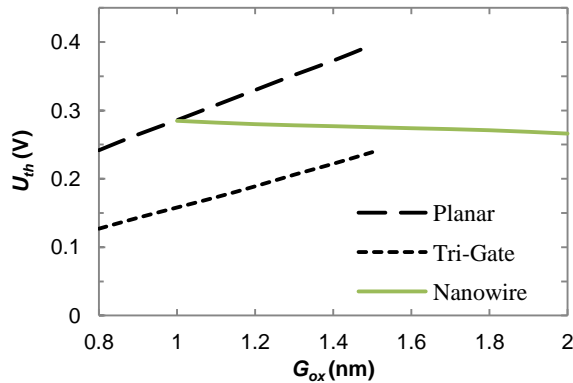


Abbildung 6.13: Berechnete Schwellspannung (Sättigung) eines planaren, Tri-Gate- (ohne Eckenimplantation zur Korrektur der Schwellspannung) sowie Nanowire-Transistors (n-MOSFET) in Abhängigkeit von der Gateoxidicke. ($L_{Gate} = 26$ nm)

Die Unterschwellsteigung der Nanowire-Transistoren liegt knapp über dem theoretischen Minimum von ca. 60 mV/dec und ist im Vergleich zu den anderen Transistorstrukturen deutlich besser (Abbildung 6.14). Wie auch bei den Schwellspannungen nimmt die Steuerfähigkeit des Gates erst unterhalb der nominellen Gatelänge ($L_{Gate} = 26$ nm) ab und lässt die Werte ansteigen. Bezüglich der linearen Unterschwellsteigung in Abhängigkeit vom Skalierungsparameter α liegen die Nanowire-Transistoren leicht unter dem Niveau der FinFETs und somit auch noch ein wenig näher an der Ideallinie. Für die Berechnung der natürlichen Länge wurde die Gleichung einer Quadrupel-Gate-Struktur mit $\lambda_4 = 3.59$ nm angenommen, da die Kanalgebiete nicht ideal rund sind, sondern bei ihrer Strukturierung aufgrund des orthogonalen Ätzens eine eher quadratische Form annehmen.

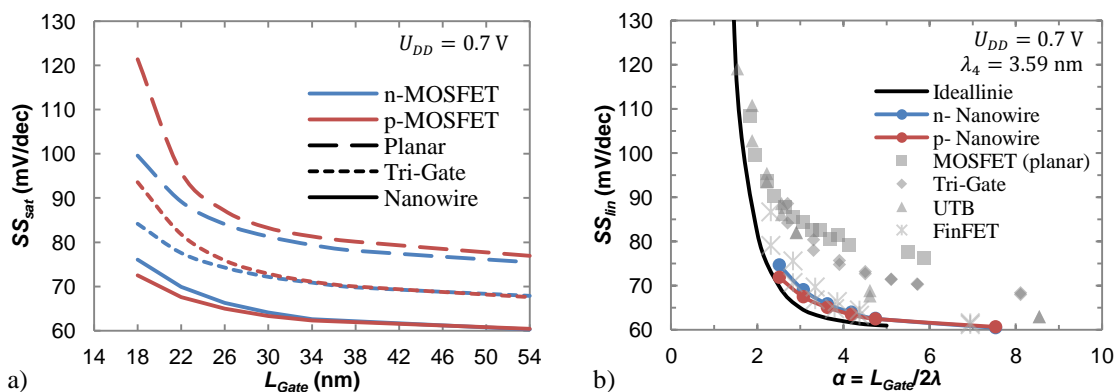


Abbildung 6.14: Berechnete Unterschwellsteigung von n- und p-Nanowire-MOSFETs im a) Sättigungsbereich in Abhängigkeit von der physikalischen Gatelänge und b) Linearbereich in Abhängigkeit von dem Skalierungsparameter α . ($t_{ox} = 1$ nm und $W_{NW} = 17$ nm)

Der Durchgangswiderstand lässt für die vertikalen Nanowire-MOSFETs ein lineares Verhalten erkennen, wobei der p-MOSFET deutlich höhere Werte aufweist (Abbildung 6.15). Eine lineare Approximation der Simulationsdaten zeigt den jeweiligen Schnittpunkt der Geraden mit der y-Achse und somit den parasitären Kontaktwiderstand R_{SD} von Source und Drain. Dieser ist beim p-MOSFET mit $570 \Omega \cdot \mu\text{m}$ fast doppelt so groß wie beim n-MOSFET mit $300 \Omega \cdot \mu\text{m}$. Einer der Gründe hierfür ist die geringe Dotierung des Sourcegebiets. Da die Diffusion von Bor wesentlich größer ist als die von Arsen, wurde beim epitaktischen Aufwachsen des Sourcegebiets vom p-MOSFET eine geringere Konzentration angesetzt, um auch nach der thermischen Ausheilung noch eine vergleichbare effektive Gatelänge von n- und p-MOSFETs zu erhalten. Ein weiterer Grund ist die höhere Grundbeweglichkeit der Elektronen. Beide Einflüsse stehen im direkten Zusammenhang mit der Leitfähigkeit der Kontaktgebiete.

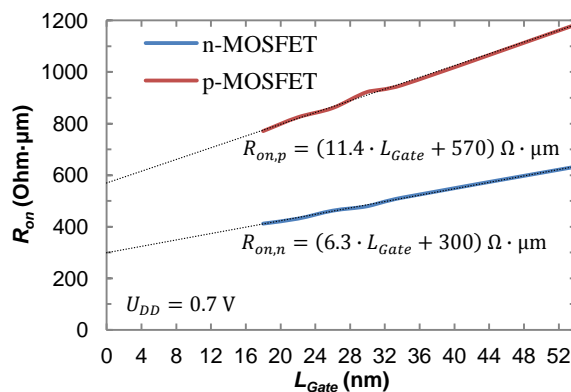


Abbildung 6.15: Berechneter Durchgangswiderstand mit linearer Approximation in Abhängigkeit von der physikalischen Gatelänge der n- und p-Nanowire-MOSFETs. ($t_{ox} = 1 \text{ nm}$ und $W_{NW} = 17 \text{ nm}$)

Aufgrund der dreidimensionalen Struktur der Nanowire-MOSFETs gibt es auch bei der Betrachtung der stationären elektrischen Größen eine größere Anzahl an Strukturparametern, die zur Charakterisierung herangezogen werden können, wie z. B. der Durchmesser des Kanalgebiets. Er kann auch als Weite des Nanowire-MOSFETs W_{NW} bezeichnet werden und verfügt über eine vergleichbare Aussagekraft wie die Gatelänge. Mit größer werdender Weite der Nanowire-Strukturen verschlechtern sich die Schalteigenschaften der Transistoren und die Schwellspannung fällt ab (Abbildung 6.16). Auffällig ist, dass die Differenz zwischen linearer und Sättigungsschwellspannung (*DIBL*) schnell ansteigt. Ein wesentlicher Grund dafür ist die sehr niedrige Kanaldotierung. Wie bereits bei den Tri-Gate-Transistoren mit dem Eckeneffekt gezeigt werden konnte (Abbildung 5.12), verlagert sich die Ladungsträgerdichte bei geringer Dotierung zur Kanalmitte, wo sie sich bei größeren Durchmessern schwerer steuern lassen.

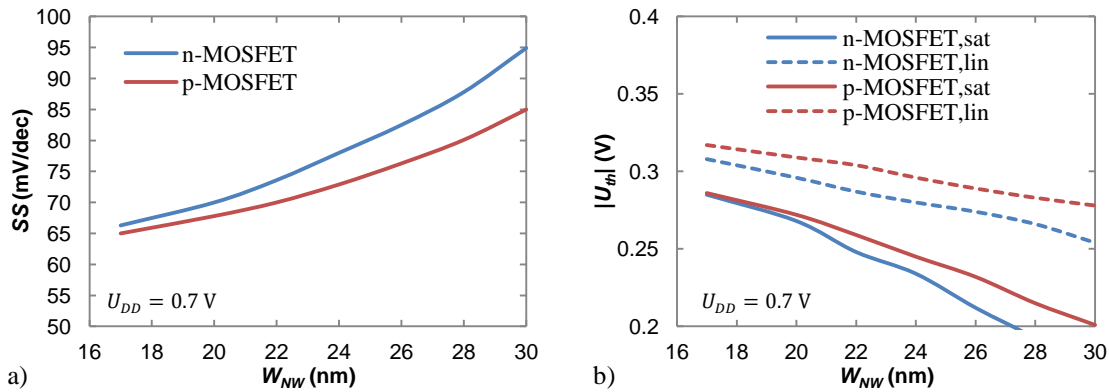


Abbildung 6.16: Berechnete a) Unterschwellsteigung und b) Schwellspannung von n- und p-Nanowire-MOSFET in Abhängigkeit von der Kanalweite. ($t_{ox} = 1$ nm, $L_{Gate} = 26$ nm)

Ein Vergleich mit Tri-Gate- bzw. planaren Transistoren macht deutlich, welche Leistungsfähigkeit ein vertikaler Nanowire-MOSFET gegenüber den derzeit in Produktion befindlichen Strukturen besitzt (Abbildung 6.17). Die Universalkennlinie zeigt bei gleichem Sperrstrom eine Steigerung des Sättigungsstromes von 28 % gegenüber dem Tri-Gate-Transistor. Auch anhand der Schwellspannung lässt sich der positive Einfluss des Rundum-Gates verdeutlichen. Während der planare Transistor unterhalb einer Gatelänge von 34 nm einen steiler werdenden Abfall der Schwellspannung und einen deutlichen Einfluss des Kurzkanaleffektes zeigt, bleibt der Verlauf des vertikalen Nanowire-MOSFETs stabil und geht erst bei einer Gatelänge $L_{Gate} \leq 26$ nm in einen Abfall der Schwellspannung über.

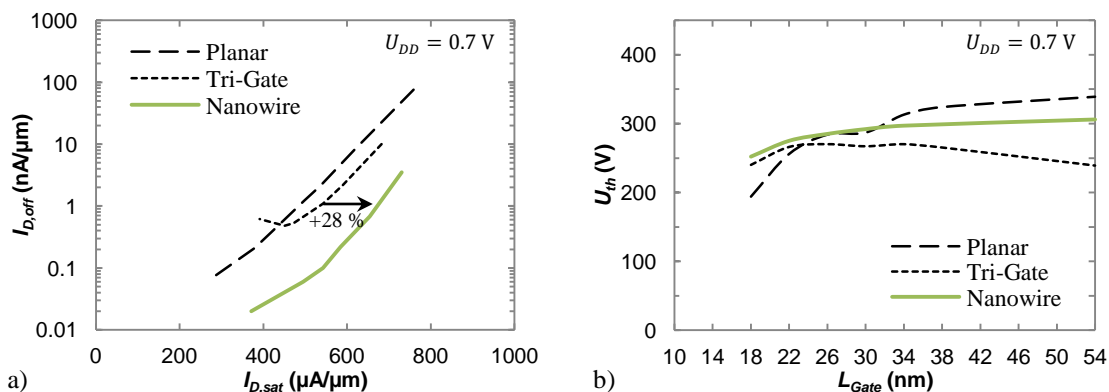


Abbildung 6.17: Berechnete a) Universalkennlinie und b) Schwellspannung (Sättigung) von planaren, Tri-Gate sowie Nanowire-MOSFETs. Kennwerte normiert auf die effektive Gateweite. (n-MOSFET, $L_{Gate} = 18 \dots 54$ nm)

Auch bei einer Normierung auf die gesamte Aktivfläche des Transistors ist der Vorteil einer vertikalen Integration des Nanowire-MOSFETs sehr deutlich erkennbar. Dies liegt jedoch nicht an einer besseren Platzausbeute der Struktur, sondern an dem effektiv

nutzbaren Kanalquerschnitt pro Aktivfläche. Als Beispiel stehen einem Nanowire-Transistor mit einem Draht von 17 nm Durchmesser und 17 nm Abstand zum nächsten ($Pitch_{NW} = 34 \text{ nm}$) 1156 nm^2 Silizium zur Verfügung, von denen der Kanalquerschnitt 227 nm^2 beansprucht (Abbildung 6.18). Hinzu kommt eine Fläche von 2040 nm^2 (64 % der Aktivfläche), die für die 30 nm breiten Grabenkontakte von Drain und Gate benötigt werden. Somit ergibt sich eine gesamte Aktivfläche von 3196 nm^2 .

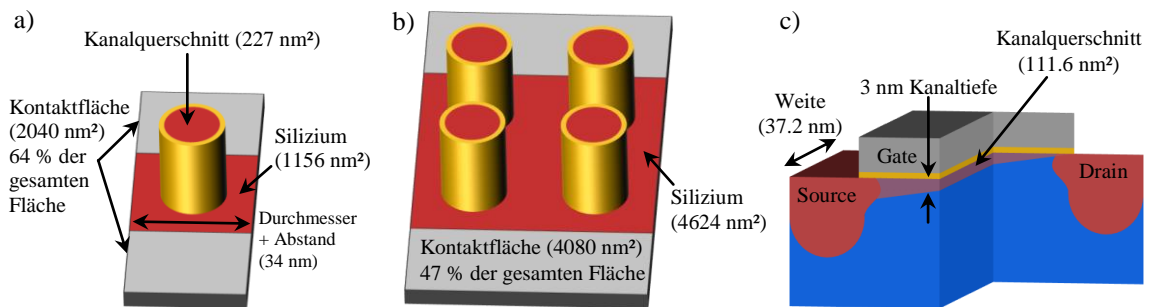


Abbildung 6.18: Berechnung der Aktivfläche und des Kanalquerschnitts von Nanowire-Transistoren mit a) einem Draht, b) zwei Drähten und c) eines planaren MOSFETs mit äquivalenter Aktivfläche zu a).

Ein planarer Transistor mit 86 nm $Pitch$ kommt bei gleicher Aktivfläche auf 37.2 nm Weite und 111.6 nm^2 Kanalquerschnitt, wenn man davon ausgeht, dass sich sein Kanal bis ca. 3 nm tief unter das Gate ausdehnt. Dies entspricht dem Faktor 2 zwischen den Kanalquerschnitten und dem damit verbundenen Strom pro Aktivfläche und resultiert in einer stark vergrößerten Leistungsfähigkeit der vertikalen Nanowire-MOSFETs. Zudem reduziert sich der prozentuale Anteil der Kontaktfläche, je mehr Drähte pro Nanowire-Transistor verwendet werden. Bei einem vernachlässigbar kleinen Anteil der Kontakte (0 %), beträgt der Faktor zwischen den Kanalquerschnitten des Nanowire- und des planaren Transistors rund fünf (Abbildung 6.19). Mit einem durchdachten Schaltungsdesign kann die benötigte Fläche für die Kontaktierung unter den Transistoren nochmals erheblich reduziert und die Leistungsfähigkeit pro Aktivfläche somit verbessert werden [181].

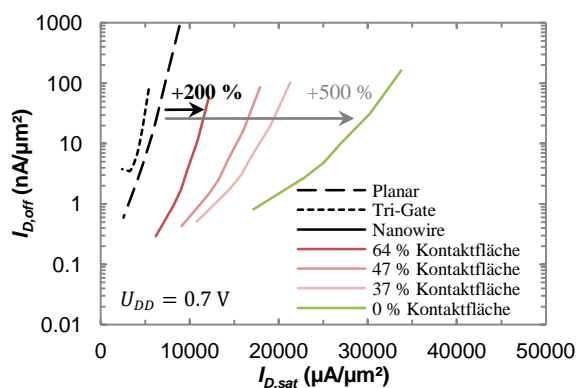


Abbildung 6.19: Berechnete Universalkennlinie von planaren, Tri-Gate sowie Nanowire-MOSFETs mit unterschiedlicher Anzahl an Drähten pro Transistor (1 Draht \rightarrow 64 % Kontaktfläche, 2x2 Drähte \rightarrow 47 % Kontaktfläche, 3x3 Drähte \rightarrow 37 % Kontaktfläche). Kennwerte normiert auf die Aktivfläche. (n-MOSFET, $L_{Gate} = 18 \dots 54 \text{ nm}$)

Auch bei den dynamischen Kenndaten der 22 nm-Nanowire-Transistoren machen sich die Vorteile der vertikalen Integration bemerkbar. Die Überlappungs- und Inversionskapazität C_{ov} bzw. C_{inv} (normiert auf die effektive Gateweite) erreichen im Vergleich zu den planaren Transistoren deutlich geringere Werte (Tabelle 6.8). Damit kann angenommen werden, dass die kapazitive Wirkung zwischen dem Gatekontakt und der Drain- bzw. Source-Schicht aufgrund der 10 ... 12 nm dicken Spacer-Schichten hinreichend gut entkoppelt ist. In Kombination mit dem hohen Sättigungsstrom erreichen die Schaltungen des dreistufigen Inverters und des fünfstufigen Ringoszillators im Vergleich zu den anderen Transistorstrukturen die kürzesten Schaltzeiten bei einer akzeptablen Stromaufnahme pro Aktivfläche. Dabei wurden Nanowire-MOSFETs mit einem Draht pro Transistor simuliert. Es ist davon auszugehen, dass sich auch die Verzögerungszeiten ändern, wenn sich die Anzahl der Drähte pro Transistor erhöht [182].

Tabelle 6.8: Berechnete dynamische Kenndaten von Nanowire-Transistoren in 22 nm-Technologie, $L_{Gate} = 26$ nm, $W_{NW} = 17$ nm, $U_{DD} = 0.7$ V.

Einzeltransistor	Typ	C_{ov} (fF/ μ m)	C_{inv} (fF/ μ m)	f_{max} (GHz)	f_i (GHz)	
	n-MOSFET	0.255	0.872	1601	466	
p-MOSFET	0.212	0.861	833	310		
Dreistufiger Inverter	t_r (ps)	t_f (ps)	Δt_{hi} (ps)	Δt_{low} (ps)	U_{max} (V)	U_{min} (V)
	3.11	2.40	2.21	2.17	0.750	-0.060
Fünfstufiger Ringoszillator	f_{RO} (GHz)	$I_{RO,eff}$ (μ A/ μ m ²)	$I_{RO,off}$ (nA/ μ m ²)	t_D (ps)	Gesamte Aktivfläche (μ m ²)	
	43.83	965	1.79	2.28	0.03196	

6.3.4 Asymmetrisches Dotierungsprofil

Das asymmetrische Dotierungsprofil ist ein Resultat des vertikalen Transistoraufbaus. Durch die unterschiedlichen Zeitpunkte, zu denen Drain- und Sourcegebiete der Transistoren gefertigt werden, sind sie auch verschiedenen thermischen Budgets ausgesetzt. Sie können demnach nicht mit dem gleichen Konzentrationsverlauf der Dotanden aufgewachsen werden, da entweder das Draingebiet (unterhalb) den Kanal durch Diffusion der überzähligen Dotanden überfluten würde oder der Kontaktwiderstand auf der Source-Seite (oberhalb) bedingt durch fehlende Ladungsträgern spürbar zu hoch wäre. Somit ist die Verwendung der Kontakte als Drain bzw. Source im Gegensatz zu symmetrisch aufgebauten horizontalen Transistoren nicht mehr frei wählbar, wenn man auf ein Maximum der Leistungsfähigkeit bedacht ist (Abbildung 6.20). Anhand der Universalkennlinie wird deutlich, dass ein n-Nanowire-Transistor mit vertauschten Source- und Draingebieten um 30 % leistungärmer ist als ein Transistor mit regulärer Beschaltung.

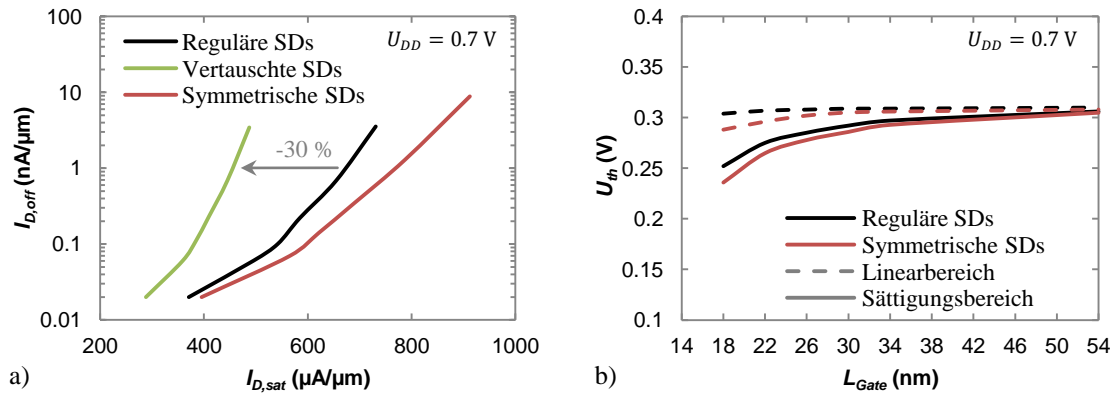


Abbildung 6.20: n-Nanowire-MOSFET mit regulärem (asymmetrischen) und symmetrischem Dotierungsprofil sowie inverser Beschaltung von Source und Drain. Berechnete a) Universalkennlinie und b) Schwellspannung in Abhängigkeit von der Gatelänge.

Zum Vergleich wurden noch Nanowire-MOSFETs mit symmetrischem Dotierungsprofil simuliert. Sie zeigen die beste Leistungsfähigkeit der drei Varianten, jedoch auch ein etwas stärkeren Kurzkanaleffekt, dargestellt durch den Verlauf der Schwellspannung über der Gatelänge. Die Ursache für den größeren Sättigungsstrom liegt zum einen bei der erhöhten Dotierungskonzentration des Draingebiets und der damit verbesserten Leitfähigkeit und zum anderen bei der höheren durchschnittlichen Feldstärke in Transportrichtung (Abbildung 6.21).

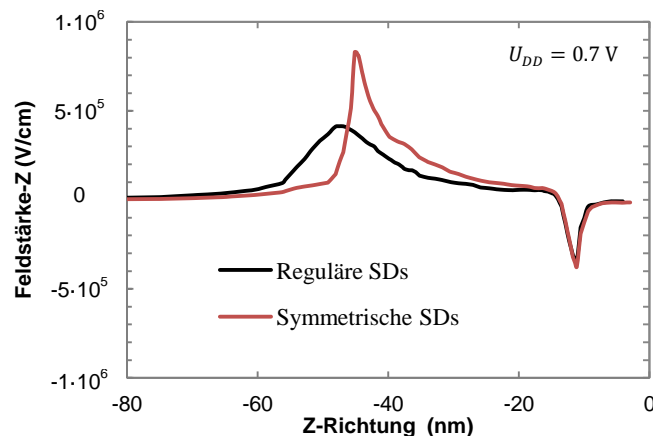


Abbildung 6.21: Elektrische Feldstärke in Transportrichtung für einen n-Nanowire-MOSFET mit regulärem (asymmetrischem) und symmetrischem Dotierungsprofil.

Die drainseitige Überlappung ist zugunsten eines besseren Kurzkanalverhaltens etwas verringert worden, wodurch die Feldstärkenspitze außerhalb des Gatekontakts liegt. Die Leistungsfähigkeit des Transistors wird dadurch nur geringfügig gemindert. Entscheidend ist für den MOSFET die sourceseitige Überlappung (Abbildung 6.22). Im negativen Bereich (keine Überlappung) besitzt sie einen deutlich steileren Anstieg von $-13\ \Omega \cdot \mu\text{m}/\text{nm}$, bis sich eine Überlappung mit dem Gatekontakt ergibt und der Verlauf im Positiven zu einem Anstieg von $-5\ \Omega \cdot \mu\text{m}/\text{nm}$ abknickt, der dem Verhalten über der

Gatelänge aus Abbildung 6.14 gleichkommt. Die Asymmetrie wirkt sich hinsichtlich der Leistungsfähigkeit jedoch nicht nur nachteilig aus, sondern ermöglicht auch eine besonders hohe Spannungsfestigkeiten des Bauteils, da das Maximum der Feldstärke nur der Hälfte von Transistoren mit symmetrischem Profil entspricht [183].

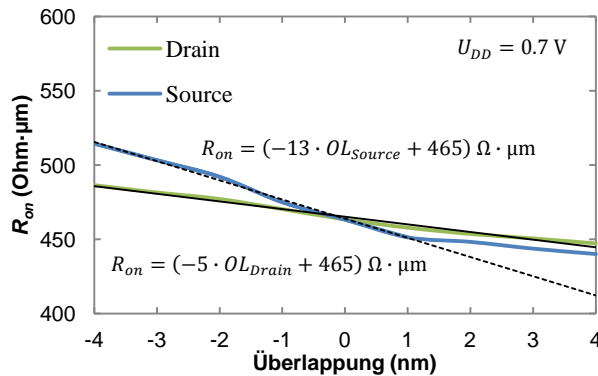


Abbildung 6.22: Durchgangswiderstand R_{on} in Abhängigkeit der Überlappung vom Drain- bzw. Sourcegebiet zum Gatekontakt. Negative Werte entsprechen einer fehlenden Überlappung.

6.3.5 Mechanische Verspannung

Wie auch bei Tri-Gate-Transistoren ermöglichen vertikale Nanowire-MOSFETs neue Konzepte für den Einsatz von Verspannungstechniken. Das prinzipielle Vorgehen zur Einbringung mechanischer Verspannungen in den Transistor entspricht dem von planaren Transistoren. So nutzt man auch bei Multi-Gate-Strukturen die Möglichkeit, verschiedene Materialien mit einer intrinsischen Verspannung auszustatten, oder eine Gitterdeformation durch die Epitaxie von Silizium-Germanium auf reinem Silizium hervorzurufen. Eine intrinsische Verspannung kann dank moderner Herstellungsverfahren bereits bei den meisten Materialien des Nanowire-Transistors erzeugt werden. Somit ist es möglich, die Spacerschichten zwischen dem Gatemetall und den Source- bzw. Drainschichten zug- oder druckverspannt zu gestalten. Ebenso bietet das Gatemetall selbst eine isotrope Druckverspannung von ca. 500 MPa. Die Simulationen für n- und p-Nanowire-Transistoren beider Varianten zeigten allerdings nur sehr geringe Verschiebungen der Universalkennlinien, sodass der Aufwand zur Prozessierung dessen Nutzen übersteigt.

Ähnlich dem Einsatz von Silizium-Germanium im Kanal des planaren p-MOSFETs eröffnet sich auch beim Nanowire-Transistor die Möglichkeit, das Kanalgebiet mit einem Kern aus Silizium und einer umgebenden Hülle aus aufgewachsenem Silizium-Germanium zu gestalten (Abbildung 6.23). Die unterschiedlichen Gitterkonstanten der beiden Materialien erzeugen eine biaxiale Verspannung entlang der Grenzfläche, wobei der Kern zug- und die Hülle, in der sich bei Inversion der Großteil der Ladungsträger sammelt, druckverspannt ist. Orthogonal zur Oberfläche des Siliziumdrahtes kommt es zur vollständigen Relaxation. Die Stärke der mechanischen Verspannung hängt auch hier vorrangig vom prozentualen Anteil des Germaniums in der Hülle ab. Für die Simu-

lationen wurde ein Germaniumgehalt von 30% angenommen, womit sich eine Verspannung im Bereich von ca. 1 GPa der jeweiligen Komponenten einstellte. Weiterhin ist das Verhältnis der Volumen beider Bereiche wichtig. Bei einer Dominanz der SiGe-Hülle würde der Si-Kern keinen ausreichenden Gegendruck aufbauen können und die Verspannung im Bereich des Ladungsträgertransports kleiner werden. Überwiegt der Kern, könnte sich die Grenze der Verspannung in das Gebiet der Inversion verlagern, womit ein Teil der Ladungsträger einer Zugverspannung und der andere Teil einer Druckverspannung ausgesetzt wäre und sich deren Wirkungen gegenseitig neutralisieren würden. Bei dem prozessierten Durchmesser des Kanals von 17 nm bot es sich an, das Volumen zu gleichen Teilen auf Kern und Hülle zu verteilen, womit sich eine Schichtdicke der Hülle von 2.5 nm ergab.

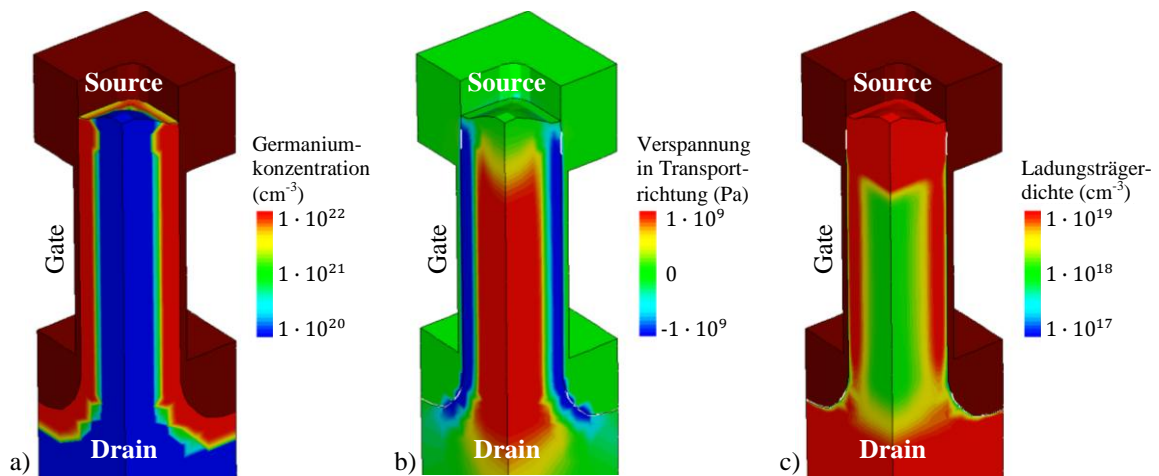


Abbildung 6.23: Mechanische Verspannung von Nanowire-Transistoren mit SiGe-Hülle. a) Verteilung des Germaniumanteils, b) erzeugte mechanische Verspannung in Transportrichtung und c) Ladungsträgerdichte.

Ebenso besteht die Möglichkeit, einen SiGe-Kern mit einer Si-Hülle zu erzeugen. Die Wirkung der mechanischen Verspannung würde sich dabei entgegengesetzt zum vorher genannten Beispiel auswirken. Für beide Transistortypen zeigte sich jedoch in den Simulationen, dass nur der Si-Kern mit der SiGe-Hülle zu einer Leistungssteigerung von 25 ... 60 % führen kann (Abbildung 6.24). Die positive Reaktion des p-MOSFETs auf die biaxiale Druckverspannung in der Hülle konnte bereits bei den planaren Transistoren der 32 nm-SOI-Technologie beobachtet werden und verhielt sich damit erwartungsgemäß [15]. Überraschend war jedoch die Reaktion des n-MOSFETs, dessen Drainstrom durch die Druckverspannung ebenfalls gesteigert werden konnte. Hierbei muss berücksichtigt werden, dass der simulierte Nanowire-Transistor nicht ideal rund, sondern durch die zwei orthogonal überlagerten Masken eher rechteckig geformt ist und abgerundete Kanten besitzt. Somit ergeben sich zwei verschiedene Oberflächenorientierungen des Gatekontakts mit einer gemeinsamen Kanalrichtung. Bei der (100)-Ober-

fläche verursacht die Druckverspannung eine Aufspaltung der Subbänder zu Ungunsten der Orbitale mit der geringeren Transportmasse, was jedoch größtenteils durch die entgegenwirkende Quantisierung der Leitungsbänder kompensiert werden kann. Die (110)-Oberfläche hingegen profitiert von den Verspannungskomponenten, wodurch auch die Hälfte der Ladungsträger im Nanowire über eine erhöhte Ladungsträgerbeweglichkeit verfügt (Anhang 6).

Die Simulationen müssen jedoch kritisch betrachtet werden, da sie mit Modellen berechnet wurden, die eine große Ausdehnung des Kristallgitters voraussetzen. Der Einsatz von Silizium-Germanium verändert zudem die Bandlücke und somit die Einsatzspannung. Es ist an dieser Stelle schwer einzuschätzen, wie stark die Schwellspannung durch den Si-Kern bzw. der SiGe-Hülle insgesamt beeinflusst wird.

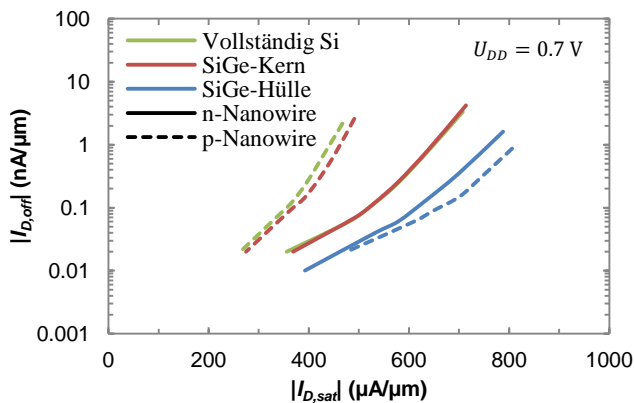


Abbildung 6.24: Berechnete Universal-kennlinien von n- und p-Nanowire-Transistoren mit mechanischer Verspannung, erzeugt durch den Einsatz von Silizium-Germanium im Kern bzw. der Hülle des Kanalgebiets. ($W_{NW} = 17 \text{ nm}$, $L_{Gate} = 18 \dots 54 \text{ nm}$)

7 Skalierung und statistische Schwankungen der Strukturen

Mit den beschriebenen Transistorstrukturen auf Basis einer 22 nm-Technologie sind nur die derzeitigen bzw. kurzfristigen Ziele der Halbleiterindustrie analysiert worden. Mittel- und langfristig (10 ... 20 Jahre) betrachtet sollten diejenigen Strukturvarianten für die Volumenproduktion favorisiert werden, die auch weitere Skalierungsschritte über die 22 nm-Technologie hinaus ermöglichen und zudem eine steigerbare Leistungsfähigkeit bieten. Hinzu kommen die Anforderungen für einen kostengünstigen Herstellungsprozesses und für die Zuverlässigkeit der Bauelemente hinsichtlich der Variabilität ihres Dotierungsprofils, der fixen Ladungen im Gateoxid, der Metall-Gate-Granularität und der Geometrie der Transistoren. Die steigende Anzahl an Transistoren bzw. die zunehmende Komplexität der Schaltkreise verlangt dabei ein hohes Maß an statistischer Sicherheit für die Funktionalität der gesamten Schaltung.

Im ersten Abschnitt 7.1 des Kapitels werden die ausgewählten Transistorstrukturen entsprechend der ITRS-Richtlinien für eine 14 nm-Technologie skaliert und bezüglich der Leistungsfähigkeit, des Kurzkanalverhaltes und der Steuerfähigkeit des Gatekontakts mit den 22 nm-Technologien verglichen. Anschließend werden im zweiten Abschnitt 7.2 alle Strukturvarianten hinsichtlich der Fluktuationen von Prozessparametern und der damit verbundenen Variation ihrer stationären Eigenschaften charakterisiert, bevor ein Vergleich aller Varianten mit kombinierten Störquellen das Kapitel abschließt.

7.1 Skalierung zur 14 nm-Technologie

7.1.1 Leistungsfähigkeit

Mit der Skalierung der verschiedenen Transistorstrukturen soll untersucht werden, welche Variante für eine Weiterentwicklung über die 22 nm-Technologie hinaus am besten geeignet ist. Die MOSFETs sind dabei hinsichtlich ihrer Dimensionen nach den Richtlinien der ITRS [22] verkleinert und an eine Sättigungsschwellspannung von ca. $U_{th,sat} = 290$ mV angepasst worden (Tabelle 7.1). Die äquivalente Gateoxiddicke EOT wird jedoch aus prozesstechnischen Gründen mit 1 nm beibehalten. Die Schritte zur Formierung der Finnen aller Multi-Gate-Strukturen stellen in der Realität die größten Herausforderungen dar. In diesem Zusammenhang ist eine Prozessverbesserung für den Schichtauf- bzw. Schichtabtrag vorausgesetzt worden, wie sie bereits in den letzten Jahren immer wieder erzielt werden konnte [184].

Tabelle 7.1: Vergleich der Technologieparameter für die 22 nm- und die 14 nm-Technologien der verschiedenen Transistorstrukturen.

Strukturvariante	Technologieparameter	22 nm	14 nm
Allgemein	L_{Gate}	26 nm	20 nm
	EOT	1.0 nm	1.0 nm
	$Pitch$	86 nm	60 nm
Planar/Tri-Gate	Spacerbreite	15 nm	10 nm
	Halowinkel	25°	20°
	n-Halodosis	$BF_2, 5.7 \cdot 10^{13} \text{ cm}^{-2}$	$BF_2, 8.5 \cdot 10^{13} \text{ cm}^{-2}$
	p-Halodosis	$As, 3.7 \cdot 10^{13} \text{ cm}^{-2}$	$As, 5.5 \cdot 10^{13} \text{ cm}^{-2}$
	Energie, n-Erweiterung	1.5 keV	1.0 keV
Tri-Gate	$Pitch_{Fin}$	50 nm	35 nm
	W_{Fin}	20 nm	15 nm
	H_{Fin}	7 nm	5 nm
	n-Eckendosis	$BF_2, 5 \cdot 10^{14} \text{ cm}^{-2}$	$BF_2, 4 \cdot 10^{14} \text{ cm}^{-2}$
	p-Eckendosis	$As, 5 \cdot 10^{14} \text{ cm}^{-2}$	$As, 2 \cdot 10^{14} \text{ cm}^{-2}$
UTB	t_{Si}	8 nm	6 nm
FinFET	$Pitch_{Fin}$	50 nm	35 nm
	W_{Fin}	10 nm	7 nm
	H_{Fin}	30 nm	30 nm
	Spacerbreite	8 nm	6 nm
Nanowire	$Pitch_{NW}$	17 nm	13 nm
	W_{NW}	17 nm	13 nm

Die Leistungsfähigkeit der unverspannten n-MOSFETs ist in allen Fällen erwartungsgemäß leicht angestiegen (Abbildung 7.1a). Dies resultiert hauptsächlich aus der Verkürzung der Struktur und der damit verbundenen Reduzierung der seriellen Widerstände der Transistoren. Auch für den p-UTB- bzw. den p-Nanowire-Transistor trifft diese Aussage zu. Im Gegensatz dazu besitzen die skalierten p-MOSFETs der planaren, Tri-Gate- und FinFET-Strukturen einen etwas geringeren Sättigungsstrom, da die mechanische Verspannung der mit Germanium versetzten Source-Drain-Gebieten aufgrund des verringerten Volumens abgenommen hat.

Die Betrachtung der auf die Aktivfläche der Transistoren normierten Sättigungsströme macht deutlich, dass sich bei einer Skalierung der FinFETs das Verhältnis von effektiver Gateweite zu Strukturweite von $70 \text{ nm} : 50 \text{ nm} \approx 1.4$ zu $67 \text{ nm} : 35 \text{ nm} \approx 1.9$ erhöht und somit die Leistungsfähigkeit der Transistoren stärker ansteigt als die der planaren Strukturen (Abbildung 7.1b). So gesehen können auch die Nanowire-Transistoren von der Halbierung ihrer Aktivfläche profitieren, da sich ihre stromdurchflossene Fläche aufgrund der Kreisgeometrie und der zunehmenden Volumeninversion nur um ca. 25 % verringert [185].

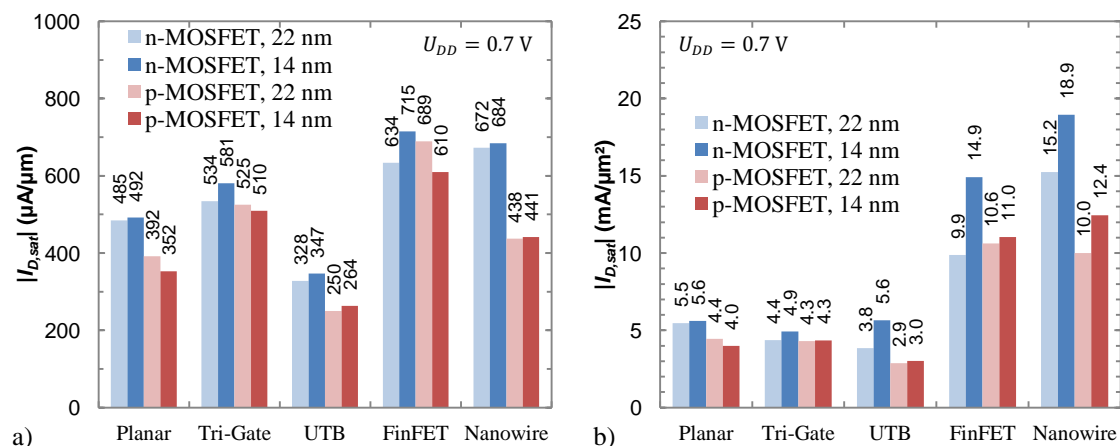


Abbildung 7.1: Berechnete Sättigungsströme von n- und p-MOSFETs der fünf untersuchten Transistorstrukturen in 22 nm- und 14 nm-Technologie. a) normiert auf die effektive Gateweite bei $I_{D,off} = 1 \text{ nA}/\mu\text{m}$ und b) normiert auf die Aktivfläche bei $I_{D,off} = 10 \text{ nA}/\mu\text{m}^2$ (Nanowire mit 4 Drähten).

7.1.2 Kurzkanalverhalten und Steuerfähigkeit

Zur Bewertung des Kurzkanalverhaltens ist für alle Strukturen der $DIBL$ der nominellen Gatelänge herangezogen worden (Abbildung 7.2). Dabei ist offensichtlich, dass die Werte der n- und p-MOSFETs in 14 nm-Technologie in den meisten Fällen durch die verkürzte Gatelänge deutlich zugenommen haben. Lediglich die UTB- und Nanowire-Strukturen zeigen eine Reduzierung des $DIBLs$, da ihrer Kanaldicken im Vergleich zu ihren Gatelängen mehr verringert wurden.

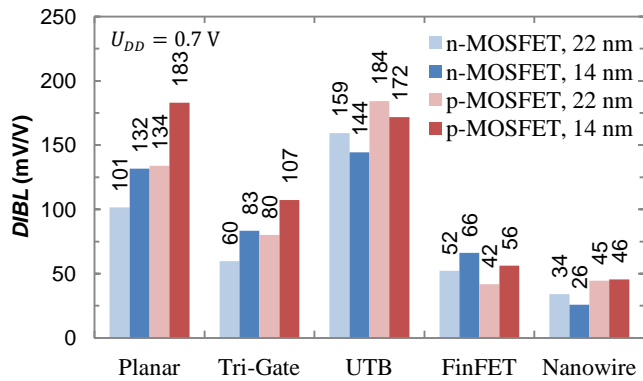


Abbildung 7.2: Berechneter $DIBL$ von n- und p-MOSFETs der fünf untersuchten Transistorstrukturen in 22 nm- ($L_{Gate} = 26$ nm) und 14 nm-Technologie ($L_{Gate} = 20$ nm).

Ebenso verhält sich die Steuerfähigkeit der Gatekontakte, welche anhand der Unterschwellsteigungen im Sättigungsbereich analysiert wurde (Abbildung 7.3). Auch hier zeigen nur die UTB- und Nanowire-Transistoren eine Verbesserung, während bei den anderen Strukturen aufgrund der kürzeren Gatelänge eine Erhöhung der Werte zu beobachten ist.

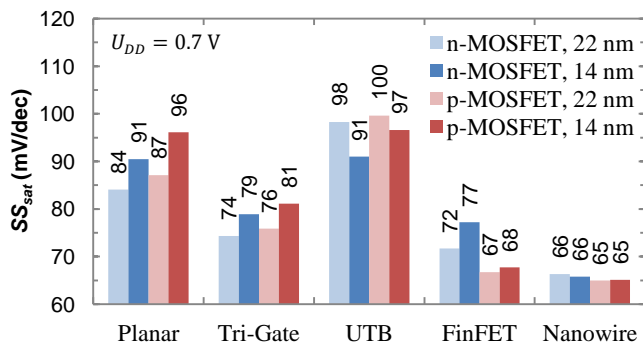


Abbildung 7.3: Berechnete Unterschwellsteigung SS_{sat} im Sättigungsbereich von n- und p-MOSFETs der fünf untersuchten Transistorstrukturen in 22 nm- ($L_{Gate} = 26$ nm) und 14 nm-Technologie ($L_{Gate} = 20$ nm).

Damit wird deutlich, dass die FinFETs unter den horizontalen Strukturen die besten Möglichkeiten für eine weitere Skalierung bieten, da sich ihre auf die genutzte Aktivfläche bezogene Leistungsfähigkeit durch eine Reduzierung des Finnenabstands bzw. Erhöhung der Finne gut steigern lässt und sie bezüglich des Kurzkanalverhaltens sowie der Steuerfähigkeit sehr gute Werte besitzen. Auch wenn die UTB-Transistoren mit dem Übergang zur 14 nm-Technologie in ihren stationären Eigenschaften durchgehend verbessert werden konnten, weisen sie im Vergleich zu den anderen Strukturen immer noch die schlechteren Werte auf. Die Nanowire-Transistoren mit ihrem vertikalen Aufbau sind in diesem Vergleich die überlegene Technologie. Die Reduzierung des Kanaldurchmessers aufgrund der skalierten Aktivfläche verbessert das Kurzkanalverhalten und die Steuerfähigkeit des Gates zunehmend. Die Leistungsfähigkeit konnte trotz des verringerten Kanalquerschnitts weiter gesteigert werden, da sich der Drainstrom erst ab einem Durchmesser von ca. 7 nm, bei dem eine vollständige Volumeninversion vorliegt, überproportional zur Aktivfläche reduziert.

7.2 Statistische Schwankungen

In den vorangegangenen Abschnitten dieser Arbeit standen die durch die Verkleinerung der Transistorstrukturen systematisch hervorgerufenen Kurzkanaleffekte im Vordergrund. Mithilfe von steilen, epitaktisch erzeugten Dotierungsprofilen und Multi-Gate-Strukturen konnte gezeigt werden, dass die Skalierung nach dem Gesetz von MOORE bis über die physikalischen Grenzen der planaren MOSFETs hinaus weiter vorangetrieben werden kann. Ein wesentlicher Aspekt der Strukturen im Bereich unterhalb von 100 nm wurde in den Simulationen jedoch bislang nicht berücksichtigt. Hierbei handelt es sich um die statistischen Schwankungen, denen die Transistoren eines Chips während des Herstellungsprozesses unterliegen und welche sich signifikant auf ihre elektrischen Eigenschaften auswirken können [186].

Bei den Störquellen der Transistorvariationen unterscheidet man zwischen der zufälligen Dotierungsfluktuation (Random Doping Fluctuation, RDF), den fixen Oxidladungen, der Metall-Gate-Granularität (MGG) und den geometrischen Fluktuationen. Letztere müssen je nach Transistorstruktur nochmals zwischen Gate-Kanten-Rauigkeit und Oxid- bzw. Silizium-Dicken-Variation differenziert werden. Dabei arbeitet der Simulator zur Ermittlung der Standardabweichung der elektrischen Transistorparameter mithilfe der Impedanz-Feld-Methode [117]. Auf die Simulation der Gate-Kanten-Rauigkeit wurde jedoch verzichtet, da die Software keine statistische Simulation für dieses Problem bietet.

7.2.1 Impedanz-Feld-Methode

Mithilfe der Impedanz-Feld-Methode von SHOCKLEY [187] können die Schwankungen der elektrischen Parameter für Transistorstrukturen simuliert werden, ohne dass eine Vielzahl an aufwendigen statistischen Prozesssimulationen (Monte-Carlo-Simulationen) erfolgen muss. Lediglich die Prozesssimulation eines Referenztransistors, der auch ein analytisch erzeugtes Dotierungsprofil besitzen kann, ist hierfür erforderlich.

Der Grundgedanke hinter der Impedanz-Feld-Methode besteht darin, dass jedem Arbeitspunkt aus der quasistationären Lösung eine kleine Abweichung, verursacht durch eine Störfunktion, überlagert wird (Kleinsignalverhalten). Der Zusammenhang zwischen der Änderung von Kontaktstrom bzw. Kontaktspannung und der lokalen Fluktuation kann dabei als linear angenommen werden. In Abhängigkeit vom Transportmodell werden am Referenztransistor zunächst für den jeweiligen Arbeitspunkt die GREENSchen Funktionen $G_{k,s}^I(\mathbf{r})$ bzw. $G_{k,s}^U(\mathbf{r})$ (Impedanzfelder) vom Simulator berechnet. Dies geschieht für jeden definierten Kontaktknoten k und die drei Variablen Elektronendichte n , Löcherdichte p und elektrisches Potenzial φ separat. Jede GREENSche Funktion bildet somit die Abhängigkeit des Kontaktstromes bzw. dessen Spannung

von der jeweiligen Variable bezüglich aller Gitterpunkte im Bauelement ab. Mit ihrer Hilfe lassen sich anschließend die Änderungen des Rauschstromes

$$\delta I_{\vartheta,k} = \sum_{s=n,p,\varphi} \int d^3r G_{k,s}^I(\mathbf{r}) \delta N_{\vartheta,s}(\mathbf{r}) \quad (7.1)$$

bei konstanter Kontaktspannung bzw. der Rauschspannung

$$\delta U_{\vartheta,k} = \sum_{s=n,p,\varphi} \int d^3r G_{k,s}^U(\mathbf{r}) \delta N_{\vartheta,s}(\mathbf{r}) \quad (7.2)$$

bei konstantem Kontaktstrom als lineare Antwort ermitteln, wobei

$$\delta N_{\vartheta,s}(\mathbf{r}) = N_{\vartheta,s}(\mathbf{r}) - N_{ref,s}(\mathbf{r}) \quad (7.3)$$

die Änderung der jeweiligen Störquelle ϑ in Abhängigkeit von einer statistischen Verteilung $N_{\vartheta,s}(\mathbf{r})$ (Störfunktion) ist [188].

Die Simulationen zur Variation können vom Simulator auf zwei Wegen mit teilweise unterschiedlichen Modellansätzen berechnet werden, der rauschartigen Impedanz-Feld-Methode (noise-like-IFM, n-IFM) und der statistischen Impedanz-Feld-Methode (s-IFM).

Bei der statistischen Impedanz-Feld-Methode werden vom Simulator für alle Gitterpunkte eines Bauteils eine definierte Anzahl an zufällig verteilten Werten unter Berücksichtigung des Referenzwertes, wie z. B. der Dotierungskonzentration, erzeugt und die daraus berechneten Abweichungen $\delta I_{\vartheta,k}$ des Kontaktstromes bei fester Kontaktspannung ermittelt. Somit ergibt sich annähernd eine Gaußverteilung der Häufigkeiten (Abbildung 7.4).

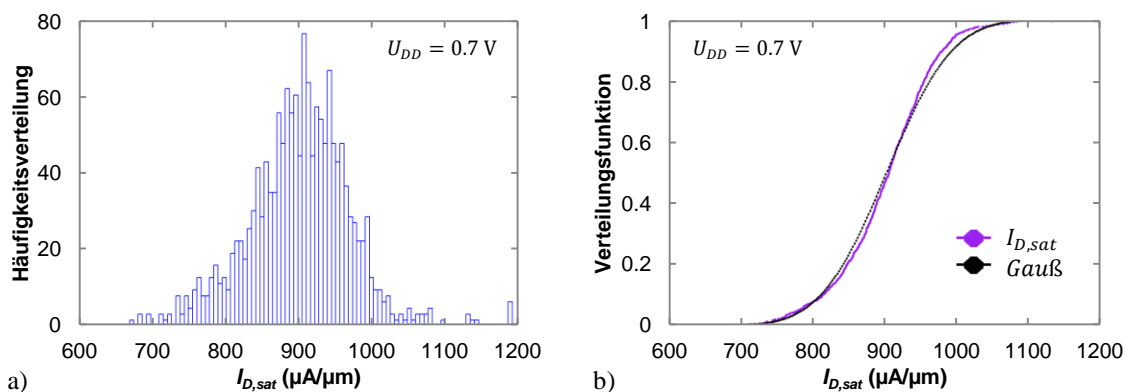


Abbildung 7.4: Statistische Schwankung (s-IFM) des Sättigungsstroms von n-FinFETs einer 22 nm-Technologie. a) Häufigkeit und b) Verteilungsfunktion der 1000 berechneten Zufallsprofile im Vergleich zur Gaußverteilung.

Aus ihr lässt sich anschließend die Standardabweichung

$$\sigma_{I_D} = \sqrt{\frac{1}{n} \sum_{\vartheta}^n (\delta I_{\vartheta,D})^2} \quad (7.4)$$

des Drainstromes für den entsprechenden Arbeitspunkt der Simulation in Abhängigkeit von den Störquelle berechnen.

Die rauschartige Impedanz-Feld-Methode (n-IFM) arbeitet anstelle individueller Abweichungen eines jeden Gitterpunktes direkt mit den statistischen Größen (Momenten). Da jedoch die Störquellen so beschaffen sind, dass sowohl positive als auch negative Abweichungen gleichermaßen auftreten können, bestimmt sich ihr statistischer Mittelwert (Erwartungswert, statistisches Moment erster Ordnung), symbolisiert durch die spitzen Klammern, hierbei über die gesamte Struktur zu $\langle \delta N_{\vartheta,s}(\mathbf{r}) \rangle = 0$. Somit ergibt sich aus den Gleichungen (7.1) bzw. (7.2) auch für die mittlere Änderung des Rauschstroms bzw. der Rauschspannung $\langle \delta I_{\vartheta,s} \rangle = \langle \delta U_{\vartheta,s} \rangle = 0$. Die Information über die stochastischen Eigenschaften einer Störfunktion lassen sich demnach nur aus ihrer Varianz (statistisches Moment zweiter Ordnung) ermitteln. Diese berechnet sich aus dem quadratischen Mittelwert aller Abweichungen zu

$$V(N_{\vartheta,s}) = \langle (\delta N_{\vartheta,s}(\mathbf{r}))^2 \rangle, \quad (7.5)$$

wodurch das Vorzeichen entfällt und sich die absoluten Abweichungen nicht mehr gegenseitig aufheben. Im Simulator wurde an Stelle der Varianz von Strom bzw. Spannung eines Kontakts jedoch die Kovarianz zwischen zwei Knoten k und l der entsprechenden Größe integriert, um gleichzeitig einen direkten Bezug zwischen den Potenzialen bzw. den Strömen von Drain, Source und Gate zu schaffen. Die Kovarianz wird hier als Spektraldichte des Rauschstromes sowie der Rauschspannung mit den Einheiten A²s bzw. V²s betrachtet [117] und berechnet sich zu

$$S_{I,kl} = \langle \delta I_k \delta I_l \rangle = \sum_{s=n,p,\varphi} \iint d^3r_1 d^3r_2 G_{k,s}^l(\mathbf{r}_1) G_{l,s}^l(\mathbf{r}_2) K_{\vartheta,s}(\mathbf{r}_1, \mathbf{r}_2), \quad (7.6)$$

wobei

$$K_{\vartheta,s}(\mathbf{r}_1, \mathbf{r}_2) = \langle \delta N_{\vartheta,s}(\mathbf{r}_1) \delta N_{\vartheta,s}(\mathbf{r}_2) \rangle \quad (7.7)$$

die explizite Störfunktion darstellt und der Kovarianz zwischen $\delta N_{\vartheta,s}(\mathbf{r}_1)$ und $\delta N_{\vartheta,s}(\mathbf{r}_2)$ entspricht. Kovarianzen nehmen immer dann einen positiven Wert an, wenn die beiden statistischen Verteilungsfunktionen einen gleichsinnigen linearen Zusammenhang besitzen und somit korrelieren [189], [190]. Die Kovarianz zwischen zwei zeitlich oder räumlich verschiedenen Messwerten eines einzelnen Merkmals wird auch als statistische Autokorrelation bezeichnet [189].

Für die n-IFM-Simulationen berechnen sich die zugehörigen Standardabweichungen von Drainstrom und Gatespannung über

$$\sigma I_D = \sqrt{S_{I,DD} \cdot f_{n-IFM}} \quad (7.8)$$

bzw.

$$\sigma U_{GS} = \frac{\sqrt{S_{I,DD} \cdot f_{n-IFM}}}{y_{D,G}}, \quad (7.9)$$

wobei die Spektraldichte mit der Kleinsignalfrequenz von $f_{n-IFM} = 1$ Hz multipliziert wird [191]. Durch eine additive bzw. subtraktive Überlagerung der Werte mit der zugehörigen Transferkennlinie erhält man eine einhüllende Kurve zum Drainstrom der Referenzstruktur (Abbildung 7.5). Die Standardabweichungen der Schwellspannung

$$\sigma U_{th} = \sigma U_{GS}|_{U_{GS}=U_{th}} \quad (7.10)$$

und anderer signifikanter elektrischer Parameter können anschließend im jeweiligen Arbeitspunkt abgelesen zu werden.

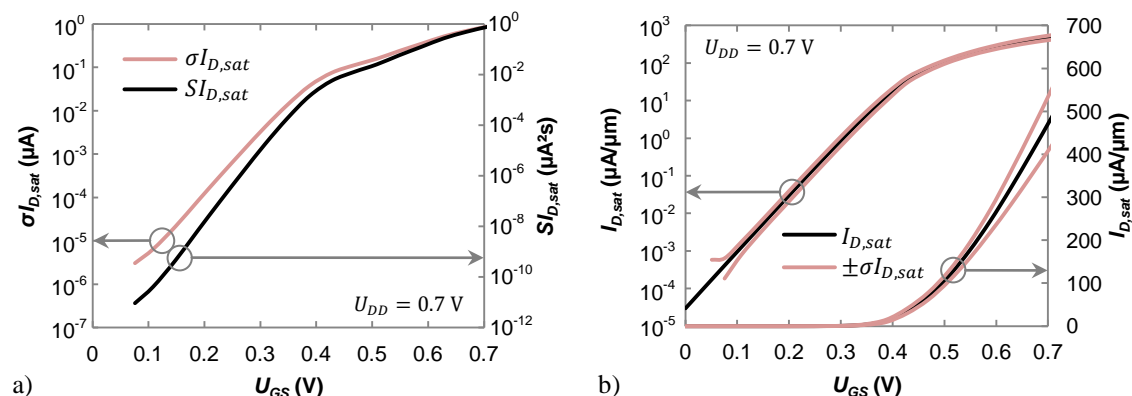


Abbildung 7.5: Simulationen (n-IFM) der statistischen Schwankung des Drainstroms durch zufällige Dotierungsfluktuation eines Nanowire-Transistors (n-MOSFET). a) Varianz und Standardabweichung in logarithmischer Darstellung und b) Transferkennlinie (Sättigung) in linearer und logarithmischer Darstellung mit der einhüllenden Standardabweichung $\pm \sigma I_{D,sat}$.

7.2.2 Zufällige Dotierungsfluktuation

Die zufällige Dotierungsfluktuation bezeichnet die statistische Verteilung von Dotanden im Siliziumkristall. Während die Implantation und Diffusion der Ionen in der Prozesssimulation in der Regel mithilfe von analytischen Modellen als konstantes Profil nachgebildet werden, sind diese Vorgänge in der Realität stochastischer Natur und können bei kleineren Strukturen mit einer großen Schwankung auftreten. Dabei variiert nicht nur die Anzahl der im Siliziumkristall aktivierten Dotanden, sondern auch ihre Position im Bauelement. Welchen Einfluss ein einzelner Dotand in einem Ultrakurzkanal-MOSFET haben kann, zeigt eine Beispielrechnung zu einem Tri-Gate-Transistor.

Über die Dichte der Siliziumatome von $5 \cdot 10^{22} \text{ cm}^{-3}$ im Kristallgitter und der Finnen-geometrie ($L_{Gate} = 26 \text{ nm}$, $H_{Fin} = 10 \text{ nm}$, $W_{Fin} = 20 \text{ nm}$) ergibt sich eine Gesamtanzahl von $n_{Si} = 260.000$ Siliziumatomen in der Finne. Ein einzelnes zusätzliches Dotieratom bewirkt demnach eine auf das gesamte Kanalgebiet bezogene Konzentrationsänderung um rund $2 \cdot 10^{17} \text{ cm}^{-3}$.

Die n-IFM-Simulation basiert auf der Annahme, dass die Akzeptoren bzw. Donatoren einer Poisson-Verteilung folgen und sich die Dotierungsfluktuation an einem Punkt \mathbf{r} unkorreliert zu der eines beliebigen anderen Punktes im Transistor verhält. Die Störfunktion für die zufällige Dotierungsfluktuation der n-IFM ist daher im Simulator als

$$K_{RDF,A}(\mathbf{r}_1, \mathbf{r}_2) = \langle \delta N_{RDF,A}(\mathbf{r}_1) \delta N_{RDF,A}(\mathbf{r}_2) \rangle = N_{A,ref}(\mathbf{r}_1) \delta(\mathbf{r}_1 - \mathbf{r}_2) \quad (7.11)$$

definiert [192]. Bei einer POISSON-Verteilung entspricht die Varianz von $\delta N_{RDF,A}(\mathbf{r})$ gleichzeitig dem Erwartungswert $N_{A,ref}(\mathbf{r})$. Da in diesem Ausdruck jedoch die Kovarianz von $\delta N_{RDF,A}(\mathbf{r}_1)$ und $\delta N_{RDF,A}(\mathbf{r}_2)$ gegeben ist, muss der Erwartungswert noch mit der Delta-Funktion verknüpft werden, die auch als ein mehrdimensionales Abtasten des Dotierungsprofils interpretiert werden kann. Die Störfunktion (Varianz) entspricht somit für $\mathbf{r}_1 = \mathbf{r}_2$ dem Wert des Dotierungsprofils der Referenzstruktur und wird für $\mathbf{r}_1 \neq \mathbf{r}_2$ zu Null.

Für die s-IFM-Simulation wird zu Beginn der Simulation, ausgehend vom Akzeptor- bzw. Donatorprofil, eine definierte Anzahl an statistisch variierten Profilen erstellt (Abbildung 7.6). Sie folgen ebenfalls einer POISSON-Verteilung mit der Form

$$P_i(k) = \frac{(N_{RDF,i} V_i)^k}{k!} e^{(-N_{RDF,i} V_i)} \quad (7.12)$$

und repräsentieren die Wahrscheinlichkeit dafür, dass für einen Gitterpunkt i mit dem umgebenden Volumen V_i und der mittleren Dotierkonzentration N_i exakt k Dotanden in dem Volumen vorzufinden sind. Das Produkt aus $N_{RDF,i}$ und V_i entspricht der mittleren Anzahl der Dotanden im Volumen.

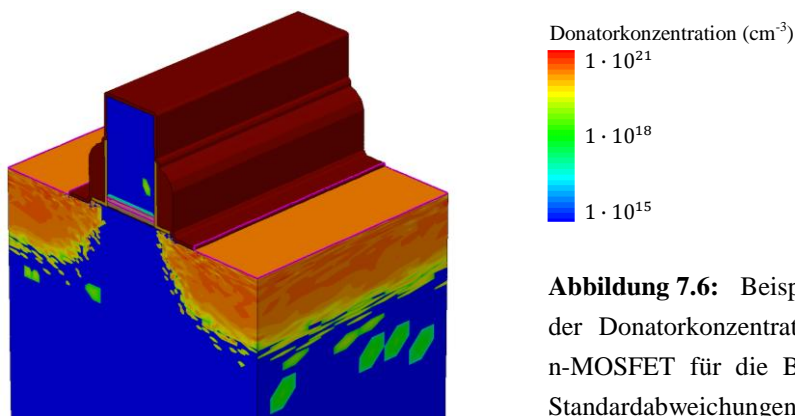


Abbildung 7.6: Beispielhaftes Zufallsprofil der Donatorkonzentration in einem planaren n-MOSFET für die Berechnung (s-IFM) der Standardabweichungen.

Für jedes statistisch variierte Profil wird anschließend mithilfe der Impedanz-Feld-Methode eine Änderung $\delta I_{RDF,k,n}$ im entsprechenden Arbeitspunkt eines Kontakts simuliert, aus denen die gesuchte Standardabweichung berechnet werden kann.

Der Vergleich der Strukturen in Abbildung 7.7 zeigt, dass die MOSFETs mit hoch dotierten Kanalgebieten und steilen pn-Übergängen von Source bzw. Drain zum Kanal, wie sie beim planaren bzw. beim Tri-Gate-Transistor auftreten, auch eine deutlich erhöhte Anfälligkeit ihrer Parameter bezüglich der zufälligen Dotierungsfluktuation aufweisen. Die drei Varianten mit vollständig verarmtem Kanal sind erwartungsgemäß robuster und somit für die Realisierung von Ultrakurzkanal-Transistoren besser geeignet.

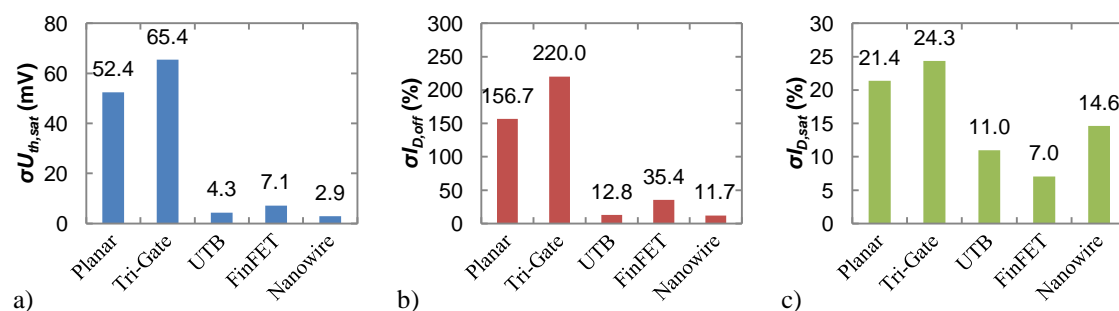


Abbildung 7.7: Standardabweichungen (s-IFM) der Transistorparameter aufgrund der zufälligen Dotierungsfluktuation der berechneten Transistorstrukturen (n-MOSFETs, 22 nm-Technologie) für a) die Sättigungsschwelspannung, b) den Sperrstrom und c) den Sättigungsstrom. Simulationen mittels n-IFM liefern ähnliche Ergebnisse. Ebenso weisen p-MOSFETs tendenziell das gleiche Verhalten auf.

7.2.3 Fixe Ladungen im Oxid

Die Variationssimulationen zu den fixen Oxidladungen umfassen zum einen fest eingebundene Ladungen im High-k-Oxid, die bei der Herstellung des Gatedielektrikums verursacht werden, und zum anderen Energiezustände für Elektronen und Löcher im Oxid und der Grenzfläche zum Silizium, die auch als Traps (Ladungsfallen) bezeichnet werden. Das Modell berücksichtigt dabei die Auswirkungen auf die Raumladung der POISSON-Gleichung und die Trap-bezogenen Generations- und Rekombinationsprozesse gemäß den Kontinuitätsgleichungen. Die Störfunktion der n-IFM ähnelt mit

$$K_{Trap,i}(\mathbf{r}_1, \mathbf{r}_2) = N_{Trap,i}(\mathbf{r}_1)\delta(\mathbf{r}_1 - \mathbf{r}_2) \quad (7.13)$$

der zufälligen Dotierungsfluktuation, wobei $N_{Trap,i}(\mathbf{r}_1)$ die Konzentration im i-ten Energiezustand ist. Unter der Annahme, dass sich auch die Traps komplett unkorreliert verhalten, wird für die s-IFM eine zufällige Verteilung gemäß der Dotierungsfluktuation aller Energiezustände verwendet. Für die n- und s-IFM-Simulationen ist als durchschnittliche Konzentration ein Wert von $N_{Trap} = 1 \cdot 10^{11} \text{ cm}^{-3}$ verwendet worden.

Die Schwankungen der Schwellspannungen und der Drainströme fallen im Vergleich zur zufälligen Dotierungsfluktuation deutlich niedriger aus (Abbildung 7.8). Bis auf den UTB-MOSFET, der die geringsten Standardabweichungen aufweist, ist keine nennenswerte Differenz unter den Strukturen, weder den planaren und Multi-Gate-Transistoren noch den dotierten und undotierten Kanalgebieten, zu erkennen. Es bleibt jedoch ein signifikanter Unterschied zwischen dem Sättigungs- und dem Sperrstrom, da jeder Transistor im Unterschwellbereich aufgrund der geringen Stromdichten wesentlich sensibler auf die fixen Oxidladungen reagiert.

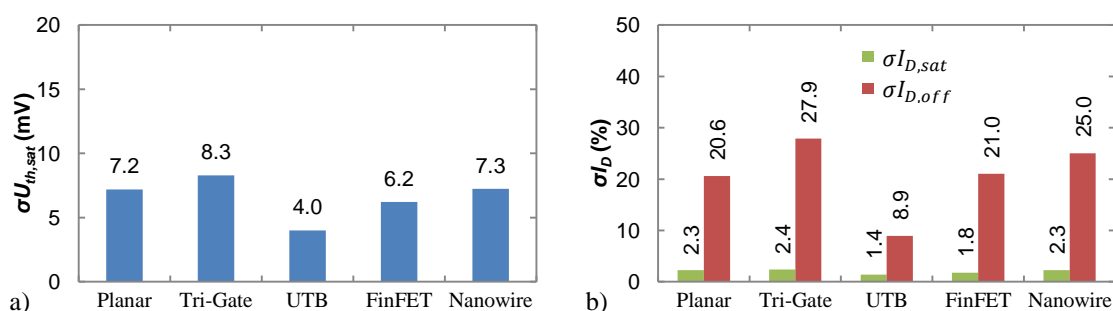


Abbildung 7.8: Standardabweichungen (s-IFM) der Transistorparameter aufgrund der fixen Ladungen im Oxid der berechneten Strukturen (n-MOSFETs, 22 nm-Technologie) für a) die Sättigungsschwelspannung und b) die Drainströme.

7.2.4 Metall-Gate-Granularität

Die Korngrenzen des Gatemetalls unterliegt in der Realität ebenfalls einer zufälligen Variation, da sie je nach ihrer Orientierung zur Grenzfläche unterschiedliche Austrittsarbeiten besitzen. Aus [193] geht hervor, dass das verwendete polykristalline Titannitrit Metallkörner mit einer (200)- bzw. (111)-Orientierung und eine maximale Abweichung der Austrittsarbeit von bis zu 200 meV besitzen kann. Die Größe der Körner hängt sehr stark vom Herstellungsprozess ab und beträgt zwischen 5 ... 20 nm. Für die n-IFM-Simulation ergibt sich daher eine Störfunktion der Form

$$K_{MGG}(\mathbf{r}_1, \mathbf{r}_2) = a(\mathbf{r}_1)a(\mathbf{r}_2)e^{-\frac{(\mathbf{r}_1 - \mathbf{r}_2)^2}{\lambda^2}}. \quad (7.14)$$

Hierbei ist a die Standardabweichung der Austrittsarbeit, die in einem Bereich von 50 ... 110 meV liegen kann, und λ die Korrelationslänge, die auch als durchschnittliche Korngröße interpretiert wird.

Auf der Basis der s-IFM-Simulation arbeitet der Simulator wieder mit einer definierten Anzahl an zufällig erzeugten Parametersätzen für die Austrittsarbeit des Metalls. Hierbei wird neben beliebig vielen Wertepaaren, bestehend aus einer Austrittsarbeit und der zugehörigen Wahrscheinlichkeit, ebenfalls eine durchschnittliche Korngröße vorgegeben. Für den oben beschriebenen Fall aus [193] kann ein Verhältnis zwischen der

(200)- und der (111)-Orientierung von 60 % : 40 % angenommen werden, womit sich, ausgehend von der nominellen Austrittsarbeit, über die Beziehungen

$$W_{A,nom} = 0.6 \cdot W_{A,(200)} + 0.4 \cdot W_{A,(111)} \quad (7.15)$$

und

$$W_{A,(200)} = W_{A,(111)} + 200 \text{ meV} \quad (7.16)$$

die jeweiligen Austrittsarbeiten bestimmen lassen.

Wie bei den berechneten Fluktuationen der fixen Oxidladungen kann auch für die Schwankungen aufgrund der Metall-Gate-Granularität keine direkte Abhängigkeit von der Transistorstruktur beobachtet werden (Abbildung 7.9). Es zeigte sich jedoch durch die erhöhten Standardabweichungen der Schwellspannungen und Sperrströme gegenüber den Sättigungsströmen, dass der Unterschwellbereich wesentlich stärker durch die Metall-Gate-Granularität beeinflusst wird.

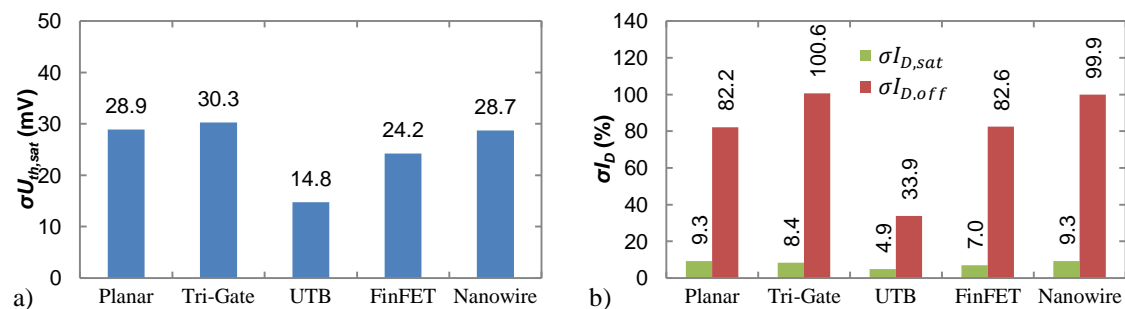


Abbildung 7.9: Standardabweichungen (s-IFM) der Transistorparameter aufgrund der Metall-Gate-Granularität der berechneten Strukturen (n-MOSFETs, 22 nm-Technologie) für a) die Sättigungsschwellschwelle und b) die Drainströme.

7.2.5 Geometrische Variationen

Die geometrischen Schwankungen von Grenzflächen und Schichten, wie dem Gateoxid, gewinnen mit zunehmender Skalierung der Strukturen an Einfluss, besonders da die Schichtdicke des Gateoxids bereits so dünn produziert wird, dass sie nur noch aus wenigen Atomlagen besteht und damit einer vergleichsweise großen Variation unterliegt. Für Multi-Gate-Strukturen, wie dem Tri-Gate-Transistor und dem FinFET, ist dies ebenfalls von entscheidender Bedeutung, da sie durch die Ätzprozesse der Finnen eine signifikante Schwankung von Finhöhe und Finweite aufweisen können.

Die geometrischen Variationsanalysen lassen sich mit der aktuellen Version des Simulators jedoch nur auf Basis der n-IFM lösen. Es muss zudem beachtet werden, dass die geometrischen Variationen wie auch die Metall-Gate-Granularität nicht mehr einer POISSON-Verteilung unterliegen, wo Erwartungswert und Varianz gleich groß sind. Die Störfunktion ergibt sich deshalb, ähnlich wie die der Metall-Gate-Granularität, zu

$$K_{Geo}(\mathbf{r}_1, \mathbf{r}_2) = \mathbf{n}(\mathbf{r}_1) \cdot \mathbf{n}(\mathbf{r}_2) \cdot [a_{iso}(\mathbf{r}_1) + |\mathbf{a}(\mathbf{r}_1) \cdot \mathbf{n}(\mathbf{r}_1)|] \cdot [a_{iso}(\mathbf{r}_2) + |\mathbf{a}(\mathbf{r}_2) \cdot \mathbf{n}(\mathbf{r}_2)|] \cdot e^{\left(-\frac{(\mathbf{r}_1 - \mathbf{r}_2)^2}{\lambda^2}\right)} . \quad (7.17)$$

Hierbei sind $\mathbf{n}(\mathbf{r})$ der Normalenvektor zur Grenzfläche am Punkt \mathbf{r} , λ die Korrelationslänge und a_{iso} sowie \mathbf{a} die isotrope bzw. vektorielle Amplitude der Grenzflächenschwankung. Damit werden nur Verschiebungen senkrecht zur Grenzfläche berücksichtigt, deren Wirkung mit der Entfernung abnimmt. Eine Anpassung des Dotierungsprofils wird jedoch bei der n-IFM-Simulation vernachlässigt, weshalb auf die Simulation der Gate-Kanten-Rauigkeit verzichtet wird [194].

Für die Variation der Gateoxiddicke wurde eine Korrelationslänge von $\lambda_{tox} = 4$ nm und eine isotrope Amplitude von $a_{tox} = 0.2$ nm angenommen [117]. Bei den Dicken Schwankungen des Siliziumkanals wurde die Korrelationslänge hingegen auf den Maximalwert von $\lambda_{tsi} = 10$ μm gesetzt, da die Amplitude der jeweiligen Schwankung über die gesamte Struktur hinweg gleich angenommen wird. Aufgrund der Toleranzen beim Schichtauf- bzw. -abtrag liegt die Amplitude des Top-Gates bei $a_{tsi,Top} = 0.5$ nm. Für die Seiten-Gates müssen zudem die Toleranzen der Maskierung mitberücksichtigt werden, wofür deren Amplitude auf $a_{tsi,Seite} = 1.5$ nm festgelegt wurde.

Die berechneten Werte der Standardabweichungen von Schwellspannungen und Drainströmen der verschiedenen Transistorstrukturen zeigen im Hinblick auf die angenommenen Parameter hohe Differenzen (Abbildung 7.10). Die sehr geringen Schwankungen aufgrund der Oxiddickenfluktuationen bei den vollständig verarmten Transistoren sind dabei gut nachvollziehbar, da bereits in Abbildung 6.13 gezeigt werden konnte, dass der Nanowire-MOSFET sich sehr stabil gegenüber einer Änderung der Gateoxiddicke verhält. Auch die geringen Werte des planaren Transistors sind vergleichbar mit denen aus [195].

Die sehr hohen Werte der Kanaldickenschwankungen müssen jedoch kritisch betrachtet werden, da die n-IFM-Simulation auch hier keine Anpassung des Dotierungsprofils berücksichtigt und nur eine gedachte mathematische Verschiebung der Grenzflächen darstellt. Dies trifft vor allem die Multi-Gate-Strukturen, die sehr anfällig gegenüber der Kanalweite und der Kanalhöhe erscheinen. Auch bei einem Vergleich der berechneten Werte für den Tri-Gate-Transistor mit denen aus Abbildung 5.21 wird deutlich, dass die Höhe der Standardabweichungen teilweise zu groß abgeschätzt wird. Dies ändert allerdings nicht die tendenzielle Aussage, dass die Multi-Gate-Strukturen (Tri-Gate, FinFET, Nanowire) im Vergleich zu den planaren Transistoren anfälliger gegenüber Schwankungen der Kanaldicke sind.

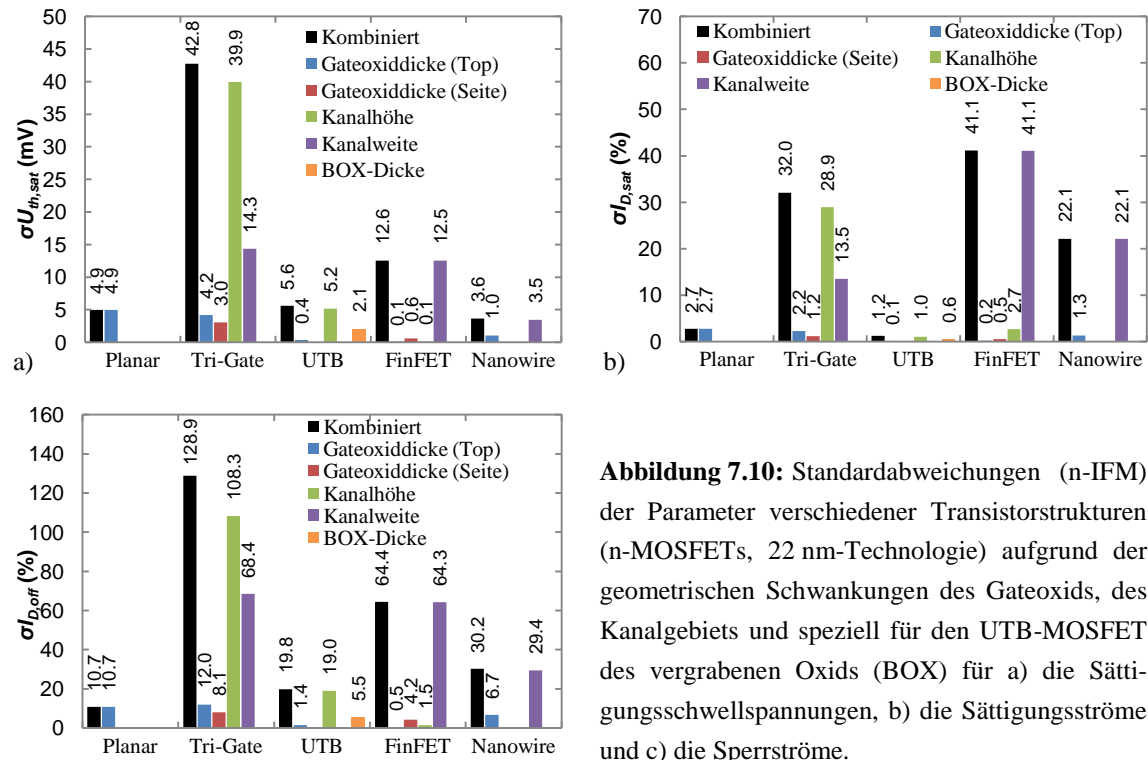


Abbildung 7.10: Standardabweichungen (n-IFM) der Parameter verschiedener Transistorstrukturen (n-MOSFETs, 22 nm-Technologie) aufgrund der geometrischen Schwankungen des Gateoxids, des Kanalgebiets und speziell für den UTB-MOSFET des vergrabenen Oxids (BOX) für a) die Sättigungsschwellspannungen, b) die Sättigungsströme und c) die Sperrströme.

7.2.6 Kombination der Störquellen

Da die verschiedenen Transistorstrukturen unterschiedlich stark auf die einzelnen Störquellen reagieren, müssen sie für einen endgültigen Vergleich in Kombination betrachtet werden. Die kombinierten Standardabweichungen der Drainströme berechnen sich über

$$\sigma I_{komb} = \sqrt{\sum_i \sigma I_i^2}. \quad (7.18)$$

Die Gegenüberstellung der Werte erfolgte für die Daten der s-IFM-Simulation, da sie für die zufällige Dotierungsfluktuation, die fixen Oxidladungen und die Metall-Gate-Granularität mit denen der n-IFM-Simulation gut vergleichbar sind. Die Ergebnisse der geometrischen Schwankungen auf Basis der n-IFM wurden bei den kombinierten Werten nicht mit einbezogen, da die Oxiddickenvariation einen vernachlässigbar kleinen Anteil einbringt und die Variation der Kanaldicken nur schlechte Schätzwerte erbracht hatte.

Bei Betrachtung der kombinierten Standardabweichungen fällt auf, dass die planaren und die Tri-Gate-Transistoren mit den hoch dotierten Kanälen für alle Parameter die höchsten Werte aufweisen (Abbildung 7.11). Diese sind maßgeblich durch die zufällige Dotierungsfluktuation bestimmt, was bei den vollständig verarmten Transistoren nur für den Sättigungsstrom zutrifft. Deren Streuungen der Schwellspannungen und Sperrströme sind hauptsächlich durch die Metall-Gate-Granularität verursacht. Die Schwan-

kungen aufgrund der fixen Oxidladungen sind im Vergleich zu den anderen beiden Störquellen gering und besitzen kaum Einfluss auf die kombinierten Werte.

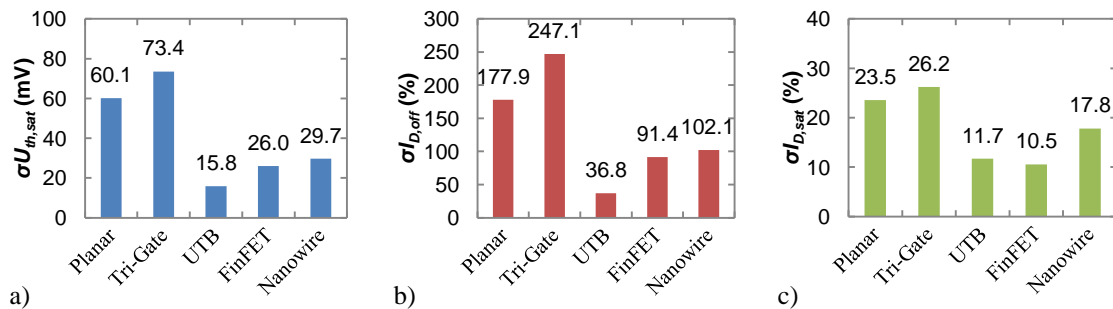


Abbildung 7.11: Standardabweichungen (s-IFM) der Transistorparameter aufgrund der Kombination der zufälligen Dotierungsfluktuation, den fixen Oxidladungen und der Metall-Gate-Granularität der verschiedenen Transistorstrukturen (n-MOSFETs, 22 nm-Technologie) für a) die Sättigungsschwellspannung, b) den Sperrstrom und c) den Sättigungsstrom.

Zusammenfassend lässt sich sagen, dass die statistischen Schwankungen mit kleiner werdendem Kanalvolumen pro Gatefläche zunehmen. Deutlich wird dies beim Vergleich zwischen planarem und Tri-Gate-Transistor bzw. zwischen FinFET und Nanowire-Transistor. Ein notwendiger Schritt für die Zuverlässigkeit einer Technologie bzw. eine hohe Ausbeute der Produktion ist der Übergang zu vollständig verarmten Strukturen, die nicht mehr von der Kanaldotierung abhängen. Der UTB-MOSFET stellt dabei mit seiner verhältnismäßig einfachen Struktur und der nahen Verwandtschaft zum klassischen SOI-MOSFET eine prozessstabile Alternative zum planaren 22 nm-Transistor dar. Seine Möglichkeiten zur weiteren Skalierung sind jedoch eingeschränkt. Der FinFET und der Nanowire-Transistor bieten hingegen auch weiterhin das Potenzial zur fortlaufenden Verkleinerung ihrer Dimensionen, besitzen jedoch als Multi-Gate-Strukturen eine erhöhte Anfälligkeit gegenüber geometrischen Schwankungen.

8 Zusammenfassung und Ausblick

Bedingt durch die physikalischen Grenzen wird eine Skalierung der planaren Transistoren über die 22 nm-Technologie hinaus immer schwieriger. Bereits bei früheren Technologieknoten ist man gezwungen gewesen, neue Verfahren zur Implantation der Dotanden und zur Ausheilung des Kristallgitters in den Herstellungsprozess zu integrieren, um die Funktionalität der MOSFETs aufrecht zu erhalten. Die Konzepte der mechanischen Verspannung des Kanalgebiets boten zudem die Möglichkeit einer deutlichen Verbesserung der Leistungsfähigkeit, ohne direkt das Kurzkanalverhalten der Transistoren zu beeinflussen. Ebenso wurden neue Materialien als High-k-Gateoxid bzw. als Gatemetall in den Herstellungsprozess aufgenommen, um die äquivalente Gateoxiddicke weiter zu reduzieren bzw. den Verarmungsproblemen des Polysiliziums vorzubeugen. Allerdings stiegen durch diese Maßnahmen die Entwicklungs- und Herstellungskosten deutlich an. Um die Verkleinerung der MOSFETs weiter voranzutreiben, ist nun eine Änderung des Aufbaus hin zu Multi-Gate-Transistoren erforderlich, die auch den Anforderungen der ITRS und einer Volumenproduktion gerecht werden muss.

A. Grundlagen und Entwicklung der CMOS-Technologie

Neben den allgemeinen Eigenschaften der MOSFETs konnte bereits in den theoretischen Grundlagen gezeigt werden, dass sich die Steuerfähigkeit des Gatekontakts erhöht, wenn sich seine Fläche bei gleichbleibendem Kanalvolumen vergrößert. Dies stärkt den Einfluss der orthogonalen Feldstärken des Gates gegenüber dem longitudinalen elektrischen Feld zwischen Source und Drain, wodurch letztendlich die Unterschwellsteigung reduziert und das Kurzkanalverhalten des Transistors verbessert werden. Dabei unterliegen die Multi-Gate-Strukturen aufgrund ihrer geringen Abmessungen sowohl Überlagerungs- also auch Quantisierungseffekten. Diese führen bei dotierten Kanalgebieten mit einer Weite von mehr als 7 nm zu einer Oberflächeninversion in den Ecken der Finne (Eckeneffekt). Undotierte Finnen ermöglichen aufgrund der großen Eindringtiefe und somit einer wirksamen Überlagerung der Gatefeldstärken eine vollständige Verarmung des Kanalgebiets. Für eine Kanalweite von weniger als 7 nm stellt sich infolge der Quantisierung im Bereich der Grenzfläche zwischen Gateoxid und Silizium eine Volumeninversion im Kanal ein, wodurch auch die Streueffekte der Ladungsträger an den Grenzflächen gemindert werden. Weiterhin kommt es zur Aufspaltung der Subbänder und zu einer Vergrößerung der Bandlücke, wodurch sich die Schwellspannung des Transistors verändert.

B. Kalibrierung der Modellparameter

Auf dem Weg der Entwicklung neuer Technologien sind Simulationen ein wichtiges Hilfsmittel, um das Transistorverhalten frühzeitig abschätzen zu können und somit aufwendige Experimente zu vermeiden. Zudem bieten sie die Möglichkeit, Parameter zu betrachten, die mit der derzeitigen Technik nicht messbar sind. Die simulierten Werte dieser Arbeit wurden mit dem Programm *Sentaurus Workbench Advanced* von SYNOPSIS berechnet, welches eine vollständige Prozess- und Bauelementesimulation ermöglicht. Da sich die betrachteten Multi-Gate-Transistoren dicht an der Grenze zur Volumeninversion bewegen bzw. einen vollständig verarmten Kanal besitzen und in ihrer Struktur untereinander sehr verschieden sind, wurden vorrangig nur die Parameter der Bauelementesimulation kalibriert. Als Referenz diente eine 32 nm-SOI-Technologie des Halbleiterherstellers GLOBALFOUNDRIES. Für die Modellparameter der Prozesssimulationen wurden stets deren Ausgangswerte übernommen. Damit konnte eine gute Übereinstimmung zwischen den experimentellen Daten und den Simulationsergebnissen erzielt und eine entsprechende Basis für die untersuchten Multi-Gate-Strukturen geschaffen werden.

C. Planare Transistoren auf Basis einer 22 nm-Technologie

Zunächst wurden planare MOSFETs einer möglichen 22 nm-Technologie simuliert und die verbliebenen Varianten der mechanischen Verspannung untersucht. Dabei zeigte sich, dass aufgrund des hohen Skalierungsgrades nur noch geringfügig Platz für die verspannten Deckschichten zwischen dem Gate- und den Source-Drain-Kontakten verbleibt und somit keine leistungsverändernde Wirkung mehr von ihnen ausgeht. Das eingebettete Silizium-Germanium in den Source-Drain-Gebieten der p-MOSFETs konnte als das wichtigste und effektivste Verspannungselement identifiziert werden. Mit der verkürzten Gatelänge und der zusätzlichen Gitterdeformation infolge des Replacement-Gate-Prozesses erhöhte sich die berechnete Leistungsfähigkeit der p-MOSFETs trotz einer schlechteren Grundbeweglichkeit der Löcher bis knapp unter das Niveau der unverspannten n-MOSFETs.

D. Tri-Gate-Transistoren

Mit den planaren Transistoren und der Einbindung einer zusätzlichen Maske zur Strukturierung von Finnen konnten funktionsfähige Tri-Gate-Transistoren mit einer verbesserten Steuerfähigkeit des Gatekontakts simuliert werden. Dies ermöglichte die Umsetzung eines hybriden Prozess, der mit geringem Aufwand sowohl planare als auch Tri-Gate-Transistoren auf einem Substrat vereint. Dabei teilen sich die beiden Strukturvarianten bis auf wenige Schritte den selben Herstellungsprozess und besitzen nachher ein ähnliches Dotierungsprofil und die gleiche Austrittsarbeit des Gatemetalls. Somit können Kosten gespart und ein harter Umbruch in der Volumenproduktion vermieden werden. Der Tri-Gate-Transistor in dieser Form stellt jedoch nur eine Zwischenstufe auf

dem Weg zum klassischen FinFET mit schmaler hoher Finne dar. Mit dem oberflächennahen Dotierungsprofil des planaren Transistors liegt die ermittelte optimale Finhöhe bei geringen 7 nm, was die effektive Gateweite stark limitiert. Damit wurde für den simulierten n-Tri-Gate-Transistor nur eine leichte Verbesserung seiner Leistungsfähigkeit (bezogen auf die effektive Gateweite) von 9 % im Vergleich zu seinem planaren Gegenstück erreicht. Der p-Tri-Gate-Transistor zeigte bei gleichem Sperrstrom einen um 35 % höheren Sättigungsstrom, da mit der Formierung der Finne, ähnlich zum Replacement-Gate-Prozess, eine weitere Steigerung der Gitterdeformation und somit eine höhere mechanische Verspannung im Kanal erreicht wird.

Das dotierte Kanalgebiet sorgt für einen starken Eckeneffekt, der hinsichtlich seiner Abhängigkeiten gegenüber der Finnengeometrie untersucht und für den ein analytisches Modell zur frühzeitigen Abschätzung der Ladungsträgerverteilung in der Finne entwickelt wurde. Zudem bewirkt er im Vergleich zu den planaren Transistoren ein Absinken der Schwellspannung um über 100 mV. Mit der beschriebenen Eckenimplantation konnte eine kostengünstige und selbstlimitierende Lösung für dieses Problem vorgeschlagen werden. Die Implantation erfolgt dabei nach der Finnenmaskierung bzw. vor dem Ätzen der Gräben zwischen den Finnen und nutzt die oberflächennahe Streuung der Ionen an den Atomrümpfen aus, um die Gegendotierung in den Eckbereichen über die gesamte Länge der Finne zu erhöhen. Dabei konnte eine gleichmäßige Verschiebung der Schwellspannungen aller Gatelängen beobachtet und die Vorteile gegenüber einer separaten Haloimplantation aufgezeigt werden.

Am Beispiel der Tri-Gate-Transistoren ist auch der Einfluss der Kristallorientierung auf die Ladungsträgerbeweglichkeit untersucht worden. Dabei zeigte sich, dass die $\langle 100 \rangle$ -Kanalrichtung in Verbindung mit der (100)-Oberflächenorientierung von Top- und Seiten-Gate die beste Leistungsfähigkeit für den n-Tri-Gate-Transistor ermöglicht. Im Gegensatz dazu ist für den p-Tri-Gate-Transistor die $\langle 100 \rangle$ -Kanalrichtung mit (110)-Oberflächenorientierungen zu bevorzugen. Von den vier betrachteten Varianten bot die $\langle 110 \rangle$ -Kanalrichtung mit (100)-Top- und (110)-Seiten-Gate-Orientierung einen guten Kompromiss für eine CMOS-Technologie auf Basis von Tri-Gate-Transistoren.

Diese Kristallorientierung eignet sich zudem am besten für den Einsatz eines verspannten Isolationsoxids zwischen den Finnen. Das vorgeschlagene Konzept zur mechanischen Verspannung von FinFETs ermöglicht mit seiner vorrangig transversalen Wirkungsweise zur Transportrichtung eine gleichgerichtete Leistungsänderung beider Transistortypen. Durch eine intrinsische Zugverspannung des Isolationsoxids von 1 GPa konnte eine positive Leistungssteigerung von n- und p-MOSFET um ca. 10 % erreicht werden. Dieser Effekt ließe sich mit einer Skalierung der Gatelänge bzw. der Finweite weiter steigern und bietet somit auch eine Einsatzmöglichkeit in zukünftigen FinFET-Technologien.

E. Transistoren mit vollständig verarmtem Kanal

Um eine optimale Steuerfähigkeit des Gatekontakts zu erreichen, ist eine Transistorstruktur erforderlich, deren Kanalgebiet vollständig verarmt werden kann. Der ultradünne SOI-MOSFET als planare Variante ist aufgrund seines einseitigen Gatekontakts und des geringen Sättigungsstromes für Kurzkanaltransistoren bzw. für die Skalierung der nächsten Transistorgenerationen jedoch weniger geeignet.

Der vorgeschlagene FinFET mit schmaler hoher Finne ist die Weiterentwicklung des Tri-Gate-Transistors. Er erfüllt alle wesentlichen Kriterien für einen leistungsfähigen Multi-Gate-Transistor, kann jedoch nicht mehr in Kombination mit einem hybriden Prozessansatz realisiert werden. Mit der Finhöhe lässt sich seine effektive Gateweite und somit auch sein Drainstrom unabhängig von der Strukturweite steigern. Damit besitzt der FinFET eine deutlich größere Leistungsfähigkeit als planare MOSFETs.

Wie auch die Tri-Gate-Transistoren profitieren n- und p-FinFETs von einer intrinsischen Zugverspannung des Isolationsoxids zwischen den Finnen, obwohl hier der Großteil der Ladungsträger entlang der Seiten-Gates mit (110)-Orientierung fließt. Mit den transienten Simulationen konnten gegenüber den planaren bzw. Tri-Gate-Transistoren schnellere Schaltzeiten und eine signifikant erhöhte Ringoszillationsfrequenz nachgewiesen werden, wobei die effektive Stromaufnahme durch die größere Gatefläche ebenfalls erhöht ist.

Ein kleiner jedoch kompensierbarer Nachteil ist der erhöhte *DIBL* und die etwas schlechtere Unterschwellsteigung des n-FinFETs aufgrund der Diffusion der Arsenionen und der fehlenden Gegendotierung im Kanal. Eine zusätzliche diffusionsblockende Implantation, ähnlich dem Germanium in den Source-Drain-Gebieten des p-FinFETs, könnte das Kurzkanalverhalten des n-FinFETs wieder verbessern.

Die beste Steuerfähigkeit und die geringsten Kurzkanaleffekte wurden mit den entwickelten vertikalen Nanowire-Transistoren erzielt. Ihre Integration beruht auf dem vorgeschlagenen Verfahren zur Multi-Spacer-Strukturierung, welches eine Modifizierung der bereits in der Volumenproduktion verwendeten Mehrfach-Strukturierung ist. Der vertikale Aufbau ermöglicht im Vergleich zu horizontalen Nanowire-Strukturen eine einfachere Realisierung des Rundum-Gates, welches bei einer zukünftigen Skalierung der Aktivfläche in der Länge unverändert bleiben kann.

Durch die allseitige Steuerung des Kanalgebiets wurden für die Nanowire-Transistoren die geringsten *DIBL*-Werte aller Strukturvarianten und eine Unterschwellsteigung nahe der theoretischen Untergrenze von ca. 60 mV/dec erreicht. Dafür muss die Gatelänge jedoch mindestens dem anderthalbfachen der Kanalweite entsprechen. Mit der Verkleinerung der Aktivfläche über den Abstand und den Durchmesser der Nanodrähte konnte sogar eine Verbesserung der elektrischen Eigenschaften und des Kurzkanalverhaltens beobachtet werden.

Der Einsatz herkömmlicher mechanischer Verspannungstechniken ist bei so kleinen Strukturen nicht mehr möglich. Dafür bietet die vertikale Struktur Möglichkeiten für alternative Ansätze. Die vorgeschlagene Epitaxie einer Hülle aus Silizium-Germanium um den Siliziumkern ist dabei eine vielversprechende Variante, um die Leistungsfähigkeit von n- und p-Nanowire-MOSFETs um 25 ... 60 % zu steigern. Mit der Verringerung der Kanalweite dürfte jedoch die Verspannung und auch ihre Wirkung auf die Ladungsträger abnehmen.

Trotz der erhöhten parasitären Kapazitäten durch den schichtartigen Aufbau der Kontakte von Source, Drain und Gate sind die Werte der dynamischen Schalteigenschaften der Nanowire-Transistoren besser als die der anderen Transistorstrukturen. Ein Grund dafür sind die deutlich höheren Sättigungsströme pro Aktivfläche, die je nach Anzahl der Drähte eines Nanowire-Transistors um das Zwei- bis Fünffache über denen der planaren MOSFETs liegen.

F. Skalierung und statistische Schwankungen der Strukturen

Hinsichtlich einer fortführenden Skalierung und der Anfälligkeit gegenüber statistischen Prozessschwankungen konnten durch Simulationen die FinFETs und die Nanowire-Transistoren als die geeignetsten Transistorstrukturen identifiziert werden. Bei beiden wurde mit dem Übergang zum 14 nm-Technologieknoten eine überproportionale Steigerung ihrer Leistungsfähigkeit beobachtet, was auf einen vergrößerten Kanalquerschnitt pro Aktivfläche zurückzuführen ist. Da sie wie auch die ultradünnen SOI-MOSFET vollständig verarmt sind, bieten sie aufgrund der niedrigen Kanaldotierung eine deutlich bessere Stabilität gegenüber der zufälligen Dotierungsfluktuation als die planaren oder Tri-Gate-Transistoren. Bezüglich ihrer geometrischen Variationen lassen sich mit den Simulationen auf Basis der rauschartigen Impedanz-Feld-Methode keine zuverlässigen Schlussfolgerungen treffen. Der Softwarehersteller hat allerdings zum gegenwärtigen Zeitpunkt eine Umsetzung mittels statistischer Impedanz-Feld-Methode angekündigt, die auch eine Berücksichtigung des Dotierungsprofils vorsieht. Es wird jedoch davon ausgegangen, dass die strukturellen Schwankungen eine entscheidende Bedeutung für Multi-Gate-Transistoren haben und mit der Skalierung der Strukturen weiter an Einfluss gewinnen werden.

G. Ausblick

Neben den aktuell produzierten planaren 28 nm-Technologien wird sich die Halbleiterindustrie in naher Zukunft auf die Produktion und Optimierung von FinFETs konzentrieren. Der Tri-Gate-Transistor bzw. der hybride Prozessansatz bieten keine mittel- oder gar langfristige Lösung der Skalierungsproblematik und stellen somit nur eine Zwischenstufe zum FinFET dar. Die Umstellung von planaren zu Multi-Gate-Strukturen ist für die Hersteller mit einem großen finanziellen Aufwand verbunden und schließt eine parallele Entwicklung von mehreren Transistorstrukturen mit hoher

Wahrscheinlichkeit aus. Voraussichtlich werden die vertikalen Nanowire-Transistoren zunächst ungenutzt bleiben, bis auch die FinFETs bei einer physikalischen Gatelänge von ca. 8 nm ihre physikalischen Grenzen erreicht haben werden. Die Verwendung neuer Kanalmaterialien mit höherer Ladungsträgerbeweglichkeit, wie Germanium oder III-V-Halbleiter, sind nur in Verbindung mit FinFET-Strukturen oder für planare Langkanaltransistoren mit höherer Spannungsfestigkeit sinnvoll.

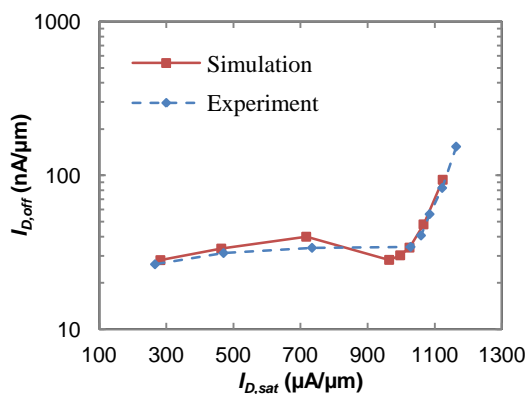
Offen bleibt die Frage, ob die Halbleiterindustrie in einem mittelfristigen Zeitraum den Schritt vom klassischen MOSFET zu neuen Transistorkonzepten mit höherer Funktionalität wagen wird [196]. Mit ihnen eröffnet sich die Möglichkeit, die Komplexität hochintegrierter Schaltkreise auf der schaltungstechnischen bzw. algorithmischen Ebene und somit unabhängig von der Skalierung der Bauelemente weiter zu steigern.

Anhang

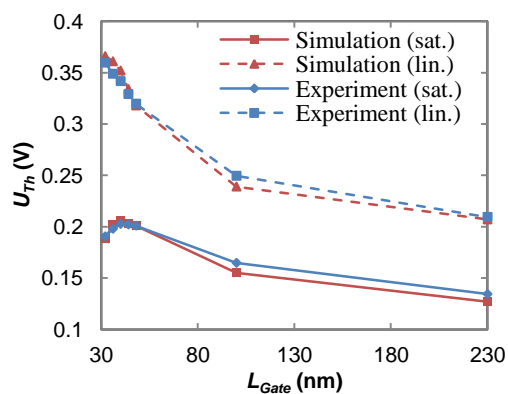
Anhang 1: Vergleich simulierter mit experimentell ermittelten Kennlinien für n-MOSFETs der 32 nm-SOI-Technologie.

n-MOSFET

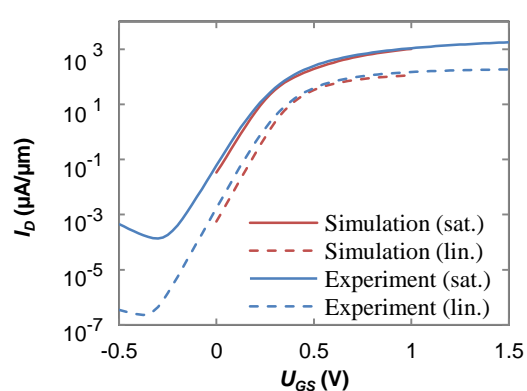
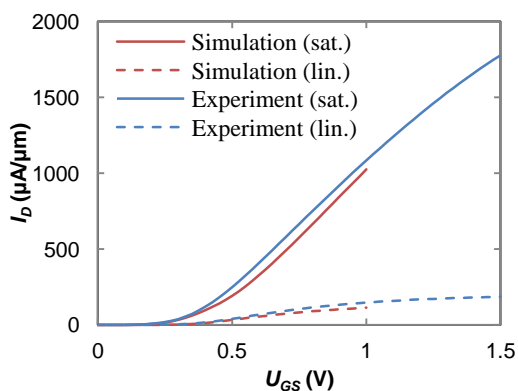
Universalkennlinie



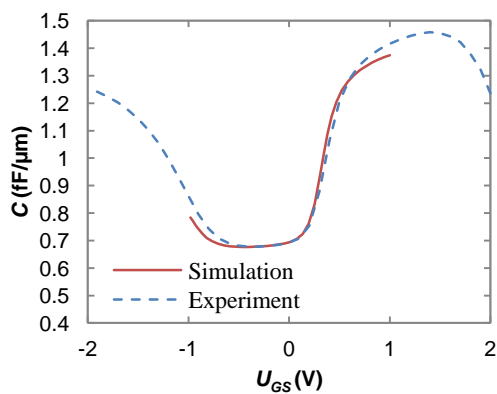
Schwellspannung versus Gatelänge



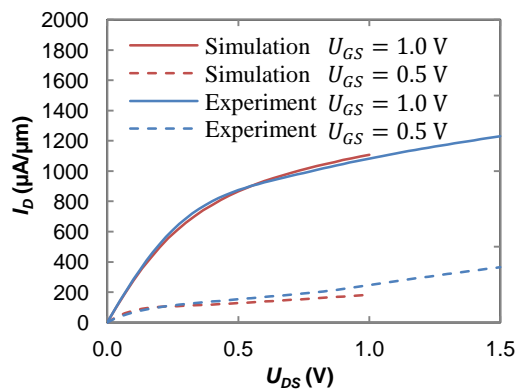
Transferkennlinie in linearer und logarithmischer Darstellung ($L_{Gate} = 40$ nm)



Gate-Source-Kapazität



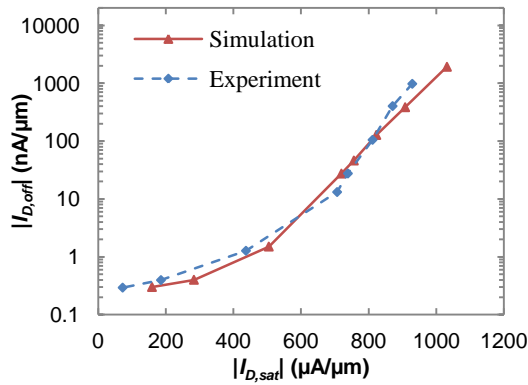
Ausgangskennlinienfeld ($L_{Gate} = 40$ nm).



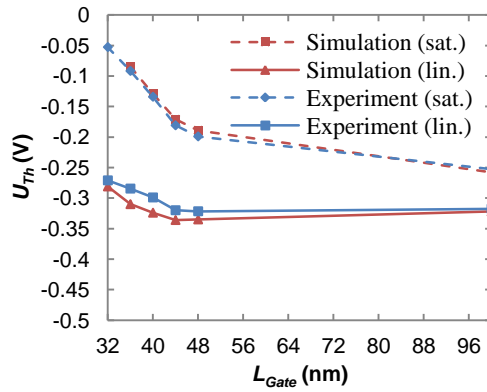
Anhang 2: Vergleich simulierter mit experimentell ermittelten Kennlinien für p-MOSFETs der 32 nm-SOI-Technologie.

p-MOSFET

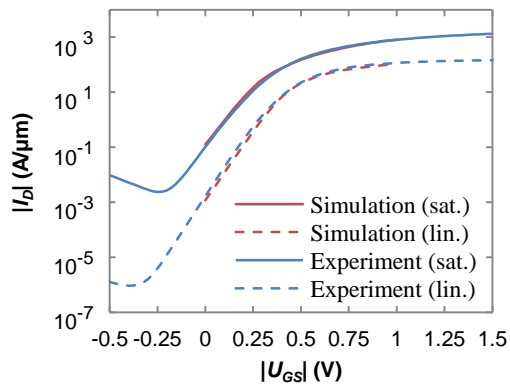
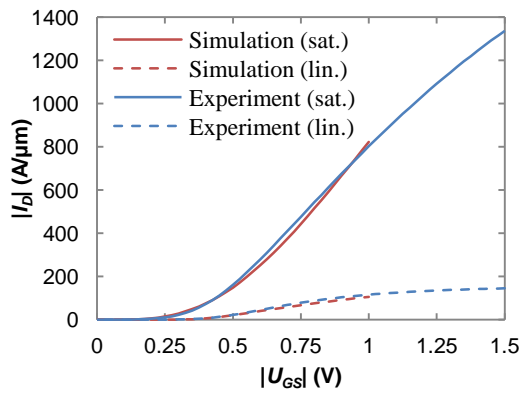
Universalkennlinie



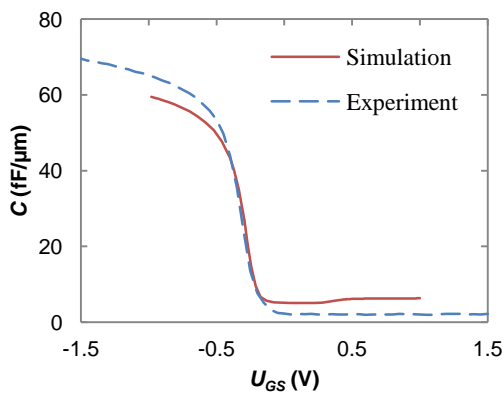
Schwellspannung versus Gatelänge



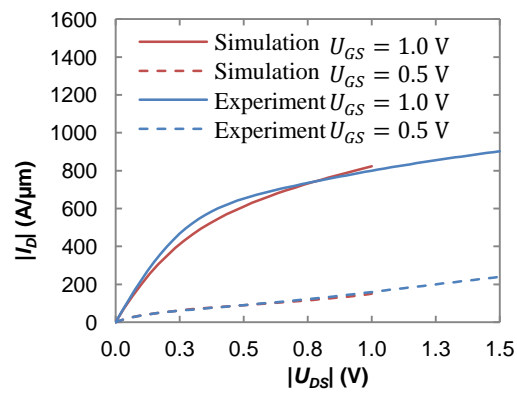
Transferkennlinie in linearer und logarithmischer Darstellung ($L_{Gate} = 40 \text{ nm}$)



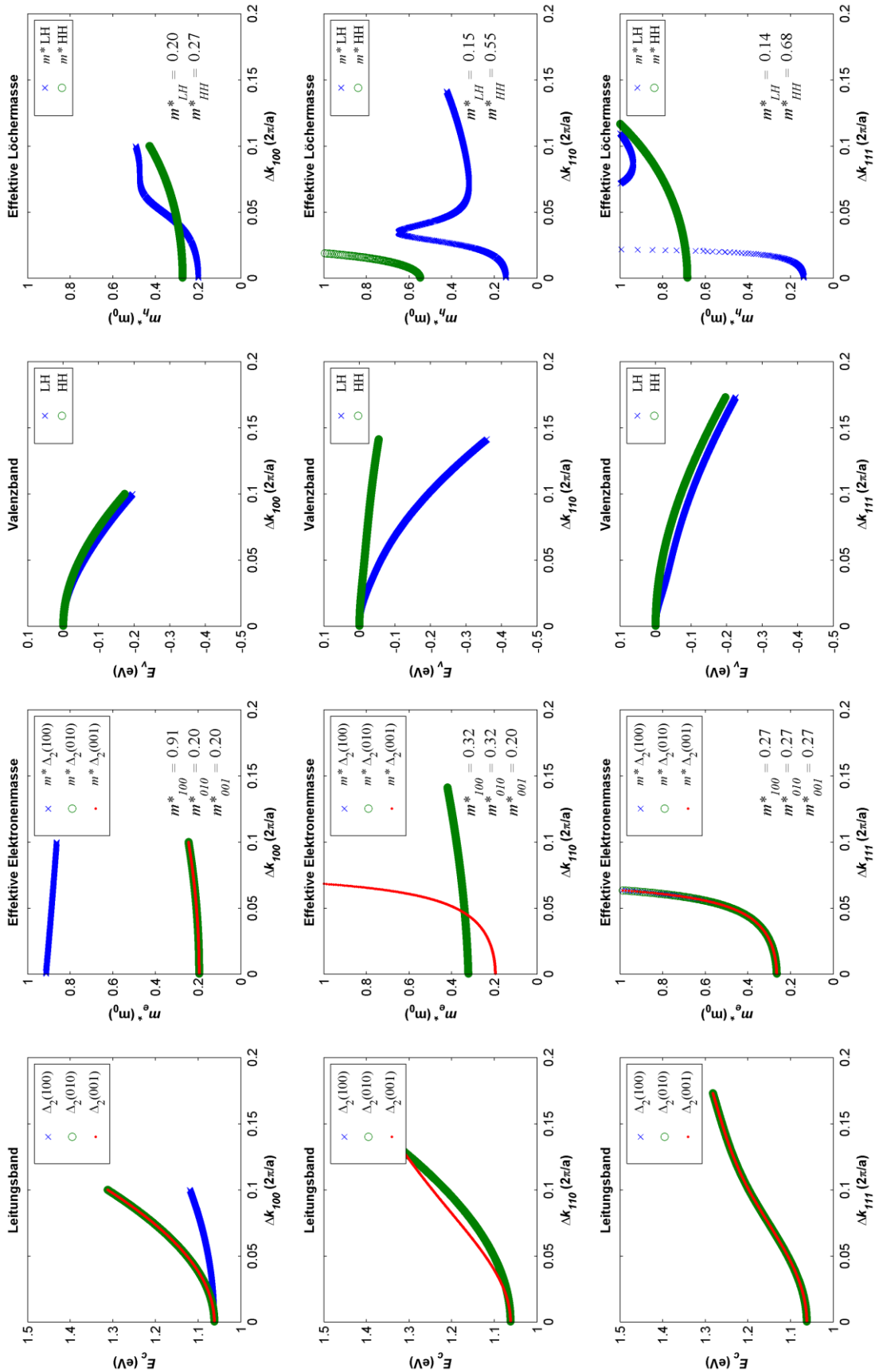
Gate-Source-Kapazität



Ausgangskennlinienfeld ($L_{Gate} = 40 \text{ nm}$).



Anhang 3: Berechneter Verlauf der unverspannten Bandstruktur und der zugehörigen effektiven Massen von Elektronen und Löchern für drei verschiedene Raumrichtungen. $\Delta k = 0$ markiert das Leitungsbandminimum bzw. Valenzbandmaximum.



Anhang 4: Effektive Massen und Energieniveaus des Leitungsbandes nach der Verschiebung bzw. Verbiegung der Subbänder durch mechanische Verspannung und Quantisierung.

	Substratorientierung																										
	(001)/(1-10)/<110> Transporthrichtung				(011)/(100)/<010> Transporthrichtung				(111)/(100)/<010> Transporthrichtung																		
	TopGate (001)	SideGate (1-10)	Corner (1-11)	pn	TopGate (011)	SideGate (100)	Corner (111)	pn	TopGate (111)	SideGate (100)	Corner (011)	pn															
unverspannt	0																										
	Stress/Tiefe (MPa)																										
	Stress/Weite (MPa)																										
	Stress/Länge (MPa)																										
	m*quantisierung (m0)	0.20	0.32	0.27		0.20	0.91	0.27		0.20	0.91	0.32															
	m*transport (m0)	0.32	0.32	0.32		0.20	0.20	0.20		0.20	0.20	0.20															
	E_min (eV)	1.062	1.062	1.062		1.062	1.062	1.062		1.062	1.062	1.062															
	AE_quant (eV)	0.290	0.248	0.262		0.290	0.175	0.262		0.290	0.175	0.248															
	m*quantisierung (m0)	0.20	0.32	0.27		0.32	0.20	0.27		0.20	0.20	0.20															
	m*transport (m0)	0.32	0.32	0.32		0.32	0.32	0.32		0.32	0.32	0.32															
	E_min (eV)	1.062	1.062	1.062		1.062	1.062	1.062		1.062	1.062	1.062															
	AE_quant (eV)	0.290	0.248	0.262		0.248	0.290	0.262		0.290	0.248	0.248															
m*quantisierung (m0)	0.20	0.20	0.27		0.32	0.20	0.27		0.91	0.20	0.32																
m*transport (m0)	0.20	0.20	0.20		0.32	0.32	0.32		0.91	0.20	0.32																
E_min (eV)	1.062	1.062	1.062		1.062	1.062	1.062		1.062	1.062	1.062																
AE_quant (eV)	0.175	0.290	0.262		0.248	0.290	0.262		1.062	1.062	1.062																
AE_Stress (meV) (+gut/-schlecht)	0	0	0		0	0	0		0	0	0																
AE_Quantisierung (meV) (+gut/-schlecht)	115	-42	0		-42	115	0		97	97	42																
AE_gesamt (meV) (+gut/-schlecht)	115	-42	0		-42	115	0		97	97	42																
1 GP tensil	Stress/Tiefe (MPa)	8	6	-11	-292	3	-111		10	0	11	-359	5	-88	15	-2	-13	-298	1	-91	0	-1	55	-308	20	-83	
	Stress/Weite (MPa)	166	831	24	347	1	522	232	788	23	347	0	557	215	769	21	346	6	548	237	890	26	342	18	572	18	572
	Stress/Länge (MPa)	288	95	105	383	248	200	513	242	144	515	278	391	412	198	110	413	242	311	400	256	154	388	260	324	324	
	m*quantisierung (m0)	0.20	0.19	0.32	0.32	0.26	0.27	0.21	0.20	0.91	0.94	0.27	0.19	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
	m*transport (m0)	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32
	E_min (eV)	1.063	1.062	1.063	1.062	1.063	1.063	1.058	1.027	1.062	1.039	1.062	1.038	1.072	1.052	1.047	1.058	1.052	1.047	1.058	1.052	1.037	1.050	1.051	1.051	1.051	
	AE_quant (eV)	0.290	0.295	0.248	0.248	0.266	0.262	0.265	0.290	0.175	0.173	0.262	0.259	0.290	0.295	0.175	0.174	0.248	0.248	0.290	0.281	0.290	0.300	0.290	0.290	0.290	
	m*quantisierung (m0)	0.20	0.19	0.32	0.32	0.26	0.27	0.32	0.32	0.20	0.20	0.26	0.27	0.19	0.19	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
	m*transport (m0)	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32
	E_min (eV)	1.043	1.019	1.056	1.002	1.051	1.020	1.058	1.027	1.062	1.039	1.062	1.038	1.036	1.017	1.055	1.000	1.015	1.047	1.058	1.056	1.037	1.050	1.051	1.051	1.051	
	AE_quant (eV)	0.175	0.174	0.290	0.290	0.262	0.266	0.248	0.248	0.248	0.290	0.266	0.262	0.175	0.175	0.290	0.290	0.248	0.248	0.248	0.248	0.248	0.248	0.248	0.248	0.248	0.248
	AE_Stress (meV) (+gut/-schlecht)	20	43	7	60	11	43	4	-60	5	-22	13	-36	29	2	10	50	22	22	22	25	-20	11	32	22	6	
AE_Quantisierung (meV) (+gut/-schlecht)	115	121	-42	-42	3	-3	-37	-42	115	117	3	3	102	102	97	98	42	43	42	33	42	52	97	97	97		
AE_gesamt (meV) (+gut/-schlecht)	135	164	-35	-18	14	40	-33	-102	120	95	16	-33	131	104	107	148	64	65	67	13	53	84	119	103	103		
1 GP kompressiv	Stress/Tiefe (MPa)	-12	-5	12	303	-4	104	-13	3	-3	359	1	81	-16	2	20	295	0	87	-1	5	-43	303	-18	81		
	Stress/Weite (MPa)	-168	-650	-22	-288	3	-504	-256	-789	-20	-330	-9	-554	-228	-764	-18	-347	-12	-545	-235	-910	-16	-354	-5	-569		
	Stress/Länge (MPa)	-292	-88	-110	-360	-252	-215	-526	-235	-145	-524	-288	-394	-416	-202	-106	-404	-244	-314	-407	-251	-149	-392	-258	-325		
	m*quantisierung (m0)	0.20	0.20	0.32	0.32	0.27	0.27	0.19	0.19	0.91	0.93	0.26	0.25	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	
	m*transport (m0)	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	
	E_min (eV)	1.050	1.030	1.058	1.029	1.054	1.037	1.056	1.000	1.063	1.034	1.065	1.025	1.053	1.001	1.061	1.030	1.063	1.022	1.036	1.056	1.052	1.027	1.041	1.048		
	AE_quant (eV)	0.290	0.290	0.248	0.248	0.262	0.262	0.295	0.295	0.175	0.174	0.266	0.269	0.290	0.290	0.175	0.175	0.248	0.248	0.290	0.300	0.290	0.285	0.290	0.290		
	m*quantisierung (m0)	0.20	0.20	0.32	0.32	0.27	0.27	0.32	0.32	0.20	0.19	0.27	0.27	0.20	0.20	0.20	0.19	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20	
	m*transport (m0)	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	
	E_min (eV)	1.050	1.030	1.058	1.029	1.054	1.037	1.055	1.061	1.059	1.053	1.062	1.036	1.036	1.052	1.053	1.024	1.042	1.043	1.063	1.041	1.063	1.062	1.063	1.057		
	AE_quant (eV)	0.92	0.94	0.19	0.19	0.26	0.27	0.32	0.32	0.20	0.19	0.27	0.27	0.92	0.93	0.20	0.19	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	
	m*quantisierung (m0)	0.20	0.18	0.20	0.20	0.20	0.19	0.32	0.32	0.32	0.32	0.32	0.32	0.19	0.20	0.20	0.19	0.19	0.19	0.19	0.19	0.20	0.20	0.19	0.19	0.20	
E_min (eV)	1.070	1.070	1.065	1.066	1.065	1.078	1.055	1.061	1.059	1.053	1.062	1.072	1.070	1.065	1.068	1.064	1.080	1.063	1.041	1.063	1.062	1.063	1.063	1.057			
AE_quant (eV)	0.174	0.173	0.295	0.295	0.266	0.262	0.248	0.248	0.290	0.295	0.262	0.262	0.174	0.174	0.290	0.295	0.248	0.248	0.248	0.248	0.248	0.248	0.248	0.248	0.248		
AE_Stress (meV) (+gut/-schlecht)	-20	-40	-7	-57	-11	-41	-1	61	-4	25	-12	37	-25	35	-10	-21	-21	6	-27	15	-11	-35	-22	-9			
AE_Quantisierung (meV) (+gut/-schlecht)	116	117	-47	-47	-3	0	-47	-47	115	121	-3	-7	98	99	87	101	42	42	42	47	52	42	37	98	97		
AE_gesamt (meV) (+gut/-schlecht)	96	77	-54	-104	-14	-41	-48	-14	111	146	-5	30	73	133	87	81	21	48	20	67	62	31	2	76	88		

Verspannung des Isolationsoxids

Anhang 5: Effektive Massen und Energieniveaus des Valenzbandes nach der Verschiebung bzw. Verbiegung der Subbänder durch mechanische Verspannung und Quantisierung.

		Substratorientierung											
		<001>/<1-10>/<110> Transportrichtung				<011>/<100>/<011> Transportrichtung				<011>/<01-1>/<100> Transportrichtung			
		TopGate	SideGate	Corner	TopGate	SideGate	Corner	TopGate	SideGate	Corner	TopGate	SideGate	Corner
		(001)	(1-10)	(1-11)	(001)	(100)	(111)	(001)	(100)	(101)	(011)	(01-1)	(010)
		Mitte	pn	Mitte	Mitte	pn	Mitte	Mitte	pn	Mitte	pn	Mitte	pn
		0											
Stress/Tiefe (MPa)		0											
Stress/Weite (MPa)		0											
m*quantisierung (m0)		0.27	0.55	0.55	0.55	0.27	0.88	0.27	0.27	0.27	0.55	0.55	0.27
E_max (eV)		0	0	0	0	0	0	0	0	0	0	0	0
AE_quant (eV)		-0.262	-0.207	-0.193	-0.207	-0.262	-0.193	-0.262	-0.262	-0.262	-0.207	-0.207	-0.262
m*quantisierung (m0)		0.2	0.15	0.14	0.15	0.2	0.14	0.2	0.2	0.2	0.15	0.15	0.2
m*transport (m0)		0	0	0	0	0	0	0	0	0	0	0	0
E_max (eV)		-0.290	-0.319	-0.326	-0.319	-0.290	-0.326	-0.290	-0.290	-0.290	-0.319	-0.319	-0.290
AE_quant (eV)		0	0	0	0	0	0	0	0	0	0	0	0
AE_Stress (meV) (-gut/+schlecht)		27.59005356	112.1281877	133.674874	112.1281877	27.59005356	133.674874	27.59005356	27.59005356	112.1281877	112.1281877	112.1281877	27.59005356
AE_Quantisierung (meV) (-gut/+schlecht)		27.59005356	112.1281877	133.674874	112.1281877	27.59005356	133.674874	27.59005356	27.59005356	112.1281877	112.1281877	112.1281877	27.59005356
AE_gesamt (meV) (-gut/+schlecht)		5	80	18	527	-11	389	3	96	9	693	-3	397
Stress/Tiefe (MPa)		-31	-599	-4	102	-14	486	-95	407	-13	386	-103	612
Stress/Weite (MPa)		-2228	-1596	-1862	-2056	-2301	-1264	-2136	-1609	-1903	-2035	-2223	-1385
Stress/Länge (MPa)		0.28	0.28	1.53	1.75	0.64	0.61	1.67	1.45	0.28	0.29	0.64	0.61
m*quantisierung (m0)		0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11
m*transport (m0)		0	0	0	0	0	0	0	0	0	0	0	0
E_max (eV)		-0.259	-0.259	-0.147	-0.141	-0.197	-0.2	-0.143	-0.15	-0.259	-0.256	-0.197	-0.2
AE_quant (eV)		0.25	0.21	0.12	0.12	0.15	0.15	0.13	0.12	0.24	0.21	0.15	0.15
m*quantisierung (m0)		1.22	0.55	1.03	1.56	1.23	0.86	1.02	1.21	1.01	0.59	1.16	0.96
m*transport (m0)		-0.086	-0.073	-0.072	-0.068	-0.09	-0.065	-0.082	-0.088	-0.073	-0.095	-0.087	-0.067
E_max (eV)		-0.269	-0.285	-0.344	-0.344	-0.319	-0.319	-0.335	-0.335	-0.273	-0.289	-0.319	-0.319
AE_quant (eV)		-86	-73	-72	-88	-90	-65	-82	-68	-73	-95	-87	-67
AE_Stress (meV) (-gut/+schlecht)		-9.974	-26.07	-196.5	-203	-122.3	-119.1	-191.7	-193.9	-133.66	-29.09	-123.3	-119.1
AE_Quantisierung (meV) (-gut/+schlecht)		95.97	99.07	868.5	291	212.3	184.1	273.7	261.9	86.66	124.1	209.3	186.1
AE_gesamt (meV) (-gut/+schlecht)		24	65	3	297	0	233	11	66	27	96	9	281
Stress/Tiefe (MPa)		154	1674	24	440	-18	1086	159	1420	18	464	21	1091
Stress/Weite (MPa)		-1715	-1478	-1714	-1529	-1890	-1018	-1499	-1383	-1856	-1445	-1934	-987
Stress/Länge (MPa)		0.28	0.28	1.45	1.56	0.62	0.6	1.31	1.32	0.28	0.28	0.62	0.57
m*quantisierung (m0)		0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11
m*transport (m0)		0	0	0	0	0	0	0	0	0	0	0	0
E_max (eV)		-0.259	-0.259	-0.15	-0.146	-0.199	-0.201	-0.155	-0.155	-0.259	-0.259	-0.199	-0.204
AE_quant (eV)		0.24	0.18	0.13	0.13	0.15	0.19	0.12	0.12	0.24	0.25	0.15	0.15
m*quantisierung (m0)		0.86	0.32	0.93	0.89	1.04	0.37	0.98	1.34	1.01	1.07	1.01	1.11
m*transport (m0)		-0.069	-0.065	-0.066	-0.073	-0.072	-0.054	-0.059	-0.063	-0.072	-0.071	-0.075	-0.054
E_max (eV)		-0.273	-0.3	-0.335	-0.335	-0.319	-0.295	-0.344	-0.344	-0.273	-0.269	-0.319	-0.319
AE_quant (eV)		-69	-65	-66	-73	-72	-54	-59	-63	-72	-71	-75	-54
AE_Stress (meV) (-gut/+schlecht)		-13.66	-41.11	-184.8	-188.4	-120.2	-93.87	-188.7	-189.1	-135.66	-9.74	-120.2	-114.6
AE_Quantisierung (meV) (-gut/+schlecht)		82.66	106.1	250.8	261.4	192.2	147.9	247.7	252.1	85.66	124.1	209.3	186.6
AE_gesamt (meV) (-gut/+schlecht)		-7	81	35	753	-21	535	-12	102	-8	990	-15	519
Stress/Tiefe (MPa)		-190	-506	-33	-208	-9	-140	-345	-650	-44	-313	-29	-302
Stress/Weite (MPa)		-2756	-1873	-2005	-2596	-2711	-1560	-2776	-1921	-1948	-2629	-2512	-1771
Stress/Länge (MPa)		0.29	0.28	1.6	1.95	0.66	0.58	1.92	1.46	0.28	0.29	0.65	0.62
m*quantisierung (m0)		0.11	0.12	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11
m*transport (m0)		0	0	0	0	0	0	0	0	0	0	0	0
E_max (eV)		-0.256	-0.259	-0.145	-0.136	-0.195	-0.203	-0.136	-0.149	-0.259	-0.256	-0.196	-0.199
AE_quant (eV)		0.25	0.26	0.12	0.11	0.14	0.14	0.13	0.18	0.24	0.19	0.14	0.19
m*quantisierung (m0)		1.59	1.16	1.14	1.97	1.43	1.21	0.99	0.39	1	0.38	1.32	0.39
m*transport (m0)		-0.104	-0.058	-0.077	-0.102	-0.108	-0.06	-0.107	-0.057	-0.074	-0.111	-0.099	-0.066
E_max (eV)		-0.269	-0.266	-0.344	-0.354	-0.326	-0.326	-0.335	-0.3	-0.273	-0.295	-0.326	-0.295
AE_quant (eV)		-104	-58	-77	-102	-108	-60	-107	-57	-74	-111	-99	-66
AE_Stress (meV) (-gut/+schlecht)		-12.99	-6.48	-198.7	-218.1	-131.7	-123.2	-198.2	-150.8	-13.66	-38.76	-130.8	-96.06
AE_Quantisierung (meV) (-gut/+schlecht)		117	64.48	275.7	320.1	239.7	183.2	305.2	207.8	87.66	149.8	229.8	162.1
AE_gesamt (meV) (-gut/+schlecht)		-11	-64.48	-275.7	-320.1	-239.7	-183.2	-305.2	-207.8	-87.66	-149.8	-229.8	-162.1

Anhang 6: Übersicht zu den effektiven Massen, Energieniveaus und Verschiebungen des Leitungsbandes durch mechanische Verspannung und Quantisierung in einem n-Nanowire-Transistor.

		Substratorientierung				
		(100) Oberflächenorientierung <011> Transportrichtung		(0-11) Oberflächenorientierung <011> Transportrichtung		
		EPM als Si	EPM als SiGe	EPM als Si	EPM als SiGe	
unverspannt	Stress/orthogonal (MPa)	0		0		
	Stress/transversal (MPa)	0		0		
	Stress/longitudinal (MPa)	0		0		
	Δ_2100	m*quantisierung (m0)	0.91	0.94	0.20	0.20
		m*transport (m0)	0.20	0.20	0.20	0.20
		E_min (eV)	1.062	1.049	1.062	1.049
		ΔE_{quant} (eV)	0.175	0.173	0.290	0.290
	Δ_2010	m*quantisierung (m0)	0.20	0.20	0.32	0.33
		m*transport (m0)	0.32	0.33	0.32	0.33
		E_min (eV)	1.062	1.049	1.062	1.049
		ΔE_{quant} (eV)	0.290	0.290	0.248	0.245
	Δ_2001	m*quantisierung (m0)	0.20	0.20	0.32	0.33
		m*transport (m0)	0.32	0.33	0.32	0.33
		E_min (eV)	1.062	1.049	1.062	1.049
		ΔE_{quant} (eV)	0.290	0.290	0.248	0.245
ΔE_{Stress} (meV) (+gut/-schlecht)		0	0	0	0	
$\Delta E_{\text{Quantisierung}}$ (meV) (+gut/-schlecht)		115	117	-42	-45	
ΔE_{gesamt} (meV) (+gut/-schlecht)		115	117	-42	-45	
SiGe-Hülle mit 30% Ge	Stress/orthogonal (MPa)	160		170		
	Stress/transversal (MPa)	-1250		-1340		
	Stress/longitudinal (MPa)	-920		-1410		
	Δ_2100	m*quantisierung (m0)	0.94	0.97	0.17	0.18
		m*transport (m0)	0.18	0.19	0.24	0.25
		E_min (eV)	1.107	1.106	0.985	0.973
		ΔE_{quant} (eV)	0.173	0.171	0.306	0.300
	Δ_2010	m*quantisierung (m0)	0.20	0.21	0.33	0.34
		m*transport (m0)	0.32	0.33	0.32	0.33
		E_min (eV)	0.995	0.982	1.059	1.056
		ΔE_{quant} (eV)	0.290	0.285	0.245	0.243
	Δ_2001	m*quantisierung (m0)	0.20	0.21	0.33	0.34
		m*transport (m0)	0.32	0.33	0.32	0.33
		E_min (eV)	0.995	0.982	1.059	1.056
		ΔE_{quant} (eV)	0.290	0.285	0.245	0.243
ΔE_{Stress} (meV) (+gut/-schlecht)		-112	-124	74	83	
$\Delta E_{\text{Quantisierung}}$ (meV) (+gut/-schlecht)		117	114	-61	-57	
ΔE_{gesamt} (meV) (+gut/-schlecht)		5	-10	13	26	

Literaturverzeichnis

- [1] <http://www.intel.de/content/www/de/de/history/museum-transistors-to-transformations-brochure.html> (20.08.2013)
- [2] <http://www.amd.com/us/aboutamd/corporate-information/Pages/timeline.aspx> (20.08.2013)
- [3] <http://www.pc-erfahrung.de/prozessor/cpu-prozessortabelle.html> (20.08.2013)
- [4] G. E. Moore: *Progress in digital integrated electronics*, IEEE International Electron Devices Meeting, 12 (1975), 11-13
- [5] A. Tilke, K. Schrüfer, C. Stapelmann: *Mikroelektronik: kein Ende der Skalierung in Sicht*, Physik Journal, 6 (2007), 35-41
- [6] M. Reisch: *Halbleiter-Bauelemente*, 2. Aufl., Springer-Verlag, 2007
- [7] Y. Taur: *Fundamentals of Modern VLSI Devices*, Cambridge University Press, Cambridge, 1998
- [8] M. Wordman: *Characterization of depletion mode MOSFET's*, IEEE International Electron Devices Meeting, 25 (1979), 26-29
- [9] F. Thuselt: *Physik der Halbleiterbauelemente*, Springer-Verlag, 2005
- [10] S. M. Sze: *Physics of Semiconductor Devices*, Wiley-Interscience, New York, 1969
- [11] C. Kittel: *Einführung in die Festkörperphysik*, 14. Aufl., Oldenbourg Verlag, 2006
- [12] M. Schmidt: *Neuartige Metallelektroden und Dielektrika für die MOS-Technologie*, Dissertation, Universität der Bundeswehr München, 2005
- [13] T. Giebel: *Grundlagen der CMOS-Technologie*, B. G. Teubner Verlag, Stuttgart, 2002
- [14] H. Göbel: *Einführung in die Halbleiter-Schaltungstechnik*, Springer-Verlag, Hamburg, 2005
- [15] S. Flachowsky: *Verspannungstechniken zur Leistungssteigerung von SOI-CMOS-Transistoren*, Dissertation, Technische Universität Dresden, 2010
- [16] J. D. Cressler: *Silicon Heterostructure Handbook*, CRC Press, Boca Raton, 2006
- [17] D. K. Schroder: *Semiconductor Material and Device Characterization*, 3. Aufl., Wiley-Interscience, 2006
- [18] W. Y. Choi, B. G. Park, J. D. Lee et al.: *Tunneling field-effect transistors (TFETs) with subthreshold swing (SS) less than 60 mV/dec*, IEEE Electron Device Letters, 28 (2007), 743-745
- [19] K. Boucart, A. M. Ionescu: *A new definition of threshold voltage in tunnel FETs*, Solid-State Electronics, 52 (2008), 1318-1323
- [20] K. Boucart, A. M. Ionescu: *Double gate tunnel FET with high- κ gate dielectric*, IEEE Transactions on Electron Devices, 54 (2007), 1725-1733
- [21] T. Herrmann: *Simulation und Optimierung neuartiger SOI-MOSFETs*, Dissertation, Technische Universität Chemnitz, 2009

- [22] ITRS: *International Technology Roadmap for Semiconductors*, Edition 2011 <http://www.itrs.net>
- [23] K. J. Kuhn: *Considerations for ultimate CMOS scaling*, IEEE Transaction on Electron Devices, 59 (2012), 1813-1828
- [24] T. H. Ning: *A perspective on the theory of MOSFET scaling and its impact*, Solid-State Circuits Newsletter, 12 (2007), 27-30
- [25] S. E. Thompson: *Source-drain series resistance: The real limiter to MOSFET scaling*, Proceedings - Electrochemical Society, 1 (2004), 412-419
- [26] V. Sverdlov: *Strain-Induced Effects in Advanced MOSFETs*, Springer-Verlag, Vienna, 2011
- [27] Y. Sun, S. E. Thompson, T. Nishida: *Physics of strain effects in semiconductor and metal-oxide-semiconductor field-effect transistors*, Journal of Applied Physics, 101 (2007), 1-22
- [28] W.-K. Yeh, J.-W. Chou: *Optimum halo structure for sub-0.1 μm CMOSFETs*, IEEE Transaction on Electron Devices, 48 (2001), 2357-2362
- [29] Globalfoundries Limited Liability Company & Co. KG, (2010)
- [30] K. Nehari, N. Cavassilas, J. Autran et al.: *Influence of band structure on electron ballistic transport in silicon nanowire MOSFETs - an atomistic study*, Solid-State Electronics, 50 (2006), 716-721
- [31] R. Wang, H. Liu, R. Huang et al.: *Experimental investigations on carrier transport in Si nanowire transistors ballistic efficiency and apparent mobility*, IEEE Transaction on Electron Devices, 55 (2008), 2960-2967
- [32] T. Skotnicki, G. Merckel, T. Pedron: *The voltage-doping transformation: A new approach to the modeling of MOSFET short-channel effects*, IEEE Electron Device Letters, 9 (1988), 109-112
- [33] T. Skotnicki: *Heading for decananometer CMOS – Is navigation among icebergs still a viable strategy?*, 30th European Solid-State Device Research Conference, (2000), 19-33
- [34] J. P. Colinge: *FinFETs and Other Multi-Gate Transistors*, Springer-Verlag, Cork, Ireland, 2008
- [35] J. P. Colinge: *Multiple-gate SOI MOSFETs*, Solid-State Electronics, 48 (2004), 897-905
- [36] Q. Chen, E. M. Harrell: *A physical short-channel threshold voltage model for undoped symmetric double-gate MOSFETs*, IEEE Transactions on Electron Devices, 50 (2003), 1631-1637
- [37] C. Auth, J. Plummer: *Scaling theory for cylindrical, fully-depleted, surrounding-gate MOSFET's*, IEEE Electron Device Letters, 18 (1997), 74-76
- [38] W. Yang, Z. Yu, L. Tian: *Scaling theory for FinFETs based on 3-D effects investigation*, IEEE Transactions on Electron Devices, 54 (2007), 1140-1147
- [39] C. W. Lee, S. R. Yun, C. G. Yu et al.: *Device design guidelines for nano-scale MuGFETs*, Solid-State Electronics, 51 (2007), 505-510
- [40] S. E. Thompson, S. Suthram, Y. Sun et al.: *Future of strained Si/semiconductors in nanoscale MOSFETs*, IEEE International Electron Devices Meeting, (2006), 1-4
- [41] C. K. Maiti, N. B. Chakrabarti, S. K. Ray: *Strained Silicon Heterostructures: Materials and Devices*, The Institution of Electrical Engineers, London, 2001

- [42] P. R. Chidambaram, C. Bowen, S. Chakravarthi et al.: *Fundamentals of silicon material properties for successful exploitation of strain engineering in modern CMOS manufacturing*, IEEE Transactions on Electron Devices, 53 (2006), 944-964
- [43] S. Flachowsky, A. Wei, R. Illgen et al.: *Understanding strain-induced drive-current enhancement in strained-silicon n-MOSFET and p-MOSFET*, IEEE Transactions on Electron Devices, 57 (2010), 1343-1354
- [44] M. Lundstrom: *Fundamentals of Carrier Transport*, 2. Aufl., Cambridge University Press, 2000
- [45] T. Skotnicki, C. Fenouillet-Beranger, C. Gallon et al.: *Innovative Materials, Devices, and CMOS Technologies for Low-Power Mobile Multimedia*, IEEE Transactions on Electron Devices, 55 (2008), 96-130
- [46] M. V. Fischetti, Z. Ren, P. M. Solomon et al.: *Six-band k_p calculation of the hole mobility in silicon inversion layers: Dependence on surface orientation, strain, and silicon thickness*, Journal of Applied Physics, 94 (2003), 1079-1095
- [47] S. Flachowsky, A. Wei, T. Herrmann et al.: *Gate length scaling trends of drive current enhancement in CMOSFETs with dual stress overlayers and embedded-SiGe*, Materials Science and Engineering, 154-155 (2008), 98-101
- [48] G. Eneman, M. Jurczak, P. Verheyen et al.: *Scalability of strained nitride capping layers for future CMOS generations*, Proceedings of ESSDERC (Grenoble), (2005), 449-452
- [49] C. Populaire, D. Villanueva, S. Orain et al.: *Strained contact etch stop layer integration: Geometry design impact*, Simulation of Semiconductor Processes and Devices, 12 (2007), 421-424
- [50] G. Eneman, P. Verheyen, A. De Keersgieter et al.: *Scalability of stress induced by contact-etch-stop layers: A simulation study*, IEEE Transaction on Electron Devices, 54 (2007), 1446-1453
- [51] S. E. Thompson, G. Sun, Y. S. Choi et al.: *Uniaxial-process-induced strained-Si: Extending the CMOS roadmap*, IEEE Transaction on Electron Devices, 53 (2006), 1010-1020
- [52] S. Orain, V. Fiori, D. Villanueva et al.: *Method for managing the stress due to the strained nitride capping layer in MOS transistors*, IEEE Transactions on Electron Devices, 54 (2007), 814-821
- [53] C. D. Sheraw, M. Yang, D. M. Fried et al.: *Dual stress liner enhancement in hybrid orientation technology*, Symposium on VLSI, (2005), 12-13
- [54] T. Baldauf, R. Illgen, S. Flachowsky et al.: *Optimization of stressor overlayer parameters for MOSFETs in "Cool Silicon"-technologies*, Proceeding, 8. International Nanotechnology Symposium (Dresden, Fraunhofer IWS), (2010)
- [55] S. Flachowsky, R. Illgen, T. Herrmann et al.: *Stress memorization technique for n-MOSFETs: Where is the stress memorized*, International Conference on Ultimate Integration on Silicon ULIS (Glasgow), (2010), 149-152
- [56] X. Wang, J. Wu: *Progress in modeling of SMT "stress memorization technique" and prediction of stress enhancement by a novel PMOS SMT process*, International Conference on Simulation of Semiconductor Processes and Devices SISPAD, (2008), 117-120

- [57] S. Flachowsky, J. Höntschel, A. Wei et al.: *Scalability of advanced partially depleted n-MOSFET devices on biaxial strained SOI substrates*, International Conference on Ultimate Integration of Silicon ULIS, (2009), 161-164
- [58] S. Yamakawa, S. Mayuzumi, J. Wang et al.: *Study of stress effect on replacement gate technology with compressive stress liner and eSiGe for pFETs*, International Conference on Simulation of Semiconductor Processes and Devices SISPAD, (2008), 109-112
- [59] S. E. Thompson, M. Armstrong, C. Auth et al.: *A 90-nm logic technology featuring strained-silicon*, IEEE Transactions on Electron Devices, 51 (2004), 1790-1797
- [60] C. T. Koch, V. B. Özdöl, P. A. van Aken: *An efficient, simple, and precise way to map strain with nanometer resolution in semiconductor devices*, Applied Physics Letters, 96 (2010), 091901-1 - 091901-3
- [61] K. Ota, T. Sanuki, K. Yahashi et al.: *Scalable eSiGe S/D technology with less layout dependence for 45-nm generation*, Symposium on VLSI Technology, (2006), 64-65
- [62] B. F. Yang, Z. Ren, R. Takalkar et al.: *Recent progress and challenges in enabling embedded Si:C technology*, ECS Transactions, 16 (2008), 317-323
- [63] R. Arghavani, Z. Yuan, N. Ingle et al.: *Stress management in sub-90-nm transistor architecture*, IEEE Transactions on Electron Devices, 51 (2004), 1740-1743
- [64] A. Steegen, K. Maex, I. De Wolf: *Local mechanical stress induced defects for Ti and Co/Ti silicidation in sub-0.25µm MOS-technologies*, Symposium on VLSI Technology, (1998), 200-201
- [65] L. J. Chen, S. Cheng, H. M. Luo et al.: *The influences of stress on the growth of Ti and Ni silicide thin films on (001) Si*, International Conference on Solid-State and Integrated Circuit Technology, (1998), 256-259
- [66] C. Auth, A. Cappellani, J. S. Chun et al.: *45nm high-k + metal gate strain-enhanced transistors*, Symposium on VLSI Technology, (2008), 128-129
- [67] J. T. Park, J. P. Colinge: *Multiple-gate SOI MOSFETs: Device design guidelines*, IEEE Transactions on Electron Devices, 49 (2002), 2222-2229
- [68] H.-S. P. Wong, D. J. Frank, P. M. Solomon et al.: *Nanoscale CMOS*, Proceedings of the IEEE, 87 (1999), 537-570
- [69] M. Horstmann, A. Wei, J. Höntschel et al.: *Advanced SOI CMOS transistor technologies for high-performance microprocessor applications*, IEEE Custom Intergrated Circuits Conference CICC, (2009), 149-152
- [70] T. Sekigawa, Y. Hayashi: *Calculated threshold-voltage characteristics of an XMOS transistor having an additional bottom gate*, Solid-State Electronics, 27 (1984), 827
- [71] K. Okoto, T. Izumida, H. Kawasaki et al.: *Process integration technology and device characteristics of CMOS FinFET on bulk silicon substrate with sub-10 nm fin width and 20 nm gate length*, IEEE International Electron Devices Meeting, (2005), 721-724
- [72] X. Sun, Q. Lu, V. Moroz et al.: *Tri-gate bulk MOSFET design for CMOS scaling to the end of the roadmap*, IEEE Electron Device Letters, 29 (2008), 491-493
- [73] S. Barraud, R. Coquand, M. Cassé et al.: *Performance of omega-shaped-gate silicon nanowire MOSFET with diameter down to 8 nm*, IEEE Electron Device Letters, 33 (2012), 1526-1528

- [74] J. W. Sleight, S. Bangsaruntip, A. Majumdar et al.: *Gate-all-around silicon nanowire MOSFETs and circuits*, Device Research Conference DRC, (2010), 269-272
- [75] J. W. Sleight: *Silicon Nanowire Devices*, IEEE Short Course International Electron Devices Meeting, (2011)
- [76] S. Maheshwaram, S. K. Manhas, G. Kaushal et al.: *Device circuit co-design issues in vertical nanowire CMOS platform*, IEEE Electron Device Letters, 33 (2012), 934-936
- [77] J. P. Colinge, J. W. Park, W. Xiong: *Threshold voltage and subthreshold slope of multiple-gate SOI MOSFETs*, IEEE Electron Device Letters, 24 (2003), 515-517
- [78] W. Xiong, J. W. Park, J. P. Colinge: *Corner effect in multiple-gate SOI MOSFETs*, IEEE International SOI Conference, (2003), 111-113
- [79] M. P. Kumar, S. K. Gupta, M. Paul: *Corner effects in SOI-Trigate FinFET structure by using 3D process and device simulations*, International Conference on Computer and Communication Technology ICCCT, (2010), 683-686
- [80] X. Baie, J. P. Colinge, V. Bayot et al.: *Quantum-wire effects in thin and narrow SOI MOSFETs*, IEEE International SOI Conference, (1995), 66-67
- [81] S. A. Pervez, H. Kim, B. G. Park et al.: *Simulation study for suppressing corner effect in a saddle MOSFET for sub-50 nm high density high performance DRAM cell transistor*, International Semiconductor Device Research Symposium ISDRS, (2009), 1-2
- [82] M. Poljak, V. Jovanovic, T. Suligoj: *Suppression of corner effects in triple-gate bulk FinFETs*, EUROCON Conference, (2009), 1219-1224
- [83] J. G. Fossum, J. W. Yang, V. P. Trivedi: *Suppression of corner effects in triple-gate MOSFETs*, IEEE Electron Device Letters, 24 (2003), 745-747
- [84] T. Vogelsang, W. Hansen, R. Kircher: *Charge transport near the Si/SiO₂ interface in MOSFET devices*, 19th European Solid State Device Research Conference, (1989), 687-690
- [85] G. Paasch, H. Übensee: *A modified local density approximation: electron density in inversion layers*, Physica Status Solidi, 113 (1982), 165-178
- [86] J. P. Colinge, J. C. Alderman, W. Xiong et al.: *Quantum-mechanical effects in trigate SOI MOSFETs*, IEEE Transactions on Electron Devices, 53 (2006), 1131-1136
- [87] B. Majkusiak, T. Janik, J. Walczak: *Semiconductor thickness effects in the double-gate SOI MOSFET*, IEEE Transactions on Electron Devices, 45 (1998), 1127-1134
- [88] M. Balaguer, J. B. Roldán, L. Donetti et al.: *Inversion charge modeling in n-type and p-type double-gate MOSFETs including quantum effects: The role of crystallographic orientation*, Solid-State Electronics, 67 (2012), 30-37
- [89] Y. Zhang, Z. Jin, G. Wang et al.: *A quantum mechanical mobility model for scaled NMOS transistors with ultra-thin high-k dielectrics and metal gate electrodes*, International Semiconductor Device Research Symposium ISDRS, (2007), 1-2
- [90] G. Lixin, J. G. Fossum, F. Gamiz: *Mobility enhancement via volume inversion in double-gate MOSFETs*, IEEE International SOI Conference, (2003), 153-154
- [91] H. Matsuoka, T. Ichiguchi, T. Yoshimura et al.: *Mobility modulation in a quasi-one-dimensional Si-MOSFET with a dual-gate structure*, IEEE Electron Device Letters, 13 (1992), 20-22

- [92] M. F. Ng, M. B. Sullivan, S. W. Tong et al.: *First-principles study of silicon nanowire approaching the bulk limit*, Nano Letters, 11 (2011), 4794–4799
- [93] S. S. Iyer, Y. H. Xie: *Light emission from silicon*, SCIENCE, 260 (1993), 40-46
- [94] M. Fujita, Y. Tanaka, S. Noda: *Light emission from silicon in photonic crystal nanocavity*, Journal of Selected Topics in Quantum Electronics, 14 (2008), 1090-1097
- [95] Y. Wang, R. Han, X. Liu et al.: *The challenges for physical limitations in Si microelectronics*, International Conference on Solid-State and Integrated Circuit Technology, (1998), 25-30
- [96] S. E. Thompson, R. S. Chau, T. Ghani et al.: *In search of “forever” continued transistor scaling one new material at a time*, Transactions on Semiconductor Manufacturing, 18 (2005), 26-36
- [97] S. Thompson, N. Anand, M. Armstrong et al.: *A 90 nm logic technology featuring 50 nm strained silicon channel transistors, 7 layers of Cu interconnects, low k ILD, and 1 μm^2 SRAM cell*, IEEE International Electron Devices Meeting, (2002), 61-64
- [98] G. G. Shahidi: *SOI technology for the GHz era*, International Symposium on VLSI Technology Systems and Applications, (2001), 11-14
- [99] H. S. Kang, Y. W. Kim, K. S. Chung et al.: *Highly stable SOI technology to suppress floating body effect for high performance CMOS device*, IEEE International Electron Devices Meeting, (2001), 11.2.1-11.2.4
- [100] M. Horstmann, M. Wiatr, A. Wei et al.: *Advanced SOI CMOS transistor technology for high performance microprocessors*, International Conference on Ultimate Integration of Silicon, (2009), 11-14
- [101] P. Timans, W. Lerch, J. Niess et al.: *Challenges for ultra-shallow junction formation technologies beyond the 90 nm node*, IEEE International Conference on Advanced Thermal Processing of Semiconductors, (2003), 17-33
- [102] T. Kelwing, A. Naumann, M. Trentzsch et al.: *Impact of nitrogen post deposition annealing on hafnium zirconate dielectrics for 32 nm high- performance SOI CMOS technologies*, Microelectronic Engineering, 88 (2011), 141-144
- [103] P. Bai, C. Auth, S. Balakrishnan et al.: *A 65nm logic technology featuring 35nm gate lengths, enhanced channel strain, 8 Cu interconnect layers, low-k ILD and 0.57 μm^2 SRAM cell*, IEEE International Electron Devices Meeting, (2004), 657-660
- [104] M. Cassé, L. Thevenod, B. Guillaumot et al.: *Carrier transport in HfO₂/metal gate MOSFETs: Physical insight into critical parameters*, IEEE Transactions on Electron Devices, 53 (2006), 759-768
- [105] G. S. Lujan, S. Kubicek, S. De Gendt et al.: *Mobility degradation in high-k transistors: The role of the charge scattering*, Conference on European Solid-State Device Research ESSDERC, (2003), 399-402
- [106] J. Pan, C. Woo, C. Y. Yang et al.: *Replacement metal-gate NMOSFETs with ALD TaN/EP-Cu, PVD Ta, and PVD TaN electrode*, IEEE Electron Device Letters, 24 (2003), 304-305
- [107] S. C. Song, Z. Zhang, C. Huffman et al.: *Highly manufacturable advanced gate-stack technology for sub-45-nm self-aligned gate-first CMOSFETs*, IEEE Transactions on Electron Devices, 53 (2006), 979-989

- [108] S. E. Thompson: *Power, cost and circuit IP reuse: The real limiter to Moore's law over the next 10 years*, International Symposium on VLSI Technology Systems and Applications VLSI-TSA, (2010), 88-89
- [109] W. Y. Lu, Y. Taur: *On the scaling limit of ultrathin SOI MOSFETs*, IEEE Transactions on Electron Devices, 53 (2006), 1137-1141
- [110] C. Shin, B. Nikolić, T.-J. K. Liu: *Tri-gate bulk CMOS technology for improved SRAM scalability*, European Solid-State Device Research Conference ESSDERC, (2010), 142-145
- [111] Heise-Online: *Intels nächster Herstellungsprozess erobert die dritte Dimension*, <http://www.heise.de/newsticker/meldung/Intels-naechster-Herstellungsprozess-erobert-die-dritte-Dimension-1238246.html>, (05.05.2011)
- [112] D.-I. Moon, S.-J. Choi, J. P. Duarte et al.: *Investigation of silicon nanowire gate-all-around junctionless transistors built on a bulk substrate*, IEEE Transaction on Electron Devices, 60 (2013), 1355-1360
- [113] X. Sun, V. Moroz, N. Damrongplasit et al.: *Variation study of the planar ground-plane bulk MOSFET, SOI FinFET, and trigate bulk MOSFET designs*, IEEE Transactions on Electron Devices, 58 (2011), 3294-3299
- [114] A. D. Franklin, A. Lin, H.-S. P. Wong et al.: *Current scaling in aligned carbon nanotube array transistors with local bottom gating*, IEEE Electron Device Letters, 31 (2010), 644-646
- [115] K. Kalna, L. Yang, A. Asenov: *High performance III-V MOSFETs: A dream close to reality?*, IEEE International Symposium on Electron Devices for Microwave and Optoelectronic Applications, (2002), 243-248
- [116] P. Avouris, J. Appenzeller, V. Derycke et al.: *Carbon nanotube electronics*, Chemical Physics, 281 (2002), 429-445
- [117] Synopsys: *Sentaurus Device User Guide*, Version G-2012.06, (2012)
- [118] Synopsys: *Sentaurus Process User Guide*, Version G-2012.06, (2012)
- [119] R. L. Puurunen: *Surface chemistry of atomic layer deposition: A case study for the trimethylaluminum/water process*, Journal of Applied Physics, 97 (2005), 121301 - 121301-52
- [120] J. P. Biersack: *Basic physical aspects of high energy implantation*, Nuclear Instruments and Methods in Physics Research, 35 (1988), 205-214
- [121] G. Hobler, S. Selberherr: *Two-dimensional modeling of ion implantation induced point defects*, IEEE Transactions on Computer-Aided Design, 7 (1988), 174-180
- [122] A. Fick: *Über Diffusion*, Annalen der Physik und Chemie von Poggendorf, 94 (1855), 59-86
- [123] W. Klix: *Numerische Simulation elektrischer Bauelemente*, Habilitation, Technische Universität Dresden, 2004
- [124] C. Moglestue: *Monte Carlo Simulation of Semiconductor Devices*, Chapman & Hall, 1993
- [125] D. B. Klaassen, J. W. Slotboom, H. C. de Graaff: *Unified apparent bandgap narrowing in n- and p-type silicon*, Solid-State Electronics, 35 (1992), 125-129
- [126] J. W. Slotboom, H. C. de Graaff: *Measurements of bandgap narrowing in Si bipolar transistors*, Solid-State Electronics, 19 (1976), 857-862

- [127] W. Shockley, W. T. Read: *Statistics of the recombinations of holes and electrons*, Physical Review, 87 (1952), 835-842
- [128] A. Schenk: *A model for the field and temperature dependence of Shockley-Read-Hall lifetimes in silicon*, Solid-State Electronics, 35 (1992), 1585-1596
- [129] W. Maes, K. De Meyer, R. Van Overstraeten: *Impact ionization in silicon: A review and update*, Solid-State Electronics, 33 (1990), 705-718
- [130] D. B. Klaassen: *A unified mobility model for device simulation - model equations and concentration dependence*, Solid-State Electronics, 35 (1992), 953-959
- [131] D. M. Caughey, R. E. Thomas: *Carrier mobilities in silicon empirically related to doping and field*, Proceedings of the IEEE, 52 (1967), 2192-2193
- [132] B. Meinerzhagen, W. Engl: *The influence of the thermal equilibrium approximation on the accuracy of classical two-dimensional numerical modeling of silicon submicrometer MOS transistors*, IEEE Transactions on Electron Devices, 35 (1988), 689-697
- [133] C. Lombardi, S. Manzini, A. Saporito et al.: *A physically based mobility model for numerical simulation of nonplanar devices*, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 7 (1988), 1164-1171
- [134] M. Yang, E. Gusev, M. Jeong et al.: *Performance dependence of CMOS on silicon substrate orientation for ultrathin oxynitride and HfO₂ gate dielectrics*, IEEE Electron Device Letters, 24 (2003), 339-341
- [135] M. Darwish, J. Lentz, M. Pinto et al.: *An improved electron and hole mobility model for general purpose device simulation*, IEEE Transactions on Electron Devices, 44 (1997), 1529-1538
- [136] H. Tanimoto, M. Kondo, T. Enda et al.: *Modeling of electron mobility degradation for HfSiON MISFETs*, International Conference on Simulation of Semiconductor Processes and Devices, (2006), 47-50
- [137] D. Esseni: *Modeling of electron mobility degradation by remote coulomb scattering in ultrathin oxide MOSFETs*, IEEE Transactions on Electron Devices, 50 (2003), 1665-1674
- [138] J. J. Wortman, R. A. Evans: *Young's modulus, shear modulus, and poisson's ratio in silicon and germanium*, Journal of Applied Physics, 36 (1965), 153-156
- [139] J. L. Egley, D. Chidambarrao: *Strain effects on device characteristics: Implementation in drift-diffusion simulators*, Solid-State Electronics, 36 (1993), 1653-1664
- [140] S. Dhar, H. Kosina, V. Palankovski et al.: *Electron mobility model for strained-Si devices*, IEEE Transactions on Electron Devices, 52 (2005), 527-533
- [141] J. Höntschel: *Simulation und Optimierung komplexer nanoelektronischer Strukturen*, Dissertation, Technische Universität Dresden, 2004
- [142] T. Feudel, M. Horstmann, M. Gerhardt et al.: *Temperature scaling for 35nm gate length high-performance CMOS*, Materials Science in Semiconductor Processing, 7 (2004), 369-374
- [143] Y. Kikuchi, Y. Tateshita, T. Kataoka et al.: *A planar transistor for the 32-nm node and beyond with an ultra-shallow junction fabricated using in-situ doped selective Si epitaxy*, IEEE Solid-State Device Research Conference, (2006), 81-84

- [144] C. Y. Kang, J.-W. Yang, J. Oh et al.: *Effects of film stress modulation using TiN metal gate on stress engineering and its impact on device characteristics in metal gate/high-k dielectric SOI FinFETs*, IEEE Electron Device Letters, 29 (2008), 487-490
- [145] W. Wang, S. Chang, J. Huang et al.: *3D simulations of width effect on performance in NMOSFETs with SiC S/D stressors and CSEL linear*, Electron Devices and Solid-State Circuits, (2007), 1071-1074
- [146] Z. Ren, G. Pei, J. Li et al.: *On implementation of embedded phosphorus-doped SiC stressors in SOI nMOSFETs*, Symposium on VLSI Technology, (2008), 172-173
- [147] S. Mujumdar, K. Maitra, S. Datta: *Layout-dependent strain optimization for p-channel trigate transistors*, IEEE Transactions on Electron Devices, 59 (2012), 72-78
- [148] C. Jan, U. Bhattacharya, R. Brain et al.: *A 22nm SoC platform technology featuring 3-D tri-gate and high-k-metal gate optimized for ultra low power, high performance and high density SoC applications*, IEEE International Electron Devices Meeting, (2012), 44-47
- [149] A. Asenov, A. R. Brown, J. H. Davies et al.: *Simulation of intrinsic parameter fluctuations in decananometer and nanometer-scale MOSFETs*, IEEE Transaction on Electron Devices, 50 (2003), 1837-1852
- [150] T. Baldauf, A. Wei, R. Illgen et al.: *Study of 22/20nm Tri-Gate transistors compatible in a low-cost hybrid FinFET/planar CMOS process*, International Semiconductor Device Research Symposium (ISDRS), (2011), 1-2
- [151] C. -. Tsai, T.-J. Liu, S. H. Tsai et al.: *Segmented tri-gate bulk CMOS technology for device variability improvement*, International Symposium on VLSI Technology Systems and Applications, (2010), 114-115
- [152] N. Xu, B. Ho, M. Choi et al.: *Effectiveness of stressors in aggressively scaled FinFETs*, IEEE Transactions on Electron Devices, 59 (2012), 1592-1598
- [153] B. Doyle, B. Boyanov, S. Datta et al.: *Tri-Gate fully-depleted CMOS transistors: fabrication, design and layout*, Symposium on VLSI Technology, (2003), 133-134
- [154] B. Ho, X. Sun, C. Shin et al.: *Design optimization of multigate bulk MOSFETs*, IEEE Transactions on Electron Devices, 60 (2013), 28-33
- [155] J. Kavalieros, B. Doyle, S. Datta et al.: *Tri-Gate transistor architecture with high-k gate dielectrics, metal gates and strain engineering*, Symposium on VLSI Technology, (2006), 50-51
- [156] T. Baldauf, A. Wei, T. Herrmann et al.: *Suppression of the corner effects in a 22 nm hybrid Tri-Gate/planar process*, Semiconductor Conference Dresden (SCD), (2011), 1-4
- [157] T. Baldauf, A. Wei, R. Illgen et al.: *Simulation and optimization of Tri-Gates in a 22 nm hybrid Tri-Gate/planar process*, 12th International Conference on Ultimate Integration on Silicon (ULIS), (2011), 1-4
- [158] V. Jovanovic, T. Suligoj, L. K. Nanver: *Bulk-Si FinFET technology for ultra-high aspect-ratio devices*, ESSDERC'09, Proceedings of the European, (2009), 241-244
- [159] T. Baldauf, R. Stenzel, W. Klix et al.: *Strained isolation oxide as novel overall stress element for Tri-Gate transistors of 22nm CMOS and beyond*, International Semiconductor Conference Dresden-Grenoble (ISCDG), (2012), 61-63

- [160] C. Gallon, G. Reibold, G. Ghibaudo et al.: *Electrical analysis of mechanical stress induced by shallow trench isolation [MOSFETs]*, Conference on European Solid-State Device Research, 33 (2003), 359-362
- [161] C. Kang, R. Choi, S. Song et al.: *A novel electrode-induced strain engineering for high performance SOI FinFET utilizing Si channel for both N and PMOSFETs*, IEEE International Electron Devices Meeting, (2006), 1-4
- [162] M. Shrivastava, M. S. Baghini, D. K. Sharma et al.: *A novel bottom spacer FinFET structure for improved short-channel, power-delay and thermal performance*, IEEE Transaction on Electron Devices, 57 (2010), 1287-1294
- [163] S. Bedell, N. Daval, A. Khakifirooz et al.: *New opportunities for SiGe and Ge channel p-FETs*, Microelectronic Engineering, 88 (2011), 324–330
- [164] D. Lizzit, P. Palestri, D. Esseni et al.: *Analysis of the performance of n-Type FinFETs with strained SiGe channel*, IEEE Transaction on Electron Devices, 60 (2013), 1884-1891
- [165] L. Chang, M. Jeong, M. Yang: *CMOS Circuit Performance Enhancement by Surface Orientation Optimization*, IEEE Transactions on Electron Devices, 51 (2004), 1621-1627
- [166] I. Tienda-Luna, F. Ruiz, A. Godoy et al.: *Influence of orientation, geometry, and strain on electron distribution in silicon gate-all-around (GAA) MOSFETs*, IEEE Transactions on Electron Devices, 58 (2011), 3350-3357
- [167] M. Yang, V. W. Chan, K. K. Chan et al.: *Hybrid-orientation technology (HOT) - opportunities and challenges*, IEEE Transaction on Electron Devices, 53 (2006), 965-978
- [168] M. Yang, M. Jeong, L. Shi et al.: *High performance CMOS fabricated on hybrid substrate with different crystal orientations*, IEEE International Electron Devices Meeting, (2003), 18.7.1-18.7.4
- [169] Synopsys: *Process and device simulation of partially depleted SOI MOSFET*, Anwendungsbeispiel, Version D-2010.03
- [170] P. Flatresse, G. Cesana, X. Cauchy: *Planar fully depleted silicon technology to design competitive SoCs at 28nm and beyond*, White paper, STMicroelectronics and Soitec (2012)
- [171] M. Shrivastava, R. Mehta, S. Gupta et al.: *Toward system on chip (SoC) development using FinFET technology: challenges, solutions, process co-development & optimization guidelines*, IEEE Transactions on Electron Devices, 58 (2011), 1597-1607
- [172] R. Illgen: *Neuartige Ausheilverfahren in der SOI-CMOSFET-Technologie*, Dissertation, Technische Universität Dresden, 2011
- [173] B. Yang, K. Buddharaju, S.-G. Teo et al.: *Vertical sSilicon-nNanowire formation and gate-all-around MOSFET*, IEEE Electron Device Letters, 29 (2008), 791-794
- [174] R. Gandhi, Z. Chen, N. Singh et al.: *CMOS-compatible vertical-silicon-nanowire gate-all-around p-Type tunneling FETs with ≤ 50 mV/decade subthreshold swing*, IEEE Electron Device Letters, 32 (2011), 1504-1506
- [175] R. Gandhi, Z. Chen, N. Singh et al.: *Vertical Si-nanowire n-type tunneling FETs with low subthreshold swing (≤ 50 mV/decade) at room temperature*, IEEE Electron Device Letters, 32 (2011), 437-439

- [176] E. Steinslanda, M. Nese, A. Hanneborg et al.: *Boron etch-stop in TMAH solutions*, Sensors and actuators A: Physical, 54 (1996), 728-732
- [177] E. Gnani, S. Reggiani, M. Rudan et al.: *Effects of high-k (HfO₂) gate dielectrics in double-gate and cylindrical-nanowire FETs scaled to the ultimate technology nodes*, IEEE Transactions on Nanotechnology, 6 (2007), 90-96
- [178] J. Goldberger, A. I. Hochbaum, R. Fan et al.: *Silicon Vertically Integrated Nanowire Field Effect Transistors*, Nano Letters, 6 (2006), 973-977
- [179] W. H. Arnold: *Lithography for the 14nm Node and Beyond*, IEEE International Electron Devices Meeting, Short Course (2011)
- [180] L. E. Greene, M. Law, D. H. Tan et al.: *General Route to Vertical ZnO Nanowire Arrays Using Textured ZnO Seeds*, Nano Letters, 5 (2005), 1231-1236
- [181] S. Maheshwaram, S. K. Manhas, G. Kaushal et al.: *Vertical silicon nanowire gate-all-around field effect transistor based nanoscale CMOS*, IEEE Electron Device Letters, 32 (2011), 1011-1013
- [182] Y. Liu, R. Huang, R. Wang et al.: *Design optimization for digital circuits built with gate-all-around silicon nanowire transistors*, IEEE Transactions on Electron Devices, 59 (2012), 1844-1850
- [183] J. Hoentschel, A. Wei, M. Wiatr et al.: *Implementation and optimization of asymmetric transistors in advanced SOI CMOS technologies for high performance microprocessors*, IEEE International Electron Devices Meeting, (2008), 1-4
- [184] H. A. William: *Lithography for the 14nm node and beyond*, IEEE Short Course International Electron Devices Meeting, (2011)
- [185] Y. Lee, K. Kakushima, K. Natori et al.: *Gate capacitance modeling and diameter-dependent performance of nanowire MOSFETs*, IEEE Transactions on Electron Devices, 59 (2012), 1037-1045
- [186] A. Asenov: *Simulation of statistical variability in nano MOSFETs*, Symposium on VLSI Technology Digest of Technical, (2007), 86-87
- [187] W. Shockley, J. A. Copeland, R. P. James: *The impedance field method of noise calculation in active semiconductor devices*, Quantum Theory of Atoms, Molecules and the Solid-State, P.-O. Lowdin, Ed. New York: Academic, (1966)
- [188] F. Bonani, G. Ghione, M. Pinto et al.: *An efficient approach to noise analysis through multidimensional physics-based models*, IEEE Transaction on Electron Devices, 45 (1998), 261-269
- [189] H.-J. Bartsch: *Taschenbuch Mathematischer Formeln*, 21. Aufl., Carl Hanser Verlag, München, 2007
- [190] N. Henze: *Stochastik für Einsteiger*, 7. Aufl., Vieweg Verlag, Wiesbaden, 2008
- [191] V. F. Synopsis: *Modeling random dopant fluctuation effects in MOSFETs using impedance field methods*, Anwendungsbeispiel, Version F-2011.09 (2011)
- [192] A. W. al.: *Random dopant fluctuation modelling with the impedance field method*, International Conference on Simulation of Semiconductor Processes and Devices, SISPAD, (2003), 91-94

- [193] X. Wang, G. Roy, O. Saxod et al.: *Simulation study of dominant statistical variability sources in 32-nm high- κ /metal gate CMOS*, IEEE Electron Device Letters, 33 (2012), 643-645
- [194] E. Baravelli, A. Dixit, R. Rooyackers et al.: *Impact of line-edge roughness on FinFET matching performance*, IEEE Transactions on Electron Devices, 54 (2007), 2466-2474
- [195] T. Mizuno, J. Okamura, A. Toriumi: *Experimental study of threshold voltage fluctuation due to statistical variation of channel dopant number in MOSFET's*, IEEE Transaction on Electron Devices, 41 (1994), 2216-2221
- [196] A. Heinzig, S. Slesazeck, F. Kreupl et al.: *Reconfigurable silicon nanowire transistors*, NanoLetters, 12 (2012), 119-124

Danksagung

Viele Menschen haben mich während der vier Jahre meiner Promotion in verschiedener Art und Weise unterstützt und trugen damit indirekt einen erheblichen Anteil zur Vollendung dieser Arbeit bei.

Mein Dank gehört vor allem meinem Doktorvater Prof. Dr.-Ing. habil. Gerald Gerlach, der die Betreuung seitens der Technischen Universität Dresden übernommen und mir die Promotion somit erst ermöglicht hat.

Großen Dank möchte ich auch meinem Betreuer Prof. Dr.-Ing. habil. Roland Stenzel von der Hochschule für Technik und Wirtschaft Dresden aussprechen, auf dessen Initiative hin die Kooperation mit GLOBALFOUNDRIES Dresden Module One LLC & Co. KG entstanden ist, welche die wesentliche Grundlage meiner Promotion darstellte. Neben seinen umfangreichen Verpflichtungen als Rektor der Hochschule nahm er sich oft Zeit für mich und lenkte meine Forschungsarbeit durch fundiertes Fachwissen sowie gezielte Fragen stets in die richtige Richtung.

Ich möchte mich auch bei Prof. Dr.-Ing. habil. Wilfried Klix (HTW-Dresden) sowie den Mitarbeitern des Device-Teams von GLOBALFOUNDRIES Dr. Thomas Feudel, Dr. Andy Wei, Dr. Jan Höntschel und dem Abteilungsleiter Dr. Manfred Horstmann für ihre kompetente Hilfe bei fachlichen Problemen bedanken. Zu dem Device-Team zählen nun auch meine ehemaligen Promotionskollegen Dr. Stefan Flachowsky, Dr. Tom Herrmann und Dr. Ralf Illgen, denen ich für ihre wertvollen Ratschläge und die zahlreichen gemeinsamen Mittagspausen sehr dankbar bin.

Weiteren Dank verdienen die Doktoranden Sören Miersch, Sebastian Günther und Martin Eckart (HTW-Dresden) für die vielen gemeinsamen Stunden der Vorbereitung auf die Zugangsprüfungen der TU-Dresden sowie André Heintzig (NaMLab) für seinen fachlichen Rat und die schönen, manchmal endlosen Nächte am Kickertisch.

Ina und Sebastian, euch danke ich aufrichtig für eure bedingungslose Unterstützung und die vielen Stunden eurer Freizeit, die ihr mir geschenkt habt. Ihr sorgtet stets für die nötige geistige Ablenkung und hattet immer ein offenes Ohr für meine Probleme.

Schließlich möchte ich mich ganz herzlich bei meinen Eltern und meiner Familie bedanken. Ihr habt mir all die nötigen Freiräume sowie eure Unterstützung gegeben, um meinen eigenen Weg zu finden und ihn auch gehen zu können.

Keineswegs vergessen möchte ich aber auch all meine hier nicht namentlich aufgeführten Freunde, Kollegen und Volleyballer, die mein Leben rundum bereichert haben.

Danke!

Acknowledgement

Die Projekte "CoolTrans/STRESSOREN", Projektnummer 13579/2323, und „CoolComputing II“, Projektnummer 5200193130, welche der Veröffentlichung zu Grunde liegen, wurden im Einklang mit der Technologischen Förderung für regionale Entwicklung der Europäischen Union (EFRE) und der sächsischen Aufbaubank finanziert. Der Autor der Veröffentlichung trägt die Verantwortung für alle veröffentlichten Inhalte.

The projects “CoolTrans/STRESSOREN”, Project Number 13579/2323, and “CoolComputing II”, Project Number 5200193130, described in this publication were funded in line with the technology funding for regional development (ERDF) of the European Union and by funds of the Free State of Saxony. The author named in the publication bears responsibility for all published contents.

