Technische Universität Dresden

NEUARTIGE AUSHEILVERFAHREN IN DER SOI-CMOSFET-TECHNOLOGIE

Ralf Illgen

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. et Ing. habil. Elfgard Kühnicke		
Gutachter:	Prof. Dr.-Ing. habil. Gerald Gerlach	Tag der Einreichung:	06.12.2010
	Prof. Dr.-Ing. habil. Roland Stenzel	Tag der Verteidigung:	20.05.2011

Kurzfassung

Thermische Ausheilprozesse werden bei der Transistorformation im Wesentlichen eingesetzt, um die durch die Ionenimplantation entstandenen Kristallschäden auszuheilen und die eingebrachten Dotanden zu aktivieren. Besonders kritisch sind dabei die finalen Aktivierungsprozesse, bei denen die Source/Drain-Gebiete der Transistoren gebildet werden. Im Zuge der kontinuierlichen Skalierung der CMOSFET-Technologie ist es außerdem erforderlich, möglichst flache, abrupte Dotierungsprofile mit maximaler elektrischer Aktivierung zu erhalten, um die bei diesen Bauelementeabmessungen immer stärker auftretenden Kurzkanaleffekte zu unterdrücken und gleichzeitig eine höhere Leistungsfähigkeit der Transistoren zu gewährleisten. Zur maximalen Aktivierung bei minimaler Diffusion der eingebrachten Dotanden müssen dazu während der finalen Ausheilung extrem kurze Ausheilzeiten bei sehr hohen Temperaturen bewerkstelligt werden. Mit dem derzeitig angewandten Ausheilverfahren, der schnellen thermischen Ausheilung (RTA), bei der die minimale Ausheilzeit im Bereich von 1 s liegt, sind diese Vorgaben nicht mehr realisierbar. Nur durch den Einsatz von neuartigen thermischen Ausheilprozessen mit Ausheilzeiten im Millisekundenbereich können diese Forderungen erreicht werden.

Das Thema der vorliegenden Arbeit ist die wissenschaftliche Untersuchung der neuartigen Ausheilprozesse und die experimentelle Realisierung von Integrationsmöglichkeiten in die planare Hochleistungs-SOI-CMOSFET-Technologie.

Dazu wird zunächst die Notwendigkeit der Einführung der neuartigen Ausheilprozesse erläutert. Anschließend wird basierend auf experimentellen Untersuchungen der Einfluss der Kurzzeitausheilung auf die Diffusion und Aktivierung der Dotierstoffe für eine p- und n-Dotierung analysiert. Des Weiteren werden zwei unterschiedliche Technologien der Kurzzeitausheilung, die Blitzlampen- und Laser-Ausheilung, und deren Einfluss auf das Transistorverhalten sowohl auf Wafer- als auch auf Mikroprozessorebene untersucht. Der Schwerpunkt der vorliegenden Arbeit liegt auf der experimentellen Untersuchung zur Integration der Kurzzeitausheilung in den Herstellungsprozess von Hochleistungs-SOI-CMOSFETs. Zwei verschiedene Ansätze werden dabei näher betrachtet. Zum Einen wird der Einfluss der Kurzzeitausheilung als zusätzlicher Ausheilschritt im Anschluss an die herkömmliche RTA und zum Anderen als alleiniger Ausheilschritt ohne RTA untersucht. Die Ergebnisse der durchgeführten Experimente zeigen, dass durch die zusätzliche Kurzzeitausheilung nach Ansatz 1 ohne eine Veränderung des Herstellungsprozesses ein verbessertes Transistorverhalten erreicht werden kann. Demgegenüber ist die Integration der Kurzzeitausheilung nach Ansatz 2 nur durch eine Anpassung der Transistorarchitektur und eine Optimierung der Implantationsparameter für die Halo-, Source/Drain-Erweiterungs- und Source/Drain-Gebiete möglich. Ein Hauptaugenmerk bei der Herstellung diffusionsarmer p-MOSFETs nach Ansatz 2 liegt in der Implementierung von $Si_{1-x}Ge_x$-Source/Drain-Gebieten, um die Erhöhung der Leistungsfähigkeit durch diese Verspannungsquelle auch bei diesen Transistortypen zu gewährleisten. Die dazu durchgeführten experimentellen Untersuchungen zeigen, dass bei diffusionsarmen p-MOSFETs mit $Si_{1-x}Ge_x$ in den Source/Drain-Gebieten des Transistors, die Wahl der richtigen Implantationsspezies von entscheidender Bedeutung ist. Abschließend erfolgt eine Gegenüberstellung der Ergebnisse von optimierten, diffusionsarmen n- und p-MOSFETs mit Transistoren der 45 nm-Technologie. Letztere basieren auf einem Prozess mit einer kombinierten Ausheilung von RTA und Kurzzeitausheilung. Dabei wird gezeigt, dass im Gegensatz zur herkömmlichen RTA-Ausheilung eine weitere Miniaturisierung der planaren Transistorstruktur mit Hilfe der Kurzzeitausheilung möglich ist.

Abstract

In a standard CMOSFET manufacturing process thermal annealing processes are basically used for defect annealing and dopant activation. Very critical is here the final annealing of the Source/Drain regions. Furthermore, in aggressively scaled CMOSFETs, very shallow pn-junctions and low sheet resistances are essential for short-channel effect control and high device performance. High annealing temperatures with very short annealing times are therefore necessary for high activation with less diffusion of the implanted dopants. The conventional rapid thermal annealing (RTA) with minimal annealing times of around 1 s does not meet these requirements due to thermal diffusion and solid solubility limitations. To overcome the problem, high temperature less-diffusive ultra fast annealing technologies with annealing times in the millisecond range have been intensively investigated. Although the millisecond anneals offers a number of advantages over RTA, there remain some challenges for successful implementation of the millisecond anneal for CMOSFET fabrication.

The purpose of this thesis is the scientific investigation of the new ultra fast annealing technologies as well as the implementation of these annealing schemes in the planar SOI-CMOSFET technology.

First, the needs to implement the ultra fast annealing process will be given. Subsequently, based on experimental research the impact of the millisecond annealing on diffusion and activation of p- and n-type dopants are analyzed. Furthermore, two different ultra fast annealing technologies, the flash and laser annealing, and their impact on transistor performance on wafer as well as on die level will be investigated. The main focus of this work lies on the integration of the new annealing technologies in a manufacturing process of high performance planar SOI-CMOSFETs. Two different approaches will be considered in more detail. First, the impact of the ultra fast annealing in combination with the conventional RTA, and second, the substitution of the RTA with the new annealing process will be investigated with the help of numerical simulations, material analysis and electrical measurements. The experimental results obtained show that according to the first approach the additional millisecond annealing improves the transistor performance without any changes in the manufacturing process. On the other hand the replacement of the RTA with the ultra fast annealing technology is only possible with an adjustment of the transistor architecture and an optimization of the halo, source/drain extension as well as deep source/drain implantations. The main focus of attention on the fabrication of diffusionless p-MOSFETs corresponding to the second approach lies on the implementation of strained $Si_{1-x}Ge_x$ in the source/drain regions. The performed experimental studies show that low damage creating source/drain extension implantations are necessary to enhance the performance of diffusionless $Si_{1-x}Ge_x$ strained SOI p-MOSFETs. Finally, a comparison of optimized diffusionless n- and p-MOSFETs with transistors of the 45 nm technology node will be done. The latter ones are based on a combined annealing process of RTA followed by ultra fast annealing. It will be shown that in contrast to the conventional RTA a further scaling of the planar transistor architecture is possible with the help of the new ultra fast annealing technologies.

Inhaltsverzeichnis

1 Einleitung 1

2 Grundlagen der Transistorherstellung 5
2.1 Herstellungsprozess eines MOSFETs 5
2.2 Charakteristische Kenngrößen 8
2.2.1 Technologische und elektrische Kenngrößen 8
2.2.2 Ladungsträgerbeweglichkeit 11
2.2.3 Kapazitäten eines SOI-MOSFETs 12
2.2.4 Leistungsparameter für MOSFETs 13
2.3 Skalierungseffekte 13
2.3.1 Kurzkanaleffekte 14
2.3.2 Externer Widerstand 17
2.4 Implantation 20
2.5 Ausheilung 20
2.5.1 Rekristallisation und Aktivierung 21
2.5.2 Diffusion 23
2.5.3 Thermisches Budget 29
2.6 Zusammenfassung 29

3 Neuartige Ausheilverfahren 31
3.1 Vorbetrachtungen 31
3.2 Verwendete Ausheiltechnologien 34
3.2.1 Blitzlampen-Ausheilung 34
3.2.2 Laser-Ausheilung 35
3.3 Einfluss auf die Dotierstoffverteilung 37
3.3.1 p-Dotierung 37
3.3.2 n-Dotierung 39
3.3.3 Defektentwicklung 40
3.4 Vergleich der Ausheiltechnologien 40
3.5 Zusammenfassung 45

4 Modelle und Kalibrierung der Prozess- und Bauelementesimulation 47
4.1 Vorbetrachtungen 47
4.2 Prozesssimulation 49
4.2.1 Implantation 49
4.2.2 Ausheilung, Diffusion und Aktivierung 50
4.2.3 Mechanische Verspannung 51
4.3 Bauelementesimulation 54
4.3.1 Modellhierarchie 54
4.3.2 Modellierung des Ladungsträgertransports 55
4.3.3 Modellierung der mechanischen Verspannungen 61
4.4 Kalibrierung der Simulationsparameter 67
4.4.1 Prozesssimulation 67
4.4.2 Bauelementesimulation 70
4.5 Zusammenfassung 75

5 Kombination von Kurzzeitausheilung und konventioneller RTA-Ausheilung ... 79
5.1 Vorbetrachtungen ... 79
5.2 Einfluss der Kurzzeitausheilung auf elektrische Transistorparameter ... 80
5.2.1 Experimentelle Ergebnisse ... 80
5.2.2 Untersuchung des Gatedielektrikums ... 83
5.2.3 Ladungsträgerbeweglichkeit ... 88
5.3 Einfluss nachfolgender Prozessschritte ... 89
5.4 Einfluss auf verspannte Silizium-Germanium-Schichten ... 90
5.5 Variation der Ausheiltemperatur der konventionellen Ausheilung ... 95
5.6 Zusammenfassung ... 101

6 Diffusionsarmer Transistor ... 103
6.1 Vorbetrachtungen ... 103
6.2 Herstellung diffusionsarmer Transistoren ... 105
6.2.1 Erste Ergebnisse ... 105
6.2.2 Einfluss der SDE-Implantation ... 108
6.2.3 Diskussion ... 111
6.3 Variation der Implantationsparameter ... 115
6.3.1 Halo-Implantation ... 115
6.3.2 SDE-Implantation ... 118
6.3.3 SD-Implantation ... 122
6.3.4 Zusammenfassung ... 125
6.4 Einfluss zusätzlicher Ausheilprozesse ... 126
6.5 Einbau von verspannten Silizium-Germanium-Schichten ... 129
6.5.1 Vorbetrachtungen ... 129
6.5.2 Einfluss auf die Leistungsfähigkeit von p-MOSFETs ... 130
6.5.3 Einfluss der Implantationsspezies ... 132
6.5.4 Einfluss der Spacer0-Ätzung ... 139
6.5.5 *In situ* bordotiertes Silizium-Germanium ... 142
6.5.6 Zusammenfassung ... 148
6.6 Einbau von verspannten Silizium-Kohlenstoff-Schichten ... 148
6.6.1 Vorbetrachtungen ... 148
6.6.2 Experimentelle Ergebnisse ... 150
6.7 Zusammenfassung ... 153

7 Zusammenfassung und Ausblick ... 157
7.1 Zusammenfassung ... 157
7.2 Ausblick ... 162

Literaturverzeichnis ... 165

Nachwort ... 175

Lebenslauf ... 177

Verzeichnis der verwendeten Abkürzungen und Formelzeichen

Abkürzungen

Begriff	Erläuterung
4PP	Vierspitzenmethode (Four-Point-Probe)
a.u.	Arbitrary Unit
A	Fremdatom (Impurity)
AI	Fremdatom-Zwischengitteratom-Paar (Impurity Interstitial Pair)
As	Arsen
AsV	Arsen-Leerstellen-Komplexe (Arsenic Vacancy Cluster)
AV	Fremdatom-Leerstelle-Paar (Impurity Vacancy Pair)
B	Bor
$B_{18}H_{22}$	Octadecaborane
BF_2	Bordifluorid
BOX	Vergrabenes Oxid (Buried Oxide)
BIC	Bor-Zwischengitteratom-Komplexe (Boron Interstitial Cluster)
C	Kohlenstoff
CMOS	Complementary Metal Oxide Semiconductor
CO_2	Kohlenstoffdioxid
CP	Charge Pumping
Die	Mikroprozessor
DIBL	Drain-induced Barrier Lowering
ECS	Electrochemical Society
EOR	End of Range
EOT	Electrical Oxide Thickness
ESSDERC	European Solid-State Device Research Conference
FCA	Freie-Ladungsträger-Absorption (Free Carrier Absorption)
FLA	Blitzlampen-Ausheilung (Flash Lamp Annealing)
Ge	Germanium
I	Silizium-Zwischengitteratom (Interstitial)
I^0, $I^{-/+}$, $I^{--/++}$	Neutrale, einfach und zweifach geladene Silizium-Zwischengitteratome
IEDM	International Electron Devices Meeting
IEEE	Institute of Electrical and Electronics Engineers
IIT	Ion Implantation Technology
IR	Infrarot
ISBD	*in situ* bordotiert

Begriff	Erläuterung
ISDRS	International Semiconductor Device Research Symposium
ISPD	*in situ* phosphordotiert
ITRS	International Technology Roadmap for Semiconductors
IWJT	International Workshop on Junction Technology
LSA	Laser-Ausheilung (Laser Annealing)
MOSFET	Metal Oxide Semiconductor Field Effect Transistor
MRS	Materials Research Society
MSA	Millisekundenausheilung
NT-RTA	Niedrigtemperatur-RTA
P	Phosphor
PAI	Prä-Amorphisierungsimplantation
PD	Teilweise verarmt (Partially Depleted)
RTA	Rapid Thermal Annealing
RTP	Advanced Thermal Processing of Semiconductors
Si	Silizium
SiC	Silizium-Kohlenstoff
SiGe	Silizium-Germanium
SD-Gebiet	Source/Drain-Gebiet
SDE-Gebiet	Source/Drain-Erweiterungsgebiet
SIMS	Sekundärionen-Massenspektroskopie
Si_3N_4	Siliziumnitrid
SiO_2	Siliziumdioxid
SiON	Siliziumoxinitrid
SOI	Silicon on Insulator
SPE	Festphasenepitaxie (Solid Phase Epitaxy)
SRH	SHOCKLEY-READ-HALL
SSDM	Solid-State Devices and Materials
STI	Shallow Trench Isolation
TCAD	Technology Computer-Aided Design
TED	Transient Enhanced Diffusion
TEM	Transmission Electron Microscopy
UFA	Ultra Fast Annealing
ULIS	Ultimate Integration of Silicon
UV	Ultraviolett
V	Leerstelle (Vakanz, Vacancy)
V^0, $V^{-/+}$, $V^{--/++}$	Neutrale, einfach und zweifach geladene Leerstellen
VLSI	Very-Large-Scale Integration
Xe	Xenon
XRD	Röntgenbeugung (X-Ray Defraction)

Physikalische Konstanten

Größe	Bedeutung	Zahlenwert
a_{Si}	Gitterkonstante von Silizium	$0.543095 \cdot 10^{-9}$ m
c	Lichtgeschwindigkeit	299792458 $m \cdot s^{-1}$
h	PLANCKsches Wirkungsquantum	$6.6262 \cdot 10^{-34}$ J·s
$\hbar$	$\hbar = h/2\pi$	$1.0546 \cdot 10^{-34}$ J·s
k_B	BOLTZMANN-Konstante	$1.381 \cdot 10^{-23}$ $J \cdot K^{-1}$
m_0	Elektronenruhemasse	$9.11 \cdot 10^{31}$ kg
q	Elementarladung	$1.602 \cdot 10^{-19}$ A·s
ε_0	Absolute Permittivität	$8.85419 \cdot 10^{-12}$ $A \cdot s \cdot V^{-1} \cdot m^{-1}$
$\varepsilon_{r,Oxid}$	Relative Permittivität des Gateoxids	3.9 (SiO_2), 4.1 (SiON)

Physikalische Größen

Die Indizes n beziehungsweise p unterscheiden zwischen Elektronen und Löchern

Größe	Bedeutung	Maßeinheit
a	Gitterparameter	m
A_G	Physikalische bzw. geometrische Gatefläche	m²
c_p	Spezifische Wärmekapazität bei konstantem Druck	$J \cdot kg^{-1} \cdot K^{-1}$
C	Konzentration	m^{-3}
C_A	Konzentration eines Fremdatoms A	m^{-3}
C_{As}	Arsenkonzentration	m^{-3}
C_B	Borkonzentration	m^{-3}
C_{AI}	Konzentration eines Fremdatoms auf einem Gitterplatz mit einem angrenzenden Silizium-Zwischengitteratomen	m^{-3}
C_{AV}	Konzentration eines Fremdatoms auf einem Gitterplatz mit einer angrenzenden Leerstelle	m^{-3}
C_{BOX}	Kapazität zwischen Wanne und Substrat	F
C_{BOXSD}	Kapazität zwischen Source- bzw. Draingebiet und Substrat	F
C_{do}	Direkte Überlappungskapazität	F
C_G	Gatekapazität	F
C''_G	Flächenbezogene Gatekapazität	$F \cdot m^{-2}$
$C_{G,inv}$	Gatekapazität im Inversionsfall	F
$C''_{G,inv}$	Flächenbezogene Gatekapazität im Inversionsfall	$F \cdot m^{-2}$
C_I, C_I^*	Konzentrationen der Silizium-Zwischengitteratome außerhalb und im thermodynamischen Gleichgewicht	m^{-3}
C_{if}	Innere umsäumende Kapazität	F
C_j	Sperrschichtkapazität zwischen Source bzw. Drain und Wanne	F

Größe	Bedeutung	Maßeinheit
C_{of}	Äußere umsäumende Kapazität	F
C_{ov}	Überlappungs- bzw. MILLERkapazität	F
C_V , C_V^*	Konzentrationen der Leerstellen außerhalb und im thermodynamischen Gleichgewicht	m^{-3}
d_{absorp}	Absorptionslänge	m
d_{Wafer}	Waferdicke	m
$\boldsymbol{D}$	Diffusionskoeffizient	$m^2 \cdot s^{-1}$
$D^0, D^{-/+}, D^{--/++}$	Diffusionskoeffizienten für die Wechselwirkung der Dotierstoffe mit neutralen, mit einfach und zweifach geladenen Punktdefekten	$m^2 \cdot s^{-1}$
D_0	Diffusionskonstante	$m^2 \cdot s^{-1}$
d_{AI}	Diffusionsvermögen eines Fremdatoms über Silizium-Zwischengitteratome	$m^2 \cdot s^{-1}$
D_{AI}	Diffusionskoeffizient für die Diffusion eines Fremdatoms über Silizium-Zwischengitteratome	$m^2 \cdot s^{-1}$
d_{AV}	Diffusionsvermögen eines Fremdatoms über Leerstellen	$m^2 \cdot s^{-1}$
D_{AV}	Diffusionskoeffizient für die Diffusion eines Fremdatoms über Leerstellen	$m^2 \cdot s^{-1}$
D_{eff}^{A}	Effektiver Diffusionskoeffizient für ein Fremdatom unter intrinsischen Bedingungen	$m^2 \cdot s^{-1}$
$D_{eff,ext}^{A,n}$, $D_{eff,ext}^{A,p}$	Effektiver Diffusionskoeffizient für ein Fremdatom unter extrinsischen Bedingungen	$m^2 \cdot s^{-1}$
D_I	Diffusionskoeffizient für die Diffusion von Silizium-Zwischengitteratomen	$m^2 \cdot s^{-1}$
DDS	Abstand der SiGe-Gebiete zum Gate	m
DIBL	Drain-induced Barrier Lowering	$mV \cdot V^{-1}$
D_{it}	Grenzflächenhaftzentrendichte	$m^{-2} \cdot eV^{-1}$
D_V	Diffusionskoeffizient für die Diffusion von Leerstellen	$m^2 \cdot s^{-1}$
E	Energie	J
$\boldsymbol{E}$	Elektrische Feldstärke	$V \cdot m^{-1}$
E_a	Aktivierungsenergie	eV
E_C	Energie des Leitungsbands	eV
E_d	Verlagerungsenergie eines Gitteratoms	eV
E_g	Bandlücke	eV
E_{trap}	Energiedifferenz zwischen der Energie der Rekombinationszentren und der FERMI-Energie des undotierten, intrinsischen Halbleiters	eV
E_V	Energie des Valenzbands	eV
E_X	Elektronenaffinität	eV
f_P	Pulsfrequenz	s^{-1}
F	Kraft	N
G	Generationsstrom	$m^{-3} \cdot s^{-1}$

Größe	Bedeutung	Maßeinheit
H_G	Physikalische bzw. geometrische Gatehöhe	m
I	Intensität	$W \cdot m^{-2}$
I_{CP}	Charge-Pumping-Strom	A
I_D	Drainstrom	A
I_{D0}	Intrinsischer Drainstrom	A
$I_{D,off}$	Sperrstrom	A
$I_{D,lin}$	Linearer Drainstrom	A
$I_{D,sat}$	Sättigungs-Drainstrom	A
J_n, J_p	Ladungsträgerstromdichte	$A \cdot m^{-2}$
J_A	Teilchenstrom der Fremdatome	$m^{-2} \cdot s^{-1}$
J_I	Teilchenstrom der Silizium-Zwischengitteratome	$m^{-2} \cdot s^{-1}$
J_T	Teilchenstrom	$m^{-2} \cdot s^{-1}$
J_V	Teilchenstrom der Leerstellen	$m^{-2} \cdot s^{-1}$
k	Wellenvektor	m^{-1}
L_D	Diffusionslänge	m
$L_{D,heat}$	Thermische Diffusionslänge	m
$L_{D,eff}$	Effektive Diffusionslänge	m
L_G	Physikalische bzw. geometrische Gatelänge	m
L_{Laser}	Laserstrahllänge	m
L_{met}	Lateraler Abstand der Erweiterungsgebiete zwischen Source und Drain	m
L_{PP}	Poly-Pitch, Abstand zweier benachbarter Polysiliziumbahnen	m
m_n^{eff} , m_p^{eff}	Effektive Ladungsträgermasse	kg
m_n^l, m_p^l	Longitudinale Ladungsträgermasse	kg
m_n^t, m_p^t	Transversale Ladungsträgermasse	kg
n	Elektronendichte	m^{-3}
n_B	Brechungsindex	1
n_i	Intrinsische Ladungsträgerkonzentration	m^{-3}
N_A	Akzeptorendichte	m^{-3}
N_D	Donatorendichte	m^{-3}
N_{sub}	Substrat- bzw. Wannendotierung	m^{-3}
p	Löcherdichte	m^{-3}
Q_f	Feste Oxidladungen	A·s
Q_{it}	Grenzflächenhaftzentren	A·s
Q_{Kanal}	Kanalladung	A·s
Q''_{Kanal}	Flächenbezogene Kanalladung	$A \cdot s \cdot m^{-2}$
Q_m	Bewegliche Ionenladungen im Gateoxid	A·s
Q_{ot}	Haftzentren im Gateoxid	A·s

Größe	Bedeutung	Maßeinheit
Q_{ox}	Oxidladungen	A·s
R	Rekombinationsrate	$m^{-3} \cdot s^{-1}$
R_{ext}	Externer Widerstand des Transistors	Ω
R_{Kanal}	Intrinsischer Widerstand bzw. Kanalwiderstand des Transistors	Ω
$R_{Kontakt}$	Widerstand der Kontaktbahnen	Ω
R_p	Projizierte Reichweite (Erwartungswert)	m
R_{Poly}	Widerstand einer Polysiliziumstruktur	Ω
$R_{Schicht}$	Schichtwiderstand	Ω
R_{SD}	Bahnwiderstand der SD-Gebiete	Ω
R_{SDE}	Bahnwiderstand der SDE-Gebiete	Ω
$R_{Silizid}$	Übergangswiderstand vom Silizium zum Silizid	Ω
R_{SRH}	SHOCKLEY-READ-HALL-Rekombinationsrate	$m^{-3} \cdot s^{-1}$
R_{Total}	Gesamtwiderstand des Transistors	Ω
S	Unterschwellsteigung	V/Dekade
S_n, S_p	Ladungsträgerenergiestromdichte	$W \cdot m^{-2}$
Si_{etch}	Tiefe der SiGe-Mulden	m
t	Zeit	s
$t_{Ausheil}$	Ausheilzeit	s
t_{Halte}	Haltezeit	s
T	Temperatur	K
$T_{Ausheil}$	Ausheiltemperatur	K
T_B	Thermisches Budget	K·s
T_L	Temperatur des Kristallgitters	K
T_{max}	Maximaltemperatur	K
T_n, T_p	Ladungsträgertemperatur	K
T_{ox}	Physikalische Dicke des Gatedielektrikums	m
$T_{ox,inv}$	Elektrische Gateoxiddicke im Inversionsfall	m
T_{Wafer}	Wafertemperatur	K
U_{DD}	Betriebsspannung	V
U_{DS}	Drain-Source-Spannung	V
U_{FB}	Flachbandspannung	V
U_{GS}	Gate-Source-Spannung	V
U_{ox}	Potenzialabfall über dem Gateoxid	V
U_{th}	Schwellspannung	V
$U_{th,lang}$	Schwellspannung von Langkanaltransistoren	V
$U_{th,lin}$	Schwellspannung im linearen Bereich	V
$U_{th,sat}$	Schwellspannung im Sättigungsbereich	V
v_n, v_p	Ladungsträgergeschwindigkeit	$m \cdot s^{-1}$
v_n^{sat} , v_p^{sat}	Sättigungsdriftgeschwindigkeit der Ladungsträger in Silizium	$m \cdot s^{-1}$
v_{Tisch}	Tischgeschwindigkeit	$m \cdot s^{-1}$

Größe	Bedeutung	Maßeinheit
W_G	Physikalische bzw. geometrische Gateweite	m
W_{Laser}	Laserstrahlweite	m
X_j	Tiefe des pn-Übergangs der SDE-Gebiete	m
X_d	Ausdehnung der Verarmungszone unter dem Gate	m
x_{Ge}	Germaniumgehalt im SiGe	%
Y_o	Distanz, wie weit die SDE-Gebiete unter das Gate reichen	m
α_{FCA}	Absorptionskoeffizient der Freien-Ladungsträger-Absorption	m^{-1}
ε	Dehnung	1
κ	Wärmeleitfähigkeit	$W \cdot m^{-1} \cdot K^{-1}$
κ_n, κ_p	Wärmeleitfähigkeit der Ladungsträger	$W \cdot m^{-1} \cdot K^{-1}$
λ	Wellenlänge	m
μ	Beweglichkeit	$m^2 \cdot V^{-1} \cdot s^{-1}$
μ_n, μ_p	Ladungsträgerbeweglichkeit	$m^2 \cdot V^{-1} \cdot s^{-1}$
$\mu_n^{C,T}$, $\mu_p^{C,T}$	Gittertemperatur- und dotierungsabhängige Ladungsträgerbeweglichkeit	$m^2 \cdot V^{-1} \cdot s^{-1}$
$\mu_n^{E,low}$, $\mu_p^{E,low}$	Ladungsträgerbeweglichkeit bei kleinen elektrischen Feldstärken in Stromflussrichtung	$m^2 \cdot V^{-1} \cdot s^{-1}$
μ_n^{HFS}, μ_p^{HFS}	Ladungsträgerbeweglichkeit bei hohen elektrischen Feldstärken in Stromflussrichtung	$m^2 \cdot V^{-1} \cdot s^{-1}$
μ_n^{SR}, μ_p^{SR}	Ladungsträgerbeweglichkeit an der Si/SiO_2-Grenzfläche bei vertikalen elektrischen Feldern	$m^2 \cdot V^{-1} \cdot s^{-1}$
μ_n^{str}, μ_p^{str}	Ladungsträgerbeweglichkeit mit Verspannung	$m^2 \cdot V^{-1} \cdot s^{-1}$
μ_n^{unstr}, μ_p^{unstr}	Ladungsträgerbeweglichkeit ohne Verspannung	$m^2 \cdot V^{-1} \cdot s^{-1}$
ξ, Ξ	Deformationspotenziale	eV
ρ	Dichte	$kg \cdot m^{-3}$
$\rho_{Silizid,n}$, $\rho_{Silizid,p}$	Spezifischer Widerstand des Metall/Halbleiter-Übergangs	$\Omega \cdot m$
σ	Verspannung	$N \cdot m^{-2}$
τ_n, τ_p	Ladungsträgerlebensdauern	s
$\bar{\tau}_n$, $\bar{\tau}_p$	Mittlere freie Flugzeit (Impulsrelaxationszeit) der Ladungsträger	s
$\tau_{\omega n}$, $\tau_{\omega p}$	Energierelaxationszeit der Ladungsträger	s
φ	Potenzial	V
φ_F	Dotierungspotenzial	V
φ_m	Austrittsarbeit im Polysilizium	eV
φ_{SD}	Potenzialbarriere zwischen Source und Drain	eV
φ_{ms}	Austrittsarbeitsdifferenz zwischen der Austrittsarbeit im Polysilizium und im Halbleiter	eV

Größe	Bedeutung	Maßeinheit
Φ_s	Austrittsarbeit im Halbleiter	eV
$\Phi_{B,n}$, $\Phi_{B,p}$	Barrierenhöhe	eV
ω_0	Ladungsträgerenergie im thermodynamischen Gleichgewicht	J
ω_n, ω_p	Ladungsträgerenergie	J

1 Einleitung

Der rasante Fortschritt in der Halbleitertechnologie während der letzten Jahrzehnte war von einer stetigen Skalierung der Bauelementeabmessungen begleitet, welche zu einer kontinuierlichen Erhöhung der Geschwindigkeit und Funktionalität der elektronischen Schaltkreise führt. Ausgangspunkt für diese Entwicklung war die Erfindung von Halbleiterbauelementen mit integrierten Schaltkreisen durch J. KILBY 1958 [1]. G. MOORE konnte 1965 aus einer Analyse des Wachstums eine jährliche Verdopplung der Anzahl der Komponenten pro integrierter Schaltung für die künftige Entwicklung ableiten (MOORESches Gesetz) [2], was zehn Jahre später auf eine Verdopplung aller zwei Jahre angepasst wurde [3]. Dieser empirische Zusammenhang stellt nach wie vor einen Leitfaden der wachsenden Halbleiterindustrie dar. Der weitere Bestand dieses Gesetzes erfolgt durch Planungen und Zielstellungen der Halbleiterindustrie selbst, festgelegt in der International Technology Roadmap for Semiconductors (ITRS) [4].

Die zentrale Technologie für die Herstellung von integrierten Schaltungen ist die CMOS- (Complementary Metal Oxide Semiconductor) Technologie. Wie kaum eine andere Technologie prägt die CMOS-Technologie die heutige Technik. Ob in Mobiltelefonen, Digitalkameras oder Computern, winzige CMOS-Transistoren auf Halbleiterkristallen bilden die Basis für unser modernes Alltagsleben. Der Name leitet sich aus dem Begriff MOSFET (Metal Oxide Semiconductor Field Effect Transistor) ab, der in den beiden komplementären Formen p-MOSFET und n-MOSFET (p- bzw. n-leitende Kanalgebiete) zum Einsatz kommt. Die Verwendung dieser Bauelemente und ihre fortschreitende Miniaturisierung, das heißt die Verkleinerung ihrer Dimensionen, bestimmen zu einem wesentlichen Anteil die Anforderungen an die Technologie und deren künftige Entwicklung. Die Miniaturisierung der CMOS-Bauelemente ist für eine weitere Erhöhung der Leistungsfähigkeit der integrierten Schaltungen erforderlich, aber auch in besonderem Maße für eine Erhöhung der Produktivität und des wirtschaftlichen Ertrages. Die Leistungssteigerung zeigt sich in erster Linie in einer kleineren Leistungsaufnahme der Schaltungen und in einer erhöhten Funktionalität durch die Integration einer größeren Anzahl von Bauelementen pro Schaltkreis. Dadurch sinken auch die Kosten je Funktion.

Die angestrebte laterale Verkleinerung des MOSFETs zur Erhöhung der Packungsdichte geht mit einer Reduzierung des Abstands zweier benachbarter Polysilizium-Bahnen, dem sogenannten Poly-Pitch, einher (Bild 1.1). Für eine gute Kontaktierung des Transistors ist die Reduzierung des Poly-Pitchs wiederum mit einer Verkleinerung der Spacer sowie der Gatelänge verbunden.

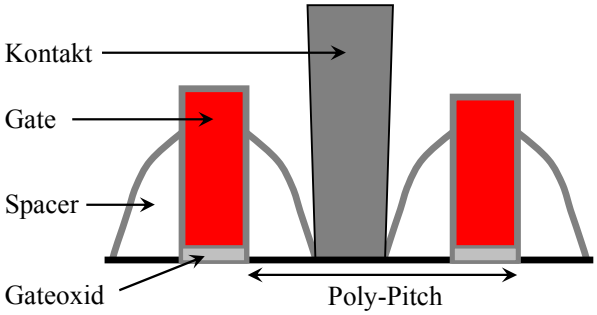


Bild 1.1: Definition technologischer Kenngrößen der planaren Halbleitertechnologie

Wenn jedoch die Gatelänge skaliert wird, müssen für eine entsprechende Funktionalität des Bauelements die anderen Dimensionen idealerweise mit verringert werden. Somit ergibt sich bereits aus den Forderungen nach Skalierung und Leistungssteigerung auch eine kleinere Gateoxiddicke. Die Verringerung der Dicke des Gateoxids kann jedoch nicht unwillkürlich fortgesetzt werden, da sich sonst der Gateleckstrom vergrößert, was sich wiederum nachteilig auf das Schaltverhalten sowie auf die Stromaufnahme integrierter Schaltkreise auswirkt. Aus diesem Grund wird seit der 90 nm-Technologie das Ziel verfolgt, die Gatelängen sowie die Spacer-Weiten der Transistoren bei nahezu gleichbleibender Gateoxiddicke zu verringern. Bei einer derartigen Miniaturisierung des Transistors ist es zwingend erforderlich, möglichst flache und abrupte pn-Übergänge mit maximaler elektrischer Aktivierung zu gestalten, um zum Einen den dabei immer stärker auftretenden Kurzkanaleffekt zu verringern und zum Anderen einen möglichst geringen parasitären Widerstand zu erreichen.

Die Dotierungsprofile werden maßgeblich durch die Implantation und die Ausheilung bestimmt. Die Ionenimplantation ist ein übliches Verfahren der oberflächennahen Dotierung in der planaren Halbleitertechnologie. Ein Nachteil dieser Form der Dotierung ist die mit der Implantation einhergehende Schädigung des Kristallgitters. Um die erzeugten Kristallfehler am Siliziumgitter zu rekristallisieren und die eingebrachten Dotanden zu aktivieren, werden bei der Transistorformation im Wesentlichen thermische Kurzzeitprozesse eingesetzt. Für möglichst flache und abrupte Dotierungsprofile mit einer maximalen Aktivierung der Dotanden, bei denen eine ungewollte Diffusion vermieden werden soll, müssen extrem kurze Ausheilzeiten (Millisekundenbereich) bei hohen Temperaturen nahe dem Schmelzpunkt realisiert werden. Die Aktivierungsvorgänge im Festkörper werden im Rahmen der erreichbaren Temperatur-Zeit-Profile in erster Linie durch die Maximaltemperatur des Ausheilschritts und nicht durch die Dauer der Temperaturbehandlung bestimmt, wohingegen Diffusionsvorgänge maßgeblich von der Prozesstemperatur und Prozesszeit während des Ausheilschrittes beeinflusst werden. Das derzeitig angewendete Ausheilverfahren, die schnelle thermische Ausheilung (Rapid Thermal Annealing, RTA), bei der minimale Pulslängen von circa 1 s bei Temperaturen im Bereich von (1050 … 1100) °C erreicht werden, stößt dabei auf einen Zielkonflikt zwischen der Diffusion der Dotanden und ihrer elektrischen Aktivierung.

Eine weitere Reduktion der Ausheilzeit lässt sich nur durch den Einsatz von neuartigen thermischen Kurzzeitprozessen, wie die Blitzlampen- oder die Laser-Ausheilung, erreichen. Bei diesen Ausheilverfahren können im Millisekundenbereich pulsartig Temperaturen von bis zu 1400 °C erreicht werden. Unter Verwendung der neuartigen Ausheiltechnologien können im Vergleich zur RTA aufgrund der höheren Ausheiltemperatur und der deutlich verkürzten Ausheilzeit bedeutend flachere Dotierungsprofile mit vergleichbaren Widerstandswerten oder vergleichbare Dotierungsprofile mit bedeutend geringeren Widerstandswerten erreicht werden.

Der Unterschied zwischen Blitzlampen- und Laser-Ausheilung liegt in der Art und Weise, wie die Energie auf den Wafer gelangt. Anlagen für die Blitzlampen-Ausheilung arbeiten mit Blitzlichtlampen im sichtbaren Wellenlängenbereich, die den gesamten Wafer auf einmal belichten. Die beiden zurzeit kommerziell erhältlichen Anlagen arbeiten entweder mit Xenon- [5] oder mit Argon-Blitzlampen [6]. Die Laser-Ausheilung funktioniert ähnlich wie die Blitzlampen-Ausheilung, wobei der schmalbandige Laserstrahl anders als das Blitzlampenfeld nicht gleichzeitig den gesamten Wafer, sondern den Wafer zeilenweise belichtet (Scannen). Anlagen für die Laser-Ausheilung kann man nach unterschiedlichen Betriebsmodi (gepulste oder ungepulste Belichtung) und verschiedenen Belichtungswellenlängen (infrarot oder sichtbar/ultraviolett) unterscheiden. Zwei der handelsüblichen Laser-Anlagen verwenden eine ungepulste Belichtung mit einem Array aus Laserdioden mit einer Wellenlänge von 808 nm [7] oder mit einem CO_2-Laser mit einer Wellenlänge im Bereich von 10.6 µm [8].

Sowohl die Blitzlampen- als auch die Laser-Ausheilung kommen bereits heutzutage bei der Herstellung von CMOSFETs als zusätzlicher Ausheilschritt zur höheren Aktivierung der Source- und Drain-Gebiete zur Anwendung. Eine der ersten Veröffentlichungen zu diesem Thema stammt von Toshiba [9]. Hier wurde unter Verwendung einer RTA in Verbindung mit einer sich anschließenden Xenon-Blitzlampen-Ausheilung für CMOSFETs mit einer Gatelänge von 50 nm überdurchschnittliche Transistorergebnisse für beide Transistortypen erreicht. Weitere Veröffentlichungen zeigen ebenfalls verbesserte Transistorparameter für die sub-65 nm-Technologie unter Verwendung der RTA in Kombination mit einer nachfolgenden Kurzzeitausheilung (entweder Blitzlampen- oder Laser-Ausheilung) [10]…[14].

Die Ausheilsequenz von RTA mit nachfolgender Ultrakurzzeitausheilung ist aber nicht der einzig mögliche Ansatz der Implementierung der neuartigen Ausheilverfahren in den Herstellungsprozess. Eine Reihe von Veröffentlichungen zeigen, dass die Verwendung der Ultrakurzzeitausheilung vor der RTA ebenfalls Vorteile bietet, um die Leistungsfähigkeit der Transistoren zu verbessern [15]…[18]. Weiterhin ist es möglich, die Kurzzeitausheilung mehrfach entweder an verschiedenen Stellen mit anderen Schritten dazwischen oder direkt nacheinander im Prozessablauf durchzuführen, um das Transistorverhalten zu verbessern [19], [20].

Der einfache Austausch der konventionellen RTA durch die neuartigen Ausheilverfahren verursacht oftmals Probleme. So kann zum Beispiel eine unzureichende Überlappung der Erweiterungsgebiete unter das Gate durch den Ausheil- und damit Aktivierungsprozess einen verringerten Drainstrom zur Folge haben. Eine Möglichkeit diese Einschränkung zu umgehen, ist die Durchführung der Implantationen der Erweiterungsgebiete unter einem Winkel, um eine ausreichende Überlappung zu gewährleisten. Dieser Ansatz hat jedoch aufgrund der am Gateoxid verursachten Implantationsschäden die potenziellen Nachteile eines erhöhten Gateleckstroms und einer schlechteren Zuverlässigkeit des Gateoxids. In der Literatur kann dennoch eine Reihe von Veröffentlichungen zu diesem Thema gefunden werden, die sich mit der Herstellung von Transistoren nur mit Kurzzeitausheilung auseinandersetzen und die Vor- sowie Nachteile dieser Form der Prozessierung untersuchen [21]…[25].

Ziel dieser Arbeit ist die wissenschaftliche Untersuchung von Integrationsmöglichkeiten der Kurzzeitausheilung in den Herstellungsprozess der Hochleistungs-SOI-CMOSFET-Technologie. Dabei wird zum Einen die Kompatibilität zum herkömmlichen Prozess mit der schnellen thermischen Ausheilung überprüft und zum Anderen die Integration der Kurzzeitausheilung ohne RTA untersucht. Das **zweite Kapitel** bietet zur Einführung in diese Arbeit eine Zusammenfassung wichtiger Grundlagen der MOSFET-Technologie sowie einen Überblick auf die mit der Skalierung der Bauelementeabmessungen einhergehenden Probleme und Möglichkeiten zu deren Minimierung. Weiterhin werden die zur Erzeugung von Dotierungsprofilen erforderlichen Prozessschritte näher betrachtet. **Kapitel 3** gibt einen Überblick auf die neuartigen Verfahren für Ausheilprozesse im Millisekundenbereich sowie deren Einfluss auf die Diffusion und Aktivierung der Dotierstoffe im Vergleich zur schnellen thermischen Ausheilung. Um ein besseres Verständnis für das innerelektronische Verhalten der Transistoren zu erhalten sowie die experimentellen Untersuchungen besser auswerten zu können, wurden weiterhin Simulationen durchgeführt. Das **vierte Kapitel** beschreibt die dabei verwendeten Modelle der Prozess- und Bauelementesimulation sowie deren Kalibrierung an einem SOI-Transistor der 65 nm-Technologie. In **Kapitel 5** werden experimentell gewonnene Ergebnisse vorgestellt, bei denen die Ausheilung im Millisekundenbereich als zusätzlicher Ausheilschritt in Kombination mit der herkömmlich verwendeten RTA durchgeführt wurde.

Das **sechste Kapitel** stellt die Herausforderungen bei der Herstellung von diffusionsarmen Transistoren und die erreichten Ergebnisse dar. Dabei wurde der Einfluss der Kurzzeitausheilung als alleiniger finaler Ausheilschritt ohne RTA in Abhängigkeit von verschiedenen Prozessparametern auf die Leistungsfähigkeit der Transistoren untersucht. In **Kapitel 7** erfolgt abschließend eine Zusammenfassung der in dieser Arbeit gewonnen Ergebnisse.

2 Grundlagen der Transistorherstellung

Dieses Kapitel fasst wichtige Grundlagen der MOSFET-Technologie zusammen. Dabei werden kurz das Funktionsprinzip und die Herstellung eines SOI-MOSFETs sowie charakteristische Kenngrößen zur Bewertung des Transistors beschrieben. Weiterhin werden die mit der Bauelementeskalierung einhergehenden Probleme sowie Möglichkeiten zu deren Minimierung betrachtet. Hauptgegenstand dieses Kapitels ist eine Einführung in die maßgeblich zur Erzeugung von Dotierungsprofilen erforderlichen Prozessschritte – die Implantation und die Ausheilung.

2.1 Herstellungsprozess eines MOSFETs

Ein Feldeffekt-Transistor ist ein Halbleiterbauelement, in dem die Leitfähigkeit des Kanalgebiets durch ein auf die Majoritätsladungsträger einwirkendes elektrisches Feld gesteuert wird. Mit steigenden Gate-Source-Spannungen verringert sich der Kanalwiderstand und der Transistor ist ab einem bestimmten Schwellwert eingeschaltet. Bei einer Gate-Source-Spannung, die unter dem Schwellwert liegt, fließt nur noch ein sehr geringer Strom und der Transistor ist ausgeschaltet. Diese beiden Betriebszustände definieren die Funktion eines Schalters, welche die Grundlage für die binäre Logik der Computertechnologie bildet.

Die Herstellung eines teilweise verarmten (Partially Depleted, PD) SOI-MOSFETs erfolgt prinzipiell wie bei einem Bulk-MOSFET. Das besondere Merkmal von PD-SOI-MOSFETs besteht in der Verwendung eines speziellen Siliziumsubstrats, das ein vergrabenes Oxid (Buried Oxide, BOX) und darüber einen dünnen monokristallinen Siliziumfilm besitzt (Silicon on Insulator, SOI) (Bild 2.1). Teilweise verarmt bedeutet dabei, dass die Dicke der Verarmungszone kleiner ist als die Dicke des Siliziumfilms. Diese Isolation in vertikaler Richtung ermöglicht einen kompakteren Schaltungsentwurf, da keine tiefen Wannen-Implantationen oder sehr tief in das Substrat hineinreichende Gräben zwischen beiden Transistortypen erzeugt werden müssen.

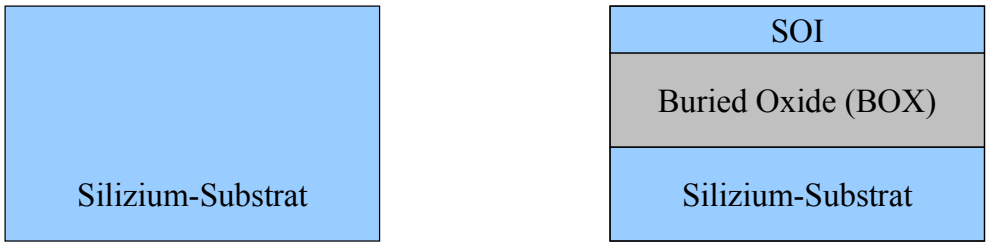


Bild 2.1: Vergleich des Ausgangsmaterials zur Herstellung: (a) eines Bulk-MOSFETs und (b) eines SOI-MOSFETs

Bild 2.2 zeigt einen vereinfachten Ablauf zur Herstellung eines konventionellen SOI-MOSFETs. Ausgehend vom SOI-Wafer als Ausgangsmaterial (a) werden zur Isolierung der einzelnen Transistoren untereinander Gräben in die SOI-Schicht geätzt und wieder mit Siliziumoxid aufgefüllt (b). Durch diese sogenannte „Shallow Trench Isolation“ (STI) werden im Unterschied zu Bulk-Transistoren die SOI-Transistoren vollständig voneinander isoliert. Nach der sich anschließenden Wannendotierung wird Siliziumdioxid (SiO_2) bzw. mit Stickstoff versetztes Siliziumoxid (SiON) thermisch aufgewachsen und darüber eine Polysiliziumschicht abgeschieden (c). Die Wannendotierung dient dabei als Grunddotierung und soll ein Durchgreifen der Raumladungszonen der tiefen Source- und Drain-Gebiete bei Kurzkanaltransistoren verhindern. SiO_2 bzw. SiON wird als Gateoxid verwendet und bildet später zusammen mit der Polysiliziumschicht die Gateelektrode. Auf diesem Schichtstapel wird mittels optischer Lithographie eine Lackschicht strukturiert. Danach folgt ein anisotroper Ätzvorgang, bei dem die verbliebene Lackmaske in die Polysiliziumschicht übertragen wird, wodurch die physikalische Gatelänge definiert wird (d).

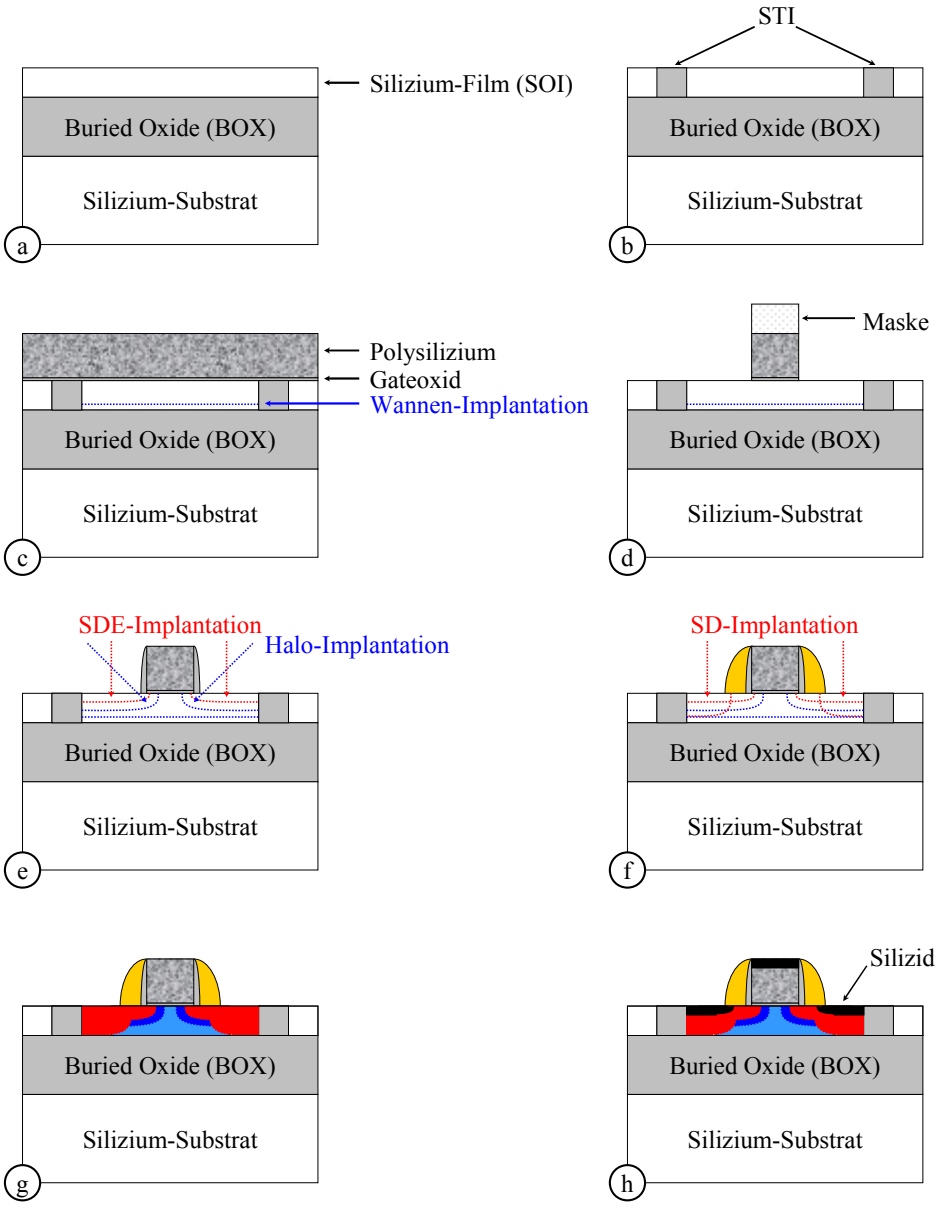


Bild 2.2: Herstellungsprozess eines PD-SOI-MOSFETs (Erläuterung im Text)

Daran anschließend wird durch Abscheiden und Rückätzen einer Oxidschicht der sogenannte Offset-Spacer (Spacer 0) strukturiert, der als Abstandshalter für die nachfolgenden Implantationen der Halo- und Erweiterungsgebiete dient. Die Implantation der Halos wird bei den hier untersuchten Kurzkanaltransistoren zur Einstellung der Schwellspannung und zur Verringerung der Kurzkanaleffekte benötigt. Die unter einem Winkel implantierten Halo-Gebiete heben lokal die Wannendotierung an und umschließen in Kanalrichtung die nachfolgend senkrecht zum Wafer implantierten Erweiterungsgebiete, wodurch sich der Name für diese Implantationen ableitet (e). Nach der Strukturierung eines weiteren Abstandshalters (Spacer 1) werden zur Verringerung des parasitären Widerstands die tiefen Source/Drain-Gebiete (SD-Gebiete) implantiert (f). Daran schließt sich ein thermischer Ausheilschritt zur Rekristallisation der durch die Implantation erzeugten Kristallfehler am Siliziumgitter und zur Aktivierung der eingebrachten Dotanden an (g). Zur Verringerung des Anschlusswiderstands zwischen Silizium und Kontakt werden die SD-Gebiete und die Gateelektrode siliziert. Die Waferoberfläche wird dabei mit einem dünnen Metallfilm überzogen. In einem weiteren Temperaturschritt bildet sich aus dem freiliegenden Silizium und dem Metall eine Metallsiliziumverbindung (Silizid). Ein anschließender Ätzvorgang entfernt den Metallfilm, lässt aber das Silizid unverändert (h). Durch Abscheiden eines Oxids, Kontaktlochätzen sowie Abscheiden und Strukturieren einer Metallschicht werden die elektrischen Anschlüsse hergestellt. Bild 2.3 zeigt den Querschnitt eines PD-SOI-MOSFETs nach der Silizierung.

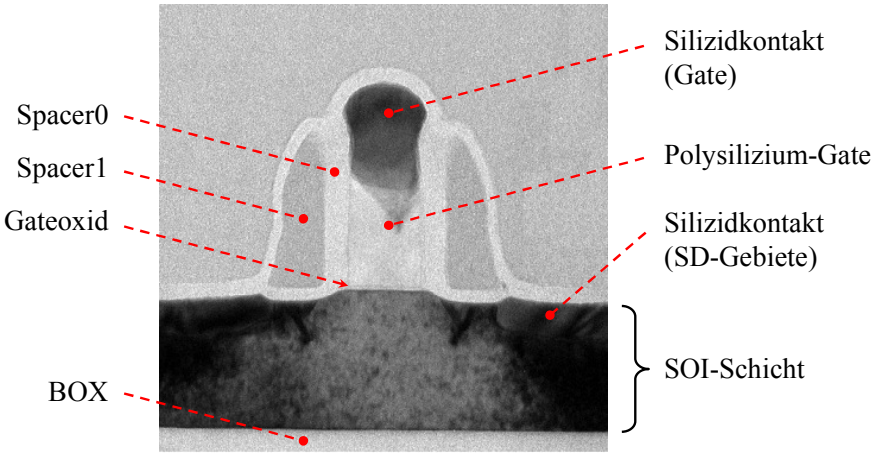


Bild 2.3: Transmissions-Elektronen-Mikroskopie- (TEM-) Querschnittsbild eines SOI-MOSFETs der 45 nm-Technologie nach der Prozessierung

Eine besondere Eigenschaft der PD-SOI-Transistoren gegenüber Bulk-Transistoren ist, dass die Wanne des Transistors nicht mit dem Substrat verbunden ist. Wird dieser Bereich auf ein Bezugspotenzial geschaltet, verhält sich der PD-SOI-MOSFET genauso wie ein Bulk-MOSFET. Ist die Wanne allerdings nicht angeschlossen, kann sich dieser ehemals neutrale Bereich in einigen Betriebszuständen aufladen und für parasitäre „Floating Body“-Effekte sorgen. Im Detail werden bei hohen Drain-Source-Spannungen durch Band-zu-Band-Tunneln oder Stoßionisation Majoritätsladungsträger generiert, die sich im Wannenbereich ansammeln, da sie aufgrund der nicht angeschlossenen (floatenden) Wanne nicht abfließen können. Dadurch steigt das Wannenpotenzial an, was zu einer Reduzierung der Schwellspannung führt, wodurch es beim Einsetzen der Stoßionisation zum plötzlichen Anstieg des Drainstroms kommt. Weiterhin kann dieser Effekt zu Leckströmen bzw. einem Verlust der Gatesteuerung und im dynamischen Betrieb, wenn die Zeit zwischen den Schaltvorgängen nicht für eine vollständige Beseitigung der Ladungsträger ausreicht, zu Hystereseeffekten führen.

2.2 Charakteristische Kenngrößen

2.2.1 Technologische und elektrische Kenngrößen

Die Bestimmung wichtiger technologischer und elektrischer Kenngrößen für die Fertigung eines Mikroprozessors erfolgt mit Hilfe von elektrischen und optischen Messungen an speziell dafür angefertigten Teststrukturen. Bei der Herstellung eines Mikroprozessors wird der Wafer in Felder eingeteilt. Ein Feld ist dabei ein Bereich, der bei einem Belichtungsschritt erfasst wird. Es enthält je nach Technologie einen oder mehrere sogenannte Dies (engl. von die = Rohchip), und Bereiche, in denen Blöcke mit verschiedenen Teststrukturen untergebracht sind, welche beim Vereinzeln der Chips zerstört werden (Bild 2.4). Diese Teststrukturen sind spezielle Anordnungen, um Eigenschaften von bestimmten Strukturen im Mikroprozessor nachzubilden und elektrisch zu überprüfen. Die Fertigung der Teststrukturen erfolgt zeitgleich mit der des eigentlichen Mikroprozessors, wodurch die gleichen Prozessbedingungen bei der Herstellung des Mikroprozessors und der Teststrukturen gegeben sind.

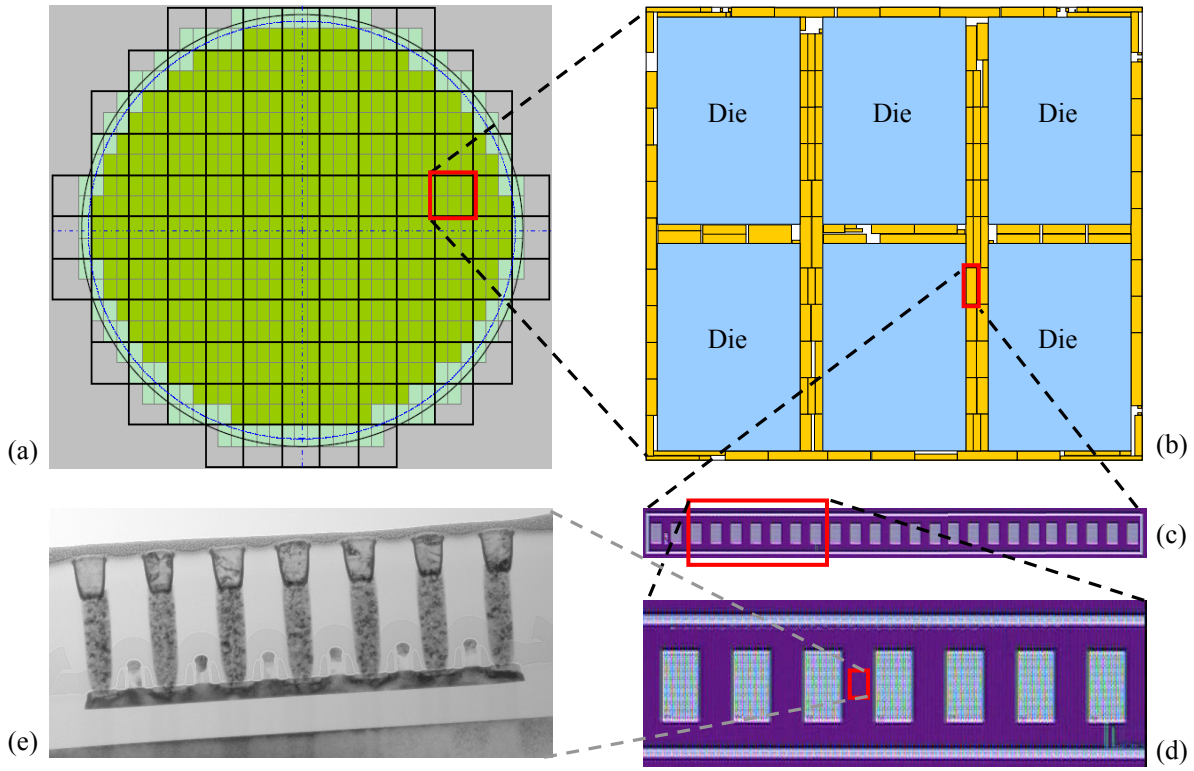
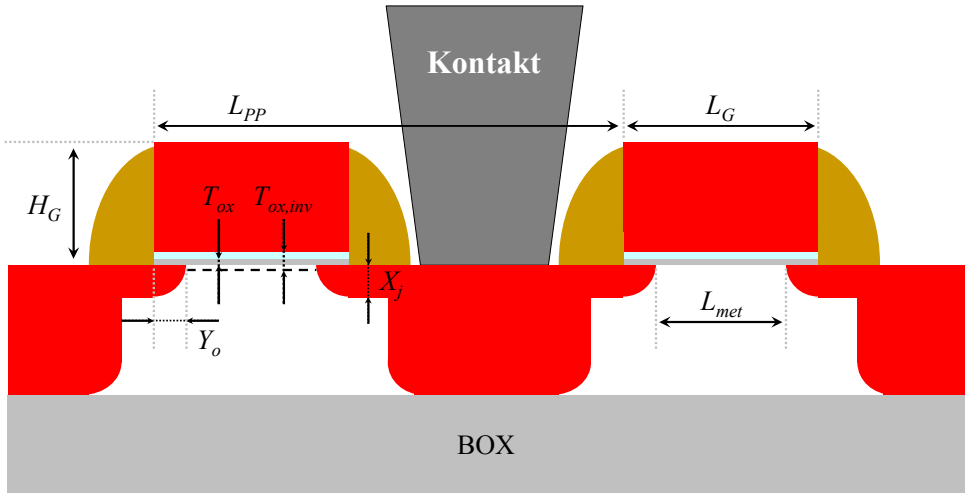


Bild 2.4: Schematische Darstellung der Teststrukturen zur Bestimmung charakteristischer Kenngrößen auf strukturierten Wafern
(a) Einteilung des Wafers in Felder;
(b) Feld mit 6 Dies sowie Blöcke mit verschiedenen Teststrukturen;
(c) Block mit Teststrukturen (helle Felder sind Kontaktpads);
(d) Vergrößerung der Teststruktur;
(e) TEM-Aufnahme der Teststruktur

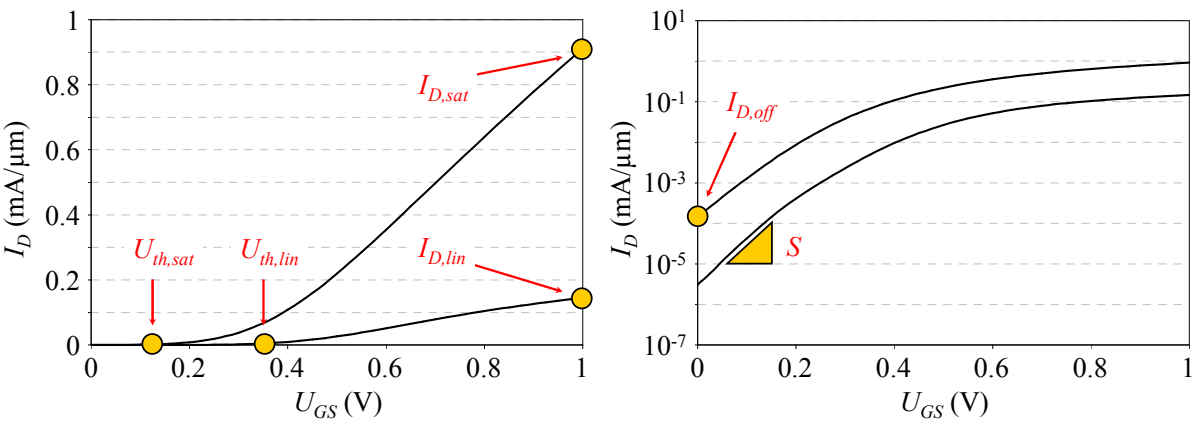
In den Tabellen 2.1 und 2.2 sind die wichtigsten technologischen bzw. elektrischen Kenngrößen eines MOSFETs definiert.

Tabelle 2.1 Technologische Kenngrößen an einem SOI-MOSFET



W_G	Physikalische bzw. geometrische Gateweite
L_G	Physikalische bzw. geometrische Gatelänge
H_G	Physikalische bzw. geometrische Gatehöhe
L_{met}	Lateraler Abstand der Erweiterungsgebiete zwischen Source und Drain
T_{ox}	Physikalische Dicke des Gatedielektrikums zwischen dem Gatematerial und dem Siliziumsubstrat
$T_{ox,inv}$	Die elektrische Gateoxiddicke im Inversionsfall $$T_{ox,inv} = \frac{\varepsilon_0 \varepsilon_{r,Oxid} \cdot A_G}{C_{G,inv}} = \frac{\varepsilon_{Oxid} \cdot A_G}{C_{G,inv}} = \frac{\varepsilon_{Oxid}}{C''_{G,inv}} \quad (2.1)$$ berücksichtigt neben der physikalischen Gateoxiddicke T_{ox} Verarmungseffekte im Polysilizium an der Gate-Isolator-Grenzfläche sowie die veränderte Ladungsverteilung in der Inversionsschicht aufgrund von quantenmechanischen Effekten. Dabei ist ε_0 die absolute Permittivität, $\varepsilon_{r,Oxid}$ die relative Permittivität des Gateoxids, A_G die Gatefläche, $C_{G,inv}$ die Gatekapazität im Inversionsfall und $C''_{G,inv}$ die flächenbezogene Gatekapazität im Inversionsfall.
X_j	Tiefe des pn-Übergangs der Source/Drain-Erweiterungsgebiete (SDE-Gebiete)
Y_o	Überlappung der SDE-Gebiete unter das Gate
L_{PP}	Poly-Pitch, Abstand zweier benachbarter Polysiliziumbahnen

Tabelle 2.2 Elektrische Kenngrößen eines MOSFET dargestellt an einer Transferkennlinie



U_{DD}	Betriebsspannung. Sie beträgt bei den untersuchten Transistoren 1 V.
U_{th}	Die Schwellspannung U_{th} ist die Gate-Source-Spannung U_{GS}, bei der der Transistor vom Aus- in den Ein-Zustand wechselt. Man unterscheidet zwischen der linearen Schwellspannung $U_{th,lin}$, die, wenn nicht anders angegeben, bei einer Drain-Source-Spannung von $U_{DS} = 0.05$ V bestimmt wird, und der Schwellspannung $U_{th,sat}$ im Sättigungsfall bei $U_{DS} = U_{DD}$. Bei den durchgeführten Experimenten und Simulationen wurden die Schwellspannungen $U_{th,lin}$ und $U_{th,sat}$ aus der Transferkennlinie anhand der Tangenten-Methode bestimmt.
$I_{D,off}$	Der Sperrstrom $I_{D,off}$ ist der Drainstrom I_D für den Aus-Zustand im Arbeitspunkt $U_{GS} = 0$ V und $U_{DS} = U_{DD}$. Je kleiner der Sperrstrom $I_{D,off}$ ist, umso geringer ist die Verlustleistung des Transistors.
$I_{D,lin}$	Linearer Drainstrom im Arbeitspunkt $U_{GS} = U_{DD}$ und $U_{DS} = 0.05$ V. Mit seiner Hilfe lassen sich Rückschlüsse auf den Widerstand R_{Total} des Transistors und die Ladungsträgerbeweglichkeit im Kanal ziehen.
$I_{D,sat}$	Der Sättigungs-Drainstrom $I_{D,sat}$ ist der Drainstrom I_D für den Ein-Zustand im Sättigungsfall im Arbeitspunkt $U_{GS} = U_{DD}$ und $U_{DS} = U_{DD}$. Er bestimmt den maximalen Treiberstrom des Transistors.
S	Die Unterschwellsteigung $$S = \left(\frac{\partial U_{GS}}{\partial \log(I_D(U_{GS}))} \right)\Bigg|_{U_{DS}=konstant} \quad (2.2)$$ beschreibt die Steigung der Transferkennlinie im Unterschwellbereich, d.h. im Bereich $U_{GS} = 0$ V bis $U_{GS} = U_{th}$. Sie ist ein indirektes Maß für die Sperrfähigkeit des Transistors und für das spannungsabhängige Umschalten zwischen AUS und EIN. Weiterhin gibt sie die Spannungsänderung ΔU_{GS} an, die an der Gateelektrode beaufschlagt werden muss, um den Drainstrom I_D um eine Dekade zu steigern.

2.2.2 Ladungsträgerbeweglichkeit

Die Messung der Ladungsträgerbeweglichkeit innerhalb des Kanals, insbesondere bei CMOS-Strukturen mit physikalischen Gatelängen L_G kleiner als 100 nm, wirft immer noch einige Probleme auf. Eine allgemein anwendbare Messmethode ist gerade in diesem Bereich von großem Nutzen, da mit fortschreitendem, immer komplexer werdendem Aufbau der Transistorstrukturen die Schwierigkeiten zunehmen, aus indirekten Messparametern auf die Ladungsträgerbeweglichkeit zu schließen.

Ein für Kurzkanalstrukturen geeignetes Messverfahren bietet die sogenannte "dR/dL-Methode" [26]. Bei diesem Verfahren geht man von der Annahme aus, dass sich der messbare Widerstand des Transistors R_{Total} aus dem eigentlichen Kanalwiderstand R_{Kanal} und dem externen Widerstand R_{ext} zusammensetzt:

$$R_{Total} = \frac{U_{DS}}{I_D} = R_{Kanal} + R_{ext} = \frac{T_{ox,inv}}{\mu \cdot W_G \cdot \varepsilon_{Oxid} \cdot \left(U_{GS} - U_{th,lin}\right)} \cdot L_G + R_{ext} \,. \tag{2.3}$$

Setzt man dies in Gleichung (2.3) und leitet die Gleichung nach dR_{Total}/dL_G ab, ergibt sich

$$\frac{dR_{Total}}{dL_G} = \frac{T_{ox,inv}}{\mu \cdot W_G \cdot \varepsilon_{Oxid} \cdot \left(U_{GS} - U_{th,lin}\right)} \tag{2.4}$$

und demnach für die Ladungsträgerbeweglichkeit

$$\mu = \frac{T_{ox,inv}}{W_G \cdot \varepsilon_{Oxid} \cdot \left(U_{GS} - U_{th,lin}\right) \cdot \dfrac{dR_{Total}}{dL_G}} \,. \tag{2.5}$$

Daraus folgt für zwei Transistoren mit gleicher geometrischer Gateweite W_G, gleicher relativer Permittivität ε_{Oxid} des Gateoxids , gleicher elektrischer Gateoxiddicke im Inversionsfall $T_{ox,inv}$ sowie gleicher Differenz zwischen Gate-Source-Spannung U_{GS} und linearer Schwellspannung $U_{th,lin}$:

$$\frac{\mu_{T1}}{\mu_{T2}} \propto \frac{\left(dR_{Total}/dL_G\right)_{T2}}{\left(dR_{Total}/dL_G\right)_{T1}} \,. \tag{2.6}$$

Anhand des Anstieges m der Kurven ist es dann möglich, eine Aussage über die Beweglichkeit der Ladungsträger eines Transistors zu einem Vergleichstransistor zu machen. Ein flacherer Anstieg m der R_{Total}-L_G-Kurve bedeutet eine größere Ladungsträgerbeweglichkeit μ im Kanal des Transistors (Bild 2.5).

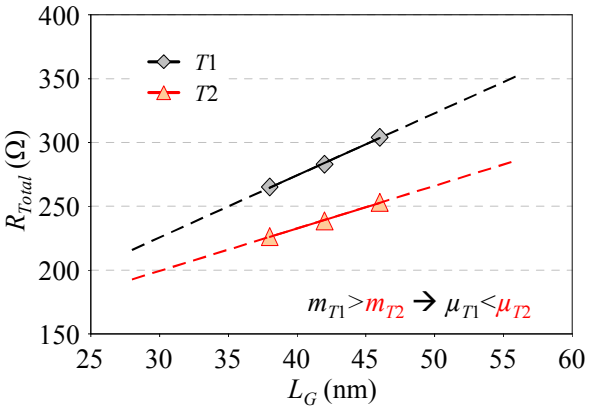


Bild 2.5: Transistorwiderstand R_{Total} in Abhängigkeit von der Gatelänge L_G zur Bewertung der Ladungsträgerbeweglichkeit μ im Kanal von Kurzkanaltransistoren. Die eingezeichneten Linien sind lineare Regressionsgeraden.

2.2.3 Kapazitäten eines SOI-MOSFETs

Weitere wichtige Größen, vor allem um Rückschlüsse auf das dynamische Verhalten des Transistors zu ziehen, sind die Kapazitäten eines MOSFETs. In Bild 2.6a sind die bei einem MOSFET wirkenden Kapazitäten vereinfacht dargestellt. Zum Einen sind das die Sperrschichtkapazitäten C_j zwischen Source bzw. Drain und Wanne zum Anderen die Kapazitäten C_{BOX} zwischen Wanne und Substrat und C_{BOXSD} zwischen Source- bzw. Draingebiet und Substrat. Unter der Annahme gleicher Dotierungsprofile ist die Sperrschichtkapazität bei SOI-MOSFETs im Vergleich zu Bulk-Transistoren durch das Aufsetzen der tiefen SD-Gebiete kleiner. Dies ist vor allem bei Anwendungen, bei denen hohe Schaltgeschwindigkeiten notwendig sind von großem Vorteil, da weniger Ladungen umgeladen werden müssen. Die Kapazitäten zwischen Wanne und Substrat sowie zwischen Source- bzw. Draingebiet und Substrat werden nicht mit in Betracht gezogen, da sie aufgrund der Dicke der Oxidschicht im Vergleich zu den anderen Kapazitäten sehr klein sind.

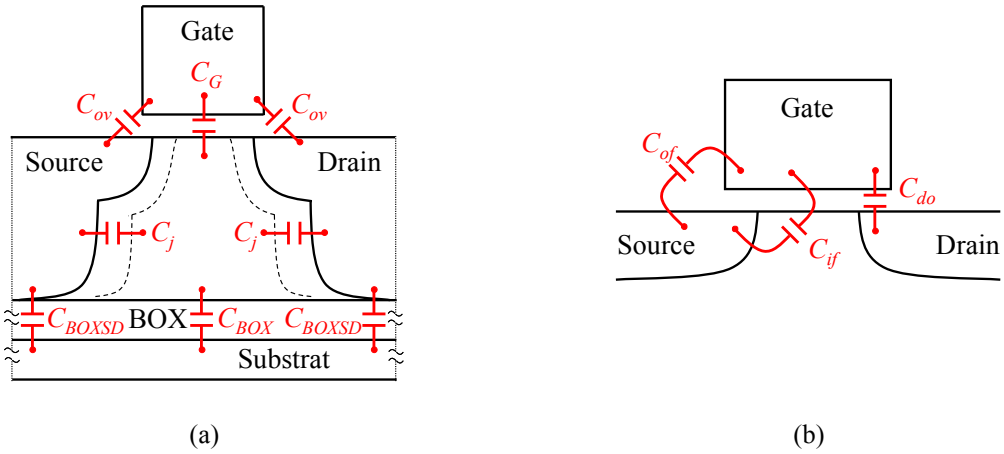


Bild 2.6: Kapazitäten an einem SOI-MOSFETs (a) bei $U_{GS} = 0$ V und (b) zwischen dem Gate und den SD-Gebieten

Von größerer Bedeutung sind dagegen die Gatekapazität C_G, die vor allem durch die Geometrie des Gates (Gatelänge L_G, Gateweite W_G) bestimmt wird, und die Überlappungs- bzw. MILLERkapazität C_{ov} (Bild 2.6b). Die MILLERkapazität beinhaltet je nach Beschaltung des Gates die äußere umsäumende Kapazität C_{of}, die innere umsäumende Kapazität C_{if} und die direkte Überlappungskapazität C_{do}. Letztere ist ein direktes Maß für die Distanz Y_o, wie weit die SDE-Gebiete unter das Gate reichen. Eine große Überlappungs- bzw. MILLERkapazität C_{ov} hat einen höheren Drainstrom I_D zur Folge, da durch die größere Überlappung Y_o ein besserer Anschluss der SDE-Gebiete an die Inversionsschicht erfolgt. Für kleine MILLERkapazitäten C_{ov} sinkt dementsprechend der Drainstrom I_D. Die maximale Frequenz von CMOS-Bauelementen ergibt sich aus den Kapazitäten und dem sie umladenden Strom. Da die MILLERkapazität C_{ov} und der Sättigungsstrom I_D ein gegenläufiges Verhalten zeigen, muss aus diesem Grund ein Optimum gefunden werden. Die MILLERkapazität C_{ov} wird bei $U_{GS} = 0$ V bestimmt, da sich die Ladungen der Erweiterungsgebiete direkt unter dem Gate befinden, sich aber noch kein Inversionskanal ausgebildet hat (Bild 2.7). Eine genaue Bestimmung der Distanz Y_o und damit der direkten Überlappungskapazität C_{do} ist daraus jedoch nicht möglich, da immer noch die äußere, umsäumende Kapazität C_{of} wirkt, in die ebenfalls der Abstand der Anschlusskontakte bzw. der ersten Metallebene zum Gate mit eingeht. Beim Vergleich zweier Transistoren gleicher Geometrie ist jedoch über die MILLERkapazität C_{ov} ein Rückschluss auf die Veränderung der Überlappung Y_o möglich.

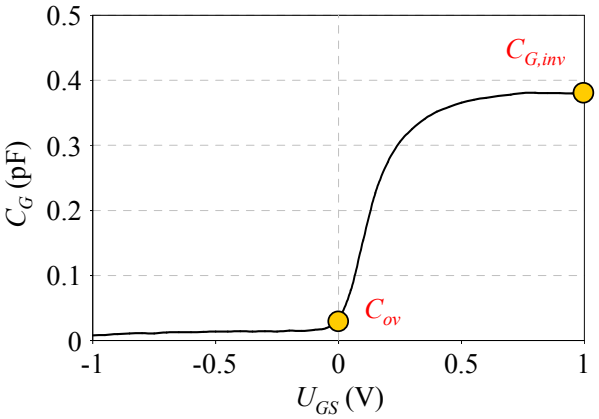


Bild 2.7: Kapazitäts-Spannungs-Kurve von SOI-n-MOSFETs

2.2.4 Leistungsparameter für MOSFETs

Die Leistungsfähigkeit eines MOSFETs wird oft durch die Universalkurve charakterisiert. Sie beschreibt die Beziehung zwischen dem Sperrstrom $I_{D,off}$ und dem linearen Drainstrom $I_{D,lin}$ bzw. dem Sättigungs-Drainstrom $I_{D,sat}$ für verschiedene Gatelängen gleicher Prozessierung (Bild 2.8a). Die verschiedenen Messpunkte für eine bestimmte Gatelänge ergeben sich aus prozessspezifischen Variationen über dem gesamten Wafer. In Bild 2.8b ist der Einfluss verschiedener technologischer und elektrischer Kenngrößen auf die Universalkurve dargestellt. Das Ziel ist dabei, durch Optimierung einzelner Prozessschritte oder durch Einführung neuer Technologien für eine gegebene Gatelänge den linearen Drainstrom $I_{D,lin}$ bzw. den Sättigungs-Drainstrom $I_{D,sat}$ bei gleichzeitiger Verringerung des Sperrstroms $I_{D,off}$ zu erhöhen. Wie man anhand von Bild 2.8b erkennt, kann vor allem durch dünnere Gateoxide die Leistungsfähigkeit des Transistors verbessert werden.

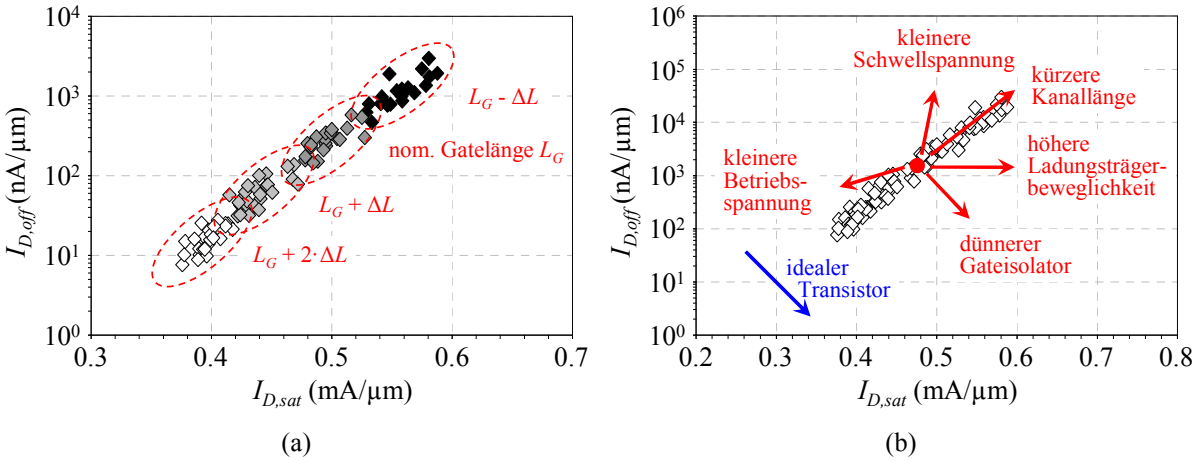


Bild 2.8: Bestimmung der Leistungsfähigkeit eines MOSFET
(a) Definition der Universalkurve
(b) Einfluss technologischer und elektrischer Kenngrößen auf die Universalkurve

2.3 Skalierungseffekte

Leider gelangt man bei der weiteren Skalierung mit den Gateoxiddicken mittlerweile an fundamentale physikalische Grenzen, so dass der in Bild 2.8b eingezeichnete Trend mit kleineren Gateoxiddicken nicht mehr aufrechterhalten werden kann.

Tunnelleckströme durch das Gateoxid, die exponentiell mit abnehmender Gateoxiddicke steigen und Größenordnungen des Sperrstroms $I_{D,off}$ erreichen können, verhindern eine weitere Reduzierung der Gateoxiddicke. Eine Alternative zu den klassischen Gateoxiden SiO_2 und SiON könnten Gateoxidmaterialien mit einer größeren Permittivität sein. Diese führen bei gleicher äquivalenter physikalischer Gateoxiddicke T_{ox} zu niedrigeren Gateleckströmen I_G oder bei gleichem Gateleckstrom I_G zu einer geringeren äquivalenten elektrischen Gateoxiddicke im Inversionsfall $T_{ox,inv}$ und damit zu einer besseren Steuerwirkung des Gates auf den Kanal. Solche Materialien sind jedoch aufgrund verschiedener Integrationsprobleme noch nicht einsatzbereit. Aus diesem Grund richtet sich seit der 90 nm-Technologie die Entwicklung auf immer kleinere Gatelängen ohne eine weitere Reduktion der Gateoxiddicke. Aufgrund dieser jetzt vorwiegend in lateraler Richtung angewandten Miniaturisierung treten sogenannte Kurzkanaleffekte auf, welche die elektrischen Eigenschaften des MOSFETs nachhaltig verschlechtern können [27].

2.3.1 Kurzkanaleffekte

a) Geometriebedingte Verringerung der Schwellspannung

Dieser Kurzkanaleffekt beschreibt die Veränderung der Schwellspannung U_{th} in Abhängigkeit von der Gatelänge L_G (Bild 2.9). Bei gleichem Aufbau und gleicher Technologie wird bei Transistoren mit kurzer Kanallänge die Raumladungszone unter dem Gate durch die Raumladung der Source- und Drain-Gebiete anteilsmäßig stärker beeinflusst als bei Langkanaltransistoren. Die vom Gate kontrollierte Ladung verringert sich mit kleiner werdender Gatelänge, während der Einfluss der Raumladungszone der SD-Gebiete auf die Ladungen im Kanal ansteigt. Aus diesem Grund haben MOSFETs mit einer kürzeren Kanallänge L_G eine niedrigere Schwellspannung U_{th}. Da weniger Ladung aus der Raumladungszone zu kompensieren ist, bildet sich schon bei einer kleineren Gate-Source-Spannungen U_{GS} eine Inversionsschicht aus. Man spricht in diesem Zusammenhang auch vom Modell der Ladungsteilung (Charge Sharing Model) [28].

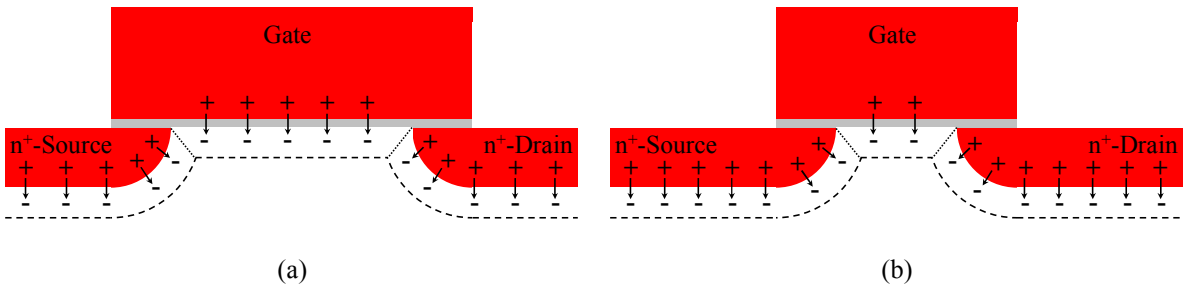


Bild 2.9: Modell zur Erklärung der geometriebedingten Verringerung der Schwellspannung für einen n-MOSFET mit (a) langer und (b) kurzer Kanallänge. Die gestrichelte Linie kennzeichnen die Grenzen der Gate-Source- und Drain-Source-Verarmungszonen. Die Pfeile repräsentieren das elektrische Feld, ausgehend von den positiven Ladungen im Gate zu den negativen Ladungen in der Verarmungszone. Die gepunkteten Linien trennen die Verarmungszone des Gates von den Verarmungszonen der SD-Gebiete ab.

b) Spannungsbedingte Verringerung der Schwellspannung

Die Diffusionsspannung der Source-Substrat- bzw. Source-Wannen-Diode erzeugt eine Potenzialbarriere, die im Fall von n-MOSFETs die Injektion von Elektronen vom Source-Gebiet in die Wanne verhindert.

Bei einem hinreichend kleinen Abstand zwischen Source und Drain erreichen die Feldlinien der drainseitigen Raumladungszone mit steigender Drain-Source-Spannung U_{DS} die sourceseitige Raumladungszone. Dies ist gleichbedeutend mit einem Abbau der Majoritätsladungsträger zwischen Source und Drain. Damit vermindert sich die Potenzialbarriere φ_{SD} zwischen Source und Drain im Kanalgebiet, wodurch die Kanalladung und der Drainstrom I_D nicht mehr vom Gate allein, sondern auch von der Drain-Source-Spannung U_{DS} beeinflusst wird (Bild 2.10). Dadurch verringert sich die Schwellspannung U_{th}, während der Sperrstrom $I_{D,off}$ (bei $U_{GS} = 0$ V) exponentiell ansteigt. Diese drainspannungsbedingte Verringerung der Schwellspannung U_{th} wird auch oft als DIBL-Effekt (Drain induced Barrier Lowering) bezeichnet:

$$DIBL = \Delta U_{th}(U_{DS}) = U_{th,lin} - U_{th,sat} = U_{th}(U_{DS} = 0.05V) - U_{th}(U_{DS} = 1V) \ . \qquad (2.7)$$

Dabei wird davon ausgegangen, dass bei einer Drain-Source-Spannung $U_{DS} = 0.05$ V noch keine Barrierenerniedrigung vorhanden ist. Hierbei ist zu beachten, dass PD-SOI-Transistoren durch den erwähnten Floating-Body-Effekt einen zusätzlichen SOI-DIBL gegenüber herkömmlichen Bulk-Transistoren besitzen, was sich insbesondere bei einer Skalierung (bei konstanter physikalischer Gateoxiddicke T_{ox}) nachteilig auf das Transistorverhalten auswirkt.

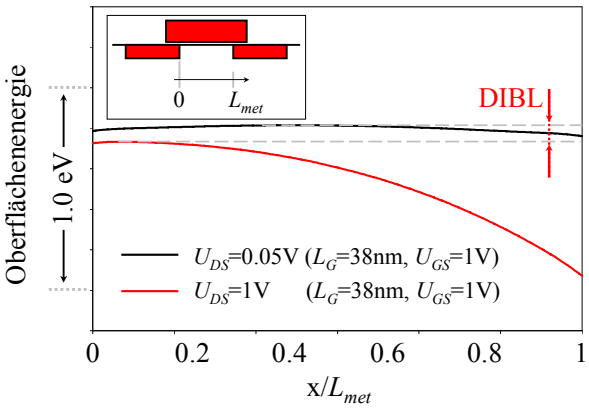


Bild 2.10: Der Bandverlauf zeigt die Barrierenerniedrigung bei steigender Drain-Source-Spannung.

c) Durchgriff

Wenn sich die Raumladungszonen von Drain und Source soweit angenähert haben, dass sie sich treffen, spricht man vom Durchgriff (Punch-through). Der physikalische Grund ist der gleiche wie beim DIBL-Effekt, nämlich die Reduzierung der Drain-Source-Potenzialbarriere φ_{SD}. Ohne die Steuerung durch das Gate können sich dadurch alternative Strompfade ausbilden. Im Gegensatz zum *DIBL* tritt dieser Effekt vor allem in tiefer liegenden, niedrig dotierten Substrat- bzw. Wannengebieten auf. Sobald der Durchgriff einsetzt, verliert das Gate die Kontrolle über den Strom. Dieser fließt nun zusätzlich durch die Berührungsfläche der vereinten Drain-Source-Raumladungszone.

d) Minimierung der Kurzkanaleffekte

Eine Möglichkeit, diese Kurzkanaleffekte zu minimieren, ist die Reduzierung der Gateoxiddicke, was aber aufgrund der schon angesprochenen Tunnelleckströme zu höheren Sperrströmen $I_{D,off}$ und damit zu einer steigenden Verlustleistung des Transistors führt. Eine Alternative zur Gateoxidskalierung besteht darin, die Dotierung im Kanalgebiet zu erhöhen, um eine größere Substratladung zu erzeugen und auf diese Weise die Ausbreitung der Raumladungszone zu reduzieren. Dies ist jedoch nicht ohne eine Verringerung der Ladungsträgerbeweglichkeit im Kanal möglich.

Bild 2.11 zeigt in diesem Zusammenhang die Abhängigkeit der Ladungsträgerbeweglichkeiten μ_n und μ_p der Elektronen bzw. Löcher von der Dotierstoffkonzentration. Durch den Einbau der Dotierstoffatome werden zusätzliche Streuzentren geschaffen, die zu einer Reduktion der Streuzeit und damit der Beweglichkeit führen. Weitere Möglichkeiten um den angesprochenen Kurzkanaleffekten entgegen zu wirken, sind retrograde Wannen- [29], [30] sowie Halo-Implantationen [31], [32], [33] (Bild 2.12). Halo-Implantationen sind lokale Dotierungen unterhalb des Kanals, um die Potenzialbarriere am source- und drainseitigen Kanalübergang zu erhöhen. Bei einer retrograden Wannen-Implantation sinkt im Gegensatz zu einer homogenen Substratdotierung die Dotierung in Richtung Kanalgebiet. Dadurch wird dem Durchgriff der Verarmungszonen der tiefen SD-Gebiete entgegengewirkt und die Ladungsträgerbeweglichkeit im Kanal durch die geringere Kanaldotierung erhöht.

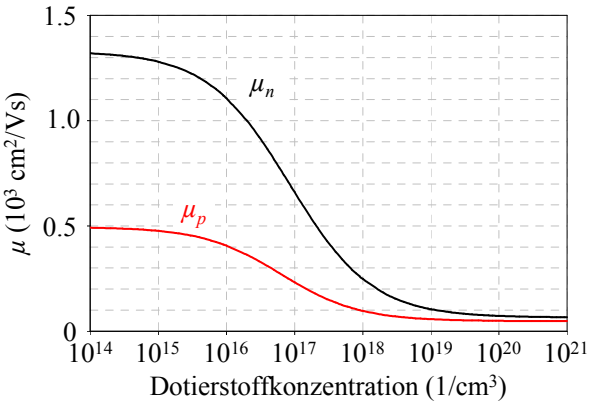


Bild 2.11: Ladungsträgerbeweglichkeit μ_n und μ_p für Elektronen und Löcher bei $T = 300$ K in Silizium in Abhängigkeit von der Dotierstoffkonzentration (nach [34])

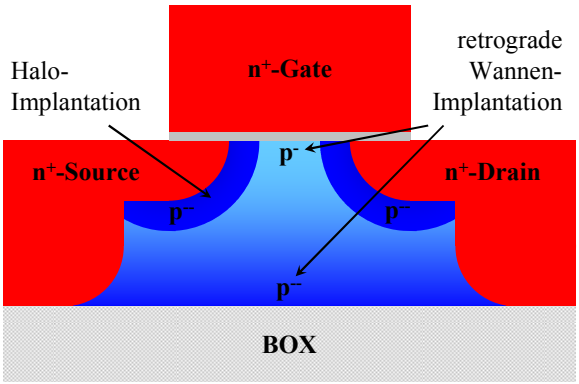
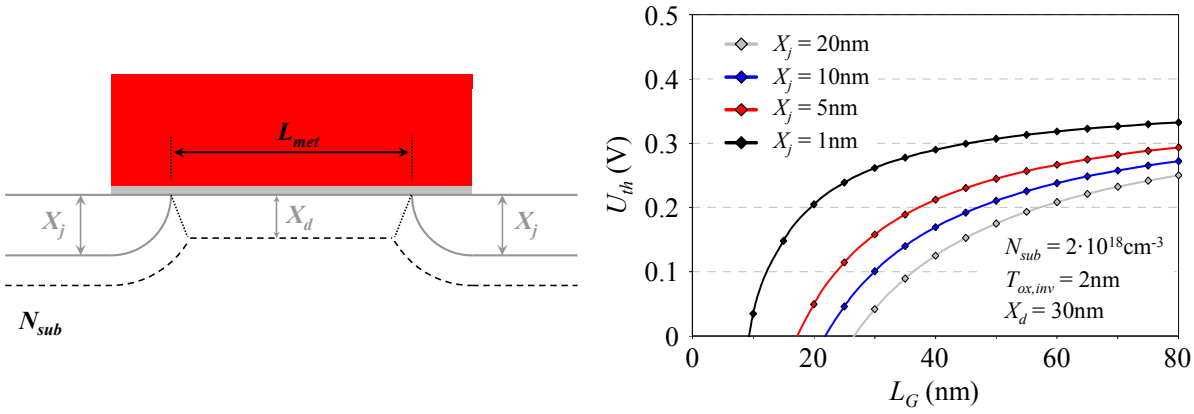


Bild 2.12: Dotierungsprofil eines n-MOSFET unter Verwendung einer Halo-Implantation und einer retrograden Wannen-Implantation

Eine weitere Möglichkeit zur Minimierung der Kurzkanaleffekte ist die Verringerung der Tiefe der pn-Übergänge. Von größter Bedeutung ist dabei die Formierung der SDE-Gebiete. Nach Bild 2.13a lässt sich zumindest für den linearen Arbeitsbereich analytisch die Abhängigkeit der geometriebedingten Verringerung der Schwellspannung U_{th} von der Tiefe X_j des pn-Übergangs der SDE-Gebiete modellieren [28]:

$$U_{th} = U_{FB} + 2 \cdot \varphi_F + \frac{q \cdot N_{sub} \cdot X_d}{C_G''} \left\{ 1 - \left(\sqrt{1 + \frac{2 X_d}{X_j}} - 1 \right) \frac{X_j}{L_{met}} \right\} \tag{2.8}$$

Dabei ist U_{FB} die Flachbandspannung, φ_F das Dotierungspotenzial, bei dem der Übergang vom Zustand der Verarmung zum Zustand der Inversion erfolgt, q die Elementarladung, N_{sub} die Substrat- bzw. im Fall von SOI-Wafern die Wannendotierung, X_d die Ausdehnung der Verarmungszone unter dem Gate und C_G'' die flächenbezogene Gatekapazität. Obwohl dieser Ansatz aufgrund der Komplexität im Sättigungsbereich nicht die veränderten Bedingungen für hohe Drain-Source-Spannungen U_{DS} berücksichtigt, lässt sich aus Bild 2.13b deutlich der Vorteil flacher SDE-Gebiete ablesen. Mit Verringerung der Tiefe X_j des pn-Übergangs der SDE-Gebiete verbessert sich die geometriebedingte Verringerung der Schwellspannung U_{th}, da der Einfluss der Raumladungszone der SDE-Gebiete auf die Ladungen im Kanal kleiner wird. Dadurch ist wiederum eine Skalierung der Gatelänge ohne eine Anpassung weiterer technologischer Kenngrößen wie der Gateoxiddicke möglich.



(a) (b)

Bild 2.13: Geometriebedingte Verringerung der Schwellspannung eines MOSFET in Abhängigkeit von der Tiefe des pn-Übergangs der SDE-Gebiete (nach [28]).
(a) Definition der Kenngrößen,
(b) Schwellspannung in Abhängigkeit von der Gatelänge für verschiedene Tiefen des pn-Übergangs der SDE-Gebiete (nach Gleichung (2.8))

2.3.2 Externer Widerstand

Ein weiteres Problem bei der lateralen Skalierung des MOSFETs ist der zunehmende Einfluss des externen Widerstands R_{ext} des Transistors. Durch die Skalierung der Gatelänge sowie den gezielten Einsatz von mechanischen Verspannungen zur Verbesserung der Ladungsträgerbeweglichkeit im Kanal kann der intrinsische Widerstand bzw. Kanalwiderstand R_{Kanal} des Transistors soweit verringert werden, dass der Einfluss des externen Widerstands R_{ext} vor allem bei zukünftigen Technologiegenerationen nicht mehr vernachlässigt werden kann. So wurde gezeigt, dass bei gleicher Technologie der maximale Drainstromgewinn durch eine mechanische Verspannung des Kanals aufgrund des Einflusses des externen Widerstands R_{ext} bei Gatelängen L_G von etwa 80 nm liegt [35]. Die Verringerung des Drainstroms I_D durch den externen Widerstand R_{ext} ist nach [36]

$$I_D = \frac{I_{D0}}{1 + \dfrac{R_{ext}}{R_{Kanal}}}, \tag{2.9}$$

wobei

$$I_{D0} = \mu \cdot \frac{W_G \cdot C''_G \cdot \left(U_{GS} - U_{th,lin}\right)}{L_{met}} \cdot U_{DS} \tag{2.10}$$

den intrinsischen Strom des Transistors im linearen Bereich darstellt. Vorangegangene Technologien haben Widerstandsverhältnisse R_{ext}/R_{Kanal} von höchstens 20 %, was nach Gleichung (2.9) in einem Drainstrom I_D nahe dem intrinsischen Strom I_{D0} des Transistors resultiert [36]. Bauelementesimulationen basierend auf den Vorgaben der ITRS zeigen jedoch für die weiteren Technologien ein stetige Erhöhung des externen Widerstands R_{ext} bei einer gleichzeitig kontinuierlichen Verringerung des intrinsischen Widerstands R_{Kanal} durch den Einsatz von mechanischen Verspannungen (Bild 2.14) Danach ist zu erwarten, dass bei der 32 nm-Technologie der externe Widerstand R_{ext} des Transistor und der Kanalwiderstand R_{Kanal} ungefähr gleich groß sind (Bild 2.15) [37]. Als Folge des hohen R_{ext}/R_{Kanal}-Verhältnisses äußert sich demnach eine Verbesserung des intrinsischen Widerstands R_{Kanal} zukünftiger Technologien nur in einer geringfügig höheren Leistungsfähigkeit des Transistors.

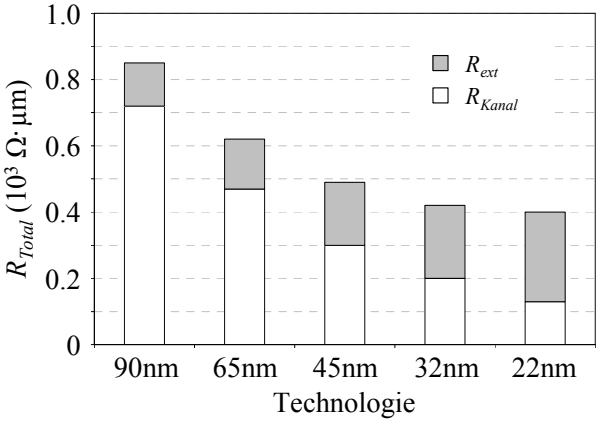


Bild 2.14: Simulierter Anteil des externen Widerstands R_{ext} und des Kanalwiderstands R_{Kanal} am Transistorwiderstand R_{Total} in Abhängigkeit von der Technologiegeneration [37]

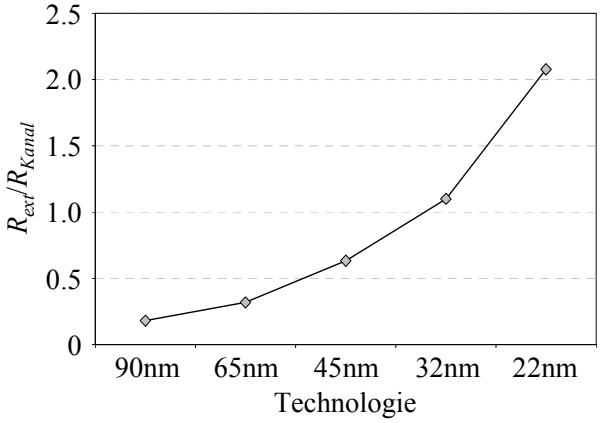


Bild 2.15: Verhältnis von externem Widerstand R_{ext} zu Kanalwiderstand R_{Kanal} in Abhängigkeit von der Technologiegeneration [37]

Der externe Widerstand R_{ext} setzt sich aus den Bahnwiderständen R_{SDE} und R_{SD} der SDE- bzw. SD-Gebiete, den Übergangswiderständen $R_{Silizid}$ vom Silizium zum Silizid und dem Widerstand $R_{Kontakt}$ der Kontaktbahnen zusammen:

$$R_{ext} = R_{SDE} + R_{SD} + R_{Silizid} + R_{Kontakt} \,. \tag{2.11}$$

In Bild 2.16 sind die Beiträge der einzelnen Komponenten am externen Widerstand R_{ext} für n- und p-MOSFETs für die 45 nm-Technologie dargestellt. Die bedeutsamsten Bestandteile am externen Widerstand R_{ext} bei n-MOSFETs sind demnach die Bahnwiderstände R_{SDE} der SDE-Gebiete (37%) sowie die Übergangswiderstände $R_{Silizid}$ vom Silizium zum Silizid (37%). Bei p-MOSFETs haben die Bahnwiderstände R_{SDE} (45%) und R_{SD} (39%) der SDE- bzw. der SD-Gebiete den größten Anteil am externen Widerstand R_{ext}. Sowohl bei n- als auch bei p-MOSFETs haben die Bahnwiderstände R_{SDE} der SDE-Gebiete einen sehr starken Einfluss auf den externen Widerstand R_{ext}. Demzufolge hat eine Verminderung der Tiefe X_j des pn-Übergangs der SDE-Gebiete zukünftiger Technologiegenerationen zur Verbesserung des Kurzkanalverhaltens nicht nur positive Auswirkungen auf die gewünschten Bauelementeeigenschaften.

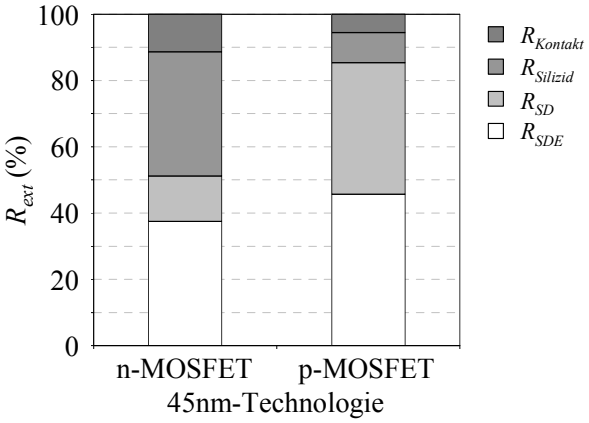


Bild 2.16: Beiträge der einzelnen Komponenten am externen Widerstand R_{ext} für n- und p-MOSFETs für die 45 nm-Technologie [37]

Obwohl sich das Kurzkanalverhalten mit flacheren SDE-Gebieten verbessert, steigt der Widerstand R_{SDE} der SDE-Gebiete mit kleinerer Tiefe X_j der pn-Übergänge entsprechend:

$$R_{SDE} = \frac{1}{\int_0^{X_j} q\left[\mu_p(x)\cdot p(x) + \mu_n(x)\cdot n(x)\right]dx} \tag{2.12}$$

und folglich auch der externe Widerstand R_{ext} des Transistors im Ein-Zustand an, was nach Gleichung (2.9) eine Verringerung des Drainstroms I_D zur Folge hat. Die Erhöhung des Widerstands R_{SDE} der SDE-Gebiete durch die Verringerung der Tiefe X_j der SDE-Gebiete kann jedoch entsprechend Gleichung (2.12) durch eine Erhöhung der Ladungsträgerbeweglichkeiten μ_p bzw. μ_n oder der Ladungsträgerdichten p bzw. n kompensiert werden.

Weiterhin hat der Übergangswiderstands $R_{Silizid}$ vom Silizium zum Silizid bei n-MOSFETs einen großen Anteil am externen Widerstand R_{ext}. Der Übergangswiderstands $R_{Silizid}$ wird maßgeblich durch die Ladungsträgerkonzentration n und p im Silizium und der Barrierenhöhe Φ_B der beteiligten Materialien bestimmt. Die Barrierenhöhe Φ_B setzt sich dabei für n-MOSFETs aus der Differenz zwischen FERMI-Energie der Metallverbindung und Leitungsband des Halbleiters und für p-MOSFETs aus der Differenz von FERMI-Energie der Metallverbindung und Valenzband des Halbleiters zusammen. Der spezifische Widerstand $\rho_{Silizid}$ des Metall/Halbleiter-Überganges kann nach [38] für n-MOSFETs zu

$$\rho_{Silizid,n} = K_1 \cdot \exp\left(K_2 \cdot \frac{q\cdot\Phi_{B,n}}{\sqrt{n}}\right) \tag{2.13}$$

und für p-MOSFETs zu

$$\rho_{Silizid,p} = K_1 \cdot \exp\left(K_2 \cdot \frac{q\cdot\Phi_{B,p}}{\sqrt{p}}\right) \tag{2.14}$$

bestimmt werden. Hier sind K_1 und K_2 Konstanten, n bzw. p die Ladungsträgerkonzentration an der Grenzfläche zwischen Silizid und Silizium sowie $\Phi_{B,n}$ und $\Phi_{B,p}$ die Barrierenhöhe beim n- bzw. p-MOSFET. Dabei gilt

$$\Phi_{B,p} = E_g - \Phi_{B,n} \tag{2.15}$$

mit E_g der Bandlücke. Zurzeit wird zur Silizierung der SD-Gebiete und des Polysiliziums eine Nickel-Platin-Legierung (NiPt) mit 5 % Platin verwendet, wodurch sich nach [39] für n-MOSFETs eine Barrierenhöhe von $\Phi_{B,n} = 0.72$ eV und dementsprechend für p-MOSFETs nach Gleichung (2.15) $\Phi_{B,p} = 0.40$ eV ergibt. Aufgrund der ungleichen Barrierenhöhe hat der Übergangswiderstand $R_{Silizid}$ bei n-MOSFETs einen größeren Einfluss auf den externen Widerstand R_{ext} als bei p-MOSFETs. Eine Verringerung des Übergangswiderstands $R_{Silizid}$ kann somit nach Gleichung (2.13) nur durch eine höhere Ladungsträgerkonzentration n bzw. p an der Grenzfläche zwischen Silizid und Silizium erreicht werden. Der Widerstand R_{SD} der SD-Gebiete, der bei p-MOSFETs einen großen Anteil am externen Widerstand R_{ext} einnimmt, kann basierend auf Gleichung (2.12) durch eine Erhöhung der Ladungsträgerbeweglichkeiten μ_p bzw. μ_n oder der Ladungsträgerdichten p bzw. n verringert werden.

Die Ladungsträgerbeweglichkeit, die Ladungsträgerkonzentration sowie die Tiefe der SDE-Gebiete hängen in einem hohen Maß vom Dotierprozess und damit von der Einbringung des Dotierstoffs bei der Implantation sowie dessen Aktivierung und Diffusion ab.

2.4 Implantation

Die gewünschten elektrischen Eigenschaften von Halbleitern lassen sich durch Einfügen von Dotier- bzw. Fremdatomen in die Kristallstruktur (Dotierung) in weiten Bereichen beeinflussen. Die Ionenimplantation ist dabei seit Mitte der 70er Jahre das übliche Verfahren der oberflächennahen Dotierung in der planaren Halbleitertechnologie. Ionisierte Atome des Dotierstoffes bzw. Moleküle, die ihn enthalten, werden in einem elektrischen Feld beschleunigt und in das Substrat hinein geschossen (implantiert). Durch Energieabgabe infolge von elastischen (Kernstöße mit Substratatomen) und unelastischen Stößen (Anregungswechselwirkungen mit Substratelektronen) kommen sie im Inneren des Festkörpers zur Ruhe. Der Vorteil dieser Art der Dotierung ist die sehr gute Kontrolle über die Verteilung und Konzentration des Dotierstoffes im Halbleiter (Dotierungsprofil). So sind die Eindringtiefe durch die Wahl der Energie und des Einfallswinkels der Ionen sowie die Dosis durch Festlegung des Ionenstroms und der Prozesszeit unabhängig voneinander einstellbar. Ein Nachteil ist die mit der Implantation einhergehende Schädigung des Kristallgitters. Defekte entstehen während der Ionenimplantation aber nur, wenn es zur Verlagerung von Targetatomen (Substratatomen) kommt. Eine solche Verlagerung findet immer dann statt, wenn es bei den elastischen Stößen zu Energieüberträgen kommt, deren Beträge größer als die Verlagerungsenergie E_d eines Gitteratoms sind. Diese Verlagerungsenergie liegt für Silizium im Bereich von etwa 15 eV [40]. Liegt die übertragene Energie im Bereich $E_d \leq E \leq 2 \cdot E_d$, wird nur ein einzelnes Punktdefektpaar (Frenkel-Paar) erzeugt, ein System aus einer Leerstelle (Vakanz, V) und einem Silizium-Zwischengitteratom (Interstitial, I). Im Fall $E >> E_d$, der für die Ionenimplantation meistens erfüllt ist, verzweigen sich die Stoßfolgen in Subkaskaden und es bleibt ein stark gestörtes Kristallgitter zurück. Die Dichte der Schädigung hängt dabei sehr stark von der Implantationsdosis und -energie und vom Massenverhältnis von Dotier- zu Targetatom ab [41]. Überschreitet die Punktdefektdichte einen bestimmten Schwellwert, führt dies zur vollständigen Zerstörung der Fernordnung im Kristall und damit zur Ausbildung von amorphen Schichten. Für Silizium liegt die kritische Punktdefektdichte bei circa 10% der Atomdichte von Silizium, also bei $5 \cdot 10^{21}$ cm^{-3} [42].

Ein weiteres Problem ist der sogenannte Channeling-Effekt. Er verursacht eine breitere Verteilung der Dotieratome. In kristallines Material können Dotanden aufgrund der regelmäßigen Struktur des Kristalls tiefer eindringen als in amorphes Material. Dieser Effekt kann durch eine Veränderung des Implantationswinkels oder durch Aufbringen eines Streuoxids gedämpft bzw. durch eine Prä-Amorphisierungsimplantation (PAI) verhindert werden.

2.5 Ausheilung

Sowohl für das Wiederherstellen (Ausheilen) der Kristallstruktur als auch für eine elektrische Aktivierung, also die Bindung der implantierten Atome an reguläre Gitterplätze, ist nach jeder Implantation ein thermischer Prozess erforderlich. Nur in Bindungen mit den Festkörperatomen können Fremdatome zur Elektronenkonfiguration des Festkörpers mit Ladungsträgern wirksam beitragen. Die Diffusion und Aktivierung der eingebrachten Dotanden während der Ausheilung sind dabei von besonderer Bedeutung, da dadurch die Leitfähigkeit sowie die Tiefe der pn-Übergänge der einzelnen Bestandteile der Source- und Drain-Gebiete bestimmt werden.

2.5.1 Rekristallisation und Aktivierung

a) Restaurierung der implantierten Schicht

Nach der Implantation ist der Kristallverband ganz oder nur teilweise zerstört. Aus diesem Grund ist ein Ausheilschritt notwendig, um den gestörten Kristall zu restaurieren, also zu rekristallisieren. Dabei hängt die Rekristallisation davon ab, wie stark die Schädigung des Kristalls ist. Die Rekristallisation einer vollständig amorphisierten Schicht geht von einer Seite aus, wohingegen bei der Restaurierung von nur teilweise amorphisierten Schichten die Rekristallisation gleichzeitig aus unterschiedlichen Richtungen erfolgt (Bild 2.17). Dadurch wird mehr Energie benötigt als bei der Restaurierung einer vollständig amorphisierten Schicht, wodurch eine perfekte Rekristallisation schwieriger zu erreichen ist.

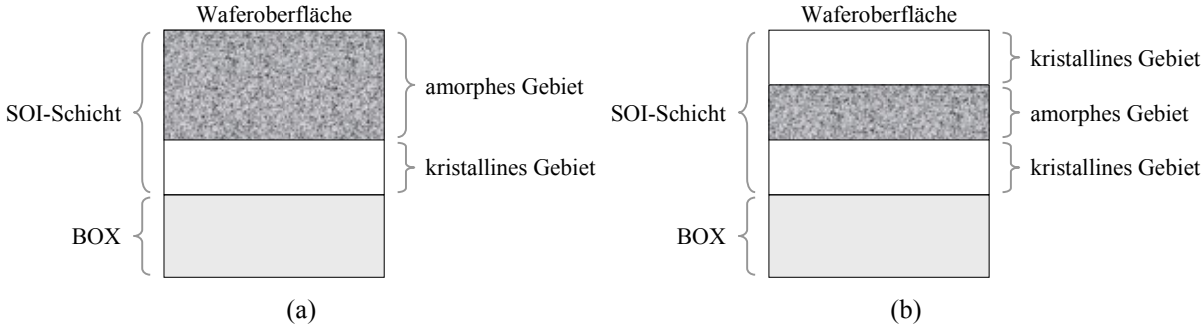


Bild 2.17: Schematische Darstellung einer (a) vollständig und (b) teilweise amorphisierten Schicht

Bei der Restaurierung einer amorphen Schicht spricht man im Zusammenhang mit einem Ausheilschritt ohne ein Aufschmelzen des Bereichs auch von einer Festphasenepitaxie (Solid Phase Epitaxy, SPE). Dabei erfolgt das Aufwachsen schon ab Temperaturen von 400 °C bis 450 °C sehr schnell Ebene für Ebene, ausgehend von der Grenzfläche zwischen amorphem und kristallinem Material. Die Aufwachsraten sind dabei stark temperaturabhängig. Mit steigender Temperatur erhöht sich die Aufwachsrate [43]. Die Bewegungsgeschwindigkeit der Grenzfläche in Richtung des amorphen Teiles hängt neben der Temperatur ebenfalls noch von der Kristallorientierung und Dotierstoffkonzentration ab [44], [45]. Typische Aufwachsraten in undotiertem Silizium bei 600 °C sind 50 nm/min in <100>-Richtung, 20 nm/min in <110>-Richtung und 2 nm/min in <111>-Richtung [42].

An der Grenzfläche zwischen amorphem und kristallinem Material ist der Siliziumkristall jedoch ebenfalls sehr stark geschädigt. Die Punktdefektdichte ist hier nur etwas geringer als der Schwellwert zur Amorphisierung. Diese Punktdefekte fügen sich während des Ausheilvorganges zu größeren Komplexen zusammen, die wiederum die Ladungsträgerbeweglichkeit beeinflussen bzw., wenn sie sich innerhalb des pn-Übergangs befinden, Leckströme verursachen können. Je nach zugeführter Energie bzw. Temperatur während des Ausheilschrittes verändern sich die nach der Implantation vorhandenen Defekte unterhalb der Grenzfläche. So bilden sich zu Beginn des Ausheilschrittes sogenannte {311}-Defekte aus Silizium-Zwischengitteratomen. <311>-Defekte werden typischerweise als eine Reihe von Silizium-Zwischengitteratom-Dimer aufgefasst, die sich in der {311}-Ebene ansammeln und sich in die <110>-Richtung des Siliziumkristalls ausbreiten, wodurch es zur Ausbildung eines Flächendefektes kommt (Bild 2.18). Unter weiterer Energiezufuhr beginnen sich die {311}-Defekte unter Abgabe von Silizium-Zwischengitteratomen von den Enden des Defektes her aufzulösen. Ist die Dichte diese Defekte unter einem bestimmten Schwellwert, lösen sich diese Defekte komplett auf, was eine vollständige Restaurierung des Kristalls zur Folge hat.

Bei einer Defektdichte über diesem Schwellwert wachsen einige dieser Defekte zu Lasten kleinerer {311}-Defekte, die sich auflösen. Größere {311}-Defekte formen wiederum bei weiterer Energiezufuhr energetisch stabilere Versetzungsringe. Diese sekundären Defekte bleiben so lange erhalten bzw. wachsen sogar, bis alle primären Defekte (Zwischengitteratome, {311}-Defekte) ausgeheilt sind. Erst danach lösen sich auch diese Defekte auf. Die größte Konzentration dieser Versetzungsringe wird dabei an der Grenzfläche zwischen amorphem und kristallinem Silizium nach der Festphasenepitaxie der amorphen Schicht durch die Ausheilung beobachtet. Diese Defekte an der ursprünglichen Grenzfläche zwischen amorphem und kristallinem Material werden auch als End of Range- (EOR-) Defekte bezeichnet (Bild 2.19) [42].

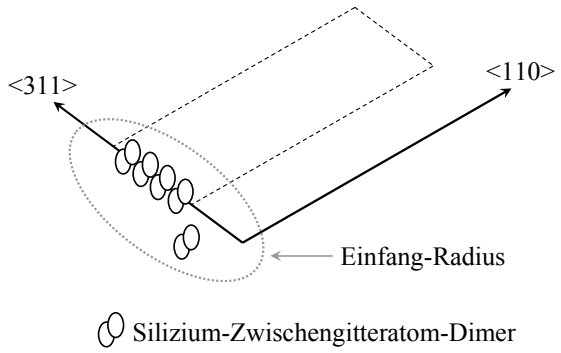


Bild 2.18: Schematische Darstellung von {311}-Defekten

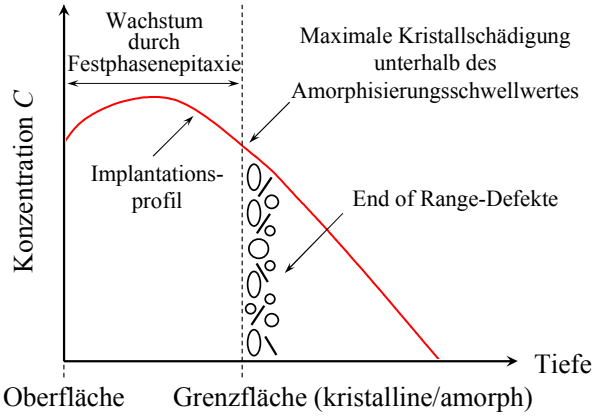


Bild 2.19: Schematische Darstellung von Defekten an der Grenzfläche zwischen amorphem und kristallinem Silizium nach der Festphasenepitaxie der amorphen Schicht (nach [42])

b) Elektrische Aktivierung

Mit der thermischen Nachbehandlung zum Zwecke der Restaurierung der geschädigten Kristallbereiche erfolgt eine elektrische Aktivierung der eingebrachten Dotierstoffe. Dabei ist in erster Linie die Maximaltemperatur des Ausheilschrittes relevant und nicht die Dauer der Temperaturbehandlung [42]. Um zur elektrischen Leitfähigkeit des Kristalls beizutragen, muss der Dotand einen regulären Gitterplatz einnehmen. Im Allgemeinen befindet sich der Dotierstoff während der Implantation nur zu einem Teil auf Gitterplätzen. Daher müssen die implantierten Dotierstoffatome, die substitutionell eingebaut werden sollen, durch einen thermischen Prozess nach der Implantation aktiviert werden. Das bedeutet, dass die eingebauten Dotierstoffatome eine kovalente Bindung mit dem Siliziumkristall eingehen, wodurch je nach Wertigkeit des Dotierstoffes (III- oder V-wertig) entweder zusätzliche Elektronen gebunden (p-Dotierung) oder Elektronen freigesetzt werden (n-Dotierung). Um eine hohe Leitfähigkeit und damit einen geringen Flächenwiderstand der SDE-Gebiete zu erreichen, ist es nach Gleichung (2.12) erforderlich, möglichst viele Dotanden zu aktivieren, um die Konzentration von Elektronen n bzw. Löchern p zu erhöhen. Die maximal mögliche Konzentration C eines Dotierstoffes, die in Silizium unter Gleichgewichtsbedingungen ohne Ausprägung einer neuen Phase gelöst sein kann, wird durch dessen Festkörperlöslichkeit begrenzt. Wird diese Konzentration überschritten, beginnen die Dotierstoffatome zu präzipieren und es bilden sich neue Phasen aus, d.h. es kommt zur Ausbildung von neuen Materialverbindungen. Die Festkörperlöslichkeit eines Stoffes verändert sich in Abhängigkeit von der Temperatur.

Die elektrische Löslichkeit der implantierten Dotanden kann aber aufgrund einer erhöhten Anzahl an Punktdefekten oder Dotierstoffatomen bedeutend geringer sein als die maximale Festkörperlöslichkeit. Oberhalb dieser elektrischen Löslichkeitsgrenze kann es zur Ausbildung von inaktiven Komplexen aus Dotierstoffatomen und Punktdefekten kommen, die elektrisch neutral sind und aus diesem Grund keine freien Ladungsträger bereitstellen können [42]. Bor z.B. geht bevorzugt eine Paarung mit Silizium-Zwischengitteratomen ein, was zur Ausbildung von Bor-Silizium-Zwischengitteratom-Komplexen führt (Boron Interstitial Cluster, BIC) [46], [47], [48]. Im Gegensatz dazu favorisiert Arsen aus energetischen Gründen eine Bindung mit Leerstellen und bildet Arsen-Leerstellen-Komplexe (Arsenic Vacancy Cluster, AsV) [49], [50], [51].

Bild 2.20 zeigt die Festkörperlöslichkeit für Arsen, Phosphor und Antimon zur n-Dotierung sowie für Bor und Gallium zur p-Dotierung in Silizium. Die Löslichkeit der Dotierstoffatome erhöht sich mit steigender Temperatur und hat je nach Dotand ein Maximum zwischen 1200 °C und 1400 °C. Störend ist für alle Dotierstoffe die mit dem Ausheilvorgang einhergehende Umverteilung der Implantationsprofile durch die gleichzeitig auftretende Diffusion im Halbleiterkristall.

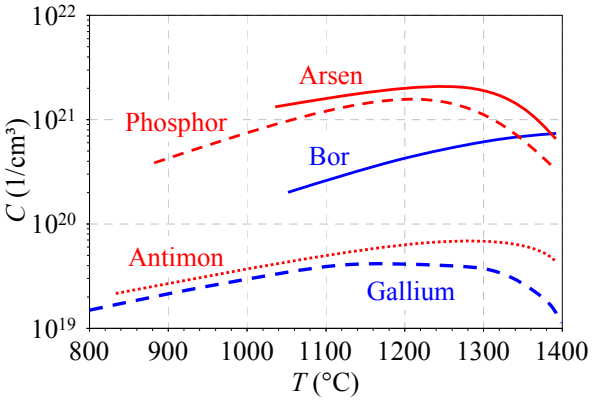


Bild 2.20: Festkörperlöslichkeit verschiedener Dotierstoffe in Silizium unter Gleichgewichtsbedingungen in Abhängigkeit von der Temperatur [52]

2.5.2 Diffusion

Diffusionsvorgänge unterscheiden sich, ob sie in gasförmiger, flüssiger oder fester Materie stattfinden. Im festen Zustand versteht man unter der Diffusion den mit den statistischen Bewegungen einer Stoffkomponente verbundenen Stofftransport infolge von örtlichen Konzentrationsgradienten in der Dotierstoffverteilung oder als Folge des Ausgleichs der kinetischen Energie der Wärmebewegung. In einem Festkörper wandern (diffundieren) die Atome durch eine thermische Anregung von einem Gitterplatz zum nächsten. Aus diesem Grund verändern sich durch die Diffusion von Dotieratomen alle Implantationsprofile im Halbleiterkristall. Höhere Ausheiltemperaturen haben dabei eine stärkere Diffusion zur Folge. Grund dafür ist, dass mit steigender Temperatur dem Gitter Zusatzenergiebeträge zugeführt werden, die eine Intensivierung der Gitterschwingungen zur Folge haben. Aufgrund ihrer jetzt größeren kinetischen Energie sind mehr Atome in der Lage, Energiebarrieren leichter zu überwinden und so zu anderen Gitterplätzen zu gelangen. Die Diffusion hält so lange an, bis entweder eine Gleichverteilung erreicht oder die Temperatur soweit abgesunken ist, dass die Dotierstoffatome „einfrieren“, also unbeweglich werden.

a) Diffusion im Kontinuum

Für den Materietransport in einem Konzentrationsgefälle gelten die FICKschen Gesetze [53], die eigentlich für das ideale Gas abgeleitet wurden, aber auch im festen Zustand gültig sind. Das erste FICKsche Gesetz stellt den Zusammenhang von Transport, also den Teilchenstrom durch Diffusion, und räumlichen Konzentrationsunterschieden dar:

$$\boldsymbol{J}_T = -\boldsymbol{D} \cdot \operatorname{grad} C \,. \tag{2.16}$$

Aufgrund der größeren Anzahl von Teilchen in Bereichen höherer Konzentration erfolgt ein Teilchenfluss $\boldsymbol{J}_T$ hin zu Bereichen mit geringerer Konzentration C. Der Proportionalitätsfaktor zwischen der Teilchenstromdichte $\boldsymbol{J}_T$ und dem Gradienten der Konzentration C ist der Diffusionskoeffizient $\boldsymbol{D}$. Er gibt an, wie rasch bei einem vorgegebenen Konzentrationsgefälle der Stofftransport in einer Substanz erfolgt. In anisotropen Medien oder unter Einfluss äußerer Felder sowie mechanischer Verspannungen ist der Diffusionskoeffizient $\boldsymbol{D}$ ein symmetrischer Tensor zweiter Stufe und dadurch richtungsabhängig. Silizium kann im unverspannten Zustand aufgrund seiner symmetrischen kubischen Kristallstruktur als isotropes Material betrachtet werden, wodurch sich der Diffusionskoeffizient sehr stark vereinfacht und als Skalar betrachtet werden kann [54].

Der Diffusionsstrom ist bei einer gegebenen Temperatur nur dann konstant, wenn sich das Konzentrationsgefälle nicht ändert. Da aber gerade die Diffusion den Konzentrationsunterschied ausgleicht, ändert sich damit auch der Diffusionsstrom. Eine zunächst große Diffusionsstromdichte $\boldsymbol{J}_T$ verringert sich in dem Maße, wie der Konzentrationsunterschied ΔC abgebaut wird:

$$\frac{\partial C}{\partial t} = -\operatorname{div} \boldsymbol{J}_T = \operatorname{div}\left(\boldsymbol{D} \cdot \operatorname{grad} C\right). \tag{2.17}$$

Diese auch zweites FICKsches Gesetz genannte Beziehung beschreibt den Zusammenhang zwischen zeitlichen und örtlichen Konzentrationsunterschieden. Gleichung (2.17) besagt, dass die Differenz von hinein- und hinausgehendem Teilchenstrom mit der Zeit innerhalb eines Volumenelementes der Konzentrationsänderung im Volumenelement entsprechen muss, wenn keine Erzeugung oder Vernichtung der Materie in dem Element auftritt.

Die Lösung der Differentialgleichung (2.17) hängt von den Anfangs- und Randbedingungen ab. Unter Annahme einer Eindiffusion aus einer begrenzten Quelle, die z.B. bei einer Diffusion nach einer Oberflächenbelegung mit Dotierstoff durch Implantation vorliegt, ergibt die Lösung eine GAUß-Funktion des Ortes und der Zeit, welche in eindimensionaler Form lautet:

$$C(x,t) = C(0,t) \cdot \exp\left(-\frac{x^2}{4\,D \cdot t}\right). \tag{2.18}$$

Ausgehend von Gleichung (2.18) wird als wichtige Größe zur Charakterisierung des Transports durch Diffusion die Diffusionslänge

$$L_D = 2 \cdot \sqrt{D \cdot t} \tag{2.19}$$

definiert, bei der die Konzentration auf das 1/e-fache der Oberflächenkonzentration gefallen ist [55].

b) Diffusion im Kristall

In einem perfekten Kristallgitter schwingt jedes Gitterteilchen um seinen festen Gitterplatz, kann diesen aber nicht verlassen. Eine notwendige Voraussetzung für Diffusion in einem kristallinen Festkörper sind daher Gitterfehler, wie z.B. Leerstellen und Silizium-Zwischengitteratome.

Nur dadurch können Platzwechsel von Atomen als Bedingung für Stofftransport stattfinden. Abhängig vom Defekt treten verschiedene Arten von Platzwechselvorgängen auf. Die wichtigsten Mechanismen der Diffusion im Festkörper sind die Diffusion mittels Leerstellen

$$A + V \leftrightarrow AV \tag{2.20}$$

und die Diffusion über Zwischengitteratome

$$A + I \leftrightarrow AI\,. \tag{2.21}$$

In den Gleichungen (2.20) und (2.21) stellt A ein Fremdatom, in diesem Fall ein Dotierstoffatom, auf einem substitutionellen Gitterplatz, V eine Leerstelle und I ein Silizium-Zwischengitteratom dar. Energetisch gesehen entspricht dabei jede Konstellation einem eigenen Zustand, worauf sich die Betrachtungsweise von Paaren als eigener Teilchenart begründet [56].

Beim Leerstellenmechanismus von Gleichung (2.20) diffundieren die Leerstellen thermisch aktiviert von einem Gitterplatz zum benachbarten Gitterplatz (unkorrelierte Diffusion) (Bild 2.21). Wenn eine Leerstelle auf ein substitutionelles Fremdatom (rot markiert) trifft, kann ein Platzwechsel zwischen diesem Atom und der Leerstelle erfolgen. Die Diffusion der Fremdatome ist damit mit den Sprüngen der Leerstellen korreliert (korrelierte Diffusion).

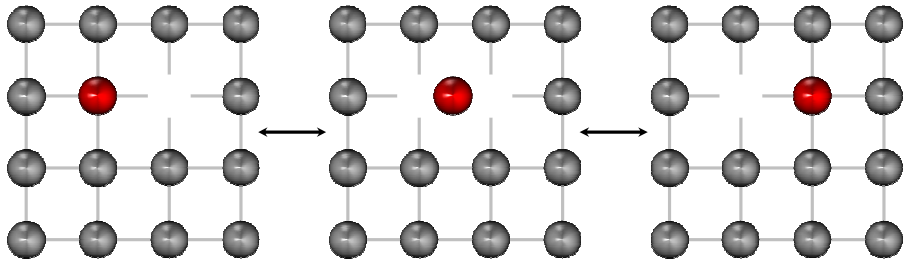


Bild 2.21: Schematische Darstellung des Leerstellendiffusionsmechanismus

Beim Zwischengitteratommechanismus von Gleichung (2.21) verdrängt ein Silizium-Zwischengitteratom ein substitutionelles Fremdatom ins Zwischengitter und nimmt selbst dessen substitutionellen Gitterplatz ein (Bild 2.22). Das Fremdatom im Zwischengitter diffundiert allerdings nicht im Zwischengitter, sondern nimmt wieder einen benachbarten substitutionellen Gitterplatz ein, indem es ein benachbartes, reguläres Gitteratom ins Zwischengitter verdrängt.

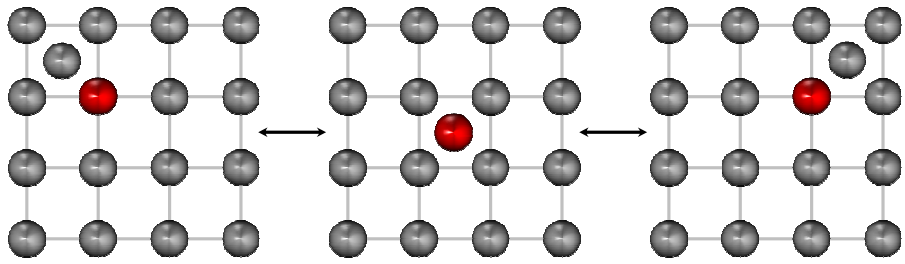


Bild 2.22: Schematische Darstellung des Zwischengitteratomdiffusionsmechanismus

Die Diffusion von Fremdatomen geschieht dabei über beide Mechanismen. Der effektive Diffusionskoeffizient D_{eff}^{A} für ein Fremdatom A lässt sich dadurch als Summe der Diffusionskoeffizienten beider Mechanismen darstellen:

$$D_{eff}^{A} = D_{AV} + D_{AI}, \tag{2.22}$$

$$D_{eff}^{A} = d_{AV} \cdot \left(\frac{C_{AV}}{C_A} \right) + d_{AI} \cdot \left(\frac{C_{AI}}{C_A} \right), \tag{2.23}$$

mit den Diffusionskoeffizienten D_{AV} und D_{AI} für die Diffusion eines Fremdatoms über Leerstellen bzw. über Silizium-Zwischengitteratome, der Konzentration C_A eines Fremdatoms A, der Konzentration C_{AV} eines Fremdatoms auf einem Gitterplatz mit einer angrenzenden Leerstelle, der Konzentration C_{AI} eines Fremdatoms auf einem Gitterplatz mit einem angrenzenden Silizium-Zwischengitteratom und dem Diffusionsvermögen d_{AV} bzw. d_{AI} eines Fremdatoms über Leerstellen bzw. über Silizium-Zwischengitteratome [56].

Je nach Dotand werden dabei unterschiedliche Diffusionsmechanismen bevorzugt. Antimon diffundiert in Silizium vorwiegend unter Wechselwirkung mit Leerstellen (≥ 97% [57]), wohingegen Bor (≥ 94% [57]), Gallium (> 88% [58]) und Phosphor (≥ 96% [57]) hauptsächlich unter Wechselwirkung mit Zwischengitteratomen wandern. Arsen hingegen diffundiert über beide Diffusionsmechanismen fast gleich stark (60% über Leerstellen und 40% über Silizium-Zwischengitteratome [57]).

Die Diffusion in Festkörpern ist die Folge vieler atomarer Platzwechselvorgänge, bei denen die diffundierenden Teilchen lokale Energiebarrieren überwinden müssen. Da diese Platzwechselvorgänge thermisch aktiviert sind, zeigt der Diffusionskoeffizient eine exponentielle Abhängigkeit von der Temperatur in der Form einer ARRHENIUS-Gleichung [59]:

$$D_{eff}^{A} = D_0 \cdot \exp\left(-\frac{E_a}{k_B \cdot T} \right). \tag{2.24}$$

Dabei ist E_a die zu überwindende Energiebarriere, die der Aktivierungsenergie für den Diffusionssprung entspricht, k_B die BOLTZMANN-Konstante und T die Temperatur. D_0 ist eine Konstante, die von der Kristallstruktur und Details des Diffusionsmechanismus abhängt. Der Faktor D_0 und die Aktivierungsenergie E_a unterscheiden sich aufgrund der unterschiedlichen Diffusionsmechanismen je nach verwendetem Dotierstoff, was zu unterschiedlichen Diffusionskoeffizienten und damit zu einer unterschiedlich starken Diffusion führt (Bild 2.23).

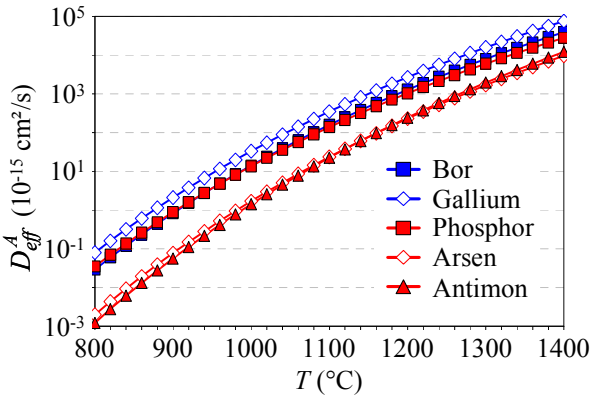


Bild 2.23: Diffusionskoeffizient D_{eff}^{A} für oft verwendete Dotierstoffe in Silizium in Abhängigkeit von der Temperatur T, berechnet mit den Werten von Tabelle 2.3. Blaue Symbole kennzeichnen Dotierstoffe für eine p-Dotierung, rote Symbole für eine n-Dotierung.

Tabelle 2.3 Diffusionskonstanten D_0 und Aktivierungsenergien E_a für oft verwendete Dotierstoffe in Silizium [60]

Dotierstoff	Bor (B)	Gallium (Ga)	Phosphor (P)	Arsen (As)	Antimon (Sb)
D_0 [cm²/s]	3.79	3.81	1.03	8.85	40.9
E_a [eV]	3.645	3.552	3.507	3.971	4.158

Da mit fortschreitender Skalierung der Bauelemente flache pn-Übergänge mit möglichst hoher Leitfähigkeit gefordert werden, ist Arsen infolge der geringeren Diffusion bzw. höheren Löslichkeit in Silizium der dominante Dotierstoff für eine n-Dotierung. Für eine p-Dotierung kommt aufgrund der hohen Festkörperlöslichkeit nur Bor in Frage. Leider hat dieser Dotierstoff ein höheres Diffusionsvermögen als Arsen, wodurch flache pn-Übergänge für eine p-Dotierung schwerer zu erreichen sind.

c) Erweiterung der FICKschen Gesetze

Die in Bild 2.23 dargestellte einfache Temperaturabhängigkeit der Diffusionskoeffizienten gilt für intrinsisches Silizium bei der entsprechenden Diffusionstemperatur, wenn eine Abhängigkeit durch andere Effekte vernachlässigt werden kann. Im Fall von intrinsischem Silizium muss bei einer entsprechenden Diffusionstemperatur für alle eingebrachten Donatoren oder Akzeptoren $N_D \leq n_i$ bzw. $N_A \leq n_i$ gelten, wobei die intrinsische Ladungsträgerkonzentration nach [61]

$$n_i = 3.9 \cdot 10^{16} \cdot \sqrt[3]{\frac{T}{K}} \cdot \exp\left(-\frac{0.605\,eV}{k_B \cdot T}\right) \mathrm{cm}^{-3} \tag{2.25}$$

beträgt und für gewöhnliche Diffusionstemperaturen von 1050 °C bereits im Bereich von mittleren Dotierstoffkonzentrationen von $9.31 \cdot 10^{18}$ cm^{-3} liegt (Bild 2.24).

Unter extrinsischen Bedingungen, wenn also bei einer entsprechenden Diffusionstemperatur die Konzentration C_A eines Fremdatoms A größer als die intrinsische Ladungsträgerkonzentration n_i ist und damit die Konzentration der Majoritätsladungsträger von der Dotierstoffkonzentration bestimmt wird, treten eine Reihe neuer Effekte in den Vordergrund, wodurch die Diffusion für alle Dotierstoffe beschleunigt wird. Maßgeblich für das Anwachsen des Diffusionskoeffizienten bei hohen Konzentrationen sind dabei die erhöhte Anzahl an Punktdefekten durch die Implantation der Dotierstoffatome und das Auftreten innerer elektrischer Felder [42].

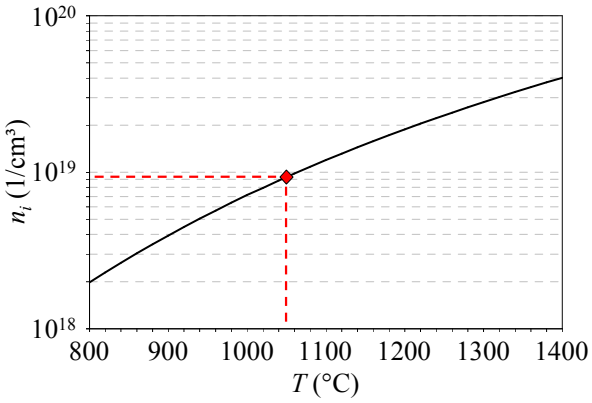


Bild 2.24: Intrinsische Ladungsträgerkonzentration n_i in Silizium in Abhängigkeit von der Temperatur T (nach [61])

Elektrische Felder entstehen bei der gekoppelten Diffusionsbewegung von verschieden geladenen Teilchen mit unterschiedlicher Beweglichkeit. Mit dem Einbau der Dotierstoffatome werden freie Ladungsträger mit dem gleichen örtlichen Gradienten erzeugt, welche aber aufgrund ihrer sehr viel höheren Beweglichkeit wesentlich schneller als die Dotierstoffatomrümpfe diffundieren. Die sich dadurch ausbildende Raumladung bewirkt ein elektrisches Feld, das einen zusätzlichen Teilchenstrom hervorruft. Unter dem Einfluss eines Potenzialgradienten erweitert sich das erste FICKsche Gesetz von Gleichung (2.16) zu:

$$\boldsymbol{J}_T = -\boldsymbol{D}\cdot\left(\operatorname{grad} C + \frac{q}{k_B\cdot T}\cdot C\cdot\operatorname{grad}\varphi\right). \tag{2.26}$$

Dementsprechend erhält man für das zweite FICKsche Gesetz von Gleichung (2.17):

$$\frac{\partial C}{\partial t} = \operatorname{div}\left(\boldsymbol{D}\cdot\left(\operatorname{grad} C + \frac{q}{k_B\cdot T}\cdot C\cdot\operatorname{grad}\varphi\right)\right). \tag{2.27}$$

Zusätzlich zeigt sich unter extrinsischen Dotierstoffkonzentrationen, dass der Diffusionskoeffizient konzentrationsabhängig ist, wodurch Gebiete mit höheren Konzentrationen schneller diffundieren, was wiederum zu boxartigen Profilen führt. Die effektiven Diffusionskoeffizienten $D_{eff,ext}^{A,n}$ und $D_{eff,ext}^{A,p}$ eines Fremdatoms A für eine n- bzw. p-Dotierung werden dann in der folgenden Form beschrieben:

$$D_{eff,ext}^{A,n} = D^0 + D^- \cdot\left(\frac{n}{n_i}\right) + D^{--}\cdot\left(\frac{n}{n_i}\right)^2 \tag{2.28}$$

bzw.

$$D_{eff,ext}^{A,p} = D^0 + D^+ \cdot\left(\frac{p}{n_i}\right) + D^{++}\cdot\left(\frac{p}{n_i}\right)^2. \tag{2.29}$$

Die Diffusionskoeffizienten D^0, D^-, D^+, D^{--} und D^{++} in den Gleichungen (2.28) und (2.29) stellen die Wechselwirkung der Dotierstoffe mit neutralen (V^0, I^0), sowie mit einfach (V^-, I^-, V^+, I^+) und zweifach (V^{--}, I^{--}, V^{++}, I^{++}) geladenen Punktdefekten dar [42]. Für den intrinsischen Fall mit $n = p = n_i$ folgt für die effektiven Diffusionskoeffizienten:

$$D_{eff,ext}^{A,n} = D^0 + D^- + D^{--} \tag{2.30}$$

bzw.

$$D_{eff,ext}^{A,p} = D^0 + D^+ + D^{++}. \tag{2.31}$$

Die einzelnen Diffusionskoeffizienten sind wiederum entsprechend einer ARRHENIUS-Gleichung (2.24) von der Temperatur abhängig.

Weiterhin ist der Einfluss mechanischer Verspannungen des Siliziumkristalls auf die Diffusion der Dotierstoffatome zu beachten. Unter dem Einfluss einer mechanischen Zug- bzw. Druckspannung verändert sich die Konzentration und Beweglichkeit von Punktdefekten, was aufgrund der angesprochenen Wechselwirkung zwischen Punktdefekten und Dotierstoffatomen wiederum Einfluss auf die Diffusion der verschiedenen Dotanden hat [62].

Ein technologisch unerwünschter Diffusionsprozess im Zusammenhang mit der Defektausheilung ist die vorübergehend verstärkte Diffusion (Transient Enhanced Diffusion, TED). Wie bereits erwähnt, entstehen während der Implantation im Siliziumkristall Gitterdefekte. Die Umwandlung und Auflösung dieser Defekte bewirkt schon bei relativ niedrigen Prozesstemperaturen eine vorübergehend verstärkte Diffusion einiger bestimmter Dotierstoffe.

Technologisch ist dieses Phänomen für die Diffusion von Bor und Phosphor von Bedeutung. Die Diffusion beider Dotierstoffe erfolgt in Silizium fast ausnahmslos über eine Wechselwirkung mit Zwischengitteratomen, so dass eine erhöhte Konzentration von Zwischengitteratomen, die bei einer Implantation prinzipiell unvermeidlich ist, zu einer verstärkten Diffusion beider Dotanden führt. Dieser Diffusionsprozess dauert so lange an, bis die Konzentration der Zwischengitteratome der Konzentration im thermodynamischen Gleichgewicht entspricht. Für flache und abrupte Profile ist dies sehr nachteilig, wodurch auch die Aktivierung dieser Dotierstoffe eine große Herausforderung ist.

2.5.3 Thermisches Budget

Wenn mehrere Temperatur- und damit Diffusionsschritte aufeinanderfolgen, was bei der Herstellung integrierter Schaltungen in der Regel der Fall ist, spricht man häufig auch von einer effektiven Diffusion, da jeder einzelne dieser Temperaturschritte das letztendliche, effektive Diffusionsprofil beeinflusst. Somit ergibt sich entsprechend Gleichung (2.18) für das Diffusionsprofil nach m-Diffusionsschritten $D_i t_i$ ($i = 1 \ldots m$)

$$C(x,t) = C(0,t) \cdot \exp\left(- \frac{x^2}{4 \cdot (Dt)_{eff}} \right) \tag{2.32}$$

mit

$$(Dt)_{eff} = \sum_{i=1}^{m} (D_i t_i) = \left(\frac{L_{D,eff}}{2} \right)^2 . \tag{2.33}$$

Die effektive Diffusionslänge $L_{D,eff}$ entspricht der Diffusionslänge L_D in Gleichung (2.19). Da der Diffusionskoeffizient D gemäß Gleichung (2.24) exponentiell von der Temperatur T abhängig ist, dominiert der Diffusionsschritt mit der höchsten Temperatur das endgültige Konzentrationsprofil.

In der Praxis wird oft das „Thermische Budget“ (*TB*) verwendet, um den Einfluss einzelner oder mehrerer Prozesse auf die Diffusion von Dotierstoffatomen zu charakterisieren [63]. Im Allgemeinen bezieht sich dabei das thermische Budget *TB* auf die Fläche unter der Temperatur-Zeit-Kurve für einen bestimmten Prozess- bzw. einer Abfolge von m-Prozessen:

$$TB = \sum_{i=1}^{m} \left(\int_{t_{i,Start}}^{t_{i,Ende}} T_i \, dt \right) . \tag{2.34}$$

2.6 Zusammenfassung

In diesem Abschnitt wurden die für diese Arbeit notwendigen Grundlagen der MOSFET-Technologie, was das Funktionsprinzip und die Herstellung eines SOI-MOSFETs sowie charakteristische Kenngrößen zur Bewertung des Transistors beinhaltet, beschrieben. Weiterhin wurden die mit der Bauelementeskalierung einhergehenden Probleme, wie die geometrie- und spannungsbedingte Verringerung der Schwellspannung U_{th}, sowie Möglichkeiten zu deren Minimierung näher betrachtet.

Eine Möglichkeit ist die Verringerung der Tiefe X_j der pn-Übergänge, was sich jedoch nachteilig auf die Schichtwiderstände R_{SDE} und R_{SD} der SDE- bzw. SD-Gebiete und somit auf den externen Widerstand R_{ext} auswirkt, dessen Einfluss bei zukünftigen Technologiegenerationen nicht mehr vernachlässigt werden kann. Durch die Erhöhung der Ladungsträgerbeweglichkeiten μ_n bzw. μ_p oder der Ladungsträgerdichten n bzw. p, die in einem hohen Maß vom Dotierprozess abhängen, können die Schichtwiderstände R_{SDE} und R_{SD} der SDE- bzw. SD-Gebiete wieder verringert werden.

Die Ionenimplantation ist das übliche Verfahren der oberflächennahen Dotierung in der planaren Halbleitertechnologie. Ein Nachteil dieser Form der Dotierung ist die mit der Implantation einhergehende Schädigung des Kristallgitters. Um die erzeugten Kristallfehler am Siliziumgitter zu rekristallisieren und die eingebrachten Dotanden zu aktivieren, werden im Wesentlichen thermische Ausheilprozesse eingesetzt, wodurch sich jedoch, aufgrund der Diffusion der eingebrachten Dotanden, alle Implantationsprofile im Halbleiterkristall verändern.

Die Diffusion der Dotierstoffatome und somit auch die Tiefe X_j der pn-Übergänge sind sowohl von der Ausheiltemperatur $T_{Ausheil}$ als auch von der Ausheilzeit $t_{Ausheil}$ abhängig. Demgegenüber werden die Aktivierungsvorgänge im Festkörper in erster Linie durch die Ausheiltemperatur $T_{Ausheil}$ des Ausheilschritts und nicht durch die Dauer der Temperaturbehandlung bestimmt. Damit lässt sich eine maximal mögliche Aktivierung der Dotierstoffe bei minimaler Diffusion nur bei sehr hohen Ausheiltemperaturen $T_{Ausheil}$ mit extrem kurzen Ausheilzeiten $t_{Ausheil}$ im Millisekundenbereich realisieren.

3 Neuartige Ausheilverfahren

Das folgende Kapitel gibt einen Überblick über das Rapid Thermal Annealing als derzeitig meist verwendetes Ausheilverfahren sowie über alternative Ausheilverfahren mit Ausheilzeiten im Millisekundenbereich. In diesem Zusammenhang wird der Einfluss dieser Kurzzeitausheilung im Vergleich zum aktuell verwendeten Ausheilverfahren auf die Diffusion und Aktivierung der Dotierstoffe für eine p- bzw. n-Dotierung dargestellt. Weiterhin werden die zwei wichtigsten Technologien der Millisekundenausheilung, die Blitzlampen-Ausheilung und die Laser-Ausheilung, näher beschrieben.

3.1 Vorbetrachtungen

Im Allgemeinen erfolgt bei der Skalierung der Bauelementeabmessungen eine Anpassung des Ausheilungsprozesses durch eine Erhöhung der Ausheiltemperatur bei gleichzeitiger Verringerung der Ausheilzeit. Diese Vorgehensweise liegt im Unterschied der thermischen Aktivierungsenergien für Diffusion und elektrische Aktivierung begründet (Bild 3.1) [64].

Die Aktivierungsenergie für die intrinsische Diffusion von Bor liegt bei 3.645 eV (Tabelle 2.3), wohingegen die Energie für die elektrische Aktivierung von Bor circa 4.7 eV beträgt [65]. Die gestrichelte Linie in Bild 3.1 stellt die Zeit dar, die bei einer bestimmten Prozesstemperatur benötigt wird, um 50% der Ladungsträger einer Borimplantation zu aktivieren [65], wohingegen die durchgehenden Linien die Zeit darstellen, um die Diffusionslängen L_D von 1 nm, 2 nm, 5 nm, 10 nm und 20 nm zu erreichen. Die Diffusionslänge L_D wurde dabei mit Gleichung (2.19) berechnet, wobei der intrinsische Diffusionskoeffizient D_{eff}^{A} für Bor nach Gleichung (2.24) mit Hilfe der Werte aus Tabelle 2.3 bestimmt wurde. Obwohl, wie in Kapitel 2 beschrieben, die Diffusion der Dotierstoffe unter realen Prozessbedingungen bedeutend größer sein kann, dienen die dargestellten Kurven zumindest als Richtwert für die zu erwartende Diffusion. Bild 3.1 zeigt, dass für Bor eine Ausheilung mit kürzerer Zeit und höheren Temperaturen von Vorteil ist, was bereits in verschiedenen Studien bestätigt wurde [65], [66]. Unter der Vorgabe einer 50%-igen Aktivierung und einer maximalen Diffusionslänge L_D von 2 nm muss nach Bild 3.1 die Ausheilzeit $t_{Ausheil}$ kürzer als 1 ms und die Ausheiltemperatur $T_{Ausheil}$ größer als 1300 °C sein.

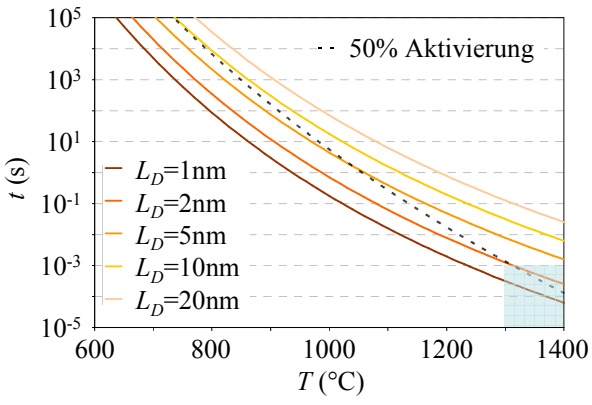


Bild 3.1: Darstellung des thermischen Budgets *TB* für verschiedene Diffusionslängen L_D von Bor sowie für eine elektrische Aktivierung von 50% für eine Borimplantation mit einer Dosis von 10^{15} cm^{-2} und einer Energie von 250 eV (nach [64])

Zum Zwecke der Restaurierung der durch die Implantation geschädigten Kristallbereiche sowie zur elektrischen Aktivierung der eingebrachten Dotierstoffe wird zurzeit bei der Transistorherstellung die schnelle thermische Ausheilung (Rapid Thermal Annealing, RTA) verwendet. Die wohl gängigste Form dieser Technologie ist die sogenannte Spike-RTA. Bei der Spike-RTA wird der Wafer auf eine bestimmte Maximaltemperatur von (900…1100)°C erwärmt, je nach gewünschter Ausheilzeit kurz auf dieser gehalten und wieder abgekühlt. Das Abkühlen des Wafers erfolgt zum Einen durch Wärmeleitung über eine aktive Kühlung des Wafer-Halters mit Hilfe von flüssigem Stickstoff und zum Anderen durch Wärmestrahlung des Wafers, wobei die Wärmeleitung über den Wafer-Halter dominiert. In Bild 3.2 sind die Temperatur-Zeit-Profile der Spike-RTA für unterschiedliche Halte- und damit Ausheilzeiten dargestellt.

Die minimale Zykluszeit bei der thermischen Ausheilung wird durch die maximal auf den Wafer abgegebene Energie und durch die minimale Reaktionszeit auf die thermische Masse der Strahlungsquelle begrenzt. Diese Einflussgrößen bestimmen die Aufheiz- und die Abkühlrate, wodurch letztendlich die Ausheilzeit $t_{Ausheil}$ definiert wird. In Bild 3.3 ist die Definition der Ausheilzeit $t_{Ausheil}$ dargestellt, die durch die Schnittpunkte der Kurve mit der Temperatur T_{max} - 50 K bestimmt wird. Bei der derzeitig angewandten Spike-RTA, welche traditionell Wolfram-Halogenlampen als Strahlungsquelle nutzen, sind aufgrund der angesprochenen Faktoren minimale Ausheilzeiten von etwa 1 s möglich [67].

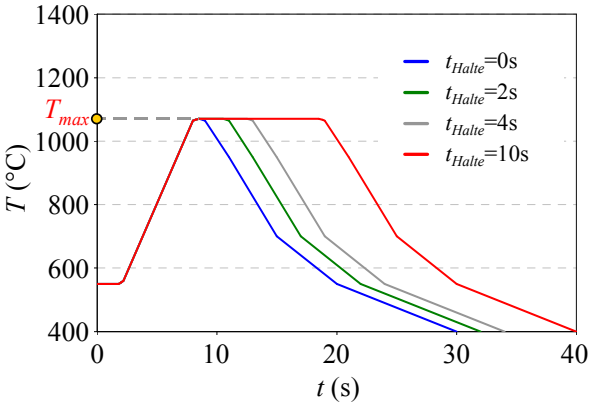


Bild 3.2: Temperatur-Zeit-Profil der Spike-RTA für unterschiedliche Halte- und damit Ausheilzeiten

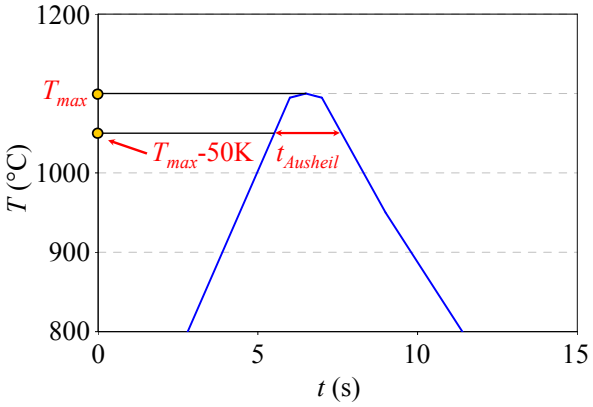


Bild 3.3: Definition der Ausheilzeit $t_{Ausheil}$

Diese Einschränkungen hinsichtlich der Ausheilzeit haben die Forschung nach neuen Verfahren mit Ausheilzeiten im Millisekundenbereich angeregt. Da bei solch schnellen thermischen Prozessen nur die oberste Schicht des Wafers (etwa 100 µm) durch einen Energieimpuls erhitzt wird und nicht wie bisher der gesamte Wafer, können so Aufheiz- und Abkühlraten im Bereich von 10^6 K/s erreicht werden (Bild 3.4). Dabei wird ausgenutzt, dass der Wafer selbst als Kühlkörper bzw. Wärmesenke dient und die Wärme bedeutend schneller und effizienter als bei einer aktive Waferkühlung von der Oberfläche abgeführt werden kann [68]. Eine nur oberflächennahe Erwärmung des Wafers tritt auf, wenn die Absorptionslänge d_{absorp} kleiner als die thermische Diffusionslänge

$$L_{D,heat} \approx \sqrt{\frac{\kappa}{\rho \cdot c_p} \cdot t_{Ausheil}} \tag{3.1}$$

ist [69]. Dabei ist ρ die Dichte, c_p die spezifische Wärmekapazität bei konstantem Druck und κ die Wärmeleitfähigkeit.

Eine schnelle Abkühlung des Wafers ist dann möglich, wenn die Waferdicke d_{Wafer} wiederum bedeutend größer als die thermische Diffusionslänge $L_{D,heat}$ ist. Unter Annahme einer Ausheilzeit $t_{Ausheil}$ von 1 ms und einer Ausheiltemperatur $T_{Ausheil}$ von 1200 °C liegt die thermische Diffusionslänge $L_{D,heat}$ für Silizium nach Gleichung (3.1) bei etwa 100 µm. Aus diesem Grund bleibt das Substrat wesentlich kühler als die Waferoberfläche und wirkt als Wärmesenke. Bild 3.5 stellt die sich daraus ergebende Temperaturverteilung über dem Wafer für eine Ausheil- und Wafertemperatur von 1100°C bzw. 450 °C für verschiedene Zeitpunkte während der Kurzzeitausheilung dar.

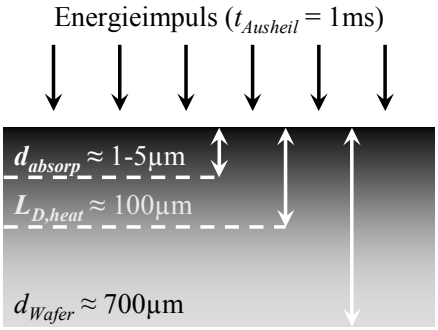


Bild 3.4: Geometrische Wärmeverhältnisse im Silizium-Substrat bei der Kurzzeitausheilung ($t_{Ausheil}$ = 1 ms)

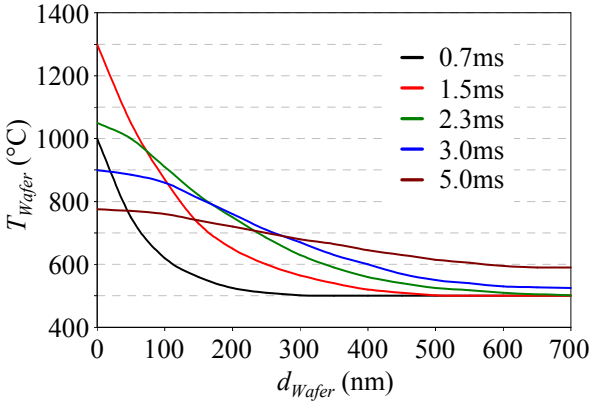


Bild 3.5: Berechnete Temperaturverteilung über dem Wafer für verschiedene Zeitpunkte während der Kurzzeitausheilung (nach [70])

Der angesprochene Temperaturgradient über dem Wafer hat aber nicht nur Vorteile. Aufgrund des großen Temperaturunterschieds zwischen Ober- und Unterseite des Wafers kommt es durch die thermische Ausdehnung der wärmeren Schichten zur Ausbildung einer thermischen Verspannung σ des Wafers (Bild 3.6). Die einseitige Erwärmung des Wafers führt zu einer Druckverspannung σ_{xx} an der Waferoberseite, die bei einer Ausheiltemperatur $T_{Ausheil}$ von 1100 °C und einer Wafertemperatur T_{Wafer} von 450 °C bis 400 MPa erreichen kann. Die maximale Zugverspannung σ_{xx} im Substrat beträgt ungefähr 100 MPa. Diese Verspannungen können Defekte am Kristall verursachen, die zu einer Verformung des Wafers und im schlimmsten Fall zum Waferbruch führen.

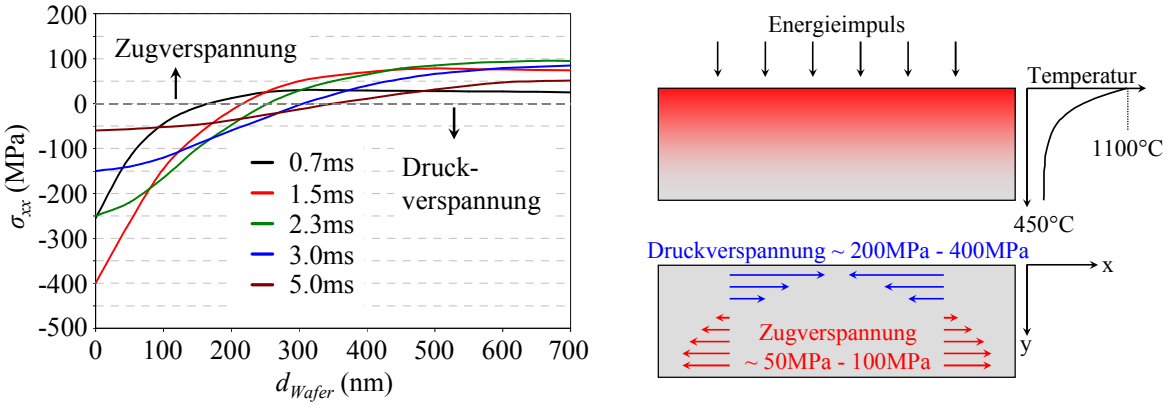


Bild 3.6: Berechnete Verspannungsverteilung σ_{xx} über dem Wafer für verschiedene Zeitpunkte während der Kurzzeitausheilung (nach [70])

Die Stärke der Verspannung σ ist sehr stark vom Temperaturgradienten über dem Wafer abhängig. Eine höhere Wafertemperatur T_{Wafer} bei gleicher Ausheiltemperatur $T_{Ausheil}$ verringert die thermische Verspannung σ des Wafers, wohingegen eine höhere Ausheiltemperatur $T_{Ausheil}$ bei gleicher Wafertemperatur T_{Wafer} die thermische Verspannung σ verstärkt.

Der Vorteil der Wärmeableitung über den Wafer geht ab Energieimpulslängen von mehr als 20 ms verloren, was dem Wert der thermischen Zeitkonstanten eines Silizium-Wafers entspricht (10…20) ms. In diesem Fall erfolgt nicht nur eine oberflächennahe Erwärmung des Wafers, sondern auch eine signifikante Temperaturerhöhung des gesamten Substrates, wodurch diffusionsarme Dotierstoffprofile mit hoher elektrischer Aktivierung schlecht realisierbar sind [71]. Bild 3.7 zeigt zum Vergleich die erreichbaren Temperatur-Zeit-Profile der derzeitig angewandten Spike-RTA und der neuartigen thermischen Kurzzeitprozesse.

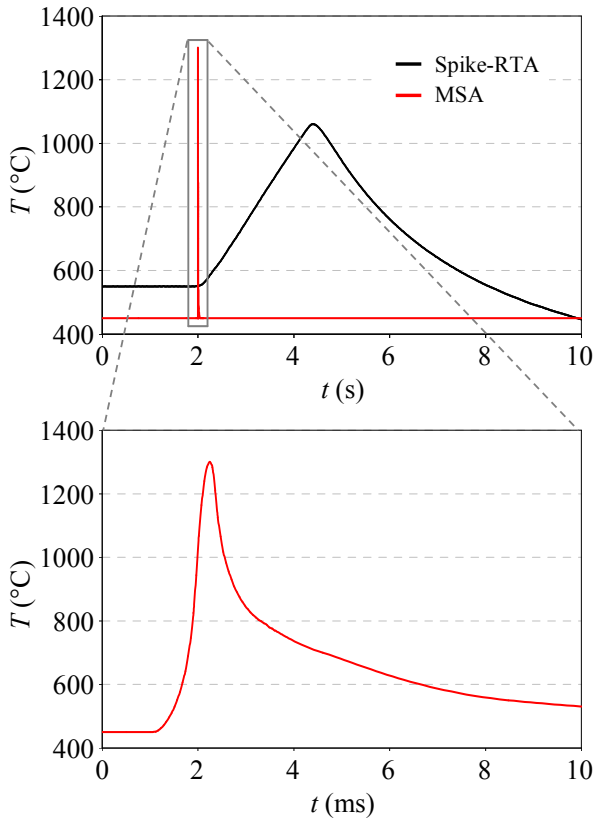


Bild 3.7: Vergleich der Temperatur-Zeit-Profile der Spike-RTA und der Kurzeitausheilung (Millisekundenausheilung, MSA) auf der Waferoberfläche

3.2 Verwendete Ausheiltechnologien

Im Zuge dieser Arbeit werden zwei unterschiedlichen Technologien der Kurzzeitausheilung, die Blitzlampen-Ausheilung (Flash Lamp Annealing, FLA) und die Laser-Ausheilung (Laser Annealing, LSA), verwendet.

3.2.1 Blitzlampen-Ausheilung

Bild 3.8 zeigt eine schematische Darstellung der verwendeten Anlage für die Blitzlampen-Ausheilung. Die vorgestellte Anlage arbeitet mit Xenon-Blitzlampen, die sich über dem Wafer am oberen Ende der Anlagenkammer befinden. Der Wafer selbst wird auf einem Wafer-Halter, dem sogenannten Chuck, gelagert. Die Prozessierung des Wafers erfolgt hauptsächlich durch zwei Prozessschritte.

Zuerst wird der Chuck mit dem Wafer auf eine Temperatur von (400…600) °C erhitzt. Die Möglichkeit der Variation der Wafertemperatur hat den Vorteil, dass die während der Kurzzeitausheilung zwischen der Oberfläche des Wafers und dem Substrat entstehende thermische Verspannung des Wafers minimiert werden kann. Im zweiten Schritt entlädt sich eine Kondensatorbank über die Blitzlampen, die sich über dem Wafer befinden. Dabei wird nur die Oberseite, also die Bauelementeseite, des Wafers belichtet. Die bei der Blitzlampen-Ausheilung verwendeten Xenon-Blitzlampen haben ein Spektrum im Wellenlängenbereich von $\lambda = (0.3…0.6)$ µm, wodurch die emittierte Strahlung im Silizium im Gegensatz zu den bei der Spike-RTA verwendeten Wolfram-Halogenlampen fast zu 100 % durch Interband-Übergänge absorbiert wird (Bild 3.9).

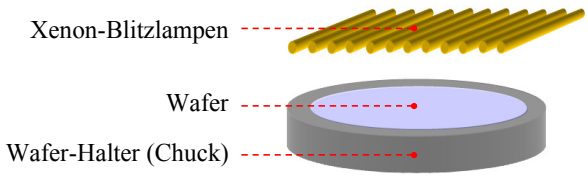


Bild 3.8: Aufbau der Anlage für die Blitzlampen-Ausheilung [5]

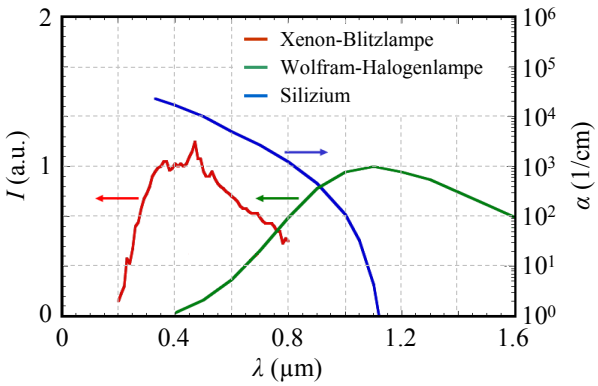


Bild 3.9: Normierte Emissionsspektren einer Xenon-Blitzlampe und einer Wolfram-Halogenlampe sowie wellenlängenabhängige Absorption a von Silizium (nach [5])

3.2.2 Laser-Ausheilung

Bild 3.10 zeigt eine schematische Darstellung der verwendeten Anlage für die Laser-Ausheilung. Bei dieser Anlage wird ein ungepulster CO_2-Laser mit einer Wellenlänge λ im mittleren infraroten Bereich von (9.8…10.4) µm verwendet. Aufgrund der Wellenlänge, die deutlich über dem Bereich des sichtbaren Lichts liegt, kann die Strahlung im Gegensatz zu Festkörperlasern oder Diodenlasern nicht in Glasfasern geleitet werden, da es keine ausreichend dämpfungsarmen Materialien für diese Wellenlänge gibt. Das Licht wird daher mit Umlenkspiegeln geführt.

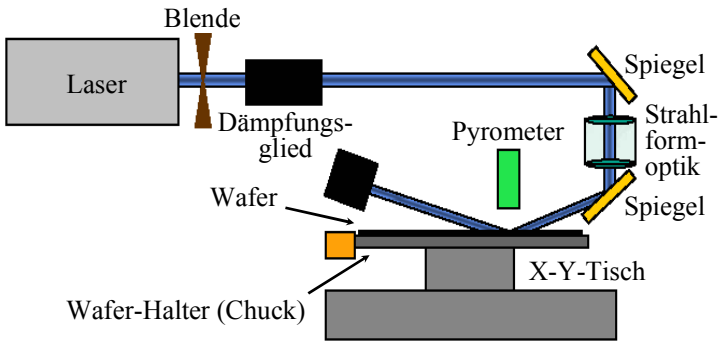


Bild 3.10: Aufbau der Anlage für die Laser-Ausheilung [8]

Nach der Fokussierung des Lasers auf den Wafer mit Hilfe eines Linsensystems (Strahlformoptik) hat der Laserstrahl ein gaußförmiges Profil mit einer Laserstrahllänge $L_{Laser} \approx 100$ µm und einer Laserstrahlweite $W_{Laser} \approx 10$ mm (Bild 3.11).

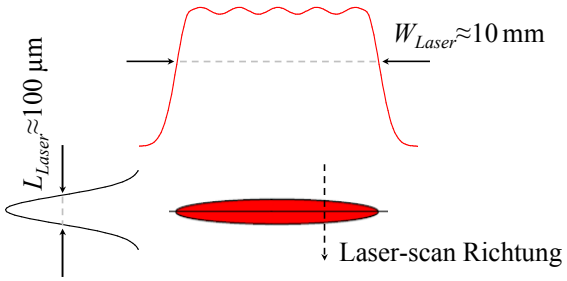


Bild 3.11: Profil des Laserstrahles auf dem Wafer nach der Fokussierung durch die Strahlformoptik

Der Chuck selbst befindet sich auf einem X-Y-Tisch. Bei einer gleichförmigen Geschwindigkeit des X-Y-Tischs wird der Wafer unter dem ortsfesten Laserstrahl abgetastet und lokal erwärmt, wodurch die belichteten Flächen ausgeheilt werden. Die Ausheiltemperatur $T_{Ausheil}$ wird durch die Laserleistung bestimmt, während die Ausheilzeit $t_{Ausheil}$ durch die Geschwindigkeit des Tischs v_{Tisch} und dem Profil des Laserstrahls definiert wird. Bei einer Tischgeschwindigkeit $v_{Tisch} = 125$ mm/s und einer Laserstrahllänge $L_{Laser} = 100$ µm ergibt sich somit eine Halte- und damit Ausheilzeit:

$$t_{Ausheil} = \frac{\text{Laserstrahllänge}}{\text{Geschwindigkeit des Tischs}} = \frac{L_{Laser}}{v_{Tisch}} \tag{3.2}$$

von 800 µs. Bild 3.12 zeigt ein berechnetes Temperaturprofil für eine Ausheilzeit von 200 µs. Im Gegensatz zu den Blitzlampen-Anlagen, bei denen der gesamte Wafer auf einmal belichtet wird, erfolgt bei der Laser-Ausheilung nur eine lokale Erwärmung der Waferoberfläche, wodurch neben dem vertikalem auch ein lateraler Temperaturgradient entsteht.

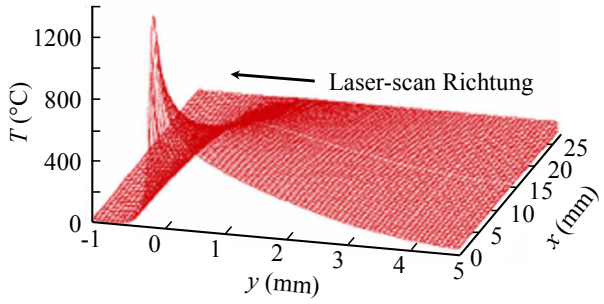


Bild 3.12: Berechnetes Temperaturprofil für eine Ausheilzeit von 200 µs (nach [8])

Aufgrund der bei der Laser-Ausheilung verwendeten Wellenlänge im infraroten Bereich erfolgt die Absorption der einfallenden Strahlung durch Intraband-Übergänge mit Hilfe der Freien-Ladungsträger-Absorption (Free Carrier Absorption, FCA). Dabei werden Photonen durch die Wechselwirkung mit freien Ladungsträgern im Leitungs- oder Valenzband absorbiert. Der Absorptionskoeffizient α_{FCA} der Freien-Ladungsträger-Absorption beträgt nach [75] für Löcher

$$\alpha_{FCA} = \frac{q^3}{4 \cdot \pi^2 \cdot \varepsilon_0 \cdot c^3 \cdot n_B \cdot m_p^{eff\,2} \cdot \mu_p} \cdot \lambda^2 \cdot p \tag{3.3}$$

mit c der Lichtgeschwindigkeit, n_B dem Brechungsindex, m_p^{eff} der effektiven Löchermasse und λ der Wellenlänge.

Unter Verwendung von n, m_n^{eff} und μ_n anstelle von p, m_p^{eff} und μ_p ist Gleichung (3.3) auch auf Elektronen anwendbar. Wesentlich für die Absorption der einfallenden Strahlung ist dabei die quadratische Abhängigkeit der Freien-Ladungsträger-Absorption von der Wellenlänge λ sowie die Proportionalität zur Ladungsträgerkonzentration n bzw. p.

Bild 3.13 zeigt den Absorptionskoeffizienten α_{FCA} für Silizium in Abhängigkeit von der Wellenlänge λ für verschiedene Löcherdichten p nach Gleichung (3.3). Die für die Absorption notwendigen freien Ladungsträger werden durch eine thermische Anregung, in diesem Fall durch die Erwärmung des Chucks, generiert. Die erzeugten Löcherdichten von $2.56 \cdot 10^{15}$ cm^{-3}, $2.01 \cdot 10^{16}$ cm^{-3} und $9.54 \cdot 10^{16}$ cm^{-3} entsprechen nach Gleichung (2.25) jeweils Temperaturen von 300 °C, 400 °C und 500 °C.

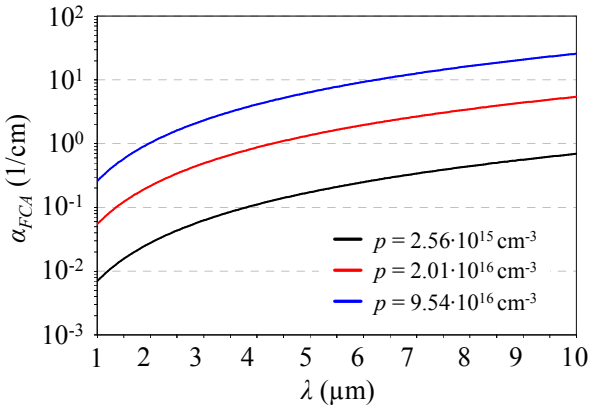


Bild 3.13: Absorptionskoeffizient α_{FCA} in Silizium nach Gleichung (3.3) in Abhängigkeit von der Wellenlänge λ und der Löcherdichte p

3.3 Einfluss auf die Dotierstoffverteilung

Im Folgenden wird der Einfluss der Kurzzeitausheilung am Beispiel der Laser-Ausheilung auf die Diffusion und Aktivierung der Dotierstoffe für eine p- und n-Dotierung dargestellt. Zu diesem Zweck wurden verschiedene Experimente auf unstrukturierten Wafern durchgeführt. Die Ausheilung erfolgte entweder mit Hilfe der konventionellen RTA, mit Hilfe der Laser-Ausheilung oder durch eine Kombination von RTA und Laser-Ausheilung. Die Ausheilzeit $T_{Ausheil}$ und -temperatur $t_{Ausheil}$ betragen bei der RTA 1.6 s bzw. 1070 °C und bei der Laser-Ausheilung etwa 1 ms bzw. 1300 °C. Das Dotierungsprofil wurde mit Hilfe der Sekundärionen-Massenspektrometrie (SIMS) bestimmt. Die Messung des Schichtwiderstands $R_{Schicht}$ erfolgte mit Hilfe der Vierspitzenmethode.

3.3.1 p-Dotierung

Bild 3.14 zeigt den Einfluss der verschiedenen Ausheilverfahren auf das Dotierungsprofil sowie den Schichtwiderstand $R_{Schicht}$ einer Borimplantation in ein mit Xenon voramorphisiertes Silizium. Die Laser-Ausheilung (LSA) hat in Verbindung mit der RTA keine weitere Diffusion der Dotanden zur Folge. Der Schichtwiderstand $R_{Schicht}$ verringert sich aber durch die zusätzliche Laser-Ausheilung ebenfalls nicht weiter, was auf eine vollständige Aktivierung der Dotanden bereits nach der RTA schließen lässt. Die Verwendung der Laser-Ausheilung als alleinigen Ausheilschritt verkürzt die Diffusionslänge von Bor bei einer Dotierstoffkonzentration C_B von $5.0 \cdot 10^{18}$ cm^{-3} im Vergleich zur RTA von 39 nm auf 13 nm bei einem nur geringfügig höheren Schichtwiderstand $R_{Schicht}$. Alle ausgeheilten Profile zeigen einen Knick in der Dotierstoffkonzentration C_B bei etwa 6 nm, der den unbeweglichen vom beweglichen Anteil trennt. Der unbewegliche Anteil der Dotierstoffkonzentration C_B umfasst in Clustern gebundene Boratome.

Unter der Annahme, dass nur substitutionell gebundene Boratome zur Diffusion beitragen, erhöht sich die Löslichkeit von Bor von ungefähr $1.0 \cdot 10^{20}$ cm^{-3} (RTA) auf etwa $5.0 \cdot 10^{20}$ cm^{-3} (LSA). Der vergleichbare Schichtwiderstand $R_{Schicht}$ in Verbindung mit der stark verringerten Diffusion für die Laser-Ausheilung lässt sich somit durch die höhere Aktivierung erklären.

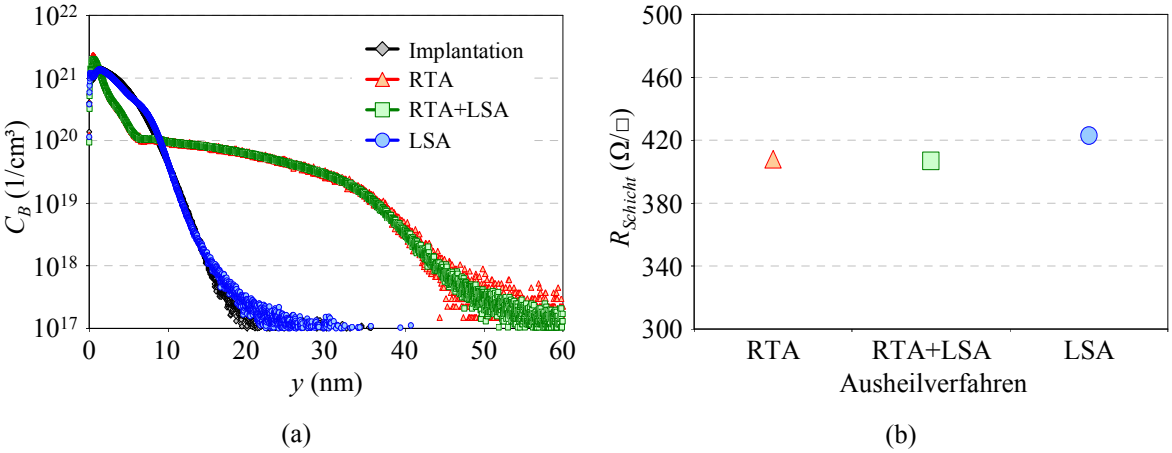


Bild 3.14: Einfluss verschiedener Ausheilverfahren auf (a) das Dotierungsprofil und (b) den Schichtwiderstand $R_{Schicht}$ einer Borimplantation (B, 0.5 keV, $1.0 \cdot 10^{15}$ cm^{-2}) in ein mit Xenon voramorphisiertes Silizium (Xe, 10 keV, $5.0 \cdot 10^{13}$ cm^{-2})

Bild 3.15 zeigt den Einfluss unterschiedlicher Ausheiltemperaturen $T_{Ausheil}$ unter Verwendung der Laser-Ausheilung als alleinigen Ausheilschritt auf das Dotierungsprofil sowie den Schichtwiderstand $R_{Schicht}$ einer Borimplantation in kristallines Silizium. Die Erhöhung der Ausheiltemperatur $T_{Ausheil}$ hat eine kontinuierliche Verringerung des Schichtwiderstands $R_{Schicht}$ bei einer gleichzeitig nur geringfügig größeren Diffusionslänge zur Folge.

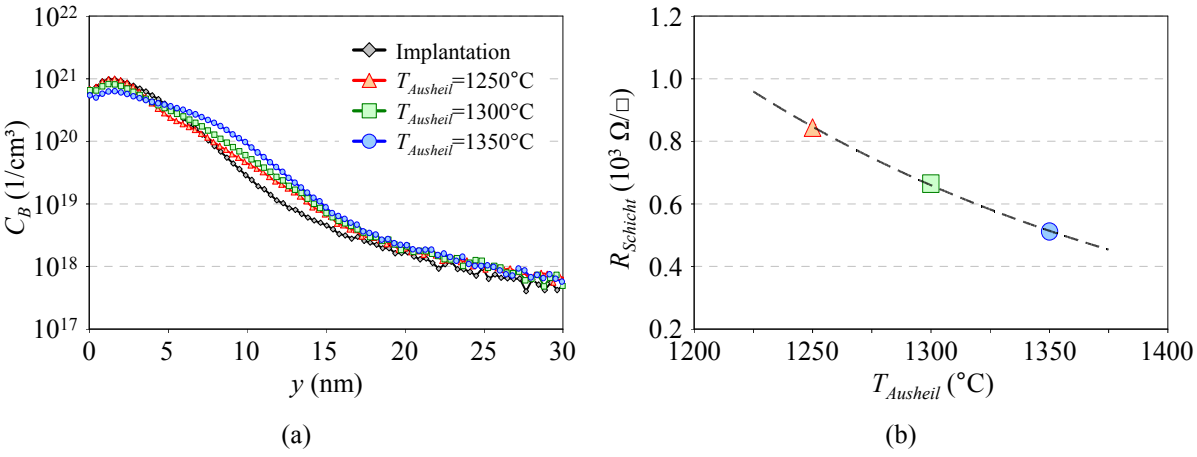


Bild 3.15: Einfluss unterschiedlicher Ausheiltemperaturen $T_{Ausheil}$ der Laser-Ausheilung auf (a) das Dotierungsprofil und (b) den Schichtwiderstand $R_{Schicht}$ einer Borimplantation (B, 0.5 keV, $0.7 \cdot 10^{15}$ cm^{-2}) in kristallines Silizium

Aufgrund des größeren thermischen Budgets TB mit steigender Ausheiltemperatur $T_{Ausheil}$ kommt es zur Auflösung der BIC, wie man anhand der sinkenden Dotierstoffkonzentration C_B im Bereich unterhalb von 5 nm erkennen kann.

Dadurch erhöht sich die Anzahl substitutionell gebundener Boratome von circa $1.5 \cdot 10^{20}$ cm^{-3} für eine Ausheiltemperatur $T_{Ausheil}$ von 1250 °C auf etwa $3.0 \cdot 10^{20}$ cm^{-3} für eine Ausheiltemperatur $T_{Ausheil}$ von 1350 °C. Ein weiterer positiver Nebeneffekt dieser Cluster-Auflösung ist die Verbesserung der Kristallqualität und die damit einhergehende höhere Ladungsträgerbeweglichkeit. Beide Effekte führen schließlich zu dem mit steigender Ausheiltemperatur $T_{Ausheil}$ sinkendem Schichtwiderstand $R_{Schicht}$.

3.3.2 n-Dotierung

Bild 3.16 zeigt den Einfluss verschiedener Ausheilverfahren auf das Dotierungsprofil sowie den Schichtwiderstand $R_{Schicht}$ einer Arsenimplantation in ein mit Xenon voramorphisiertes Silizium. Ähnlich der p-Dotierung hat die Laser-Ausheilung und damit Kurzzeitausheilung in Verbindung mit der RTA keine weitere Diffusion der Dotanden zur Folge. Im Gegensatz zur p-Dotierung verringert sich jedoch der Schichtwiderstand $R_{Schicht}$ durch die zusätzliche Ausheilung um etwa 15 % von 353 Ω/□ auf 299 Ω/□ infolge der höheren elektrisch aktivierten Dotierstoffkonzentration C_{As} von circa $7.0 \cdot 10^{20}$ cm^{-3} für den Prozess RTA+LSA im Vergleich zu $6.0 \cdot 10^{20}$ cm^{-3} für den RTA-Prozess. Die Verwendung der Laser-Ausheilung als alleinigen Ausheilschritt verkürzt die Diffusionslänge von Arsen bei einer Dotierstoffkonzentration C_{As} von $5.0 \cdot 10^{18}$ cm^{-3} im Vergleich zur RTA von 25 nm auf 16 nm. Dabei erhöht sich jedoch der Schichtwiderstand $R_{Schicht}$ um etwa 14 % von 353 Ω/□ auf 404 Ω/□. Im Vergleich zum Prozess von RTA+LSA erhöht sich der Schichtwiderstand $R_{Schicht}$ sogar um circa 35 %. Diese Ergebnisse unterscheiden sich zu denen der p-Dotierung, bei der man einen stärkeren Einfluss der Laser- und damit Kurzzeitausheilung auf die Diffusionslänge bei einem vergleichbaren Schichtwiderstand $R_{Schicht}$ feststellen konnte.

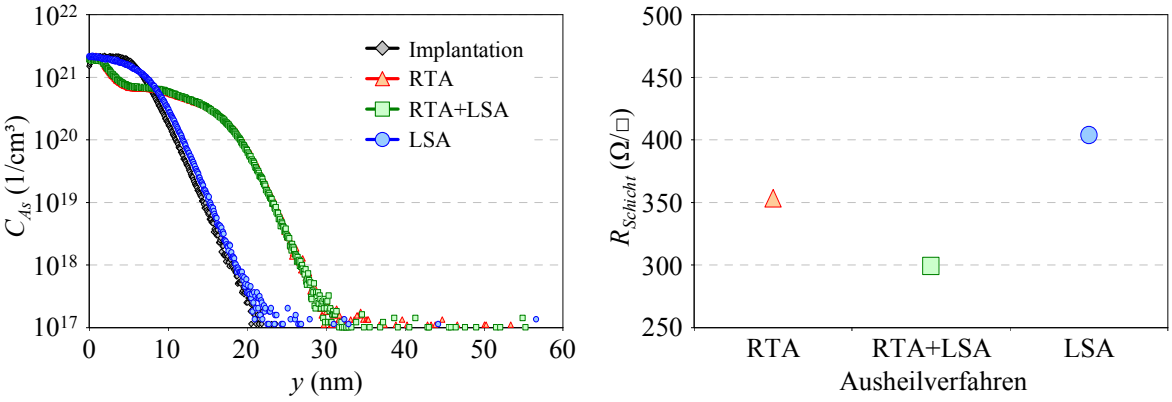
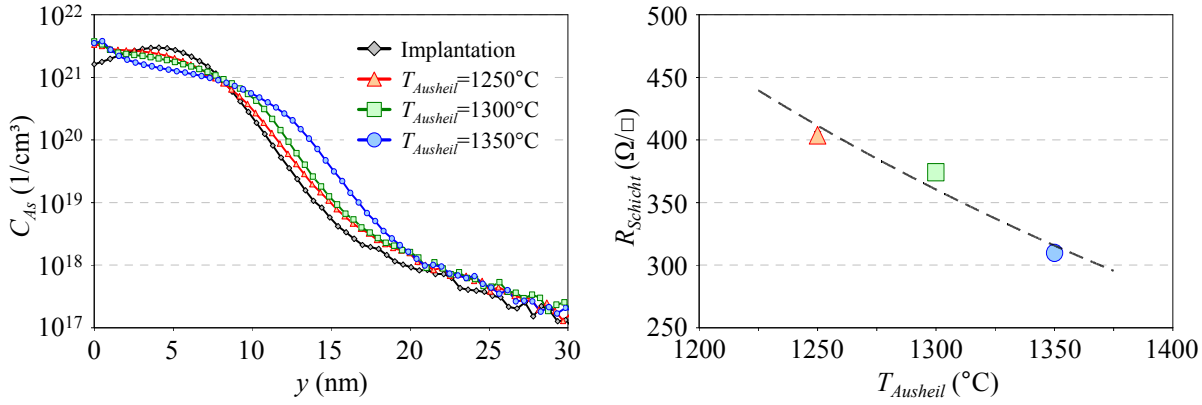


Bild 3.16: Einfluss verschiedener Ausheilverfahren auf (a) das Dotierungsprofil und (b) den Schichtwiderstand $R_{Schicht}$ einer Arsenimplantation (As, 3 keV, $2.0 \cdot 10^{15}$ cm^{-2}) in ein mit Xenon voramorphisiertes Silizium (Xe, 45 keV, $1.3 \cdot 10^{14}$ cm^{-2}).

Bild 3.17 zeigt den Einfluss unterschiedlicher Ausheiltemperaturen $T_{Ausheil}$ unter Verwendung der Laser-Ausheilung als alleinigen Ausheilschritt auf das Dotierungsprofil sowie den Schichtwiderstand $R_{Schicht}$ einer Arsenimplantation in kristallines Silizium. Alle ausgeheilten Profile zeigen die gleichen Merkmale. Eines davon ist der signifikante Dosisverlust, was auf ein Ausgasen von Arsen während der Kurzzeitausheilung hindeutet. Die Verschiebung des Maximums des Dotierungsprofils in Richtung Waferoberfläche kann durch den Schneepflug-Effekt während der Rekristallisation der durch den Implantationsprozess amorphisierten Schicht erklärt werden [72].

Weiterhin hat die stetige Erhöhung der Ausheiltemperatur $T_{Ausheil}$ eine kontinuierliche Verringerung des Schichtwiderstands $R_{Schicht}$ bei einer gleichzeitig nur geringfügig größeren Diffusionslänge zur Folge. Nach [72] ist der verringerte Schichtwiderstand $R_{Schicht}$ mit steigender Ausheiltemperatur $T_{Ausheil}$ die Folge einer verbesserten Kristallqualität aufgrund des höheren thermischen Budgets *TB* und der damit einhergehenden höheren Ladungsträgerbeweglichkeit.



(a) (b)

Bild 3.17: Einfluss unterschiedlicher Ausheiltemperaturen $T_{Ausheil}$ der Laser-Ausheilung auf (a) das Dotierungsprofil und (b) den Schichtwiderstand einer Arsenimplantation (As, 3 keV, $2.0 \cdot 10^{15}$ cm^{-2}) in kristallines Silizium

3.3.3 Defektentwicklung

Untersuchungen mit Hilfe des TEM ergaben, dass unabhängig vom verwendeten Dotierstoff in Verbindung mit einer PAI selbst nach einer Kurzzeitausheilung bei 1300 °C das thermische Budget *TB* dieser Ausheiltechnologie nicht ausreicht, um alle EOR-Defekte vollständig bzw. entsprechend einem RTA-Prozess in energetisch stabilere Konfigurationen aufzulösen [73]. Selbst Proben ohne vorherige PAI zeigen, dass trotz der hohen Ausheiltemperatur während der Kurzzeitausheilung Defekte, hier vor allem am Übergang vom unbeweglichen zum beweglichen Anteil der Dotierstoffverteilung, nicht vollständig aufgelöst sind [74].

3.4 Vergleich der Ausheiltechnologien

Die Absorption der einfallenden Strahlung ist stark von den Eigenschaften und der Geometrie der verwendeten Schichten auf der Waferoberfläche abhängig. Siliziumoxid z.B. erhöht die optische Kopplung durch seine geringe Reflexion der einfallenden Strahlung im Vergleich zu Silizium. Bild 3.18 zeigt Beispiele für die bei einem Mikroprozessor verwendeten Schichtstapel.

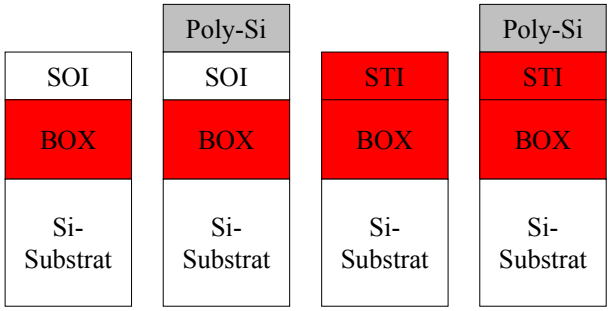


Bild 3.18: Beispiele für die bei einem Mikroprozessor verwendeten Schichtstapel

Im Fall von Ausheilungstechniken, die auf Lichteinstrahlung basieren, was die Blitzlampen- und Laser-Ausheilung mit einschließt, können die unterschiedlichen optischen Eigenschaften von Bereichen mit verschiedenen Schichtstapeln erhebliche lokale Temperaturunterschiede über dem Die verursachen. Diese auch als „Pattern"-Effekt bezeichnete strukturbedingte Temperaturvariation kann wiederum eine größere Variation der Transistorparameter über jeden einzelnen Mikroprozessor zur Folge haben, was letztendlich nachteilig für das Produkt sein kann.

Zur Quantifizierung der Temperaturvariation in Abhängigkeit von verschiedenen Waferoberflächen wurden bereits spezielle Teststrukturen entwickelt, welche die unterschiedlichen Bereiche innerhalb eines Mikroprozessors abbilden sollen [76], [77], [78]. Dabei konnte gezeigt werden, dass in Abhängigkeit von der verwendeten Ausheiltechnologie unterschiedlich starke Temperaturschwankungen auftreten. In [76] wird zur Ausheilung ein Laser mit einer Wellenlänge von 808 nm verwendet, wodurch die Ausheiltemperatur um 320 °C variiert. In [77] und [78] werden unter Verwendung der im vorherigen Abschnitt beschriebenen Xenon-Blitzlampen- (Xe-FLA) und CO_2-Laser-Ausheilung (CO_2-LSA) Temperaturvariationen von 100 °C bzw. 10 °C festgestellt. Der Grund für diese unterschiedliche Variation zwischen den einzelnen Ausheiltechniken ist in der verwendeten Wellenlänge zu finden. Durch die in [76] verwendete Wellenlänge des Lasers kommt es zu konstruktiven und destruktiven Interferenzen infolge von Reflexionen an Grenzflächen und zu Reflexionen zwischen benachbarten Polysiliziumbahnen. Diese Effekte sind vor allem dann sehr ausgeprägt, wenn die Dimensionen der verwendeten Strukturen kleiner als die verwendete Wellenlänge sind. Dieses Phänomen führt dann zu einer lokalen Absorptionserhöhung, welche Werte von fast Eins erreichen kann (Bild 3.19). Sind die Strukturbreiten jedoch größer als die einfallende Wellenlänge, verhält sich optisch gesehen, die Oberfläche des Wafers wie ein unstrukturierter Wafer mit den entsprechenden Materialparametern.

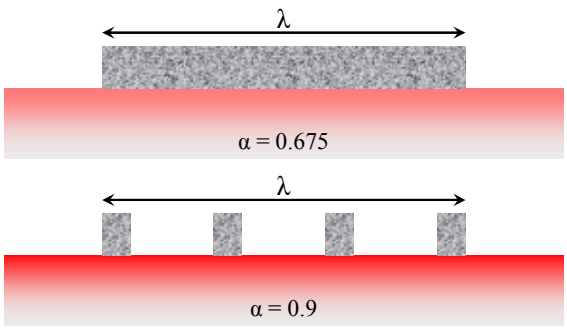


Bild 3.19: Veränderung der Absorption α als Funktion der Strukturdimensionen und Periodizität (nach [76])

Die breitbandige Lichtquelle der Xe-FLA verringert diesen Effekt zwar, kann ihn aber nicht vollständig unterdrücken. Im Gegensatz dazu werden bei der CO_2-LSA durch die Verwendung der Wellenlänge im mittleren infraroten Bereich diese Interferenzprobleme mit der vorhandenen Topographie fast vollständig unterdrückt. Weiterhin werden durch die Einkopplung des Laserstrahles unter dem BREWSTER-Winkel die Unterschiede im Absorptionsvermögen der verschiedenen Schichten (Si, SiO_2, Polysilizium) verringert, wodurch die Sensibilität gegenüber Prozessvariationen verkleinert wird [79].

Ziel ist es jedoch, unabhängig von der verwendeten Ausheiltechnologie diese lokalen Temperaturvariationen zu reduzieren und im besten Fall vollständig zu beseitigen. Dazu ist es notwendig, die gesamte Waferoberfläche so zu gestalten, dass keine bzw. nur eine sehr geringe Variation in der Lichtabsorption aufgrund unterschiedlicher optischer Eigenschaften der Waferoberfläche stattfindet.

Zur Lösung dieses Problems existieren zwei Möglichkeiten:

1. Verwendung einer Absorptionsschicht mit hoher thermischer Leitfähigkeit,
2. Entwicklung spezieller Entwurfsregeln zur Integration mehrerer zusätzlicher Polysiliziumbahnen.

Im Fall der Verwendung einer Absorptionsschicht wird vor der Kurzzeitausheilung eine Schicht mit einer hohen thermischen Leitfähigkeit über dem gesamten Wafer abgeschieden und nach der Ausheilung wieder vollständig entfernt. Obwohl diese Form der Verringerung der Temperaturvariation leicht zu bewerkstelligen ist, sind die aufgrund der zusätzlichen Prozessschritte für die Abscheidung und Entfernung der Schicht höheren Herstellungskosten nicht zu vernachlässigen. Demgegenüber werden die Produktionskosten durch die Implementierung zusätzlicher Polysiliziumbahnen nicht signifikant erhöht, wohingegen diese Möglichkeit komplexer in Bezug auf Entwicklung und Entwurfsregeln ist. Unter Verwendung der Kurzzeitausheilung wurde zur Reduzierung der Variation der Lichtabsorption die zweite Möglichkeit verwendet. Zur Bewertung von Prozessänderungen, was in diesem Fall die zusätzliche Kurzzeitausheilung ist, werden im Allgemeinen die in Bild 2.4 dargestellten Teststrukturen genutzt, wodurch sich die Variation bestimmter Transistorparameter über dem Wafer analysieren lässt (Waferebene). Aussagen über die Parametervariation innerhalb eines Mikroprozessors bzw. über dem Die sind damit nicht möglich.

Die Bilder 3.20 und 3.21 zeigen den Einfluss der Blitzlampen- und Laser-Ausheilung auf die Universalkurve sowohl in Kombination mit der RTA als auch als alleinigen Ausheilschritt. Beide Ausheilverfahren, die CO_2-LSA und die Xe-FLA, weisen nach Einstellung der Anlagenparameter die gleiche Leistungsfähigkeit für n- und p-MOSFETs unabhängig von der verwendeten Ausheilsequenz auf. Beim Schichtwiderstand R_{SDE} der SDE-Gebiete von n-MOSFETs, der sehr stark auf eine Temperaturvariation reagiert, ist ebenfalls kein Unterschied zwischen beiden Ausheilverfahren zu erkennen (Bild 3.22). Diese Ergebnisse lassen zumindest auf Waferebene auf vergleichbare Ausheiltemperaturen von der CO_2-LSA und der Xe-FLA schließen.

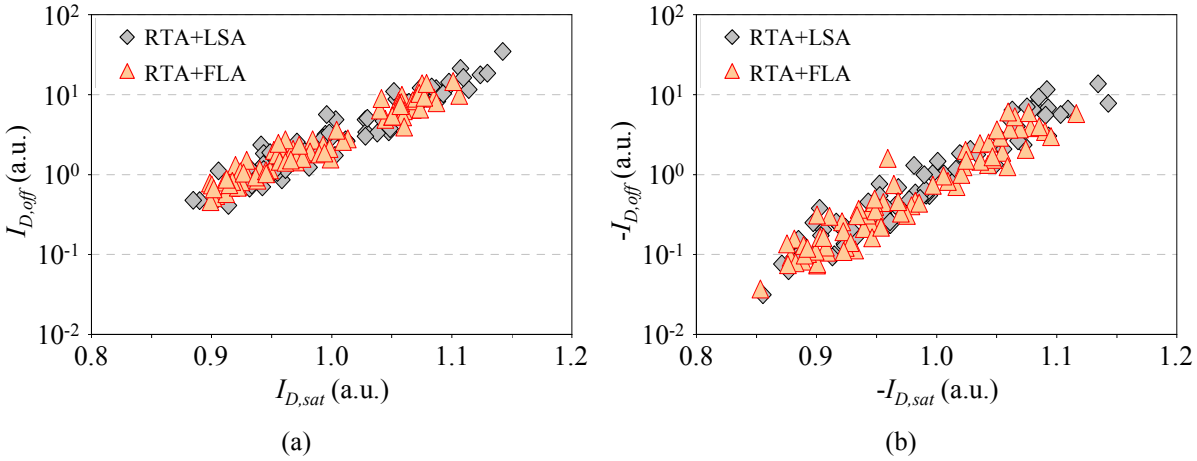


Bild 3.20: Universalkurve im Sättigungsbereich für (a) n- und (b) p-MOSFETs für die CO_2-LSA und Xe-FLA in Kombination mit der RTA

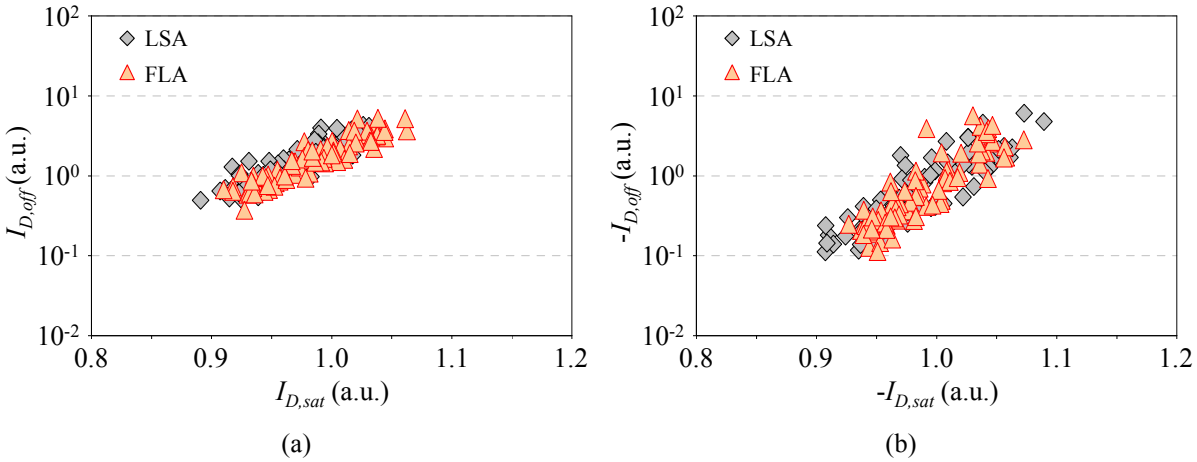


Bild 3.21: Universalkurve im Sättigungsbereich für (a) n- und (b) p-MOSFETs für die CO_2-LSA und Xe-FLA ohne RTA

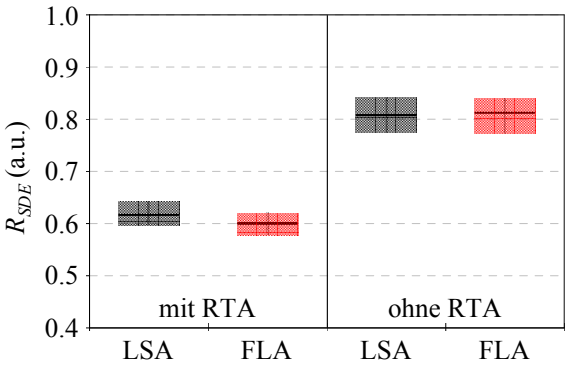


Bild 3.22: Schichtwiderstand R_{SDE} der SDE-Gebiete von n-MOSFETs für die CO_2-LSA und Xe-FLA mit und ohne RTA

Um den Einfluss der beiden Ausheilverfahren innerhalb eines Mikroprozessors zu untersuchen, wurde ein spezieller Maskensatz verwendet, mit dessen Hilfe es möglich ist, eine große Anzahl von beliebig über den Mikroprozessor verteilten Transistoren in den Logik- und Cache-Zellen zu kontaktieren und zu messen. Dadurch sind statistisch relevante Rückschlüsse auf Parametervariationen von Transistoren im Mikroprozessor möglich. Um eine Überlagerung mit der über dem Wafer auftretenden Parametervariation zu vermeiden, wurden nur Dies aus der Wafermitte gemessen (Bild 3.23).

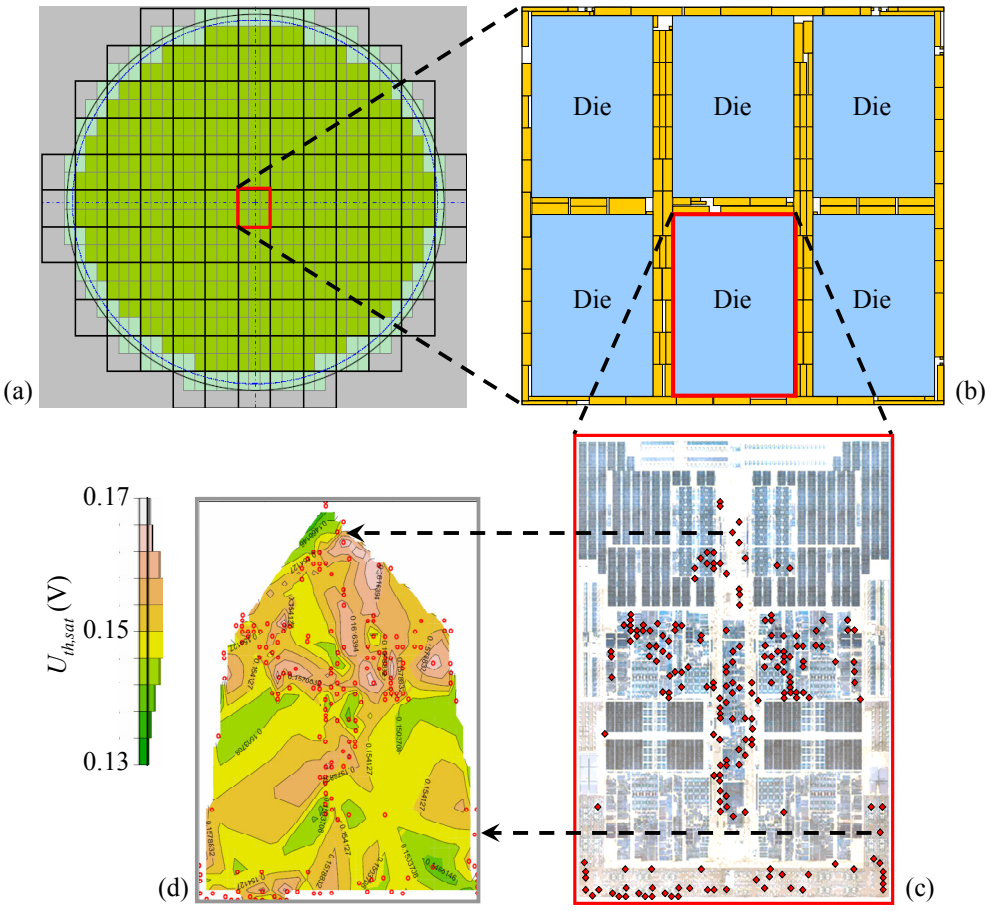


Bild 3.23: Schematische Darstellung der Messung der Parametervariation innerhalb eines Mikroprozessors
(a) Zur Bestimmung der Parametervariation wird nur das Feld in der Wafermitte verwendet;
(b) Gemessenes Feld mit 6 Dies;
(c) Position der gemessenen Transistoren auf dem Die (rote Vierecke);
(d) Sättigungsschwellspannung $U_{th,sat}$ in Abhängigkeit von der Position auf dem Die

Bild 3.24 zeigt die Variation des Sättigungs-Drainstroms $\Delta I_{D,sat}$ innerhalb eines Mikroprozessors für die CO_2-LSA und die Xe-FLA mit und ohne RTA.

Daraus sind verschiedene Sachverhalte ersichtlich:

1. Wird zusätzlich zur Kurzzeitausheilung eine RTA verwendet, dann sind keine Unterschiede zwischen beiden Ausheilverfahren in der Variation des Sättigungsstroms $\Delta I_{D,sat}$ erkennbar.
2. P-MOSFETs zeigen eine stärkere Parametervariation als n-MOSFETs.
3. Die CO_2-LSA ohne RTA weist im Gegensatz zur Xe-FLA ohne RTA eine geringere Streuung auf.

Alle drei Sachverhalte lassen sich durch Sensitivität des Transistors auf die verwendete Ausheilsequenz und Ausheiltechnologie erklären.

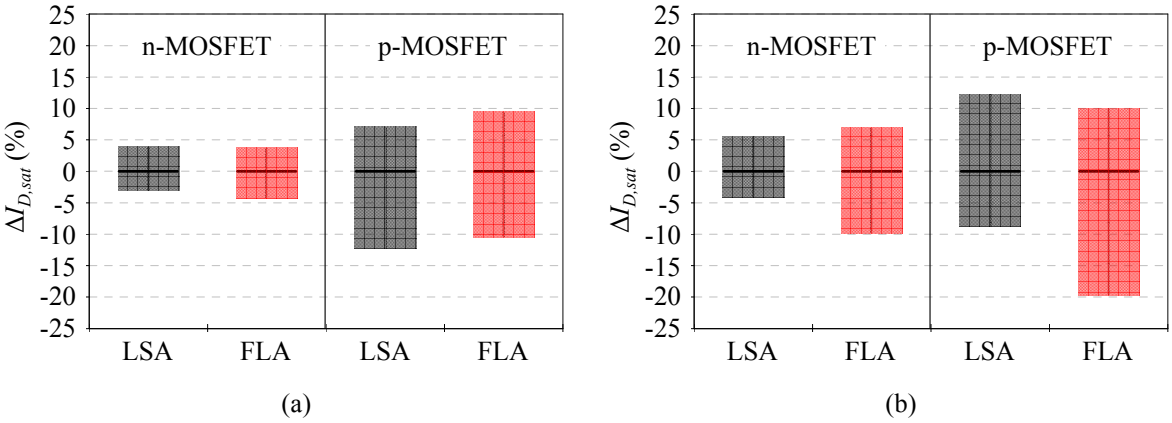


Bild 3.24: Variation des Sättigungs-Drainstroms $\Delta I_{D,sat}$ innerhalb eines Mikroprozessors für CO_2-LSA und Xe-FLA (a) in Kombination mit der RTA und (b) ohne RTA

Der Transistor reagiert aufgrund der erhöhten Dotandendiffusion bedeutend sensitiver auf einen RTA-Prozess als auf einen Ausheilprozess, der nur durch die Kurzzeitausheilung bestimmt wird [76]. Aus diesem Grund sind keine Unterschiede zwischen der CO_2-LSA und der Xe-FLA feststellbar, wenn außerdem eine RTA stattgefunden hat. Die stärkere Parametervariation von p-MOSFETs ist auf die stärkere Diffusion der Dotanden für eine p-Dotierung zurückzuführen. Die kleinere Streuung der CO_2-LSA im Vergleich zur Xe-FLA bei der Ausheilung ohne RTA ist durch die geringere Interaktion der verwendeten Wellenlänge mit der vorhandenen Waferoberfläche zu erklären.

Da aber alle weiteren Betrachtungen an den herkömmlichen Teststrukturen durchgeführt werden, bei denen beide Ausheilverfahren die gleiche Leistungsfähigkeit für n- und p-MOSFETs unabhängig von der verwendeten Ausheilsequenz zeigen, wird im Folgenden nicht zwischen der CO_2-LSA und der Xe-FLA unterschieden.

3.5 Zusammenfassung

Bei Kurzzeit-Ausheilverfahren wird im Unterschied zur konventionellen RTA ausgenutzt, dass nur die oberste Schicht des Wafers durch einen Energieimpuls erhitzt wird. Dadurch können Aufheiz- und Abkühlraten im Bereich von 10^6 K/s und damit Pulslängen von weniger als 10 ms erreicht werden. Demgegenüber sind bei der derzeitig verwendeten Spike-RTA minimale Pulslängen von höchstens 1 s möglich. Der Nachteil der Kurzzeitausheilung ist der große Temperaturunterschied zwischen Ober- und Unterseite des Wafers während des Ausheilprozesses. Durch die thermische Ausdehnung σ der wärmeren Schichten kommt es zur Ausbildung einer thermischen Verspannung σ des Wafers. Die Stärke der Verspannung σ ist dabei sehr stark vom Temperaturgradienten über dem Wafer abhängig.

Weiterhin wurde der Einfluss der Kurzzeitausheilung auf die Diffusion und Aktivierung der Dotierstoffe für eine p- und n-Dotierung untersucht. In Verbindung mit der RTA hat die Kurzzeitausheilung sowohl auf Bor als auch auf Arsen als Dotierstoff für eine p- bzw. n-Dotierung keine weitere Diffusion der Dotanden zur Folge. Der Schichtwiderstand $R_{Schicht}$ der Bordotierung bleibt trotz der zusätzlichen Ausheilung unverändert. Im Gegensatz dazu verringert sich der Schichtwiderstand $R_{Schicht}$ für Arsen durch die zusätzliche Ausheilung im Anschluss an die RTA. Die Verwendung der Kurzzeitausheilung ohne RTA führt im Vergleich zur RTA vor allem für eine p-Dotierung zu einer erheblichen kürzeren Diffusionslänge bei einem nur geringfügig höheren Schichtwiderstand $R_{Schicht}$.

Die Erhöhung der Ausheiltemperatur $T_{Ausheil}$ bei der Kurzzeitausheilung hat aufgrund einer verbesserten Kristallqualität für beide Dotierungsarten eine kontinuierliche Verringerung des Schichtwiderstands $R_{Schicht}$ bei einer gleichzeitig nur geringfügig größeren Diffusionslänge zur Folge. Der Nachteil der Kurzzeitausheilung ist neben der starken Schädigung der Kristallstruktur, dass unabhängig vom Dotierstoff selbst nach einer Kurzzeitausheilung bei 1300 °C das thermische Budget *TB* dieser Ausheiltechnologie nicht ausreicht, um alle EOR-Defekte vollständig bzw. in energetisch stabilere Konfigurationen aufzulösen.

Im Rahmen dieser Arbeit wurden zwei unterschiedliche Technologien der Kurzzeitausheilung, die Blitzlampen- und Laser-Ausheilung, untersucht. Aufgrund der verwendeten Wellenlängen kann es bei strukturierten Wafern zu Unterschieden in der Ausheiltemperatur $T_{Ausheil}$ zwischen beiden Ausheiltechnologien kommen, da die Absorption der einfallenden Strahlung stark von den Eigenschaften und der Geometrie der verwendeten Schichten auf der Waferoberfläche abhängig ist. Auf Waferebene sind nach Abgleich der Anlagenparameter zwischen Blitzlampen- und Laser-Ausheilung keine Unterschiede in der Leistungsfähigkeit der Transistoren erkennbar, was auf vergleichbare Ausheiltemperaturen $T_{Ausheil}$ schließen lässt. Auf Mikroprozessorebene weist jedoch die Laser-Ausheilung in Bezug auf mögliche Strukturierungseffekte Vorteile gegenüber der Blitzlampen-Ausheilung auf, da die Laser-Ausheilung aufgrund der verwendeten Wellenlänge eine geringere Interaktion mit der vorhandenen Waferoberfläche im Vergleich zur Blitzlampen-Ausheilung aufzeigt.

4 Modelle und Kalibrierung der Prozess- und Bauelementesimulation

Dieses Kapitel gibt einen Überblick über die verwendeten Modelle der Prozess- und Bauelementesimulation sowie deren Kalibrierung an einem SOI-Transistor der 65 nm-Technologie. Das Hauptaugenmerk der Prozesssimulation liegt dabei auf den Modellen für die Implantation, die Ausheilung und mechanischer Verspannungen, da diese Prozessschritte zum Einen die Dotierungsprofile und zum Anderen die Leistungsfähigkeit der Transistoren wesentlich beeinflussen. Bei der Bauelementesimulation wird die Modellierung des Ladungsträgertransports, der Ladungsträgerbeweglichkeit, der Generations-Rekombinations-Prozesse und sowie der mechanischen Verspannungen betrachtet, um den Einfluss eines veränderten Dotierungs- bzw. Verspannungsprofils auf die elektrischen Eigenschaften der Transistoren genauer untersuchen zu können.

4.1 Vorbetrachtungen

Die Modellierung und Simulation von Halbleiterbauelementen fördert das Verständnis für die physikalischen Vorgänge während des Herstellungsprozesses, gibt einen Einblick in das innerelektronische Verhalten der Bauelemente und bietet darüber hinaus die Möglichkeit, neuartige Transistorstrukturen bereits vor der Fertigung zu bewerten. Modellierung und Simulation bestehen dabei vorteilhaft aus zwei Schritten, der Prozesssimulation und der sich daraus ergebenden Bauelementesimulation.

Die Prozesssimulation besteht aus einzelnen, aufeinander abgestimmten Modellierungsschritten, die wesentliche Elemente des Herstellungsprozesses rechnerisch nachbilden. Aufgrund der Komplexität des Herstellungsprozesses eines MOSFETs ist aber eine Simulation, bei der jeder reale Prozessschritt in ein virtuelles, physikalisch-mathematisches Modell übertragen wird, kaum realisierbar. Die Simulation muss sich daher auf wesentliche Prozessschritte beschränken, die die Funktion des fertigen Bauelements bestimmen. Dazu gehören vor allem die Implantation der Dotierstoffe und deren Ausheilung sowie auftretende mechanische Verspannungen, da dies die Dotierprofile und die Leistungsfähigkeit des Transistors maßgeblich beeinflussen. Andere Prozesse wie das Ätzen, Maskieren und der Schichtauftrag verändern demgegenüber meist nur die Geometrie der Bauelementekomponenten. Auf sie soll deshalb im Folgenden bei der Prozesssimulation nicht näher eingegangen werden. Eine ausführliche Beschreibung zur Modellierung dieser Prozessschritte ist in [42] enthalten.

Mit Hilfe der Bauelementesimulation ist es dann möglich, neue Erkenntnisse und Aussagen über das elektrische und innerelektronische Verhalten von Halbleiterstrukturen zu erhalten. Ein wesentlicher Vorteil liegt in der Darstellung innerelektronischer Kenngrößen und Parameter, welche nur mit großem Aufwand oder überhaupt nicht messbar sind. Der Bauelementesimulator benötigt dabei aus der Prozesssimulation die vollständige Geometrie des Bauelements einschließlich der Dotierungs- und Verspannungsprofile.

Die Kalibrierung der in der Prozess- und Bauelementesimulation verwendeten Modellparameter ist ein notwendiger Schritt, bevor mit der Simulation einer bestimmten Transistorgeometrie begonnen werden kann. Dabei ist zu beachten, dass die verwendeten Modelle, welche auf physikalischen Beschreibungen basieren, meist Vereinfachungen und Annahmen nutzen, um die Vorgänge analytisch zu beschreiben. Außerhalb bestimmter Parameterbereiche verlieren diese Modelle jedoch ihre Gültigkeit, da verschiedene Vernachlässigungen dann nicht mehr zulässig sind. Aus diesem Grund ist eine Anpassung der Modellparameter an den zu untersuchenden Prozess notwendig. Mit Hilfe von derartig kalibrierten Modellen lässt sich dann der Einfluss einer Variation verschiedener Parameter auf das Transistorverhalten überprüfen bzw. können damit neue Transistorstrukturen untersucht werden.

Die Prozess- und Bauelementesimulation ist als gekoppelte Simulation durchgeführt worden. Dazu wurde das kommerziell erhältliche Technology Computer-Aided Design- (TCAD-) Softwarepaket der Firma Synopsys verwendet [80]. Gekoppelte Simulation bedeutet dabei, dass eine durch die Prozesssimulation erzeugte Struktur zur Bauelementesimulation verwendet werden kann.

Bild 4.1 zeigt eine schematische Darstellung des Simulationsflusses sowie notwendige Angaben für die Prozess- und Bauelementesimulation. Die Prozesssimulation ist angewiesen auf Aussagen über den Prozessablauf und die dabei anzuwendenden physikalischen Modelle. Die sich nach der Prozesssimulation ergebende Struktur des Transistors wird dann mit den Dotierungs- und Verspannungsprofil an die Bauelementesimulation übergeben. Diese benötigt weiterhin eine Zusammenstellung von physikalischen Modellen, die bei der Rechnung berücksichtigt werden sollen, sowie Angaben über das anzuwendende, numerische Lösungsverfahren und nicht zuletzt die Details über die Art der Simulation. Als Ergebnis erhält man elektrische, thermische oder auch optische Eigenschaften des Bauelements als Funktion der eingeprägten Spannungen oder Ströme.

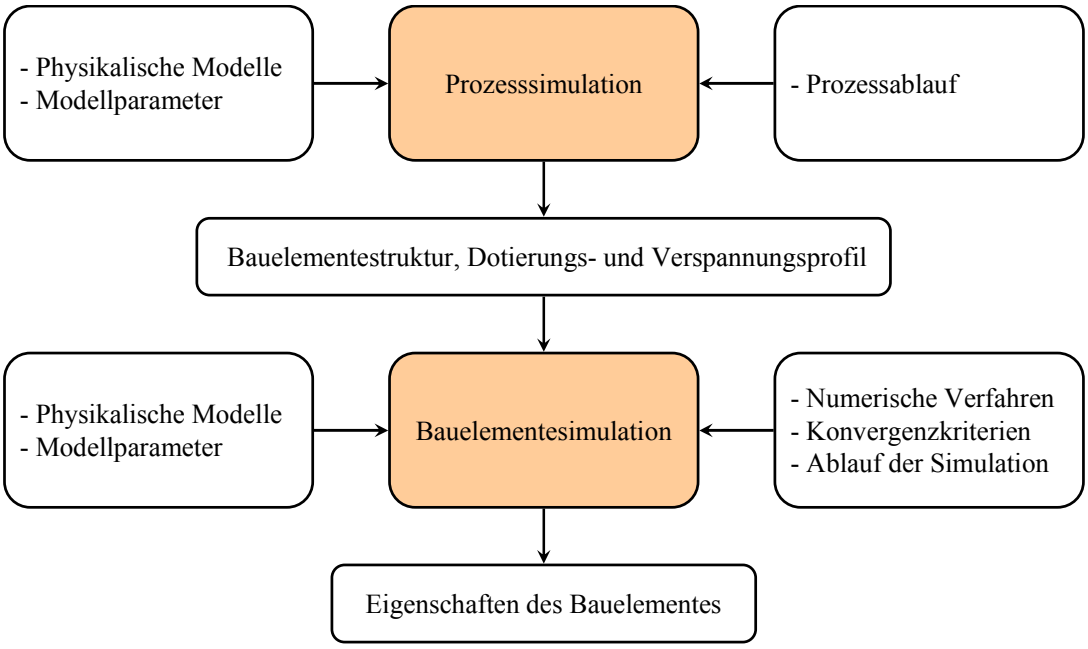


Bild 4.1: Schematische Darstellung des Simulationsflusses sowie notwendige Angaben für die Prozess- und Bauelementesimulation

4.2 Prozesssimulation

4.2.1 Implantation

Der Prozesssimulator benutzt zur Bestimmung der Verteilungen der implantierten Ionen sowohl analytische als auch numerische (Monte-Carlo-) Verfahren, wobei die analytischen Methoden auf statistischen Verteilungsdichtefunktionen basieren. Aufgrund der sehr zeitintensiven Berechnungen wurde auf Monte-Carlo-Simulationen verzichtet und dafür analytische Verfahren zur Bestimmung der Dotierstoffverteilung mit einem auf Monte-Carlo-Simulationen basiertem Parametersatz verwendet. Dieser Parametersatz enthält die zur Beschreibung der Dotierstoffverteilung notwendigen Momente:

- Projizierte Reichweite R_p (erstes Moment, Erwartungswert),
- Standardabweichung ΔR_p (zweites Moment, Varianz),
- Neigungsstärke (drittes Moment, Schiefe),
- Spitzigkeit (viertes Moment, Wölbung).

Eine gute Beschreibung der Verteilung einer großen Anzahl von implantierten Ionen als Funktion der Tiefe liefert die GAUß-Funktion, die zur Beschreibung des Dotierungsprofils nur das erste und zweite Moment, also die projizierte Reichweite R_p und die Standardabweichung ΔR_p verwendet. Eine bessere Profilbeschreibung, insbesondere für sehr tiefe Implantationen, liefern die PEARSON-Funktionen, die neben der Reichweite und der Standardabweichung auch die Schiefe und die Wölbung verwenden. Dadurch kann das Auslaufen, d.h. der sogenannte Verteilungsschwanz für sehr leichte Ionen (Bor) in Richtung Oberfläche bzw. für sehr schwere Ionen (Antimon) in Richtung Substrat besser beschrieben werden [42]. Dieses Phänomen ist beispielhaft in Bild 4.2 für eine Borimplantation dargestellt. Man erkennt deutlich, wie das Dotierungsprofil für die PEARSON-Verteilung zur Oberfläche hin ausläuft, wohingegen die GAUß-Verteilung nur den Bereich um den Scheitelpunkt der Verteilung gut beschreibt. Ursache dafür ist, dass leichtere Ionen wie Bor bei der Implantation zurückgestreut werden und dadurch mehr Ionen in der Nähe der Waferoberfläche verbleiben.

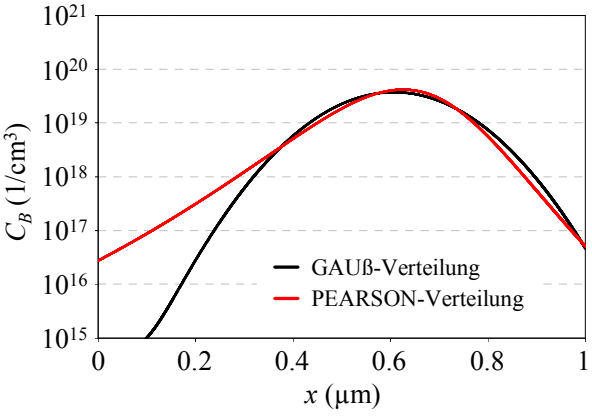


Bild 4.2: Simulierte Profile einer 200 keV Borimplantation mit einer Dosis von 10^{15} cm^{-2} unter Verwendung einer GAUß- und einer PEARSON-Verteilung

Zur Beschreibung sehr flacher Implantationen, vor allem von Bor in kristallinem Silizium, wird meist die Dual-PEARSON-Funktion verwendet:

$$f_p(x) = z \cdot f_{p,1}(x) + (1-z) \cdot f_{p,2}(x). \tag{4.1}$$

Dabei sind die Funktionen $f_{p,1}(x)$ und $f_{p,2}(x)$ zwei unabhängige PEARSON-Funktionen und z ein Faktor zur Wichtung der jeweiligen Funktion.

Durch die Überlagerung dieser beiden Funktionen kann die durch den Channeling-Effekt verursachte breitere Verteilung der Dotieratome besser modelliert bzw. nachgebildet werden [81]. Die PEARSON-Funktion $f_{p,2}(x)$ berücksichtigt dabei den Teil der implantierten Ionen, die durch den Channeling-Effekt tiefer in das Substrat gelangen.

4.2.2 Ausheilung, Diffusion und Aktivierung

Die Ausheilung und damit die Diffusion und Aktivierung der eingebrachten Dotierstoffe werden durch Transportmodelle beschrieben. Jedes der Transportmodelle hat einen gewissen Gültigkeitsbereich, der davon abhängt, welche physikalischen Effekte einbezogen werden.

Naturgemäß ergibt sich daraus eine gewisse Hierarchie in den Modellen, welche sich auch in dem zur numerischen Lösung notwendigen Aufwand widerspiegelt. Das Spektrum dieser Modelle reicht vom äußerst einfachen Modell, welches nur die Diffusion der einzelnen Dotanden aufgrund deren Konzentrationsunterschiede betrachtet, bis hin zu komplexen Modellen, welche die Diffusion der Punktdefekte und die paarweise Diffusion als separate Gleichungen mit einbeziehen (Tabelle 4.1).

Tabelle 4.1 Hierarchie der Diffusionsmodelle im Programmsystem Sentaurus Process

Komplexität	Diffusionsmodell	Berücksichtigte physikalische Effekte
gering	*Constant*	- keine Wechselwirkung zwischen Dotand und Punktdefekt - kein Einfluss des elektrischen Feldes
mittel	*Fermi*	- Wechselwirkung zwischen Dotand und Punktdefekt - Dotand-Defekt-Paar im lokalen Gleichgewicht - Punktdefekte im lokalen Gleichgewicht - Beachtung des elektrischen Feldes
hoch	*Pair*	- Wechselwirkung zwischen Dotand und Punktdefekt - Dotand-Defekt-Paar im lokalen Gleichgewicht - Punktdefekte außerhalb des lokalen Gleichgewichts - Beachtung des elektrischen Feldes
sehr hoch	*React*	- Wechselwirkung zwischen Dotand und Punktdefekt - Dotand-Defekt-Paar außerhalb des lokalen Gleichgewichts - Punktdefekte außerhalb des lokalen Gleichgewichts - Beachtung des elektrischen Feldes

Für die Simulation wurde im Rahmen dieser Arbeit das *Pair*-Diffusionsmodell verwendet, da es ein Optimum zwischen Genauigkeit und Rechenaufwand aufweist. Dieses Modell berücksichtigt die Diffusion der Dotanden über Punktdefekte durch einen von deren Konzentration abhängigen Diffusionskoeffizienten. Weiterhin wird davon ausgegangen, dass die Dotand-Defekt-Paare im lokalen Gleichgewicht mit der Dotanden- und Defektkonzentration stehen, wohingegen die Punktdefekte selbst nicht im Gleichgewicht sind.

Aus diesen Forderungen ergeben sich für den Teilchenstrom eines Fremdatoms A

$$J_A = -D_{AI} \cdot \mathrm{grad}\left(C_A \cdot \frac{C_I}{C_I^*}\right) - D_{AV} \cdot \mathrm{grad}\left(C_A \cdot \frac{C_V}{C_V^*}\right), \tag{4.2}$$

sowie für die Teilchenströme der Silizium-Zwischengitteratome I

$$J_I = -D_I \cdot C_I^* \cdot \mathrm{grad}\left(\frac{C_I}{C_I^*}\right) \tag{4.3}$$

und der Leerstellen V

$$J_V = -D_V \cdot C_V^* \cdot \mathrm{grad}\left(\frac{C_V}{C_V^*}\right). \tag{4.4}$$

Hier sind C_A die Konzentration des Fremdatomtyps A, C_I und C_I^* die Konzentrationen der Silizium-Zwischengitteratome (Interstitial) außerhalb und im thermodynamischen Gleichgewicht sowie C_V und C_V^* die Konzentrationen der Leerstellen außerhalb und im thermodynamischen Gleichgewicht. Die Parameter D_I und D_V sind die Diffusionskoeffizienten für die Diffusion der Defekte, wohingegen D_{AI} und D_{AV} die Diffusionskoeffizienten für die Diffusion der Dotanden unter Wechselwirkung mit Silizium-Zwischengitteratomen bzw. Leerstellen darstellen, die wiederum mit Hilfe der Gleichungen (2.28) und (2.29) beschrieben werden. Die mit diesem Modell zu lösenden Differentialgleichungen sind:

$$\frac{\partial C_A}{\partial t} = -\mathrm{div}\,\boldsymbol{J}_A - R_A^{clus}, \tag{4.5}$$

$$\frac{\partial C_I}{\partial t} = -\mathrm{div}\,\boldsymbol{J}_I - \mathrm{div}\,\boldsymbol{J}_A - R_{IV} - R_I^{clus}, \tag{4.6}$$

$$\frac{\partial C_V}{\partial t} = -\mathrm{div}\,\boldsymbol{J}_V - \mathrm{div}\,\boldsymbol{J}_A - R_{IV} - R_V^{clus}, \tag{4.7}$$

wobei R_A^{clus} die Aktivierung und Cluster-Bildung der Dotanden, R_I^{clus} und R_V^{clus} die Bildung und Auflösung von Defekt-Clustern und R_{IV} die Rekombination der Defekte beschreiben.

4.2.3 Mechanische Verspannung

Die stetige Verkleinerung der Bauelementeabmessungen ist neben einer größeren Packungsdichte mit höheren mechanischen Verspannungen der Aktivgebiete verbunden. So können ungewollt, z.B. durch unterschiedliche thermische Ausdehnungskoeffizienten verschiedener aneinander liegender Materialien oder durch Materialwachstum während der Transistorherstellung, mechanische Verspannungen in das Siliziumgebiet induziert werden, die ab einem bestimmten Grenzwert sowohl den Herstellungsprozess als auch die elektrischen Eigenschaften des Transistors negativ beeinflussen können. Neben diesen ungewollten verspannungserzeugenden Mechanismen wird zunehmend in der Halbleiterindustrie gezielt die Technik der mechanischen Verspannung von Silizium eingesetzt, um die Ladungsträgerbeweglichkeit im Kanal zu erhöhen, was im Wesentlichen durch eine Veränderung der Bandstruktur des Siliziums verursacht wird.

Verspannungen lassen sich beispielsweise gezielt erzeugen, indem andere Halbleitermaterialien mit einer zum Silizium abweichenden Gitterkonstanten in die SD-Gebiete des Transistors eingebettet werden, wodurch umliegende Gebiete verformt und damit verspannt werden. Die Ausgangsmaterialen einer solchen sich dabei ergebenden Heterostruktur müssen dieselbe Kristallstruktur besitzen. Aufgrund der unterschiedlichen Gitterparameter zwischen Epitaxieschicht und Substrat wächst der epitaktische Film bis zu einer kritischen Schichtdicke elastische verspannt auf. In der Schichtebene erfolgt eine elastische Gitteranpassung des Films an das Substrat (Bild 4.3). Dieser Vorgang wird als gitterangepasstes bzw. pseudomorphes Wachstum bezeichnet. Hat der epitaktische Film einen größeren intrinsischen Gitterparameter a als das Siliziumsubstrat a_{Si} ($a > a_{Si}$), wie z.B. Silizium-Germanium (SiGe), so wächst die Schicht druckverspannt auf, wohingegen eine Epitaxieschicht mit einer kleineren Gitterkonstanten a ($a < a_{Si}$), z.B. Silizium-Kohlenstoff (SiC), zugverspannt auf das Substrat aufwächst.

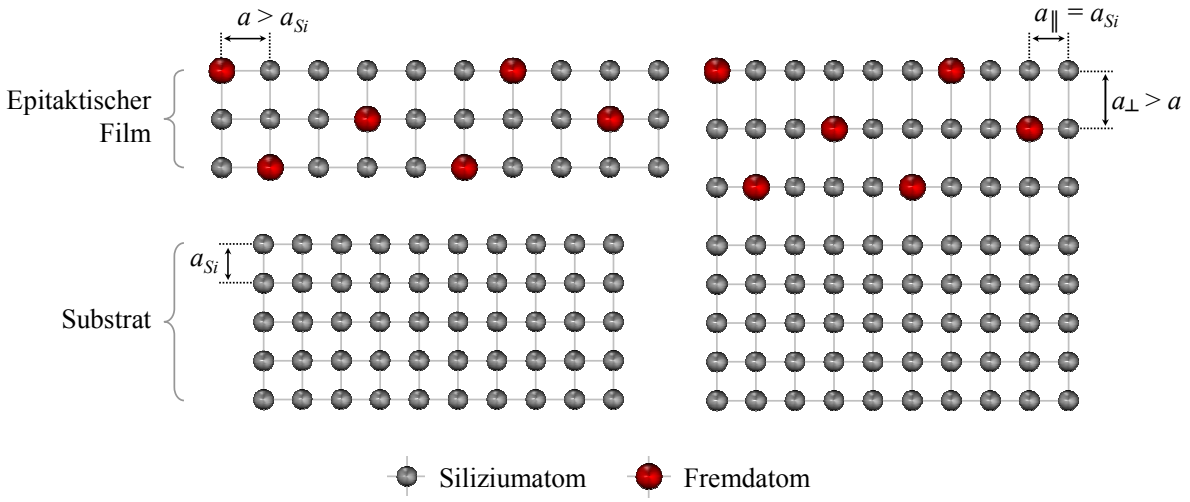


Bild 4.3: Schematische Darstellung der Verformung bzw. Deformation pseudomorpher Schichten. $a_{\perp}$ Gitterparameter der verspannten Schicht in Wachstumsrichtung, $a_{\parallel}$ dem Gitterparameter senkrecht dazu

Eine Druckverspannung entsteht, wenn ein Atomgitter durch äußere Einflüsse zusammengestaucht wird. Dabei versuchen die Bindungen der Atome den relaxierten Gitterabstand wieder herzustellen und drücken permanent gegen die beteiligten Atome. Bei einer Zugverspannung ist der normale Gitterabstand durch externe Kräfte erweitert und die Bindungskräfte ziehen an ihren zugehörigen Atomen, um den nominalen Gitterzustand wieder herbei zu führen. In der Regel ist eine Zugverspannung mathematisch positiv und eine Druckverspannung negativ definiert. Durch das Einbetten von verspannten Schichten in die SD-Gebiete des Transistors erfolgt eine uniaxiale Verspannung in Kanalrichtung, was bei den durchgeführten Untersuchungen die <110>-Richtung auf (001)-orientiertem Substrat ist (Bild 4.4). Eine uniaxiale Verspannung besitzt nur eine dominante Verspannungskomponente in einer der drei kristallographischen Achsen. In die SD-Gebiete eingebettetes SiC induziert eine uniaxiale Zugverspannung in Kanalrichtung (Bild 4.4b), wodurch die Elektronenbeweglichkeit erhöht wird, wohingegen eingebettetes SiGe eine uniaxiale Druckverspannung in Kanalrichtung erzeugt (Bild 4.4c), was die Löcherbeweglichkeit verbessert [82], [83].

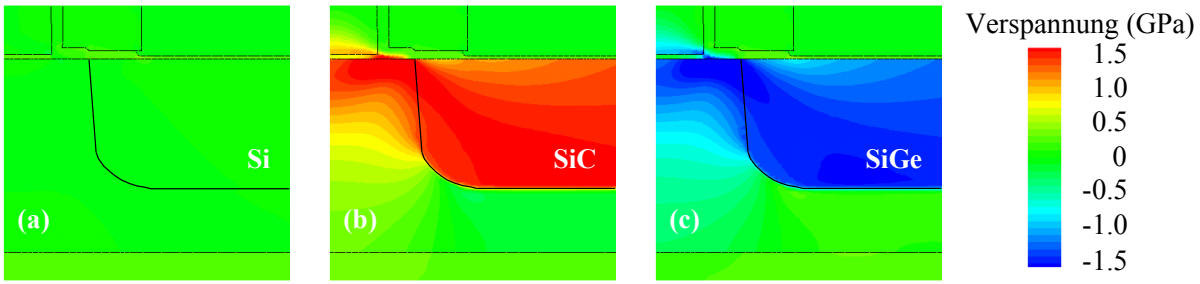


Bild 4.4: Simuliertes Verspannungsprofil eines Transistors (a) im unverspannten Fall, (b) mit eingebettetem SiC und (c) mit eingebettetem SiGe. Der Germaniumanteil im SiGe und der Kohlenstoffgehalt im SiC beträgt 23 % bzw. 2.84 %, was nach Gleichung (4.8) einer äquivalenten Gitterverformung entspricht.

Die Verspannung σ durch verschiedene Gitterkonstanten der im Transistor befindlichen Materialien wird über deren Dehnung ε berechnet. Diese ergibt sich im Prozesssimulator aus dem Unterschied der Gitterkonstanten a_{Si} und $a_{Si_{1-x}HL_x}$ von Silizium bzw. dem eingebrachten Verbindungshalbleiter zu

$$\varepsilon = \frac{a_{Si} - a_{Si_{1-x}HL_x}}{a_{Si}} . \tag{4.8}$$

Die Gitterkonstante $a_{Si_{1-x}HL_x}$ des eingebrachten Materials wird wiederum über die lineare Näherung

$$a_{Si_{1-x}HL_x} = a_{Si} \cdot (1-x) + a_{HL} \cdot x \tag{4.9}$$

zwischen den beiden Gitterkonstanten a_{Si} und a_{HL} von Silizium bzw. dem einzubauenden Halbleiter über die Konzentration x berechnet. Die Verspannung σ und die Dehnung ε sind wiederum über den Elastizitätstensor $\boldsymbol{C}$ durch das HOOKEsche Gesetz:

$$\sigma_{ij} = \boldsymbol{C}_{ijkl} \cdot \varepsilon_{kl} \tag{4.10}$$

verknüpft, was das elastische Verhalten von Festkörpern beschreibt. Die Indizes i und k definieren die Fläche, auf die die Verspannung σ wirkt bzw. die Dehnung ε erfolgt, wohingegen j und l deren Richtung angeben. Die grundlegende Gleichung, die dann in der Prozesssimulation zur Berechung von mechanischen Verspannungen gelöst wird, ist das Kräftegleichgewicht, wodurch die Balance zwischen den inneren Verspannungen σ und den von außen einwirkenden Kräften F beschrieben wird. Durch Lösen der Gleichgewichtsbedingung ergeben sich für die einzelnen Elemente interne Verspannungen σ, aus denen dann mit Hilfe von Gleichung (4.10) die resultierende Dehnung ε ermittelt wird.

Die am Ende der Prozesssimulation entstandene Verspannungs- und Deformationsmatrix wird neben dem Dotierungsprofil an die Bauelementesimulation übergeben.

4.3 Bauelementesimulation

4.3.1 Modellhierarchie

Die für die Simulation von Halbleiterbauelementen notwendigen physikalischen Modelle lassen sich einteilen in:

- Mikroskopische Modelle
  - Quantenmodelle (SCHRÖDINGER-Gleichung)
  - Teilchenmodelle (BOLTZMANNsche Transportgleichung)
  - Gemischte Modelle
- Makroskopische Modelle
  - Hydrodynamische Modelle
  - Drift-Diffusions-Modelle
- Gemischte mikroskopisch/makroskopische Modelle [84].

Mikroskopische Modelle beschreiben das Verhalten einzelner Ladungsträger, ausgehend vom Wellen- bzw. Teilchencharakter. Die Welleneigenschaften lassen sich mit Hilfe der SCHRÖDINGER-Gleichung modellieren [85]. Transportvorgänge im mikroskopischen Bereich können durch die ursprünglich aus der kinetischen Gastheorie stammenden BOLTZMANNschen Transportgleichung beschrieben werden [86]. Deren Lösung ist jedoch sehr aufwendig und erfolgt meist nur unter Annahme verschiedener Näherungen durchgeführt. Eine übliche Methode ist die Monte-Carlo Simulation [87]. Die gekoppelte Lösung von SCHRÖDINGER-Gleichung und BOLTZMANNscher Transportgleichung führt zu gemischten Modellen im mikroskopischen Bereich [88]. Mikroskopische Modelle sind für eine möglichst exakte Beschreibung von quantenmechanischen Effekten in Nanostrukturen unumgänglich. Sie erfordern allerdings einen erheblichen numerischen Aufwand, so dass vielfach nur grundlegende Untersuchungen an ausgewählten Strukturen möglich sind.

Makroskopische Modelle, bei denen im Gegensatz zu den mikroskopischen Verfahren mit durchschnittlichen, also makroskopischen Größen wie Ladungsträger-, Impuls- und Energiedichten gearbeitet wird, lassen sich in hydrodynamische und Drift-Diffusions-Modelle einteilen (Tabelle 4.2).

Tabelle 4.2 Hierarchie der makroskopischen Modelle

Modell	Berücksichtigte Effekte
Allgemeines hydrodynamisches Modell	- Kontinuitätsgleichungen für die Ladungsträger-, Impuls- und Energiedichten - POISSON-Gleichung
Vereinfachtes hydrodynamisches Modell (Energie-Balance-Modell)	- Kontinuitätsgleichungen für die Ladungsträger- und Energiedichten - POISSON-Gleichung
Drift-Diffusions-Modell	- Kontinuitätsgleichungen für die Ladungsträgerdichten - POISSON-Gleichung

Ausgehend von der BOLTZMANNschen Transportgleichung lassen sich mittels der Momenten-Methode die Kontinuitätsgleichungen für die Ladungsträger, Impuls- und Energiedichten ableiten [89]. Zusammen mit der POISSON-Gleichung

$$\mathrm{div}\left(\varepsilon \cdot \mathrm{grad}(\varphi)\right) = -q \cdot \left(p - n + N_D - N_A\right), \tag{4.11}$$

welche den Zusammenhang zwischen dem elektrostatischen Potenzial φ, der Permittivität ε, den Ladungsträgerdichten p und n sowie den Dichten der ionisierten Störstellen N_D und N_A beschreibt, bilden diese Gleichungen ein gekoppeltes Differenzialgleichungssystem, welches als allgemeines hydrodynamisches Modell bezeichnet wird. Unter Vernachlässigung der Impulserhaltung ergibt sich das vereinfachte hydrodynamische bzw. Energie-Balance-Modell [90].

Eine weitere Vereinfachung führt zum Drift-Diffusions-Modell, bestehend aus

- der POISSON-Gleichung (4.11),
- den Kontinuitätsgleichungen für Elektronen und Löcher

$$\mathrm{div}\left(\boldsymbol{J}_n\right) = q \cdot \left(R - G + \frac{\partial n}{\partial t}\right) \tag{4.12}$$

$$\mathrm{div}\left(\boldsymbol{J}_p\right) = -q \cdot \left(R - G + \frac{\partial p}{\partial t}\right), \tag{4.13}$$

  welche die zeitliche Veränderung der Ladungsträgerkonzentration in einem Volumen im Zusammenhang mit der Stromdichte J_n bzw. J_p und der Ladungsträger-Generation G bzw. -Rekombination R beschreiben,

- sowie den zugehörigen Transportgleichungen

$$\boldsymbol{J}_n = -q \cdot \mu_n \cdot n \cdot \left(\mathrm{grad}\left(\varphi\right) - \frac{1}{n} \cdot \frac{k_B \cdot T_L}{q} \cdot \mathrm{grad}\left(n\right)\right) \tag{4.14}$$

$$\boldsymbol{J}_p = -q \cdot \mu_p \cdot p \cdot \left(\mathrm{grad}\left(\varphi\right) + \frac{1}{p} \cdot \frac{k_B \cdot T_L}{q} \mathrm{grad}\left(p\right)\right), \tag{4.15}$$

  welche die Stromdichten J_n und J_p der Elektronen und Löcher mit den Ladungsträgerdichten und den elektrischen Feldern verknüpfen. Der Parameter T_L stellt die Temperatur des Kristallgitters dar.

4.3.2 Modellierung des Ladungsträgertransports

a) Transportmodell

Mit der immer weiter fortschreitenden Skalierung von Halbleiterstrukturen in den Nanometerbereich ist eine Beschreibung des innerelektronischen Verhaltens mit dem Drift-Diffusions-Modell oft nicht ausreichend. In modernen Halbleiterbauelementen befinden sich die Ladungsträger aufgrund der hohen elektrischen Feldstärken nicht mehr im thermodynamischen Gleichgewicht mit dem Gitter. Im Drift-Diffusions-Modell wird diese Aufheizung der Ladungsträger nicht mit berücksichtigt, so dass es überhöhte Driftgeschwindigkeiten nicht reproduzieren kann und Stoßionisationsraten überbewertet. Um diesen Anforderungen gerecht zu werden, wird eine Erweiterung des Modellgehaltes vorgenommen, indem man neben den Erhaltungsgleichungen für die Ladungen auch die Erhaltungsgleichungen für die Energie ω_n und ω_p der Elektronen bzw. Löcher mit in das Gleichungssystem einbezieht.

Unter Vernachlässigung des Anteils der kinetischen Energie der Ladungsträger, d.h. mit

$$\omega_n = \frac{3}{2}\cdot k_B \cdot T_n + \frac{m_n^{eff}}{2}\cdot v_n^2 \approx \frac{3}{2}\cdot k_B \cdot T_n \tag{4.16}$$

sowie

$$\omega_p = \frac{3}{2}\cdot k_B \cdot T_p + \frac{m_p^{eff}}{2}\cdot v_p^2 \approx \frac{3}{2}\cdot k_B \cdot T_p\,, \tag{4.17}$$

erhält man die Energiebalancegleichungen für die Elektronen- und Löcherenergie

$$\mathrm{div}\left(\boldsymbol{S}_n\right) = \boldsymbol{E}\cdot\boldsymbol{J}_n - \left(R-G\right)\cdot\omega_n - n\cdot\frac{\omega_n-\omega_0}{\tau_{\omega n}} - \frac{\partial\left(n\cdot\omega_n\right)}{\partial t} \tag{4.18}$$

$$\mathrm{div}\left(\boldsymbol{S}_p\right) = \boldsymbol{E}\cdot\boldsymbol{J}_p - \left(R-G\right)\cdot\omega_p - p\cdot\frac{\omega_p-\omega_0}{\tau_{\omega p}} - \frac{\partial\left(p\cdot\omega_p\right)}{\partial t} \tag{4.19}$$

sowie die Gleichungen für die Energiestromdichten

$$\boldsymbol{S}_n = -\kappa_n\cdot\mathrm{grad}\left(T_n\right) - \frac{5}{2}\cdot\frac{k_B\cdot T_n}{q}\cdot\boldsymbol{J}_n \tag{4.20}$$

$$\boldsymbol{S}_p = -\kappa_p\cdot\mathrm{grad}\left(T_p\right) + \frac{5}{2}\cdot\frac{k_B\cdot T_p}{q}\cdot\boldsymbol{J}_p\,, \tag{4.21}$$

mit der Energie ω_0 der Ladungsträger im thermodynamischen Gleichgewicht

$$\omega_0 = \frac{3}{2}\cdot k_B\cdot T_L\,. \tag{4.22}$$

Dabei bezeichnen in den Gleichungen (4.18) bzw. (4.19) $\tau_{\omega n}$ und $\tau_{\omega p}$ die Energierelaxationszeiten sowie in den Gleichungen (4.20) bzw. (4.21) κ_n und κ_p die thermischen Leitfähigkeiten der Elektronen und Löcher. Die zusätzlichen Größen der Elektronen- und Löchertemperaturen T_n und T_p, sind mit der in den Gleichungen (4.16) und (4.17) genannten Näherung ein Maß für die Energie der Elektronen bzw. Löcher. Sie werden in die Transportgleichungen des Drift-Diffusions-Modells durch einen zusätzlichen Gradiententerm eingebunden:

$$\boldsymbol{J}_n = -q\cdot\mu_n\cdot n\cdot\left(\mathrm{grad}\left(\varphi\right) - \frac{1}{n}\cdot\frac{k_B\cdot T_n}{q}\cdot\mathrm{grad}\left(n\right) - \frac{k_B}{q}\cdot\mathrm{grad}\left(T_n\right)\right) \tag{4.23}$$

$$\boldsymbol{J}_p = -q\cdot\mu_p\cdot p\cdot\left(\mathrm{grad}\left(\varphi\right) + \frac{1}{p}\cdot\frac{k_B\cdot T_p}{q}\cdot\mathrm{grad}\left(p\right) - \frac{k_B}{q}\cdot\mathrm{grad}\left(T_p\right)\right). \tag{4.24}$$

Zusammen mit der POISSON-Gleichung (4.11) sowie den Kontinuitätsgleichungen (4.12) und (4.13) bilden die Gleichungen (4.16) bis (4.24) das vereinfachte hydrodynamische bzw. Energie-Balance-Modell, welches in dieser Arbeit für die Bauelementesimulation verwendet wurde, da es ein Optimum zwischen Rechenzeit, Stabilität und Genauigkeit darstellte.

Die Verringerung der Bauelementeabmessungen führt weiterhin zu einer Beeinflussung der Ladungsträgerverteilung durch Quantisierungseffekte. Beim MOSFET entstehen im Kanalgebiet an der SiO_2/Si-Grenzfläche durch Bandverschiebung Potenzialmulden, in denen es bei Inversion zur Ausbildung eines 2D-Elektronengases kommt. Innerhalb dieser Potenzialmulden nehmen die Ladungsträger in Abhängigkeit von ihrer Energie diskrete Energiezustände, sogenannte Subbänder, ein, wodurch die Dichteverteilung der Ladungsträger sowie Streuprozesse zwischen den einzelnen Subbändern beeinflusst wird. Die Berücksichtigung dieser Quantisierungseffekte erfolgte unter Verwendung von Quantenkorrekturpotenzialen für Elektronen und Löcher, welche sich über die Momentenmethode aus der WIGNER-BOLTZMANN-Transportgleichung oder der SCHRÖDINGER-Gleichung herleiten lassen [91].

Unter Einbeziehung der Quantenkorrekturpotenziale in die Transport-, Energiestromdichte und Energiebilanzgleichungen des Energie-Balance-Modells bilden sie zusammen mit der POISSON- und den Kontinuitätsgleichungen das Quanten-Energie-Balance-Modell. Der Nachteil dieses Modells ist, dass die zusätzlichen Terme für die Einbeziehung der Quantenkorrekturpotenziale die Rechenzeit erhöhen sowie zu schlechteren Konvergenzeigenschaften für das gesamte Gleichungssystem führen. Aus diesem Grund wurde das Quanten-Energie-Balance-Modell nur für die Simulation des Bereichs unter dem Gate verwendet, wohingegen alle anderen Transistorbereiche mit dem Energie-Balance-Modell simuliert wurden (Bild 4.5).

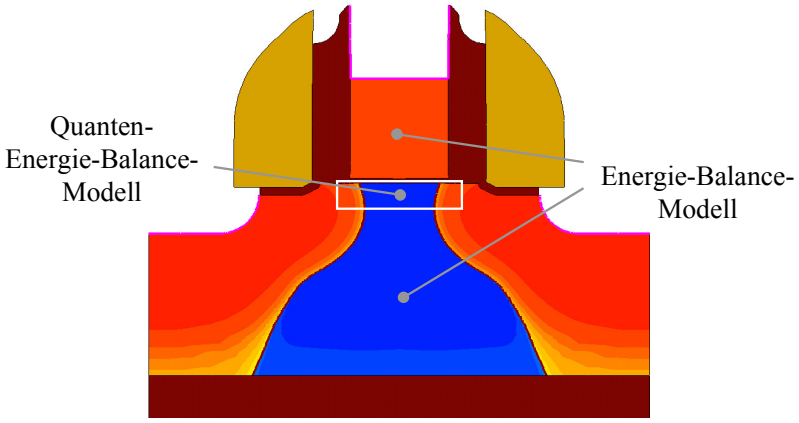


Bild 4.5: Darstellung des Bereichs für die Simulation mit Hilfe des Quanten-Energie-Balance-Modells

b) Ladungsträgerbeweglichkeit

Die Bewegungen der Ladungsträger im Halbleiter werden durch Streuprozesse gestört, wodurch die Ladungsträger auf ihrem Weg durch den Kristall abgebremst werden. Die klassische Beschreibung der Ladungsträgerbeweglichkeit erfolgt nach dem Drude-Modell gemäß

$$\mu_{n,p} = q \cdot \frac{\overline{\tau}_{n,p}}{m_{n,p}^{eff}}, \tag{4.25}$$

wobei q die Elementarladung, $\overline{\tau}_{n,p}$ die mittlere freie Flug- bzw. der Impulsrelaxationszeit und $m_{n,p}^{eff}$ die effektive Ladungsträgermasse ist. Der Einfluss der effektiven Masse auf die Ladungsträgerbeweglichkeit, der in Abschnitt 4.3.3 näher beschrieben ist, wird maßgeblich durch die Beschaffenheit der Energiebänder im Halbleiter bestimmt.

Die mittlere freie Flugzeit $\bar{\tau}_{n,p}$ der Ladungsträger stellt die durchschnittliche Zeit zwischen zwei aufeinanderfolgenden Streuereignissen dar, bei denen ein Energieübertrag stattfindet. Mit zunehmender Streuwahrscheinlichkeit nimmt die freie Flugzeit $\bar{\tau}_{n,p}$ und folglich auch die Ladungsträgerbeweglichkeit $\mu_{n,p}$ ab. Die mittlere freie Flugzeit $\bar{\tau}_{n,p}$ der Ladungsträger wird durch drei verschiedene Mechanismen bestimmt [92]:

- COULOMB-Streuung: Streuung an geladenen und ungeladenen Störstellen im Kristall sowie an geladenen Grenzflächenzuständen der SiO_2/Si-Grenzfläche und festen Oxid-Ladungen,
- Phononen-Streuung: Streuung der Ladungsträger an optischen und akustischen Phononen, die sich mit steigender Temperatur erhöht,
- Oberflächen-Streuung: Streuung an den Mikro-Rauheiten der Si/SiO_2-Grenzfläche.

Welcher dieser Mechanismen dominiert, hängt dabei von der Feldstärke $\boldsymbol{E}_\perp$ senkrecht zur Si/SiO_2-Grenzfläche, also vom vertikalen elektrischen Feld, ab [93] (Bild 4.6). Im Bereich der Unterschwellsteigung überwiegt neben der Phononen-Streuung die COULOMB-Streuung. Bei starker Inversion ($U_{GS} = U_{DD}$) kann die COULOMB-Streuung vernachlässigt werden, da die Streuzentren aufgrund der hohen Anzahl an Ladungsträgern abgeschirmt sind, und es überwiegt neben der Phononen-Streuung vor allem die Streuung an den Mikro-Rauheiten der Si/SiO_2-Grenzfläche.

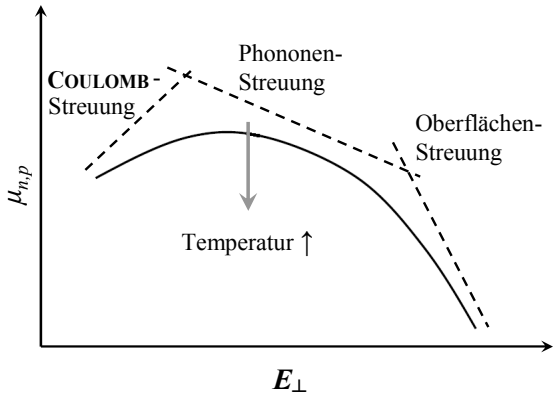


Bild 4.6: Ladungsträgerbeweglichkeit $\mu_{n,p}$ in Abhängigkeit vom vertikalen elektrischen Feld $\boldsymbol{E}_\perp$ [92]

Die Veränderung der Ladungsträgerbeweglichkeit infolge von thermischen Gitterschwingungen sowie durch Streuung an geladenen und ungeladenen Störstellen im Kristall infolge der Dotierung des Halbleiters lässt sich nach [94] modellieren:

$$\mu_{n,p}^{C,T} = \mu_{n,p}^{\min} + \frac{\mu_{n,p}^{d}}{1+\left[\left(N_A+N_D\right)/N_{n,p}^{0}\right]^{A_{n,p}}}, \tag{4.26}$$

wobei N_A und N_D die Akzeptoren- bzw. Donatorenkonzentration ist und die temperaturabhängigen Parameter $\mu_{n,p}^{\min}$, $\mu_{n,p}^{d}$, $N_{n,p}^{0}$ und $A_{n,p}$ zur Kalibrierung des Modells dienen.

Die Veränderung der Beweglichkeit an der Si/SiO_2-Grenzfläche durch vertikale elektrische Felder wird durch das Modell nach [95] beschrieben:

$$\mu_{n,p}^{SR} = \left\{ \left[\left(B_{n,p}^{SR} \cdot \frac{T_L}{E_\perp} + C_{n,p}^{SR} \cdot \frac{1\mathrm{K}}{E_\perp^{1/3}} \right) \cdot \frac{1}{T_L} \right]^{-1} + \left[\frac{\delta_{n,p}^{SR}}{\left(E_\perp / E_{ref}\right)^2} \right]^{-1} \right\}. \tag{4.27}$$

Der erste Term beschreibt dabei die Beweglichkeit unter Berücksichtigung der akustischen Oberflächenphononen, während der zweite Term den Anteil der Oberflächenrauheiten berücksichtigt. $B_{n,p}^{SR}$, $C_{n,p}^{SR}$ und $\delta_{n,p}^{SR}$ sind Parameter zum Anpassen des Modells an experimentelle Daten.

Weiterhin führen hohe elektrische Felder in Stromflussrichtung zu einem Aufheizen der Ladungsträger. Die Ladungsträger werden nicht unbegrenzt durch das elektrische Feld beschleunigt, sondern ihre Geschwindigkeit als Produkt von Beweglichkeit und elektrischem Feld nähert sich der Sättigungsdriftgeschwindigkeit an. Diese sogenannten „heißen“ Ladungsträger besitzen eine Abhängigkeit ihrer Beweglichkeit vom elektrischen Feld in einer solchen Weise, dass sie sich bei steigendem Feld verringert.

Der Einfluss hoher elektrischer Feldstärken in Stromflussrichtung auf die Ladungsträgerbeweglichkeit ist nach [96] gegeben durch

$$\mu_{n,p}^{HFS} = \frac{\mu_{n,p}^{E,low}}{\left[1 + \left(\left(\mu_{n,p}^{E,low} \cdot E_{n,p}^{HFS} \right) / v_{n,p}^{sat} \right)^{\beta_{n,p}^{HFS}} \right]^{1/\beta_{n,p}^{HFS}}}. \tag{4.28}$$

Dabei ist

$$v_{n,p}^{sat} = v_{n,p}^{sat,0} \cdot \left(\frac{300\mathrm{K}}{T} \right)^{v_{n,p}^{sat,\exp}} \tag{4.29}$$

die Sättigungsdriftgeschwindigkeit, welche in Silizium bei 300 K für Elektronen $1.07 \cdot 10^7$ cm/s und für Löcher $0.84 \cdot 10^7$ cm/s beträgt. Zur Anpassung des Modells an experimentelle Daten können die Parameter $\beta_{n,p}^{HFS}$ und $v_{n,p}^{sat,\exp}$ genutzt werden. Die Nutzung des hydrodynamischen Modells ermöglicht weiterhin die Verwendung eines von den Ladungsträgertemperaturen T_n und T_p abhängigen elektrischen Felds

$$E_{n,p}^{HFS} = \frac{\left(\omega_{n,p} - w_0\right)}{q \cdot v_{n,p}^{sat} \cdot \tau_{\omega n,\omega p}} = \frac{3}{2} \cdot \frac{k_B \cdot \left(T_{n,p} - T_L\right)}{q \cdot v_{n,p}^{sat} \cdot \tau_{\omega n,\omega p}}. \tag{4.30}$$

Der Parameter $\mu_{n,p}^{E,low}$ in Gleichung (4.28) ist die Ladungsträgerbeweglichkeit bei kleinen elektrischen Feldstärken in Stromflussrichtung. Die wird unter Berücksichtigung der bereits angesprochenen Streumechanismen und unter der Voraussetzung der Unabhängigkeit der einzelnen Beiträge nach der MATHIESSENschen-Regel bestimmt zu

$$\frac{1}{\mu_{n,p}^{E,low}} = \frac{1}{\mu_{n,p}^{C,T}} + \frac{1}{\mu_{n,p}^{SR}}. \tag{4.31}$$

c) Generation/Rekombination

Im thermodynamischen Gleichgewicht werden durch thermische Anregung gleich viele Elektronen vom Valenz- ins Leitungsband angeregt wie Elektronen mit Löchern im Valenzband rekombinieren. Auf diesem Weg bleibt die Dichte der Elektronen und Löcher konstant. Abweichungen der Ladungsträger von ihrem Wert im thermodynamischen Gleichgewicht führen zu einem Überwiegen der Generation oder Rekombination, da der Halbleiter bestrebt ist, wieder in den Zustand des thermischen Gleichgewichts zu gelangen. Sowohl bei der Generation als auch bei der Rekombination von Elektron-Loch-Paaren muss die Energie- und Impulserhaltung gewährleistet sein. Dies kann über verschiedene Mechanismen sichergestellt werden, erfordert aber stets die Wechselwirkung mit einem dritten Teilchen (Photonen, Störstellen oder Ladungsträger), dass die frei werdende Energie aufnimmt.

Die beim MOSFET entscheidenden Generations-Rekombinations-Prozesse sind die SHOCKLEY-READ-HALL- (SRH-) Rekombination, die AUGER-Rekombination, das Band-zu-Band-Tunneln und die Stoßionisation. Die Rekombination muss im Silizium aufgrund der Beschaffenheit des Leitungs- und Valenzbands auf indirektem Weg stattfinden. Das bedeutet, dass die Ladungsträger beim Übergang von einem in das andere Band in ein Rekombinationszentrum springen, welches sich energetisch in der Bandlücke befindet. Die Rekombinationsrate infolge der SRH-Rekombination wird im Bauelementesimulator beschrieben durch

$$R_{SRH} = \frac{n \cdot p - n_i}{\tau_p \cdot (n + n_1) + \tau_n \cdot (p + p_1)} \tag{4.32}$$

mit

$$n_1 = n_i \cdot \exp\left(\frac{E_{trap}}{k_B \cdot T} \right) \tag{4.33}$$

und

$$p_1 = n_i \cdot \exp\left(-\frac{E_{trap}}{k_B \cdot T} \right). \tag{4.34}$$

E_{trap} ist dabei die Energiedifferenz zwischen der Energie der Rekombinationszentren und der FERMI-Energie des undotierten, intrinsischen Halbleiters und n_i die intrinsische Ladungsträgerdichte. Die Ladungsträgerlebensdauern τ_n und τ_p werden mit Hilfe des Modells nach [97] in Abhängigkeit von der Dotierung betrachtet, da mit höherer Dotierung die Anzahl an Defekten und somit an Rekombinationszentren größer wird.

Die Rekombinationsrate der AUGER-Rekombination wird durch das Modell nach [98] berücksichtigt. Diese Form der Rekombination stellt den dominierenden Mechanismus in stark dotierten indirekten Halbleitern dar [99]. Die AUGER-Rekombination beruht auf einem Dreiteilchenprozess. Neben dem rekombinierenden Elektron-Loch-Paar ist noch ein drittes Loch oder Elektron beteiligt, auf das ein Teil der umgesetzten Energie in Form von kinetischer Energie übertragen wird. Der AUGER-Prozess erfolgt bei Silizium wie bei der SRH-Rekombination aufgrund der indirekten Bandübergänge unter Beteiligung von Rekombinationszentren innerhalb der Bandlücke.

Die Ladungsträgergeneration infolge der Stoßionisation wird durch das Modell nach [100] berücksichtigt. Bei hohen Stromdichten und großen lateralen elektrischen Feldern kommt es durch Stoßionisation zur Generation von Elektron-Loch-Paaren, wenn die Ladungsträger über einen kritischen Wert beschleunigt werden.

Die Generationsrate steigt mit der Anzahl von freien Ladungsträgern und kann zu einem Lawinendurchbruch führen. Die Stoßionisation ist beim PD-SOI-MOSFET für den ausgeprägten Floating-Body-Effekt verantwortlich.

Weiterhin können Band-zu-Band-Tunnelvorgänge aufgrund hoher elektrischer Felder bzw. steiler pn-Übergänge nicht mehr vernachlässigt werden. An pn-Übergängen mit stark abfallenden Dotierungsprofilen sind die Bandkanten so stark verbogen, dass die Möglichkeit des direkten Tunnelns von Elektronen aus dem Valenz- in das Leitungsband besteht. Ein elektrisches Feld in Sperrrichtung des pn-Übergangs führt zu einer weiteren Verbiegung der Bänder, was die Wahrscheinlichkeit eines Tunnelvorgangs vergrößert. Wird das elektrische Feld erhöht oder der Bandabstand verringert, steigt die Wahrscheinlichkeit für Tunnelvorgänge und somit wächst auch die Ladungsträgergenerationsrate. Die Generation von Ladungsträgern findet bei hohen Drain-Source-Spannungen am drainseitigen pn-Übergang statt und verursacht neben der Stossionisation ebenfalls einen Floating-Body-Effekt.

Die Generation durch direkte Band-zu-Band-Tunnelübergänge wird bei der Simulation durch das Modell nach [101] beschrieben, welches sowohl die Wahrscheinlichkeit der Ladungsträgergeneration mit Bezug auf die elektrische Feldstärke als auch infolge der temperaturbedingten Änderung der Bandlücke berücksichtigt.

4.3.3 Modellierung der mechanischen Verspannungen

Die mechanische Verspannung infolge von Deformationen des Siliziumkristalls verändert den Abstand zwischen den Atomen des Kristalls und damit dessen Bandstruktur. Als Folge davon ändern die Bänder in den Kristallrichtungen ihre energetische Lage, die Bandkanten heben oder senken sich. Weiterhin kommt es zur Aufspaltung der entarteten Bänder sowie zu einer Bandverbiegung. Ersteres hat eine Umbesetzung der Ladungsträger zwischen den Bändern zu Folge, wodurch neben der Interband- und Intraband-Streuung die effektive Masse der Ladungsträger verändert wird. Demgegenüber beeinflusst die Bandverbiegung in erster Linie die effektive Ladungsträgermasse.

Die effektive Masse der Ladungsträger $m_{n,p}^{eff}$ ist über die Beziehung

$$m_{n,p}^{eff} = \hbar^2 \cdot \left(\frac{d^2 E}{dk^2} \right)^{-1} \tag{4.35}$$

mit der Krümmung des Bandes gekoppelt. Eine stärkere Biegung des Bandes hat eine Verringerung der effektiven Masse der Ladungsträger zur Folge und umgekehrt. Folglich kann eine mechanische Verspannung, hervorgerufen durch eine Verformung des Siliziumkristalls, die Ladungsträgerbeweglichkeit durch eine Verringerung der effektiven Masse und gemäß Gleichung (4.25) durch eine Verringerung der Streurate verbessern.

a) Unverspanntes Silizium

Die im Folgenden gezeigten Bänderdiagramme wurden mit der empirischen Pseudopotenzialmethode gemäß [102] erstellt.

Bild 4.7 stellt die Bandstruktur bzw. die erste BRILLOUIN-Zone für unverspanntes Silizium dar. Die beiden signifikanten Punkte, das Valenzbandmaximum im Γ-Punkt und das Leitungsbandminimum in der Nähe des X-Punkts, bestimmen die Größe der Bandlücke E_g. Der Wellenvektor $k_{[110]}$ vom Γ- zum K-Punkt entspricht der Kanalrichtung der untersuchten Transistoren.

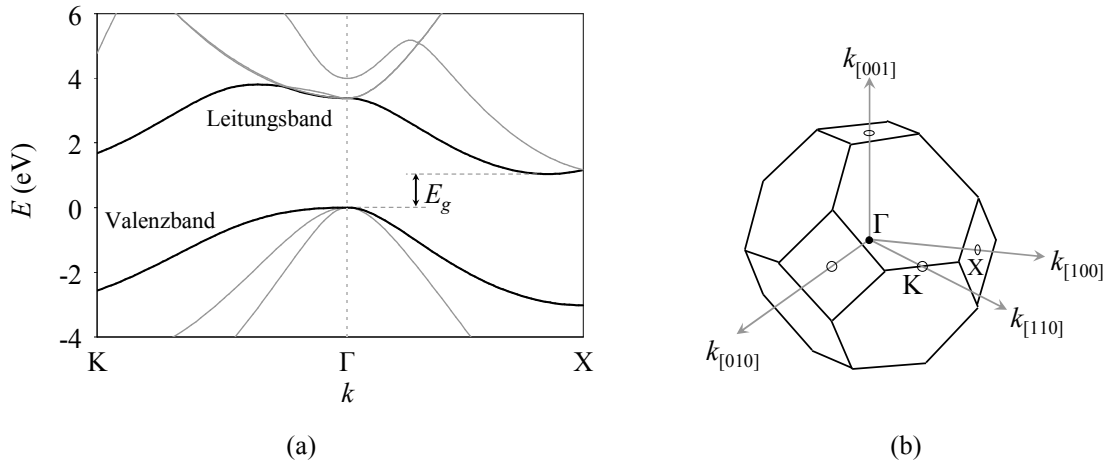


Bild 4.7: (a) Bandstruktur und (b) erste BRILLOUIN-Zone für unverspanntes Silizium

Bild 4.8 stellt die Isoenergieflächen für das Leitungsband dar. Innerhalb eines solchen von einer Isoenergiefläche aufgespannten Volumens besitzen die Ladungsträger eine hohe Aufenthaltswahrscheinlichkeit. Die Isoenergieflächen der Elektronen können annähernd als Rotationsellipsoide betrachtet werden, welche auch als Täler bezeichnet werden. Da diese sechs Täler (zwei in jede Raumrichtung) für unverspanntes Silizium alle energetisch identisch sind, werden sie als entartete Täler bezeichnet.

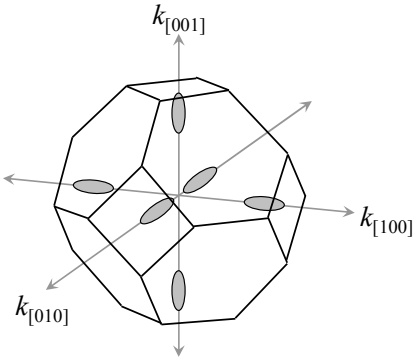


Bild 4.8: Erste BRILLOUIN-Zone mit Isoenergieflächen der Elektronen für unverspanntes Silizium

Die effektive Masse m_n^{eff} dieser Täler lässt sich durch zwei verschiedene Massen, eine longitudinale und eine transversale Masse, beschreiben. Die effektive longitudinale Masse m_n^l (parallel zu den Hauptachsen) hat einen Wert von etwa $0.98 \cdot m_0$, mit der Elektronenruhemasse $m_0 = 9.11 \cdot 10^{31}$ kg, während die effektive transversale Masse m_n^t (senkrecht zu den Hauptachsen) einen Wert von circa $0.19 \cdot m_0$ hat [103].

Der Unterschied in den Massen ist auf die verschiedene Krümmung des Leitungsbandes in den einzelnen Richtungen des *k*-Raums zurückzuführen Die resultierende (isotrope) effektive Masse m_n^{eff} der Elektronen berechnet sich aus dem Mittelwert der einzelnen Massen der sechs Täler zu

$$m_n^{eff} = \frac{1}{6} \cdot \left[\left(\frac{2}{m_n^l} \right) + \left(\frac{4}{m_n^t} \right) \right]^{-1} . \tag{4.36}$$

Bild 4.9 zeigt die Bandstruktur in Kanalrichtung bzw. die Isoenergieflächen für das Valenzband. Man unterscheidet zwischen den schweren Löchern des Bandes mit geringer Krümmung und den leichten Löchern des Bandes mit starker Krümmung. Die effektiven Massen der Löcher ergeben sich aufgrund der Krümmung für schwere Löcher zu circa $0.49 \cdot m_0$ und für leichte Löcher zu etwa $0.16 \cdot m_0$ [103]. Im unverspannten Zustand besetzen die Ladungsträger vorwiegend das obere Valenzband der schweren Löcher. Dieses ist im Γ-Punkt mit dem Valenzband der leichten Löcher entartet. Das sogenannte Spin-Orbit-Band befindet sich circa 44 meV unterhalb der beiden Bänder.

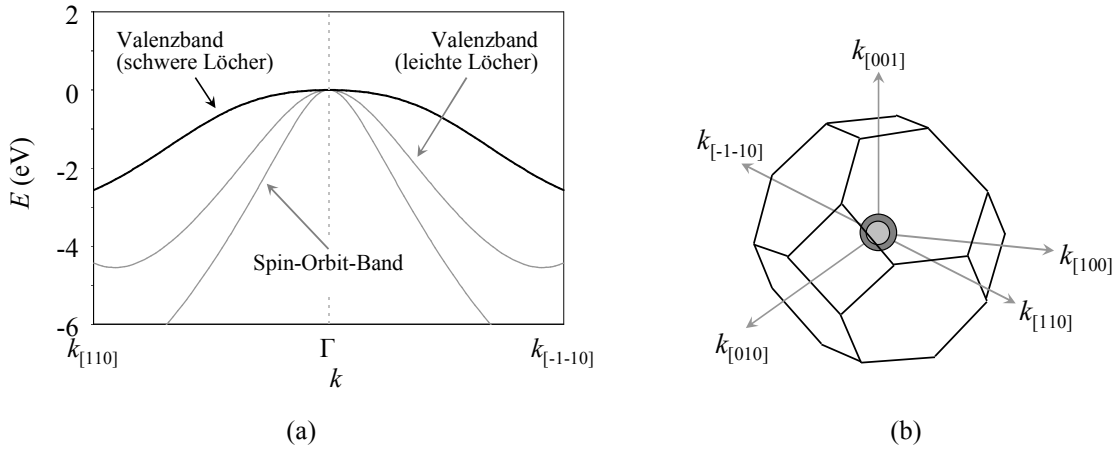


Bild 4.9: (a) Valenzbandstruktur in Kanalrichtung und (b) erste BRILLOUIN-Zone mit den Isoenergieflächen der Löcher für unverspanntes Silizium

b) Verspanntes Silizium

Die folgenden Betrachtungen beziehen sich auf eine uniaxiale Verspannung in Kanalrichtung, welche durch das in Abschnitt 4.2.3 beschriebene Einbetten von verspannten Schichten in die SD-Gebiete des Transistors erreicht wird.

Bild 4.10 zeigt die Isoenergieflächen der Elektronen für unverspanntes und für uniaxial zugverspanntes Silizium. Im Fall einer uniaxialen Zugverspannung wird die 6-fache Entartung des Leitungsbandes aufgehoben, indem die Energie der 2-fach entarteten Täler (Δ2) senkrecht zur Waferoberfläche im Vergleich zu den 4-fach entarten Tälern (Δ4) in der Ebene abgesenkt wird. Die Elektronen besetzen nun bevorzugt diese beiden Täler mit der geringeren effektiven Masse in Kanalrichtung. Zusätzlich wird durch die Aufspaltung der Bänder die Streuung der Elektronen zwischen den Tälern (Interband-Streuung) reduziert und somit die Beweglichkeit weiter verbessert [104]. Weiterhin führt eine in [110]-Kanalrichtung wirkende Zugverspannung zu einer leichten Verbiegung der 2-fach entarteten Täler, was in einer leichteren transversalen effektiven Masse in Transportrichtung resultiert [105]. Wird in dieser Kristallrichtung eine Druckverspannung angenommen, entsteht ein gegensätzliches Verhalten.

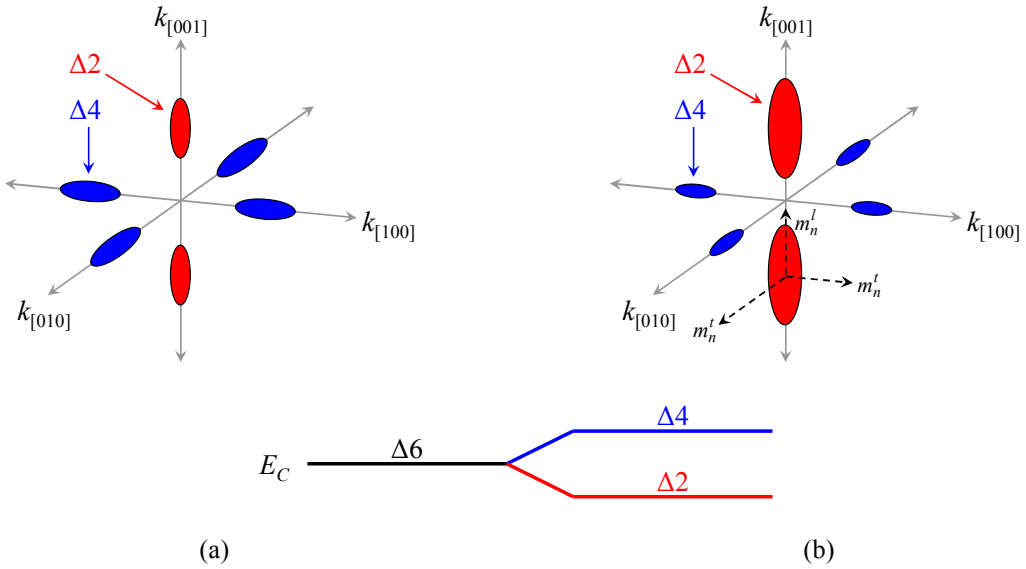


Bild 4.10: Isoenergieflächen der Elektronen für (a) unverspanntes und (b) uniaxial zugverspanntes Silizium in [110]-Kanalrichtung

Für Löcher ist eine uniaxiale Druckverspannung in Kanalrichtung von Vorteil. Aufgrund dieser Art der Verspannung kommt es zur Aufhebung der Entartung der leichten und schweren Löcherbänder im Γ-Punkt des *k*-Raums (Bild 4.11). Die Besetzungswahrscheinlichkeit für das oberste Band nimmt mit stärkerer Aufspaltung zu und die Streuprozesse der Löcher zwischen den beiden oberen Valenzbändern werden verringert. Diese Streuprozesse sind im unverspannten Silizium der Hauptgrund für die schlechte Löcherbeweglichkeit im Vergleich zur Elektronenbeweglichkeit [104]. Zusätzlich wird das obere Band so verbogen, dass in Kanalrichtung die effektive Masse der Löcher abnimmt. Eine Zugverspannung in Kanalrichtung führt zu einer Verringerung der Löcherbeweglichkeit.

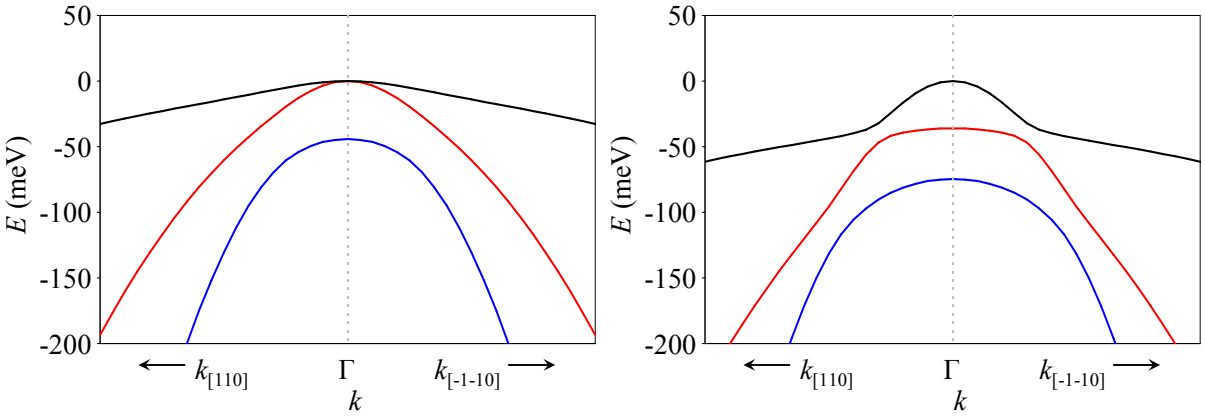


Bild 4.11: Valenzbandstruktur für (a) unverspanntes und für (b) mit 1 GPa in Kanalrichtung druckverspanntes Silizium

c) Spannungsabhängige Beweglichkeitsmodelle

Die Beschreibung der Bandverschiebung durch eine Verformung des Kristalls erfolgt bei der hier durchgeführten Bauelementesimulation durch das Deformationspotenzialmodell [106]. Dabei wird von einer linearen Verschiebung der Bänder ausgegangen, welche über sogenannte Deformationspotenziale mit der Verspannung gekoppelt sind.

Die energetische Verschiebung des Leitungsbands bzw. Valenzbands berechnet sich nach [107] zu

$$\Delta E_{C,i} = \Xi_1 \cdot \left(\varepsilon_{11} + \varepsilon_{22} + \varepsilon_{33}\right) + \Xi_2 \cdot \varepsilon_{ii} \qquad i = 1, 2, 3 \tag{4.37}$$

bzw.

$$\Delta E_{V,i} = -\xi_1 \cdot \left(\varepsilon_{11} + \varepsilon_{22} + \varepsilon_{33}\right) \pm \delta E \qquad i = 1, 2 \tag{4.38}$$

mit

$$\delta E = \sqrt{\frac{\xi_2^2}{2} \cdot \left[\left(\varepsilon_{11} - \varepsilon_{22}\right)^2 + \left(\varepsilon_{22} - \varepsilon_{33}\right)^2 + \left(\varepsilon_{11} - \varepsilon_{33}\right)^2\right] + \xi_3^2 \cdot \left(\varepsilon_{12}^2 + \varepsilon_{13}^2 + \varepsilon_{23}^2\right)} \,. \tag{4.39}$$

Dabei sind ε_{ii} der Dehnungstensor und Ξ_1, Ξ_2, ξ_1, ξ_2 und ξ_3 die Deformationspotenziale. Der Index i bezeichnet beim Leitungsband die drei verschiedenen, jeweils zweifach entarteten Täler in den drei Raumrichtungen und beim Valenzband die beiden oberen Löcherbänder.

Die Mittelwerte

$$\Delta E_C = -k_B \cdot T_L \cdot \ln\left[\frac{1}{3} \cdot \sum_{i=1}^{3} \exp\left(-\frac{\Delta E_{C,i}}{k_B \cdot T_L}\right)\right] \tag{4.40}$$

und

$$\Delta E_V = k_B \cdot T_L \cdot \ln\left[\frac{1}{2} \cdot \sum_{i=1}^{2} \exp\left(\frac{\Delta E_{V,i}}{k_B \cdot T_L}\right)\right] \tag{4.41}$$

der einzelnen Subbandverschiebungen werden anschließend zur Bestimmung der neuen Lage für das Leitungs- bzw. Valenzband verwendet, woraus schließlich die Bandlücke E_g neu berechnet wird. Dieses Modell erfasst aber nur den Einfluss einer Verspannung auf die Anzahl der Ladungsträger sowie die Veränderung der Schwellspannung aufgrund der Änderung der Bandlücke. Mit diesem Modell wird allerdings nicht der Einfluss eine mechanische Verspannung auf die Ladungsträgerbeweglichkeit berücksichtigt.

Das in der Simulation genutzte Modell nach [108] zur Beschreibung des Einflusses einer mechanischen Verspannung auf die Elektronenbeweglichkeit basiert auf der Umbesetzung der Elektronen zwischen den einzelnen Tälern aufgrund der Aufhebung der Entartung. Mit der Änderung der Leitungsbandenergie nach dem Deformationspotenzialmodell ergibt sich die Elektronenbeweglichkeit zu

$$\mu_n^{str} = \mu_n^{unstr} \cdot \left\{1 + \frac{1 - m_n^l / m_n^t}{1 + 2 \cdot \left(m_n^l / m_n^t\right)} \cdot \left[\exp\left(\frac{\Delta E_C - \Delta E_{C,i}}{k_B \cdot T}\right) - 1\right]\right\}, \tag{4.42}$$

wobei μ_n^{unstr} die Elektronenbeweglichkeit im unverspannten Zustand, m_n^l und m_n^t die effektive longitudinale und transversale Masse der Elektronen im jeweiligen Tal, ΔE_C nach Gleichung (4.40) der Mittelwert der einzelnen Subbandverschiebungen des Leitungsbandes und $\Delta E_{C,i}$ nach Gleichung (4.37) die Energie eines Subbandes entsprechend der Koordinatenrichtung i ist.

Das verwendete Modell nach [109] zur Beschreibung der verspannungsabhängigen Veränderung der Löcherbeweglichkeit basiert auf der Veränderung der effektiven Masse aufgrund der Umverteilung der Ladungsträger und der Verbiegung der Bänder. In diesem Modell wird davon ausgegangen, dass sich 75 % der Ladungsträger im Band der schweren Löcher befinden und somit hauptsächlich am Stromfluss beteiligt sind. Die verspannungsabhängige Besetzung dieses Band wird dabei durch zwei Ellipsen beschrieben, welche im unverspannten Fall symmetrisch sind (Bild 4.12).

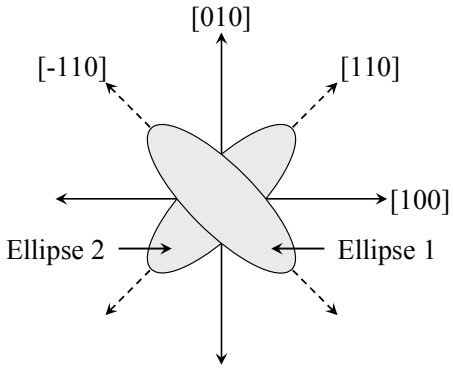


Bild 4.12: Definition der Ellipsen zur Modellierung der verspannungsabhängigen Löcherbeweglichkeit nach [109]

Jede Ellipse wird durch eine longitudinale und eine transversale Masse beschrieben, aus denen anschließend eine gemittelte Beweglichkeit bestimmt wird. Durch eine Verspannung ändert sich die Lage der beiden Ellipsen zueinander, wodurch eine Umverteilung der Ladungsträger stattfindet, was wiederum eine Veränderung der mittleren effektiven Masse und somit die Beweglichkeit zur Folge hat. Die verspannungsabhängige Löcherbeweglichkeit in [110]-Kanalrichtung berechnet sich nach [109] zu

$$\mu_{p,[110]}^{str} = \mu_p^{unstr} \cdot \frac{2 \cdot m_p^{l_0} \cdot m_p^{t_0}}{m_p^{l_0} + m_p^{t_0}} \cdot \left(\frac{f_1}{m_p^{t_1}} + \frac{f_2}{m_p^{l_2}} \right), \tag{4.43}$$

wobei μ_p^{unstr} die Löcherbeweglichkeit im unverspannten Zustand, m_p^l und m_p^t die longitudinale bzw. transversale Masse und f die Besetzungswahrscheinlichkeit der Ellipsen ist. Der Index 0 bezieht sich auf die effektiven Massen im unverspannten Zustand und die Indizes 1 und 2 gehören zu den beiden Ellipsen. Die Besetzungswahrscheinlichkeiten der ersten und zweiten Ellipse ergeben sich zu

$$f_1 = \frac{1}{1 + \exp\left(-E_\Delta / k_B \cdot T_L\right)}, \tag{4.44}$$

$$f_2 = 1 - f_1 . \tag{4.45}$$

Hier ist E_Δ die Energieaufspaltung zwischen den beiden Ellipsen, welche letztendlich die Besetzung der beiden Ellipsen bestimmt. Der Unterschied in der energetischen Lage zwischen den Ellipsen wird durch

$$E_\Delta = d \cdot \sigma_s \tag{4.46}$$

bestimmt. Der Parameter d stellt dabei den Zusammenhang zwischen Energie E_Δ und Scherkomponente σ_s des Verspannungstensors

$$\sigma_{ij} = \begin{bmatrix} \sigma_s & \sigma_b \\ \sigma_b & \sigma_s \end{bmatrix} \tag{4.47}$$

her. σ_b ist dabei die biaxiale Verspannungskomponente.

Die Scherkomponente ist für die Verformung des Valenzbandes und der damit einhergehenden Umbesetzung innerhalb des obersten Valenzbandes verantwortlich [110]. Eine Scherverspannung entsteht bei einer uniaxialen Deformation, die nicht entlang einer der Hauptachsen verläuft, sondern wie die [110]-Richtung mindestens zwei Richtungskomponenten besitzt. Eine Scherverspannung in Kanalrichtung entsteht z.B. durch das Einbetten von verspannten Schichten in die SD-Gebiete des Transistors.

Zusätzlich kommt es bei einer Verspannung in [110]-Kanalrichtung zu einer Veränderung der transversalen effektiven Masse $m_p^{t_1}$, welche durch

$$\frac{1}{m_p^{t_1}} = \frac{1}{m_p^{t_0}} \cdot \left(1 + s_{t,1} \cdot \sigma_s + s_{t,2} \cdot \sigma_s^2 + b_{t,1} \cdot \sigma_b + b_{t,2} \cdot \sigma_b^2\right) \tag{4.48}$$

beschrieben wird. Die zusätzlichen Parameter $s_{t,1}$, $s_{t,2}$, $b_{t,1}$ und $b_{t,2}$ können zur Anpassung des Modells an experimentelle Daten genutzt werden. Die longitudinale Masse wird als verspannungsunabhängig betrachtet, da sich nur die Länge der Ellipsennebenachse, nicht aber die der Hauptachse verändert.

4.4 Kalibrierung der Simulationsparameter

Grundlage für die Kalibrierung der Prozess- und Bauelementesimulation war ein SOI-Transistor der 65 nm-Technologie mit Polysilizium als Gatematerial und ein mit Stickstoff versetztes Siliziumoxid als Gateoxid. Weiterhin wurde für die Ausheilung die Spike-RTA verwendet.

4.4.1 Prozesssimulation

Bild 4.13 zeigt eine Methodik zur Kalibrierung der Modelle für die Prozesssimulation. Die Kalibrierung beginnt mit der eindimensionalen Simulation der Dotierungsprofile der Wannen-, Halo-, SDE- und SD-Gebiete mit Hilfe der Implantations- und Diffusionsmodelle. Die entstandenen Ergebnisse werden mit gemessenen SIMS-Profilen verglichen. Eine Plausibilitätsprüfung der simulierten Dotierungsprofile erfolgt mit den Ergebnissen der Schichtwiderstandsmessung.

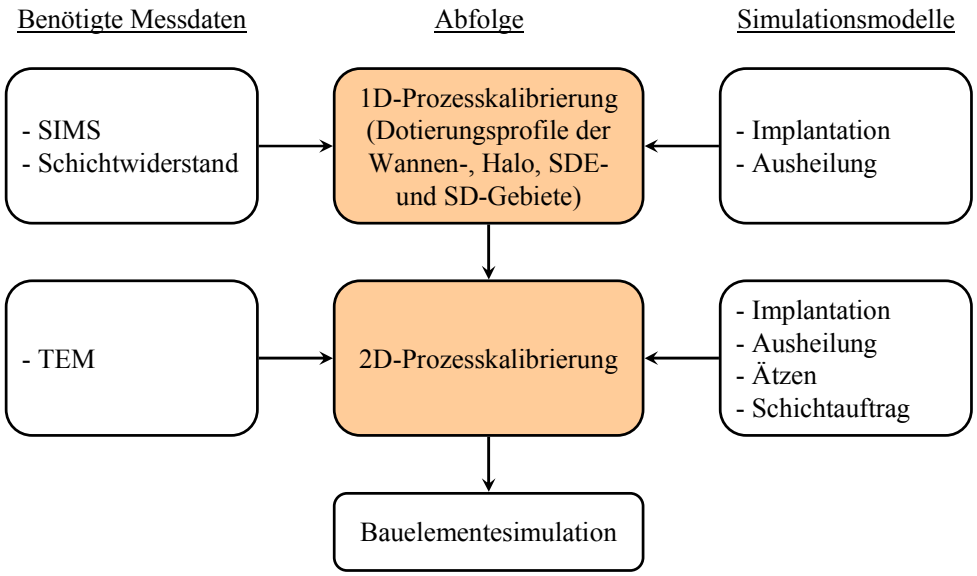


Bild 4.13: Methodik zur Kalibrierung der Modelle der Prozesssimulation

Der nächste Schritt ist die Überführung der erhaltenen Parameter für die Modelle der Implantation und Ausheilung in die zweidimensionale Prozesssimulation, wo mit Hilfe der Modelle für das Ätzen und den Schichtauftrag die Struktur des Transistors nachgebildet wird. Die dafür notwendigen Geometrieparameter werden anhand von TEM-Aufnahmen ermittelt. Die sich nach der Prozesssimulation ergebende Struktur des Transistors ist dann mit Dotierungsprofil sowie Verspannungs- und Deformationsmatrix Ausgangspunkt für die Bauelementesimulation.

Die Bilder 4.14 und 4.15 zeigen die gemessenen und kalibrierten Bor- (B-) und Arsen- (As-) Dotierungsprofile der Halo- und SDE-Gebiete nach der Implantation bzw. nach der Ausheilung für n- und p-MOSFET.

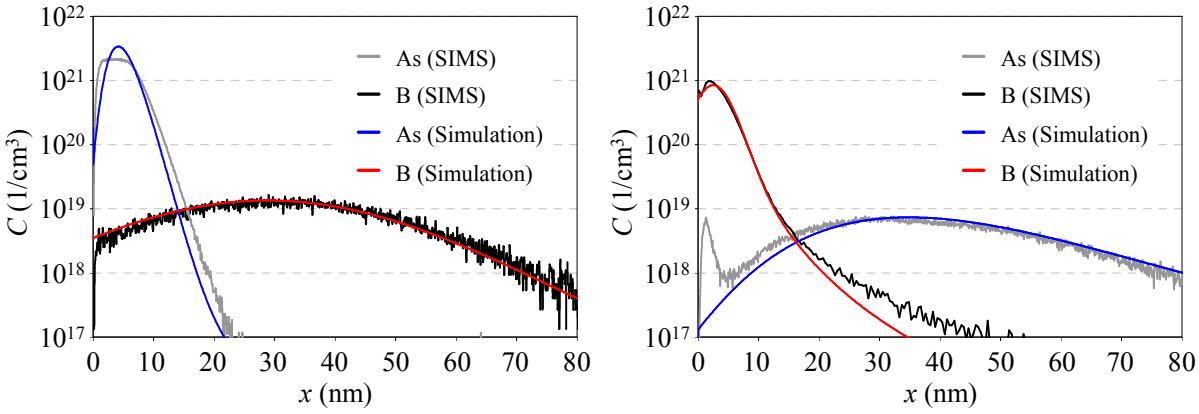


Bild 4.14: Dotierungsprofile der Halo- und SDE-Gebiete nach der Implantation für (a) n- und (b) p-MOSFET

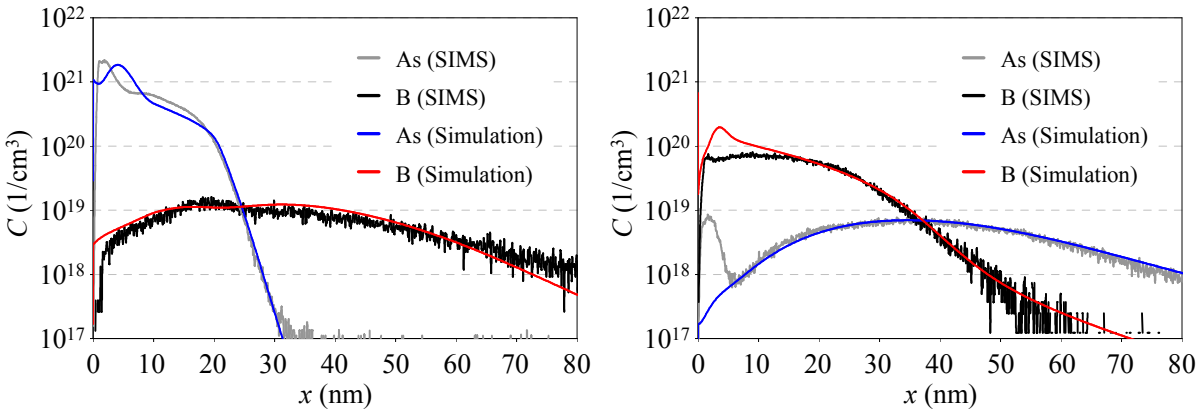


Bild 4.15: Dotierungsprofile der Halo- und SDE-Gebiete nach der Ausheilung für (a) n- und (b) p-MOSFET

Die Plausibilitätsprüfung anhand der Schichtwiderstände der SDE-Gebiete zeigt, dass für beide Transistortypen die simulierten Werte innerhalb der gemessenen Standardabweichung ΔR_p liegen und die Simulation somit vertrauenswürdig erscheint (Bild 4.16).

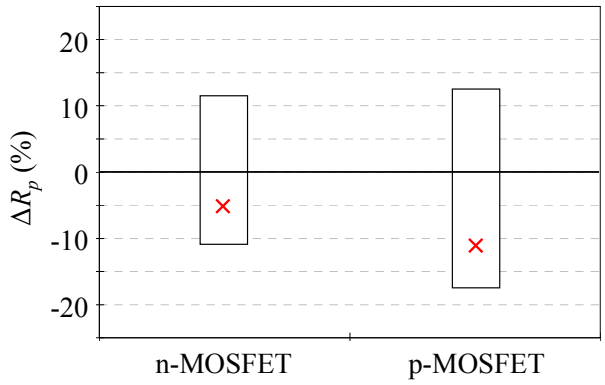


Bild 4.16: Vergleich von gemessenen und simulierten Widerstandswerten der SDE-Gebiete. Boxen: Standardabweichung ΔR_p der gemessenen Schichtwiderstände, Kreuze: simulierte Schichtwiderstände

Bild 4.17 zeigt die TEM-Aufnahme des Transistors, anhand dessen die zur Nachbildung der Struktur des Transistors notwendigen Geometrieparameter ermittelt wurden. Die wichtigsten Größen sind in Tabelle 4.3 angegeben.

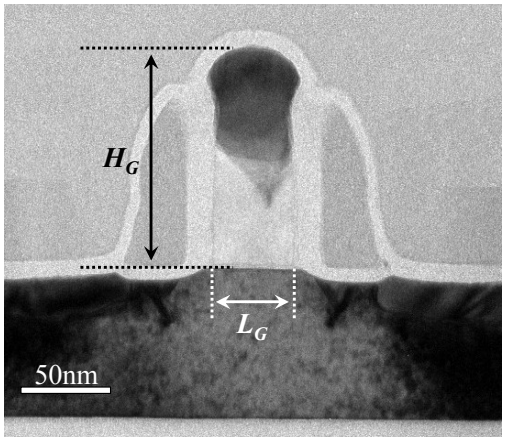


Bild 4.17: TEM-Aufnahme des Transistors, dessen Modellparameter in Abschnitt 4.4.1 kalibriert wurden

Tabelle 4.3 Geometrieparameter der simulierten Transistoren

Variable	Parameter	Wert
L_{PP}	Poly-Pitch-Abstand	190 nm
H_G	Gatehöhe (physikalisch)	120 nm
T_{ox}	Gateoxiddicke (physikalisch)	1.6 nm
L_G	Gatelänge (physikalisch)	45 nm
L_{Sp0}, L_{Sp1}	Spacer0- und Spacer1-Breite	10 nm, 30 nm

Die sich aus der Prozesssimulation ergebende geometrische Struktur ist für den n- und p-MOSFET in Bild 4.18 gegeben. Dabei kennzeichnen rote Bereiche eine überwiegende n-Dotierung, wohingegen blaue Bereiche eine vorherrschende p-Dotierung charakterisieren.

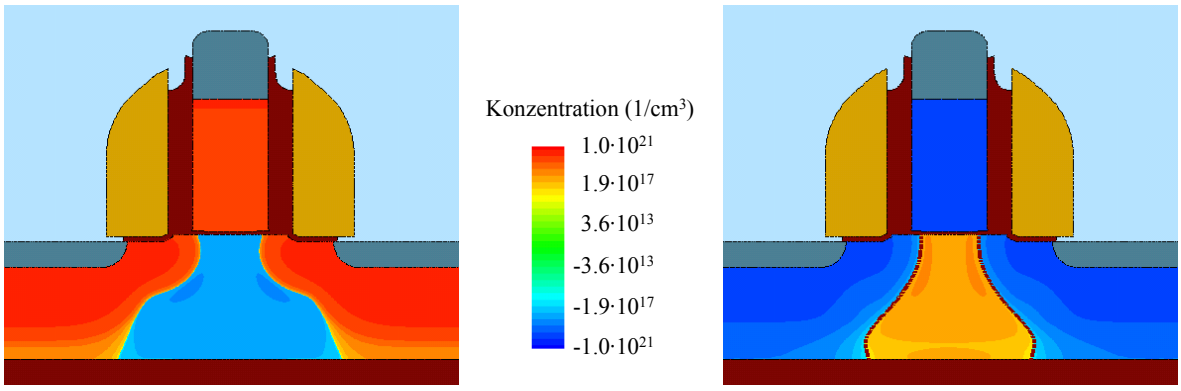


Bild 4.18: Geometrische Struktur des (a) n-MOSFETs und (b) p-MOSFETs nach der Prozesssimulation mit Darstellung des Dotierungsprofils

4.4.2 Bauelementesimulation

Bild 4.19 zeigt die in dieser Arbeit verwendete Methodik zur Kalibrierung der Modelle der Bauelementesimulation. Ausgehend von den Ergebnissen der Prozesssimulation beginnt die Kalibrierung der Bauelementesimulation mit der Simulation der Transferkennlinie eines Langkanaltransistors. Durch Vergleich der Transfercharakteristik und der Schwellspannung mit gemessenen Daten kann die Dotierungskonzentration des Polysiliziumgates kalibriert werden.

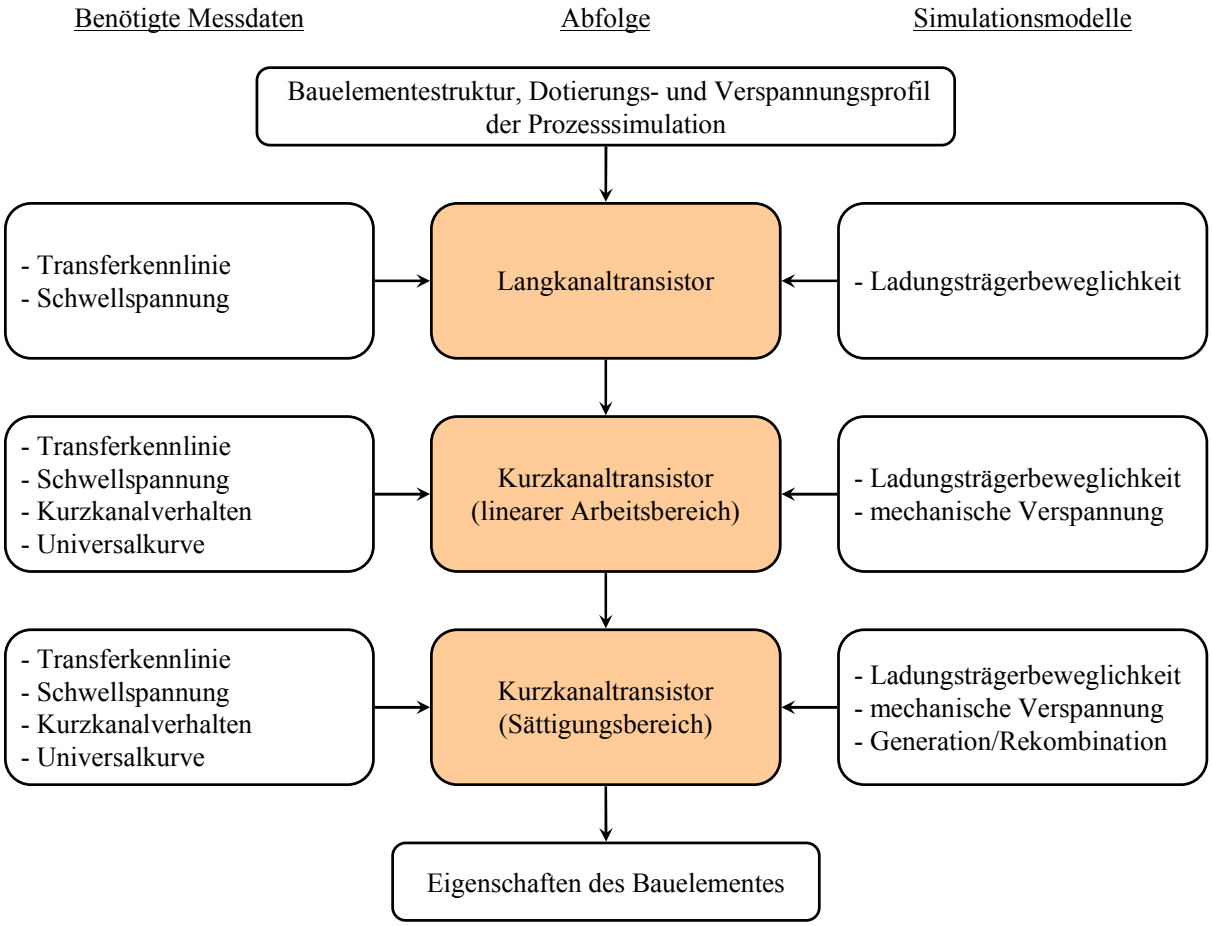


Bild 4.19: Methodik zur Kalibrierung der Modelle der Bauelementesimulation

Der nächste Schritt ist die Simulation der Transferkennlinie im linearen Bereich, da in diesem Arbeitsbereich die Auswirkungen eines hohen lateralen Feldes sowie der Stossionisation und des Band-zu-Band-Tunnelns nicht mit berücksichtigt werden müssen. Einen großen Einfluss haben hingegen die Ladungsträgerbeweglichkeit $\mu_{n,p}$ im Kanalbereich sowie der externe Widerstand R_{ext} des Transistors, wodurch eine Überprüfung und Anpassung der Parameter in den Gleichungen (4.26), (4.27) und (4.31) für die gewählten Beweglichkeitsmodelle sowie ein Abgleich des Widerstandswerts möglich ist. Nach der Kalibrierung des Transistors im linearen Arbeitsbereich erfolgt eine Anpassung der Transferkennlinie im Sättigungsbereich unter Berücksichtigung der Beweglichkeitsmodelle für die Hochfeldsättigung gemäß den Gleichungen (4.26) bis (4.31) sowie der Stossionisation und des Band-zu-Band-Tunnelns. Im Allgemeinen ergibt sich nach einem ersten Simulationslauf für die Bauelementesimulation eine schlechte Übereinstimmung zwischen Simulation und Experiment, so dass oft noch Veränderungen in der Prozesssimulation vorgenommen werden müssen. So kann z.B. eine vom Experiment abweichende Unterschwellsteigung S durch eine Variation der physikalischen Gateoxiddicke T_{ox} oder eine Veränderung der Parameter (Dosis, Energie, Winkel) für die Haloimplantation angepasst werden. Eine generelle Methode zur Kalibrierung kann aber aufgrund der großen Anzahl von Modellen, Modellparametern, Standardwerten und vereinfachenden Annahmen nicht gegeben werden.

a) n-MOSFET

Bei der Bauelementesimulation wurde für das Polysiliziumgate eine feste Dotierungskonzentration n_D von $1.5 \cdot 10^{20}$ cm^{-3} verwendet. Die Bestimmung des Silizid- bzw. Übergangswiderstands $R_{Silizid}$ erfolgte unter Berücksichtigung der Dotierstoffkonzentration im Silizium und der Barrierenhöhe zwischen Silizium und Silizid. Für die Barrierenhöhe wurde entsprechend Abschnitt 2.3.2 ein Wert von $\Phi_{B,n} = 0.72$ eV verwendet. Die Bilder 4.20 und 4.21 stellen die angepasste Transferkennlinie im linearen und im Sättigungsbereich bzw. den Verlauf der Schwellspannung U_{th} in Abhängigkeit von der Gatelänge L_G und die Universalkurve für Experiment und Simulation dar.

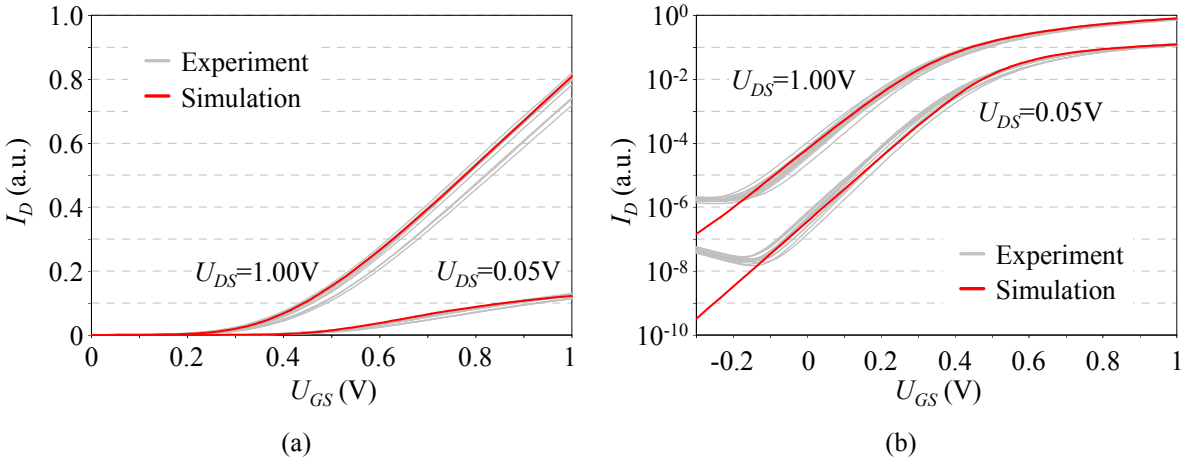


Bild 4.20: Transferkennlinien des n-MOSFETs in (a) linearer und (b) halblogarithmischer Darstellung

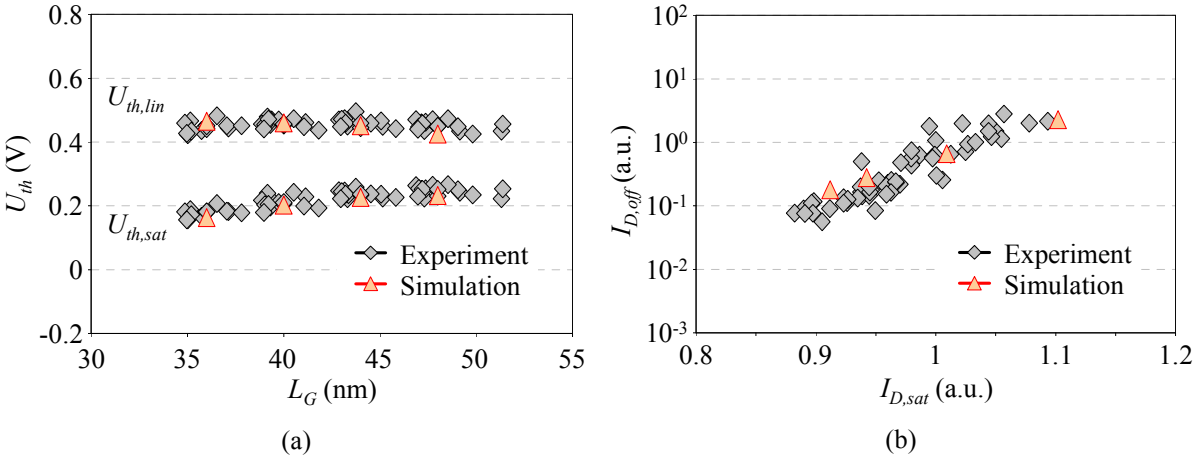


Bild 4.21: (a) Schwellspannung U_{th} in Abhängigkeit von der Gatelänge L_G und (b) Universalkurve des n-MOSFETs

b) p-MOSFET

Beim p-MOSFET wurde für das Polysiliziumgate eine feste Dotierungskonzentration n_A von $0.8 \cdot 10^{20}$ cm^{-3} verwendet. Als Barrierenhöhe $\Phi_{B,p}$ zur Bestimmung des Silizid- bzw. Übergangswiderstands $R_{Silizid}$ wurde entsprechend Abschnitt 2.3.2 ein Wert von 0.40 eV angenommen. Die Bilder 4.22 und 4.23 zeigen die sich aus der Simulation ergebenden Transferkennlinien, den Verlauf der Schwellspannung U_{th} in Abhängigkeit von der Gatelänge L_G und die Universalkurven für den kalibrierten p-MOSFET.

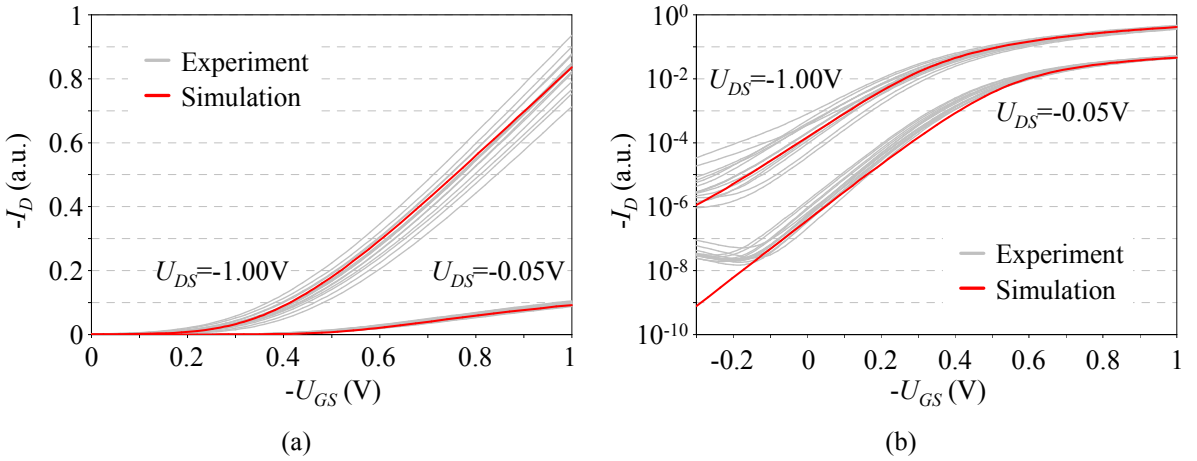


Bild 4.22: Transferkennlinien des p-MOSFETs in (a) linearer und (b) halblogarithmischer Darstellung

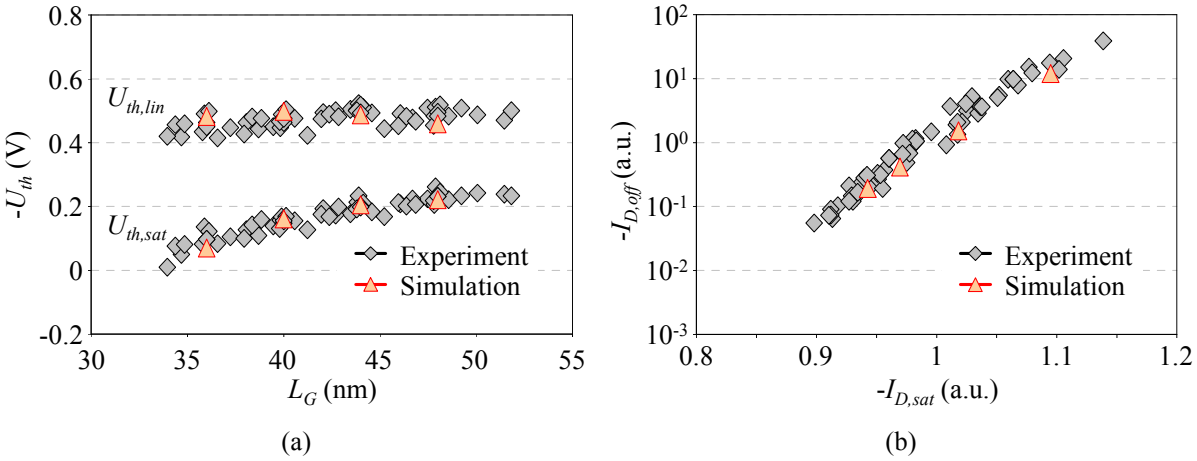


Bild 4.23: (a) Schwellspannung U_{th} in Abhängigkeit von der Gatelänge L_G und (b) Universalkurve des p-MOSFETs

c) p-MOSFET mit eingebetteten SiGe

Bild 4.24 zeigt die Universalkurve von p-MOSFETs mit und ohne SiGe. Wie man daran erkennen kann, wird der Sättigungs-Drainstrom $I_{D,sat}$ bei einem konstanten Sperrstrom $I_{D,off}$ von 100 nA/µm durch das SiGe um circa 20 % verbessert. Dieser deutlich erhöhte Drainstrom ist hauptsächlich auf die uniaxiale Druckverspannung im Kanal zurückzuführen, welche zu der in Abschnitt 4.3.3 beschriebenen Erhöhung der Löcherbeweglichkeit führt.

Neben der Ladungsträgerbeweglichkeit μ_p wird durch das in die SD-Gebiete eingebettete SiGe der Übergangswiderstand $R_{Silizid}$ zum Silizid positiv beeinflusst. Durch die kleinere Bandlücke E_g von SiGe im Vergleich zu Silizium verringert sich die Barrierenhöhe $\Phi_{B,p}$, was sich wiederum nach Gleichung (2.14) in einem kleineren Übergangswiderstand $R_{Silizid}$ niederschlägt und die Eigenschaften von p-MOSFETs weiter verbessert [83]. Ferner ist bekannt, dass es zur Bildung von GeB-Paaren kommt, die zwar elektrisch aktiv, aber nicht beweglich sind und dadurch den Schichtwiderstand aufgrund der höheren Löslichkeit von Bor in SiGe verringern sowie die Diffusion von Bor in SiGe hemmen [111]. Aufgrund dieser Ergebnisse wird SiGe im Gegensatz zu SiC, für dessen Anwendungen noch wichtige technische Probleme gelöst werden müssen, schon in der Massenproduktion verwendet.

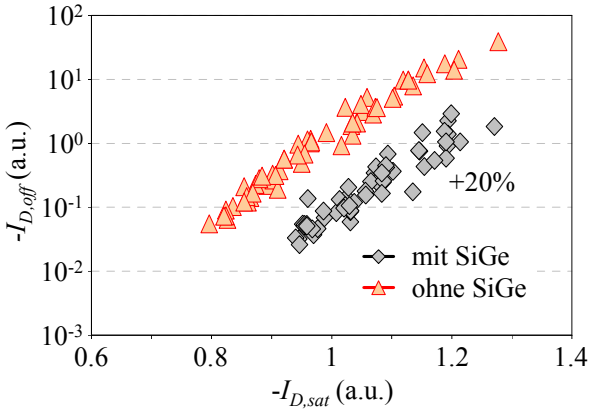


Bild 4.24: Gemessene Universalkurven für p-MOSFETs mit und ohne SiGe

Die einzige Besonderheit bei der Herstellung von SiGe ist, dass hier nach der Gatestrukturierung und vor der Formierung des Spacer0 das SiGe-Modul eingefügt wird (Bild 4.25).

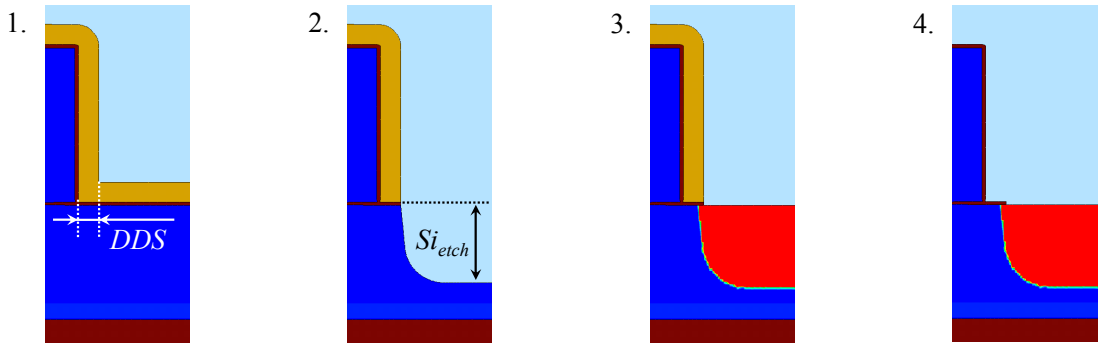


Bild 4.25: Schematische Darstellung der Prozessschritte bei der Herstellung des SiGe-Moduls im p-MOSFET-Prozess
1. Abscheidung eines Abstandshalters (Differential Disposable Spacer, *DDS*), der den Abstand des SiGe von der Gatekante definiert,
2. Ätzung von Mulden in die SD-Gebiete,
3. Epitaktisches Auffüllen der Mulden mit SiGe,
4. Entfernung des Abstandshalters.

Die für die Simulation des p-MOSFETs mit SiGe notwendigen Parameter des SiGe-Moduls wurden anhand von TEM-Aufnahmen ermittelt und sind in Tabelle 4.4 aufgelistet.

Tabelle 4.4 Geometrieparameter der simulierten Transistoren

Variable	Parameter	Wert
DDS	Abstand SiGe - Gatekante	15 nm
Si_{etch}	Tiefe der Mulden für das SiGe	50 nm
x_{Ge}	Germaniumgehalt im SiGe	20 %

Die verspannungsabhängige Veränderung der Ladungsträgerbeweglichkeit μ_p erfolgte mit dem in Abschnitt 4.3.3 vorgestellten Modell für Löcher. Zur Bestimmung des Übergangswiderstands $R_{Silizid}$ vom SiGe zum Silizid wurde für die bei der Silizierung entstehende NiSiGePt-Verbindung nach [112] eine Barrierenhöhe von $\Phi_{B,p} = 0.235$ eV verwendet. Die Bilder 4.26 und 4.27 zeigen die sich daraus ergebenden Transferkennlinien, den Verlauf der Schwellspannung U_{th} in Abhängigkeit von der Gatelänge L_G und die Universalkurven für den kalibrierten p-MOSFET mit SiGe.

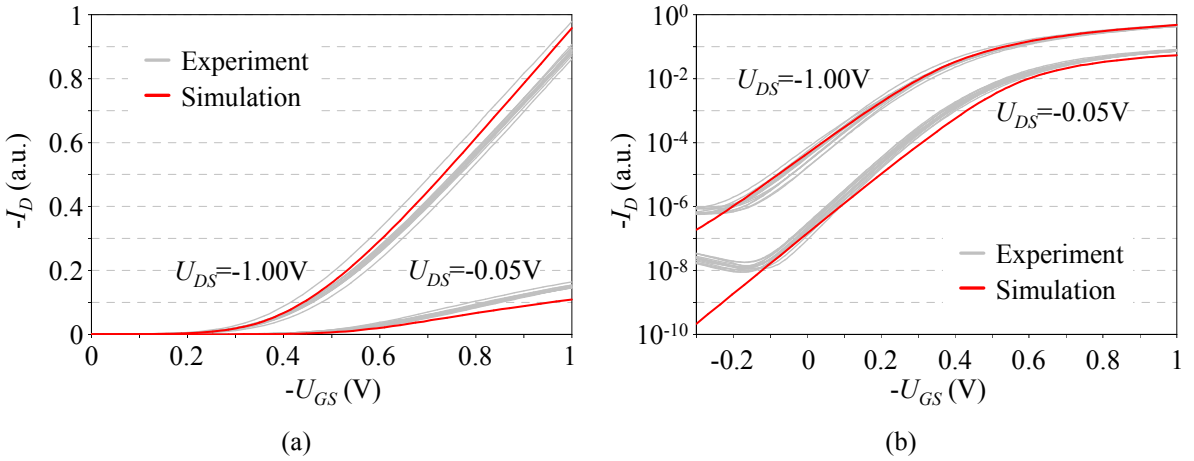


Bild 4.26: Transferkennlinien des p-MOSFETs mit SiGe in (a) linearer und (b) halblogarithmischer Darstellung

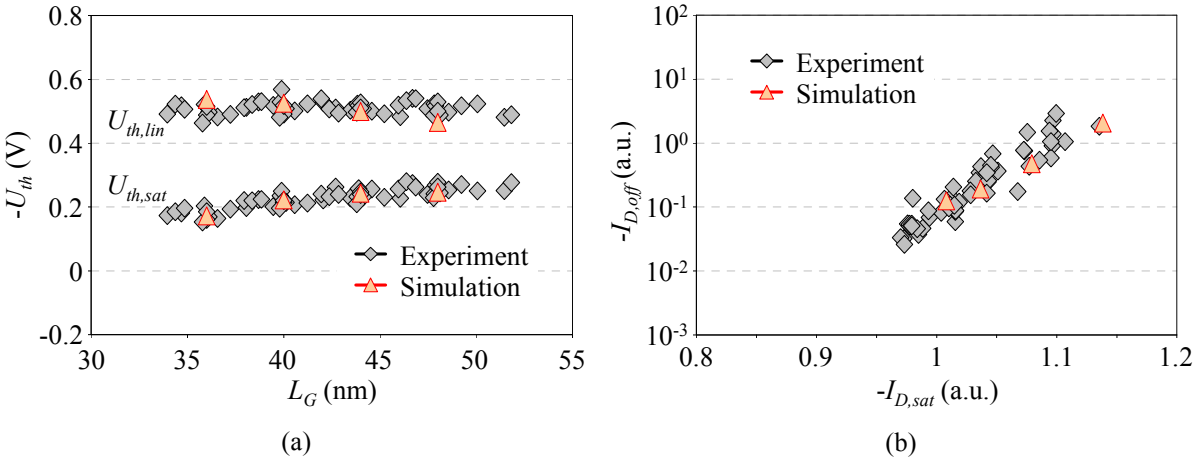


Bild 4.27: (a) Schwellspannung U_{th} in Abhängigkeit von der Gatelänge L_G und (b) Universalkurve des p-MOSFETs mit SiGe

4.5 Zusammenfassung

In diesem Kapitel erfolgte eine detaillierte Betrachtung der bei der Prozess- und Bauelementesimulation verwendeten Modelle. Das Hauptaugenmerk der Prozesssimulation lag dabei auf den Modellen der Implantation, der Ausheilung und mechanischer Verspannungen, da diese Prozessschritte zum Einen die Dotierungsprofile und zum Anderen die Leistungsfähigkeit der Transistoren wesentlich beeinflussen. Die Bestimmung des Implantationsprofils erfolgte mit Hilfe der Dual-PEARSON-Funktion, welche einen auf Monte-Carlo-Simulationen basierenden Parametersatz verwendet. Für die sehr komplexe Simulation der Diffusion und Aktivierung der Dotanden wurde das *Pair*-Diffusionsmodell verwendet, da es ein Optimum zwischen Genauigkeit und Rechenaufwand aufweist. Weiterhin wurde die in der Halbleiterindustrie gezielt eingesetzte Technik der mechanischen Verspannung von Silizium in die Prozesssimulation einbezogen. Der Schwerpunkt lag dabei auf der gezielten Erzeugung von Verspannungen basierend auf dem Einbetten anderer Halbleitermaterialen mit einer zum Silizium abweichenden Gitterkonstanten in die SD-Gebiete des Transistors, wodurch eine Verspannung bzw. Verformung der umliegenden Gebiete verursacht wird.

Für die Bauelementesimulation wurde die Modellierung des Ladungsträgertransports, der Ladungsträgerbeweglichkeit, von mechanischer Verspannungen und der Generations/Rekombinations-Prozesse genauer betrachtet, um den Einfluss eines veränderten Dotierungs- bzw. Verspannungsprofils auf die elektrischen Eigenschaften der Transistoren genauer untersuchen zu können. Die Beschreibung des Ladungsträgertransports erfolgte mit Hilfe des vereinfachten hydrodynamischen bzw. Energie-Balance-Modells, dass neben der POISSON-Gleichung und den Erhaltungsgleichungen für die Ladungsträgerdichten auch die Erhaltungsgleichungen für die Energie der Elektronen und Löcher mit in die Berechnungen einbezieht. Die Beachtung von Quantisierungseffekten an der Si/SiO_2-Grenzfläche unter dem Gate erfolgte unter Verwendung des Quanten-Energie-Balance-Modells. Die Ladungsträgerbeweglichkeit wurde unter Berücksichtigung von thermischen Gitterschwingungen sowie der Streuung an geladenen und ungeladenen Störstellen im Kristall infolge der Dotierung des Halbleiters modelliert. Weiterhin wurde die Veränderung der Ladungsträgerbeweglichkeit an der Si/SiO_2-Grenzfläche durch vertikale elektrische Felder und durch hohe elektrische Felder in Stromflussrichtung durch entsprechende Modelle beachtet. Außerdem wurden bei der Bauelementesimulation die beim MOSFET entscheidenden Generations/Rekombinations-Prozesse und die Ladungsträgerbeweglichkeit infolge einer mechanischen Verspannung in die Berechnungen einbezogen.

Tabelle 4.5 gibt einen Überblick über die verwendeten Modelle der Prozess- und Bauelementesimulation.

Tabelle 4.5 In dieser Arbeit verwendete Modelle für die Prozess- und Bauelementesimulation

Simulation		n-MOSFET	p-MOSFET ohne und mit SiGe
Prozess-simulation	Implantation	- Dual-PEARSON-Funktion unter Verwendung von einem auf Monte-Carlo-Simulationen basierendem Parametersatz	
	Diffusion	- *Pair*-Diffusionsmodell	
Bauelemente-simulation	Transportmodell	- Energie-Balance-Modell - Quanten-Energie-Balance-Modell	
	Ladungsträger-beweglichkeit	- Abhängigkeit von der Temperatur und Dotierung (nach [94]) - Abhängigkeit vom vertikalen elektrischen Feld (nach [95]) - Abhängigkeit vom lateralen elektrischen Feld (nach [96])	
	Generation-Rekombination	- SRH-Rekombination (nach [97]) - AUGER-Rekombination (nach [98]) - Stossionisation (nach [100]) - Band-zu-Band-Tunneln (nach [101])	
	Spannungs-abhängige Ladungsträger-beweglichkeit	- Deformationspotenzial (nach [107]) - Elektronenbeweglichkeit (nach [108])	- Deformationspotenzial (nach [107]) - Löcherbeweglichkeit (nach [109])

Mit Hilfe der Modelle der Prozess- und Bauelementesimulation und deren Parameter erfolgte eine Anpassung der Ergebnisse der Simulation an experimentelle Daten. Zuerst wurde eine Kalibrierung der Modelle der Prozesssimulation anhand von SIMS-Messungen, Schichtwiderstandsmessungen sowie TEM-Aufnahmen der zu kalibrierenden Transistorstruktur durchgeführt. Die daraus entstandene Struktur des Transistors war dann zusammen mit dem Dotierungsprofil sowie der Verspannungs- und Deformationsmatrix Ausgangspunkt für die Bauelementesimulation. Die Kalibrierung der Modellparameter für die Bauelementesimulation wurde anschließend für den unverspannten n- und p-MOSFET sowie den mit eingebetteten SiGe verspannten p-MOSFET anhand der Transferkennlinie, der Universalkurve sowie dem Schwellspannungsverlauf über der Gatelänge durchgeführt.

5 Kombination von Kurzzeitausheilung und konventioneller RTA-Ausheilung

In diesem Kapitel werden die experimentellen Ergebnisse vorgestellt, bei denen die Kurzzeitausheilung als zusätzlicher Ausheilschritt in Kombination mit RTA eingeführt wurden. Diese Experimente sollen sowohl die Kompatibilität zum herkömmlichen Ausheilprozess überprüfen als auch den Einfluss der Kurzzeitausheilung auf verschiedene Transistorparameter untersuchen.

5.1 Vorbetrachtungen

Ein kritisches Problem bei der Kurzzeitausheilung ist die verwendete Blitzlampen- bzw. Laser-Leistung. Obwohl die Schmelztemperatur von Silizium je nach Dotierung bei maximal 1412 °C liegt, ist ein Ausheilschritt mit einer Ausheiltemperatur $T_{Ausheil}$ um 1400 °C nicht möglich. Grund dafür ist die Verwendung von anderen Materialien mit geringeren Schmelzpunkten. So kommt Silizium in den unterschiedlichsten Formen zur Anwendung (kristallin, polykristallin, amorph). Die maximale Ausheiltemperatur wird deshalb von dem Material mit dem geringsten Schmelzpunkt bestimmt.

Bild 5.1a zeigt das Verhalten des Schichtwiderstands R_{SDE} der SDE-Gebiete des n-MOSFETs bei einer Variation der Ausheiltemperatur $T_{Ausheil}$ der Kurzzeitausheilung (MSA). Trotz der stetigen Verringerung des Schichtwiderstands R_{SDE} der SDE-Gebiete mit höheren Temperaturen steigt ab einer bestimmten Ausheiltemperatur $T_{Ausheil}$ der Gateleckstrom I_G über mehrere Größenordnungen sehr stark an (Bild 5.1b). Dieses Verhalten deutet auf eine Zerstörung des Gateoxids aufgrund einer zu hohen Ausheiltemperatur hin. Untersuchungen der entsprechenden Transistorstrukturen mit Hilfe der TEM ergaben ein Aufschmelzen des Polysiliziumgates. Anhand von Experimenten wurde schließlich eine optimale Ausheiltemperatur $T_{Ausheil}$ von 1300 °C festgestellt, womit alle weiteren Experimente durchgeführt wurden. Die Ausheilzeit $t_{Ausheil}$ betrug in allen Experimenten 1 ms. Aufgrund der Ergebnisse von Abschnitt 3.4 wird im Weiteren nicht zwischen Blitzlampen- und Laser-Ausheilung unterschieden.

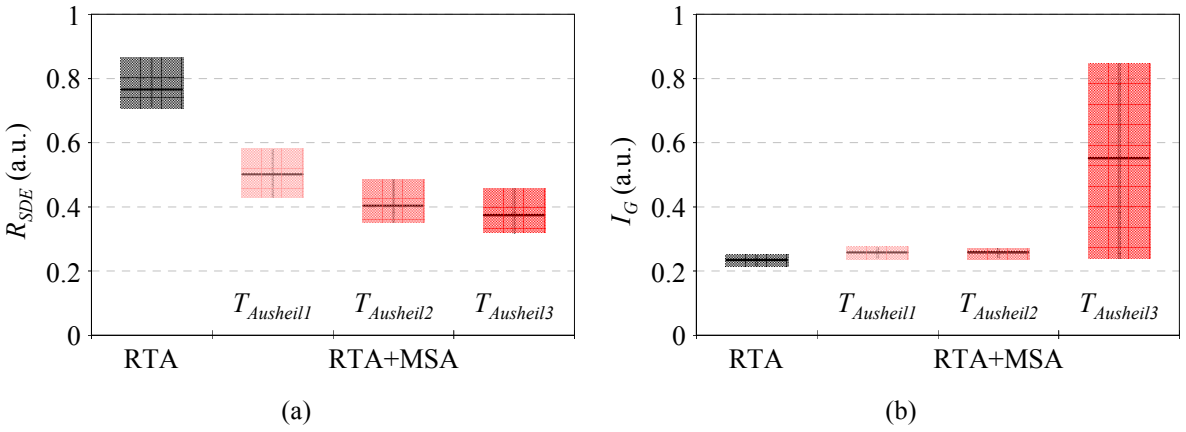


Bild 5.1: (a) Schichtwiderstand R_{SDE} der SDE-Gebiete und (b) Gateleckstrom I_G für n-MOSFETs bei variierter Ausheiltemperatur $T_{Ausheil}$ der MSA ($T_{Ausheil1} < T_{Ausheil2} < T_{Ausheil3}$)

5.2 Einfluss der Kurzzeitausheilung auf elektrische Transistorparameter

5.2.1 Experimentelle Ergebnisse

In allen folgenden Experimenten wurde die Kurzzeitausheilung als zusätzlicher Ausheilschritt direkt im Anschluss an die Spike-RTA durchgeführt wurde. Die Experimente erfolgten auf strukturierten (100)-orientierten Wafern. Als Vergleich dient ein Transistor mit gleicher Prozessierung, der jedoch nur einer RTA, aber keiner Kurzzeitausheilung unterzogen wurde (Bild 5.2).

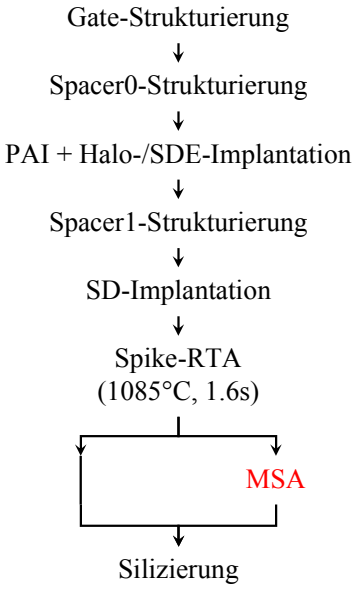


Bild 5.2: Schematische Darstellung des Prozessablaufs der untersuchten MOSFETs mit und ohne Kurzzeitausheilung (MSA)

Bild 5.3 zeigt die Kapazitäts-Spannungs- (CV-) Kennlinien mit und ohne zusätzlicher Kurzzeitausheilung. Durch die Kurzzeitausheilung erhöht sich die Gatekapazität im Inversionsfall $C_{G,inv}$, wodurch sich die elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$ sowohl bei n- als auch bei p-MOSFETs entsprechend Gleichung (2.1) um 0.9 Å verringert (Bild 5.4a).

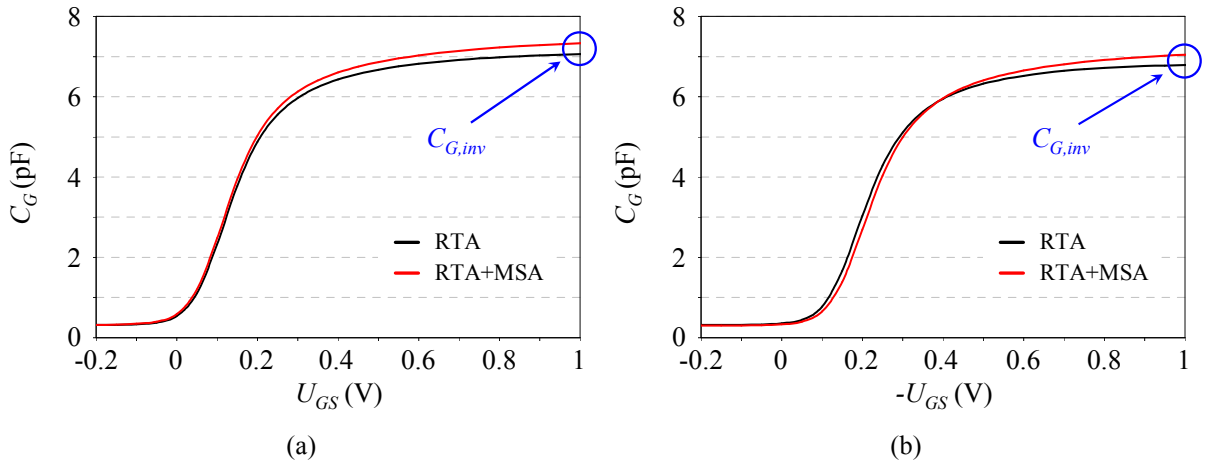


Bild 5.3: Einfluss der Kurzzeitausheilung (MSA) auf die CV-Kennlinie für (a) n- und (b) p-MOSFETs

Ursache für dieses Verhalten ist die Verringerung der Verarmungszone im Polysiliziumgate aufgrund der höheren Aktivierung der Dotanden im Polysilizium. Als Nachweis dafür dient der Widerstand R_{Poly} einer Polysiliziumstruktur (Bild 5.4b). Durch die zusätzliche Kurzzeitausheilung verringert sich der Widerstand R_{Poly} des Polysiliziums bei n- und bei p-MOSFETs um 30 % bzw. um 10 %.

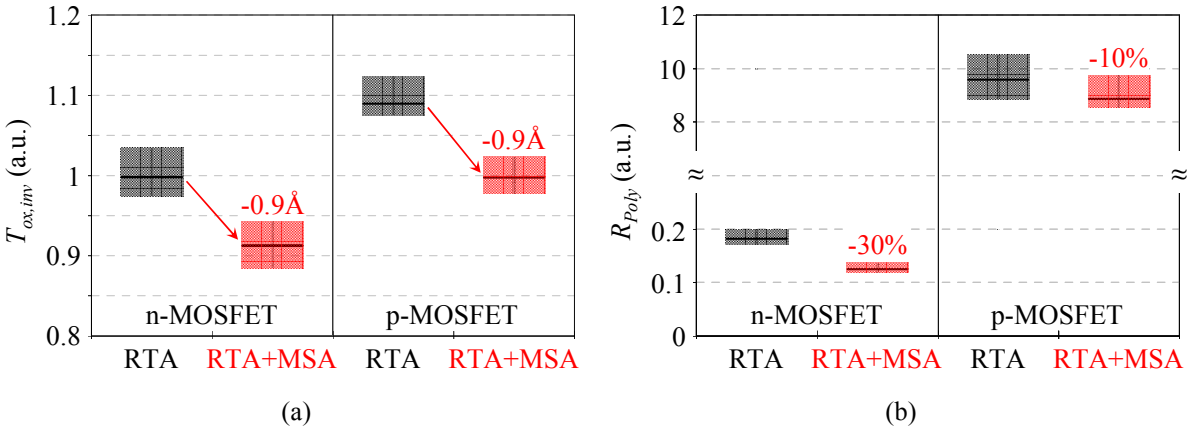


Bild 5.4: Einfluss der Kurzzeitausheilung (MSA) auf (a) die elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$ und (b) den Widerstand R_{Poly} einer Polysiliziumstruktur für n- und p-MOSFETs

Um die elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$ um 0.9 Å zu verringern, würde ein äquivalenter Ausheilschritt mit der RTA entweder eine höhere Ausheiltemperatur $T_{Ausheil}$ oder eine längere Ausheilzeit $t_{Ausheil}$ erfordern, was wiederum eine zu starke Diffusion der implantierten Dotanden der SD- und SDE-Gebiete zur Folge hätte. Die Verringerung der elektrischen Gateoxiddicke im Inversionsfall $T_{ox,inv}$ um 0.9 Å verursacht eine 30 %ige Erhöhung des Gateleckstroms I_G, was sich durch ein höheres elektrisches Feld erklären lässt. Demgegenüber resultiert eine 2 Å kleinere physikalische Gateoxiddicke T_{ox} aufgrund der höheren Tunnelströme in einem 5.5-mal höheren Gateleckstrom I_G (Bild 5.5).

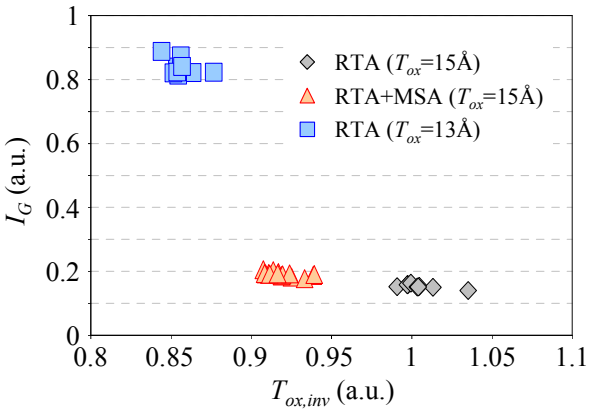


Bild 5.5: Einfluss der Kurzzeitausheilung (MSA) auf den Gateleckstrom I_G in Abhängigkeit von der elektrischen Gateoxiddicke im Inversionsfall $T_{ox,inv}$ für unterschiedliche physikalische Gateoxiddicken T_{ox} für n-MOSFETs

Bild 5.6 zeigt den Einfluss der Kurzzeitausheilung auf die Sättigungsschwellspannung $U_{th,sat}$ in Abhängigkeit von der Gatelänge L_G für n- und p-MOSFETs. Dabei lassen sich zwei Fakten feststellen. Erstens ist aufgrund der kleineren elektrischen Gateoxiddicke im Inversionsfall $T_{ox,inv}$ ein verbessertes Kurzkanalverhalten festzustellen. Dadurch kann die physikalische Gatelänge L_G verkleinert werden, was wiederum eine kleinere Gatekapazität C_G zur Folge hat und damit die parasitäre Kapazität des Transistors verringert.

Zweitens ist eine gegenläufige Verhaltensweise beider Transistortypen auf den zusätzlichen Ausheilschritt zu erkennen. Bei n-MOSFETs sinkt die Sättigungsschwellspannung $U_{th,sat}$ unter den verwendeten Messbedingungen um etwa 13 mV, wohingegen man bei p-MOSFETs betragsmäßig ein Ansteigen der Sättigungsschwellspannung $U_{th,sat}$ um fast 25 mV feststellen kann. Die Ursache für dieses Verhalten wird im Abschnitt 5.2.2 näher beschrieben.

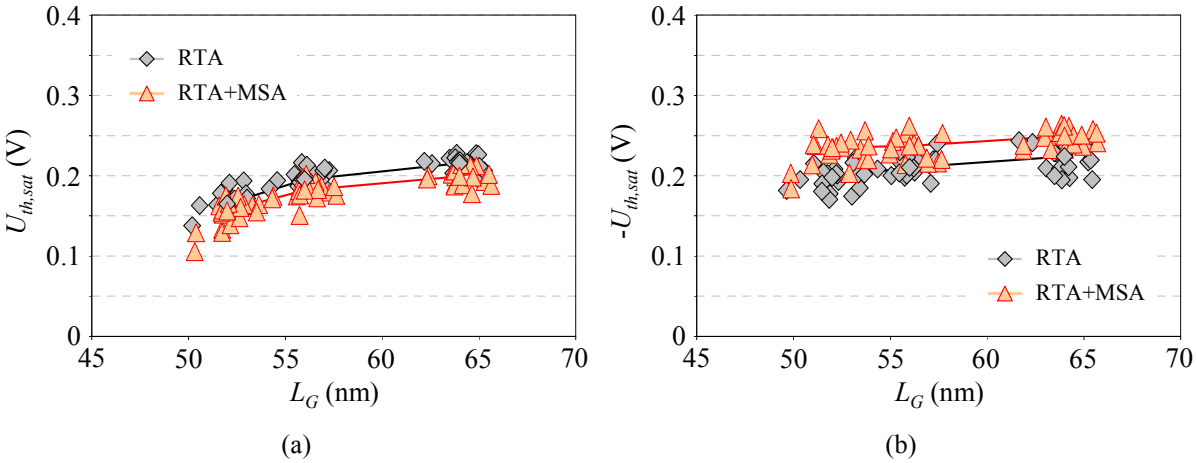


Bild 5.6: Einfluss der Kurzzeitausheilung (MSA) auf die Sättigungsschwellspannung $U_{th,sat}$ in Abhängigkeit von der Gatelänge L_G für (a) n- und (b) p-MOSFETs

Bild 5.7 zeigt den Einfluss der kombinierten Ausheilung auf den Schichtwiderstand R_{SDE} der SDE-Gebiete. Bei n-MOSFETs verringert sich der Schichtwiderstand R_{SDE} der SDE-Gebiete um circa 18 %, was vor allem auf die höhere Aktivierung der Dotanden der SDE-Gebiete zurückzuführen ist. Bei p-MOSFETs hingegen ist durch die zusätzliche Kurzzeitausheilung nur ein geringfügig kleinerer Schichtwiderstand R_{SDE} der SDE-Gebiete von (1…2) % zu erkennen, was auf eine vollständige Aktivierung der Dotanden bereits nach der RTA schließen lässt.

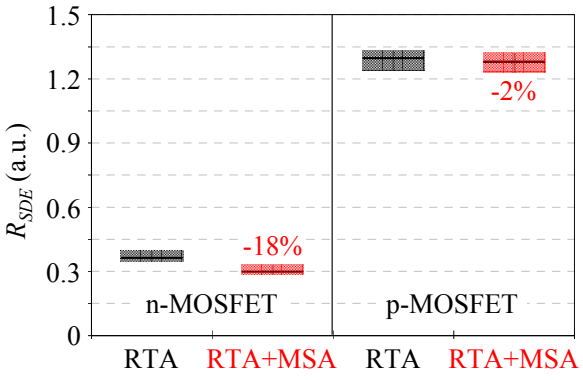


Bild 5.7: Einfluss der Kurzzeitausheilung (MSA) auf den Schichtwiderstand R_{SDE} der SDE-Gebiete für n- und p-MOSFETs

Durch die zusätzliche Kurzzeitausheilung konnte der Drainstrom $I_{D,lin}$ im linearen Arbeitsbereich bei konstantem Sperrstrom $I_{D,off}$ der n- und p-MOSFETs ohne gravierende Änderungen im Prozessablauf um 7 % bzw. 3 % erhöht werden. Die Verbesserung lässt sich auf die kleinere elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$ und, im Fall der n-MOSFETs, auf den verringerten Schichtwiderstand R_{SDE} der SDE-Gebiete zurückführen. Weiterhin zeigten Simulationen, dass sich aufgrund der höheren Aktivierung der SDE- und SD-Gebiete der Übergangswiderstand $R_{Silizid}$ vom Silizium zum Silizid verringert.

Der Gewinn in der Leistungsfähigkeit geht für n-MOSFETs im Sättigungsbereich auf 4 % zurück, da in diesem Arbeitsbereich der Einfluss des Schichtwiderstands R_{SDE} der SDE-Gebiete kleiner ist, wohingegen der Gewinn bei p-MOSFETs im linearen und im Sättigungsbereich unverändert 3 % beträgt (Bild 5.8).

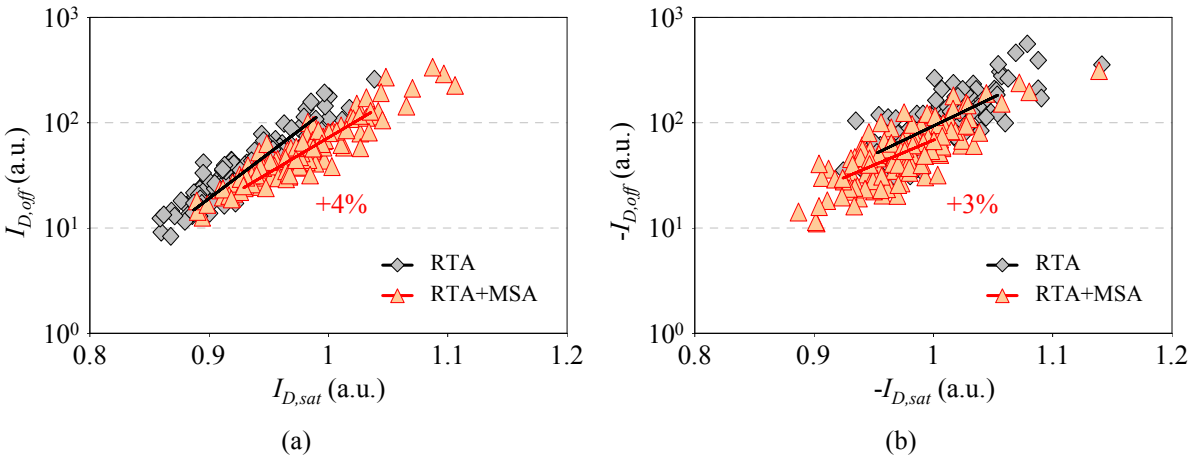


Bild 5.8: Einfluss der Kurzzeitausheilung (MSA) auf die Universalkurve für (a) n- und (b) p-MOSFETs

5.2.2 Untersuchung des Gatedielektrikums

Wie im vorherigen Abschnitt 5.2.1 beschrieben, zeigt der zusätzlichen Ausheilschritt einen gegenläufigen Einfluss auf die Schwellspannung U_{th} von n- und p-MOSFETs. Die Ursache dafür soll an einem Langkanaltransistor aufgeklärt werden, um den Einfluss der SDE-Gebiete und Halo-Implantationen auf die Schwellspannung U_{th} auszuschließen. Bild 5.9 zeigt die Sättigungsschwellspannung $U_{th,sat}$ für Langkanal-n- und p-MOSFETs mit und ohne Kurzzeitausheilung. Die Schwellspannungsänderung bei den Langkanaltransistoren ist zwar nicht so stark ausgeprägt wie bei den Kurzkanaltransistoren, zeigt aber den gleichen Trend. Bei n-MOSFETs sinkt die Schwellspannung durch den zusätzlichen Ausheilschritt um ca. 7 mV, wohingegen sie bei p-MOSFETs um etwa 12 mV ansteigt.

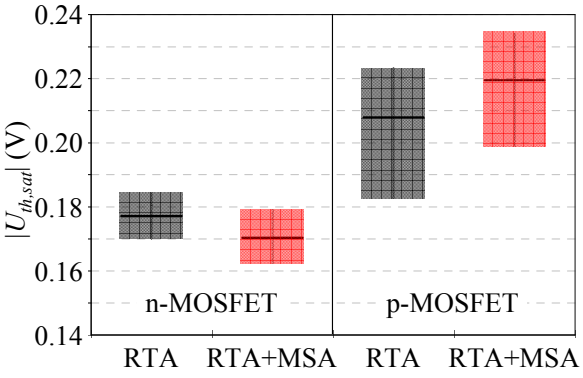


Bild 5.9: Einfluss der Kurzzeitausheilung (MSA) auf die Sättigungsschwellspannung $U_{th,sat}$ für Langkanal-n- und p-MOSFETs ($W_G = L_G = 3$ µm)

Für die Schwellspannung U_{th} von Langkanaltransistoren gilt näherungsweise nach [27] für n-MOSFETs

$$U_{th} = U_{FB} + 2 \cdot \varphi_F + \frac{\sqrt{2 \cdot q \cdot \varepsilon_{Si} \cdot N_{A,sub}^{-} \cdot 2 \cdot \varphi_F}}{C_G''} \tag{5.1}$$

bzw. für p-MOSFETs

$$U_{th} = U_{FB} - 2 \cdot \varphi_F - \frac{\sqrt{2 \cdot q \cdot \varepsilon_{Si} \cdot N^{+}_{D,sub} \cdot 2 \cdot \varphi_F}}{C''_G} \tag{5.2}$$

mit dem Dotierungspotenzial für ein p-dotiertes Substrat

$$\varphi_F = \frac{k_B \cdot T}{q} \cdot \ln\left(\frac{N^{-}_{A,sub}}{n_i}\right) \tag{5.3}$$

und für ein n-dotiertes Substrat

$$\varphi_F = \frac{k_B \cdot T}{q} \cdot \ln\left(\frac{N^{+}_{D,sub}}{n_i}\right). \tag{5.4}$$

Geht man von der Annahme aus, dass die Wannendotierung aufgrund der sehr geringen Konzentration C und die flächenbezogene Gatekapazität C''_G beim Übergang vom Zustand der Verarmung zum Zustand der Inversion unabhängig von der verwendeten Ausheilsequenz sind, kann nur eine veränderte Flachbandspannung U_{FB} die Ursache für das beobachtete Schwellspannungsverhalten sein. Die in Bild 5.10 dargestellten CV-Kennlinien bestätigen diese Theorie. Durch den zusätzlichen Ausheilschritt kommt es zu der für eine veränderte Flachbandspannung U_{FB} charakteristischen Parallelverschiebung der Kurven um die entsprechende Schwellspannungsänderung.

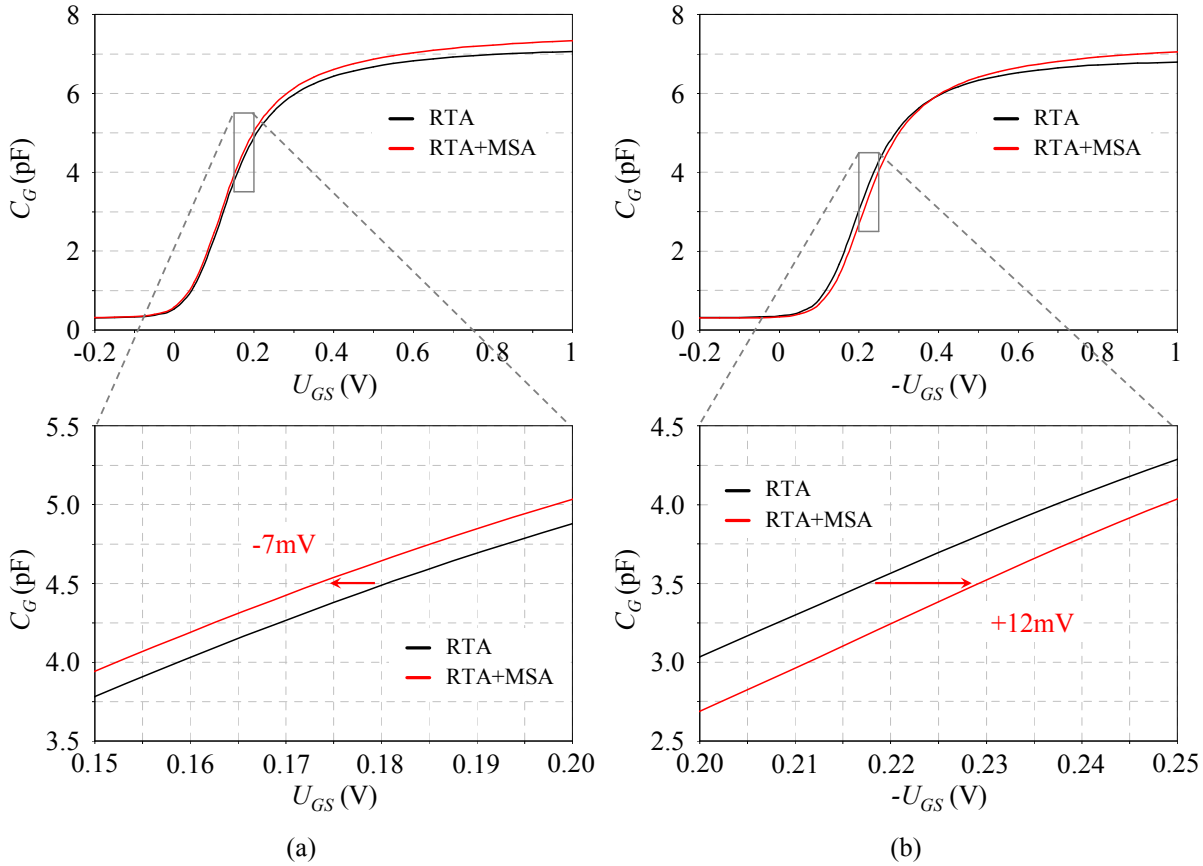


Bild 5.10: Einfluss der Kurzzeitausheilung (MSA) auf die Flachbandspannung U_{FB} für (a) n- und (b) p-MOSFETs

Für reale MOS-Kapazitäten ist die Flachbandspannung U_{FB} definiert nach [27] als

$$U_{FB} = \varphi_{ms} - U_{ox} \,. \tag{5.5}$$

Dabei ist φ_{ms} die Austrittsarbeitsdifferenz zwischen der von der Konzentration C der Dotierstoffe abhängigen Austrittsarbeit φ_m im Polysilizium und φ_s im Halbleiter ($\varphi_{ms} = \varphi_m - \varphi_s$). Der zweite Term U_{ox} beschreibt die aufgrund von Oxidladungen abfallende Spannung über dem Gateoxid. Unter der Bedingung gleicher Elektronenaffinitäten E_X und Bandlücken E_G im Polysilizium und im Substrat ergibt sich für ein n-dotiertes Polysiliziumgate auf einem p-dotierten Substrat die Austrittsarbeitsdifferenz

$$\varphi_{ms} = -\frac{k_B \cdot T}{q} \cdot \ln\left(\frac{N^+_{D,poly} \cdot N^-_{A,sub}}{(n_i)^2}\right) \tag{5.6}$$

und dementsprechend für ein p-dotiertes Polysiliziumgate auf einem n-dotierten Substrat

$$\varphi_{ms} = \frac{k_B \cdot T}{q} \cdot \ln\left(\frac{N^-_{A,poly} \cdot N^+_{D,sub}}{(n_i)^2}\right). \tag{5.7}$$

Geht man von einer höheren Aktivierung der Dotanden $N^-_{A,poly}$ und $N^+_{D,poly}$ im Polysilizium durch den zusätzlichen Ausheilschritt aus, ergibt sich für beide Transistortypen betragsmäßig eine höhere Austrittsarbeitsdifferenz φ_{ms}. Das wiederum hätte nach Gleichung (5.5) eine betragsmäßig größere Flachbandspannung U_{FB} und daher nach den Gleichungen (5.1) und (5.2) für beide Transistortypen eine kleinere Schwellspannung U_{th} zur Folge.

Bild 5.11 zeigt dazu simulierte CV-Kennlinien in Abhängigkeit von der Polysiliziumdotierung. Die Dotierung des Polysiliziumgates wurde dabei so gewählt, dass sich entsprechend den experimentellen Ergebnissen eine Veränderung der elektrischen Gateoxiddicke im Inversionsfall $T_{ox,inv}$ von 0.9 Å einstellt. Im Gateoxid wurden keine weiteren Oxidladungen beachtet. Durch die höhere Dotierung des Polysiliziumgates verschiebt sich die CV-Kennlinie des n- und p-MOSFETs in Richtung einer kleineren Gate-Source- und damit Schwellspannung, was mit den zuvor gemachten Aussagen übereinstimmt.

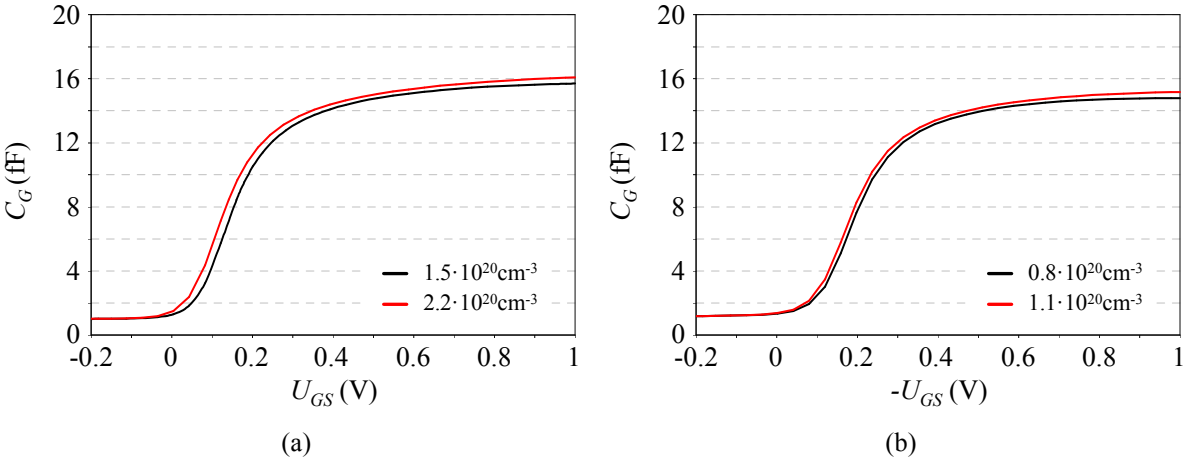


Bild 5.11: Berechnete CV-Kennlinien in Abhängigkeit von der Polysiliziumdotierung für (a) n- und (b) p-MOSFETs

Die höhere Polysiliziumdotierung hat für n- und p-MOSFETs eine betragsmäßig um etwa 20 mV bzw. 11 mV größere Austrittsarbeitsdifferenz φ_{ms} und somit Flachbandspannung U_{FB} zur Folge. Beide Transistortypen zeigen aber eine unterschiedliche Reaktion auf die Kurzzeitausheilung hinsichtlich der Schwellspannungsverschiebung, was nach Gleichung (5.5) nur durch zusätzlich Ladungen im Gateoxid oder an der Grenzfläche verursacht werden kann. Aufgrund von Verunreinigungen im Gateoxid und der imperfekten Grenzfläche zwischen Silizium und Siliziumoxid entstehen verschiedenartige Defekte im Oxid beziehungsweise an der Grenzfläche. Diese Defekte sind grundsätzlich in allen Oxiden vorhanden, die Konzentration der verschiedenen Defekte ist jedoch stark vom Siliziumausgangsmaterial, der Herstellung des Gateoxids und der weiteren Prozessierung abhängig.

Bild 5.12 gibt eine Übersicht auf die unterschiedlichen Defekte im Gateoxid. Bewegliche Ionenladungen Q_m bestehen in der Regel aus ionisierten Metallatomen wie Natrium oder Kalium. Das Vorkommen ist stark vom Herstellungsprozess des Gateoxids abhängig. Bei der verwendeten Technologie kann man aber davon ausgehen, dass die Anzahl dieser Fremdatome sehr gering ist [114]. Geladene Haftzentren im Oxid Q_{ot} sind im Grundzustand neutral. Sie können jedoch durch Injektion heißer Elektronen und Löcher, ionisierende Strahlung oder Ionenimplantation positiv bzw. negativ geladen werden. Feste Oxidladungen Q_f sind ortsfeste und nicht umladbare Ladungen, welche hauptsächlich in der Nähe zur Si/SiO_2-Grenzfläche lokalisiert sind. Sie resultieren aus einem Überangebot von Silizium während der Oxidation. Grenzflächenhaftzentren Q_{it} befinden sich unmittelbar an der Si/SiO_2-Grenzfläche. Ursache ist die Unterbrechung der Gitterstruktur an der Oberfläche des Siliziumkristalls, wodurch sich aufgrund von nicht abgesättigten Siliziumvalenzen aktive Defekte direkt an der Grenzfläche bilden. Während der Kurzzeitausheilung unterliegt vor allem die Grenzfläche aufgrund der dabei vorherrschenden hohen Prozesstemperatur gegenüber der konventionellen Ausheilung einer großen mechanischen Beanspruchung, wodurch es zur Bildung von zusätzlichen Grenzflächenhaftzentren kommen kann.

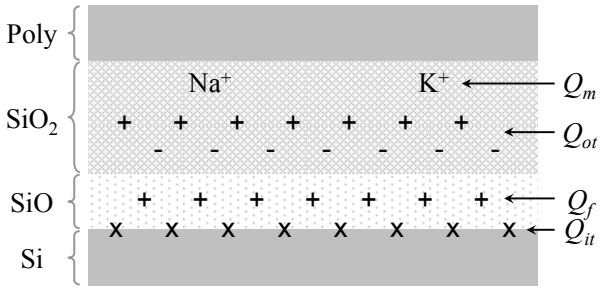


Bild 5.12: Defekte im Gateoxid [113]

Die drei Ladungsarten Q_m, Q_{ot} und Q_f werden in der Regel unter Oxidladungen Q_{ox} zusammengefasst und verursachen in Verbindung mit den Grenzflächenladungen Q_{it}, die elektrisch negativ als auch positiv geladen sein können, einen Spannungsabfall U_{ox} über dem Gateoxid, was nach Gleichung (5.5) eine Verschiebung der Flachbandspannung U_{FB} zur Folge hat. Für n-MOSFETs beträgt der Spannungsabfall U_{ox} über dem Gateoxid aufgrund der unterschiedlichen Defekte im Gateoxid

$$U_{ox} = \frac{Q_{ox} - Q_{it}}{C''_G}, \tag{5.8}$$

für p-MOSFETs

$$U_{ox} = \frac{Q_{ox} + Q_{it}}{C''_G}. \tag{5.9}$$

Während aus einem Vergleich der beiden gemessenen CV-Kennlinien (Bild 5.10) nur eine Aussage über die Summe aus Austrittsarbeitsdifferenz sowie dem Beitrag aus Oxidladungen und Grenzflächenzuständen gezogen werden kann, ist die Extraktion der Grenzflächenhaftzentrendichte (Interface trap density, D_{it}) mittels einer zusätzlichen Messung möglich. Eine sehr verbreitete Möglichkeit dafür ist die Charge Pumping- (CP-) Methode [115]. Dabei werden nur die Grenzflächenzustände mit Ladungsträgern gefüllt und wieder entleert, was zu einem gerichteten Ladungstransport von den Source- und Drain-Gebieten zum Substratkontakt führt. Dazu wird an das Gate ein Spannungsimpuls angelegt, der den Transistor von Akkumulation in Inversion und wieder zurückschaltet (Bild 5.13). In starker Inversion werden die Grenzflächenhaftzentren mit Minoritätsladungsträgern des Inversionskanals gefüllt. Beim Übergang von Inversion zu Akkumulation wird der Inversionskanal abgebaut. Aufgrund der sehr kurzen Anstiegs- bzw. Abfallzeit des Spannungsimpulses können jedoch die in den Grenzflächenhaftzentren „gefangenen“ Minoritätsladungsträger diesem Abbau nicht folgen. Diese gefangenen Ladungsträger rekombinieren im Akkumulationsfall mit den jetzt vorherrschenden Majoritätsladungsträgern, was einen messbaren Rekombinationsstrom, den sogenannten CP-Strom I_{CP} zur Folge hat, der proportional zur Dichte der Grenzflächenladungen ist.

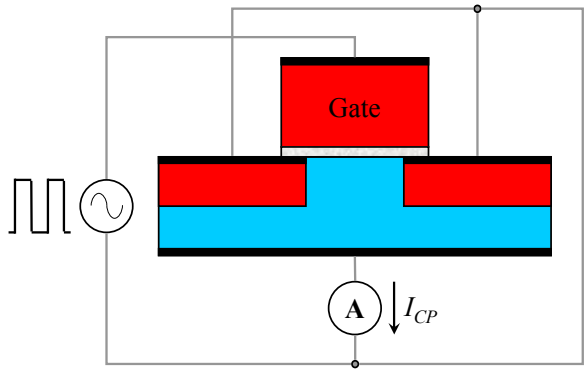


Bild 5.13: Schematische Darstellung der Messanordnung zur Bestimmung des Charge Pumping-Stroms I_{CP}

Die Bestimmung der Dichte der Grenzflächenhaftzentren erfolgt durch

$$D_{it} = \frac{I_{CP}}{q \cdot A_G \cdot f_P \cdot \Delta E}. \tag{5.10}$$

Dabei ist f_P die Pulsfrequenz und ΔE die Differenz des Oberflächenpotenzials zwischen Inversions- und Akkumulationsfall. Bild 5.14 zeigt den Einfluss der Kurzzeitausheilung auf die Grenzflächenhaftzentrendichte D_{it}. Durch den zusätzlichen Ausheilschritt erhöht sich die Dichte der Haftzentren bei n- und p-MOSFETs um etwa 43 %, was bei n-MOSFETs $\Delta D_{it} \approx 1.2 \cdot 10^{10}$ cm^{-2}eV^{-1} und bei p-MOSFETs $\Delta D_{it} \approx 5.1 \cdot 10^{10}$ cm^{-2}eV^{-1} entspricht.

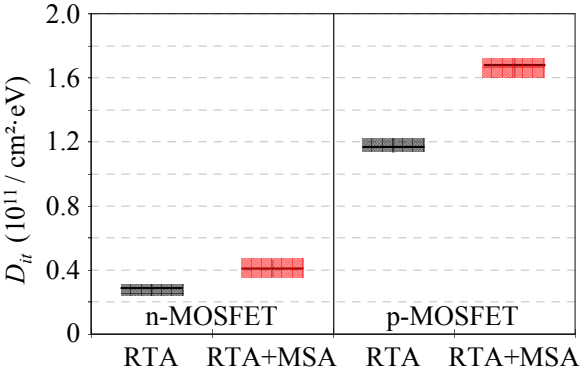


Bild 5.14: Einfluss der Kurzzeitausheilung (MSA) auf die Grenzflächenhaftzentrendichte D_{it} für n- und p-MOSFETs

Die allgemein höhere Haftzentrendichte der p-MOSFETs lässt sich durch die Bor-Penetration erklären [116]. Der sich in Abhängigkeit von der höheren Haftzentrendichte ergebende Spannungsabfall über dem Gateoxid

$$\Delta U_{ox}(Q_{it}) = \frac{\Delta Q_{it}}{C_G''} \tag{5.11}$$

beträgt etwa 4 mV für n-MOSFETs und circa 11 mV für p-MOSFETs. Nach Gleichung (5.5) ergeben diese Werte mit den aus Bild 5.11 bestimmten Spannungen eine Verschiebung der Flachbandspannung U_{FB} um -16 mV für n-MOSFETs bzw. 0 mV für p-MOSFETs.

Im Rahmen der Messungenauigkeiten und den vereinfachten Annahmen in der Simulation stimmen diese Werte mit den gemessenen Werten der beobachteten Schwellspannungsverschiebung von -7 mV für n-MOSFETs und +12 mV für p-MOSFETs überein. Neben der höheren Grenzflächenhaftzentrendichte können ebenfalls zusätzlich erzeugte Oxidladungen eine Ursache für die Unterschiede zwischen berechneten und gemessenen Werten sein. Aufgrund von Wafer-zu-Wafer-Variationen, Abweichungen der Prozessierung innerhalb eines Wafers, der vorhandenen Messtechnik, Messungenauigkeiten sowie von Messfehlern ist jedoch eine genaue Bestimmung jedes einzelnen Anteils ($\Delta\varphi_{ms}$, $\Delta U_{ox}(Q_{ox})$, $\Delta U_{ox}(Q_{it})$) an der Flachbandspannungsverschiebung nicht möglich. Man kann aber aus den Ergebnissen ableiten, dass durch die Kurzzeitausheilung zusätzliche Oxid- und Grenzflächenladungen generiert werden, die in Kombination mit der veränderten Austrittsarbeitsdifferenz φ_{ms} zu der beobachteten Verschiebung der Schwellspannung U_{th} führen.

5.2.3 Ladungsträgerbeweglichkeit

Oxidladungen und Grenzflächenhaftzentren beeinflussen die Ladungsträgerbeweglichkeit und demzufolge die Leistungsfähigkeit der Transistoren. Aus diesem Grund wurde der Einfluss der Kurzzeitausheilung auf die Ladungsträgerbeweglichkeit untersucht. Bei deren Bestimmung geht man von der Kennliniengleichung des Drainstroms $I_{D,lin}$ im linearen Arbeitsbereich aus:

$$I_{D,lin} = \mu \cdot \frac{W_G}{L_{met}} \cdot \frac{Q_{Kanal}(U_{GS})}{A_G} \cdot U_{DS} \,. \tag{5.12}$$

Damit ergibt sich für die Beweglichkeit μ:

$$\mu = \frac{L_{met}}{W_G} \cdot \frac{A_G}{Q_{Kanal}(U_{GS})} \cdot \frac{I_{D,lin}}{U_{DS}} \,. \tag{5.13}$$

Die im Kanal angereicherte Ladung Q_{Kanal} wird aus der messtechnisch ermittelten Gatekapazität C_G berechnet:

$$Q_{Kanal} = \int_{-\infty}^{U_{GS}} C_G \, dU_{GS} \,. \tag{5.14}$$

Die dazu notwendigen Messungen der Gatekapazität C_G und des Drainstroms $I_{D,lin}$ erfolgten an Langkanaltransistoren, um eine messbare Gatekapazität zu erhalten und das Problem der Bestimmung der metallurgischen Gatelänge L_{met} zu umgehen. Für Langkanaltransistoren kann man annehmen, dass die metallurgische Gatelänge L_{met} der physikalischen Gatelänge L_G entspricht, da die Überlappung der SDE-Gebiete unter dem Gate vernachlässigbar ist.

Bild 5.15 zeigt die Ladungsträgerbeweglichkeit μ_n und μ_p in Abhängigkeit von der Gate-Source-Spannung U_{GS} für beide Transistortypen. Sowohl bei n- als auch bei p-MOSFETs sind Abweichungen in der Beweglichkeit zwischen der Ausheilung mit und ohne Kurzzeitausheilung zu erkennen, welche sehr gut mit den Ergebnissen der CP-Messung aus Bild 5.14 korrelieren. Die zusätzliche Kurzzeitausheilung verursacht eine höhere Grenzflächenhaftzentrendichte D_{it}, was wiederum eine geringere Ladungsträgerbeweglichkeit μ zur Folge hat.

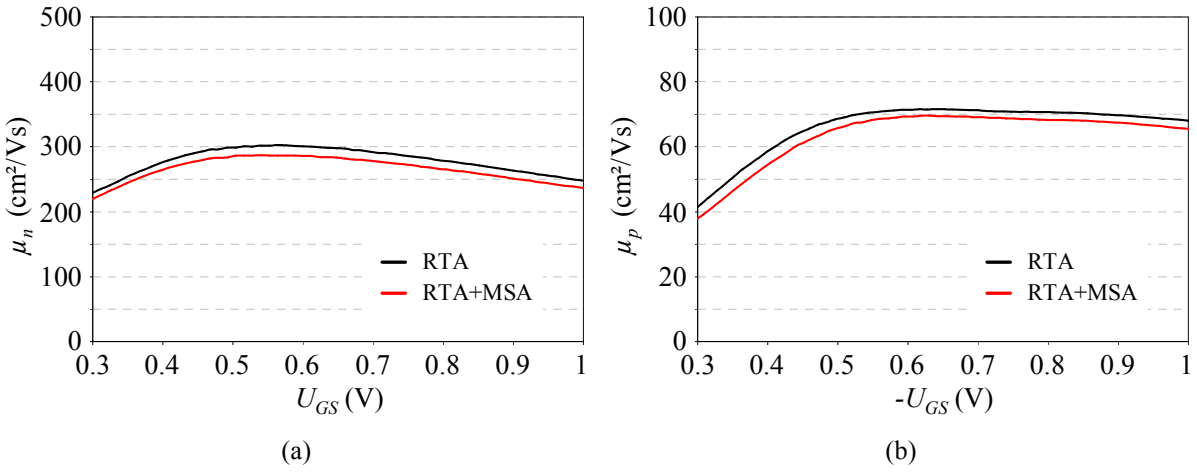


Bild 5.15: Einfluss der Kurzzeitausheilung (MSA) auf die Langkanal-Ladungsträgerbeweglichkeit μ in Abhängigkeit von der Gate-Source-Spannung U_{GS} für (a) n- und (b) p-MOSFETs

Trotz dieser negativen Beeinflussung der Ladungsträgerbeweglichkeit wird aber die Leistungsfähigkeit der n- und p-MOSFETs durch den zusätzlichen Ausheilschritt insgesamt verbessert (Abschnitt 5.2.1). Das bedeutet, dass durch die kleinere elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$ und, im Fall der n-MOSFETs, durch den verringerten Widerstand dieser negative Einfluss der Kurzzeitausheilung mehr als kompensiert wird.

5.3 Einfluss nachfolgender Prozessschritte

Eine weitere Herausforderung bei der Integration der Kurzzeitausheilung in den Herstellungsprozess ist die Empfindlichkeit der aktivierten Dotanden gegenüber nachfolgenden Temperaturschritten. Nach der Ausheilung der Implantationsschäden müssen noch weitere thermische Prozesse einschließlich der Silizidbildung durchgeführt werden, was aufgrund der durch die Kurzzeitausheilung erzeugten metastabilen Zustände eine Deaktivierung der Dotanden zur Folge haben kann [117]. Maßgeblich ist dabei der Prozess mit dem größten thermischen Budget *TB*.

Bei dem hier zugrunde liegenden Herstellungsprozess ist das die Abscheidung der Stoppschicht für die Kontaktlochätzung (Bild 5.16). Diese aus Siliziumnitrid (Si_3N_4) bestehenden Schichten werden nach der konventionellen Silizidbildung über den kompletten Transistor abgeschieden. In Abhängigkeit von den Prozessbedingungen (Druck, Temperatur und Wasserstoffgehalt im Si_3N_4) während der Nitridabscheidung sind diese Deckschichten unverspannt (neutral) oder intrinsisch druck- bzw. zugverspannt. Eine zugverspannte Nitridschicht verbessert die Leistungsfähigkeit von n-MOSFETs, eine druckverspannte Nitridschicht die von p-MOSFETs. Eine größere intrinsische Verspannung der Deckschicht sowie eine dickere Nitridschicht führen zu einer höheren Leistungsfähigkeit der Transistoren [118].

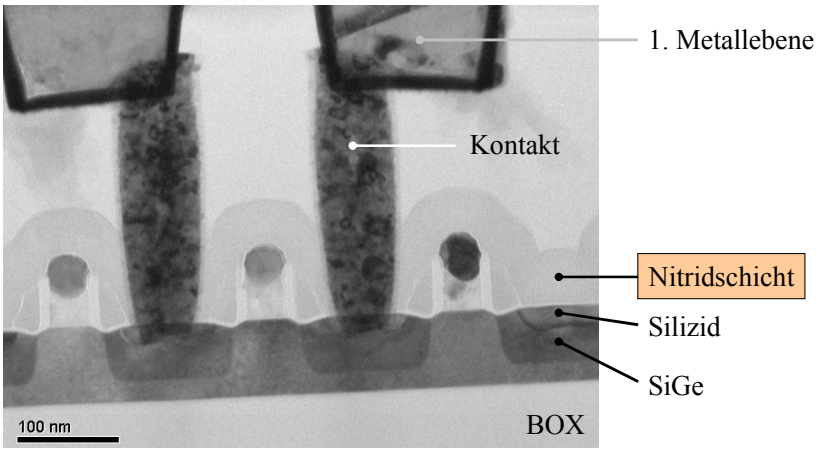


Bild 5.16: TEM-Aufnahme der Teststruktur zur Bestimmung der Leistungsfähigkeit des Transistors mit Darstellung der als Stoppschicht für die Kontaktlochätzung dienenden Nitridschicht.

Eine größere Verspannung erfordert jedoch eine höhere Prozesstemperatur und eine dickere Nitridschicht eine längere Prozesszeit, wodurch wiederum das thermische Budget dieses Prozessschritts und somit die Gefahr einer möglichen Deaktivierung der eingebrachten Dotanden erhöht wird.

Bild 5.17 zeigt dazu die Universalkurven unterschiedlich ausgeheilter MOSFETs in Abhängigkeit von der Abscheidetemperatur der Nitridschichten. Weder bei n- noch bei p-MOSFETs wird die Leistungsfähigkeit der Transistoren durch das höhere thermische Budget *TB* bei der Abscheidung der Nitridschicht beeinflusst. Daraus lässt sich schließen, dass zumindest bis 500°C durch nachfolgende Prozessschritte noch keine Deaktivierungsvorgänge zu befürchten sind.

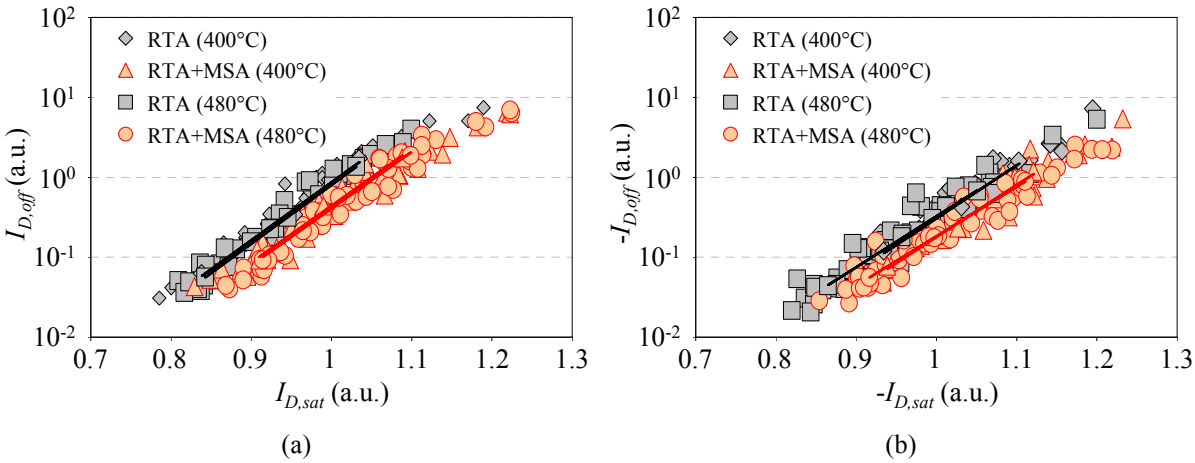


Bild 5.17: Einfluss der Kurzzeitausheilung (MSA) auf die Universalkurve von (a) n- und (b) p-MOSFETs in Abhängigkeit von der Abscheidetemperatur der Siliziumnitrid-Deckschichten

5.4 Einfluss auf verspannte Silizium-Germanium-Schichten

Wie man in Abschnitt 4.4.2 sehen konnte, ist die Implementierung von eingebettetem SiGe in die SD-Gebiete eine sehr effektive Methode zur Verbesserung der Leistungsfähigkeit von p-MOSFETs. Allerdings kann diese Verspannung während der nachfolgenden Kurzzeitausheilung relaxieren und somit die Effektivität dieser Verspannungstechnik reduzieren.

Aus diesem Grund wird im Folgenden der Einfluss der Kurzzeitausheilung auf die verspannte SiGe-Schicht untersucht. Zu diesem Zweck wurde entsprechend Bild 5.18 zunächst auf unstrukturierten Silizium-Wafern eine 40 nm dicke SiGe-Schicht mit einem Germaniumgehalt x_{Ge} von 20 % ($Si_{0.8}Ge_{0.2}$) epitaktisch abgeschieden. Danach erhielten alle Wafer eine dem realen Prozess entsprechende Halo-, SDE- und SD-Implantation und wurden anschließend entweder mit der Spike-RTA, einer Kombination aus RTA und Kurzzeitausheilung oder nur mittels Kurzzeitausheilung ausgeheilt. Bei der Kurzzeitausheilung wurde weiterhin der Einfluss der Wafertemperatur T_{Wafer} von 600 °C und 700 °C untersucht. Die Beurteilung der verschiedenen Prozessschritte auf die Verspannung der SiGe-Schicht erfolgte durch die Messung der Waferverbiegung, was ein direktes Maß für die Verspannung ist, sowie mit Hilfe der Röntgenbeugungsmessung (X-Ray Defraction, XRD), wodurch der Relaxationsgrad des SiGe-Films bestimmt werden kann.

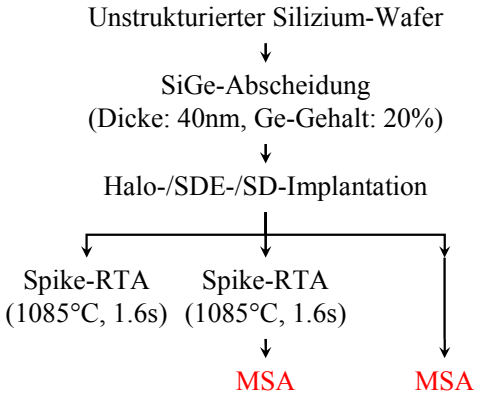


Bild 5.18: Prozessablauf zur Untersuchung des Einflusses der Kurzzeitausheilung (MSA) auf verspannte SiGe-Schichten

Bild 5.19 zeigt die Änderung Δs der Waferverbiegung als Differenz nach der SiGe-Abscheidung und nach der Ausheilung in Abhängigkeit von der Ausheiltemperatur $T_{Ausheil}$ für verschiedene Ausheilregime. Mit zunehmender Ausheiltemperatur $T_{Ausheil}$ der Kurzzeitausheilung ist eine sehr starke Veränderung Δs in der Waferverbiegung zu erkennen. Dabei spielt es keine Rolle, ob die Kurzzeitausheilung in Kombination mit der RTA oder als alleiniger Ausheilschritt durchgeführt wird. Eine Erhöhung der Wafertemperatur T_{Wafer} von 600 °C auf 700 °C reduziert diese Verspannungsänderung, so dass höhere Ausheiltemperaturen $T_{Ausheil}$ bei gleicher Waferverbiegung möglich sind. Die Waferverbiegung beschränkt zumindest auf unstrukturierten Silizium-Wafern die maximal möglichen Temperatursprünge von Wafer- zu Ausheiltemperatur, da Probleme bei der Bearbeitung entstehen (Handling-Probleme, Probleme bei nachfolgenden Lithographie-Prozessen) oder aber im schlimmsten Fall Waferbrüche auftreten können.

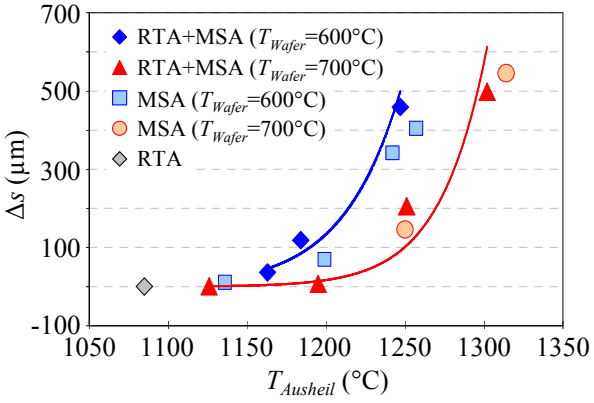


Bild 5.19: Änderung Δs der Waferverbiegung in Abhängigkeit von der Ausheiltemperatur $T_{Ausheil}$ für verschiedene Ausheilregime (Linien dienen zur Illustration des Einflusses)

Die Relaxation Δa der SiGe-Schicht, wenn entsprechend Bild 4.3 $a_{\parallel} > a_{Si} \leq a_{SiGe}$ und $a_{\perp} \geq a_{SiGe}$ gilt, zeigt ein ähnliches Verhalten in Abhängigkeit von den unterschiedlichen Ausheilregimes (Bild 5.20a). Mit zunehmender Ausheiltemperatur $T_{Ausheil}$ der Kurzzeitausheilung wird auch die Relaxation Δa der SiGe-Schicht größer, was wie im Fall der Waferverbiegung unabhängig von der Spike-RTA ist. Eine Erhöhung der Wafertemperatur T_{Wafer} von 600 °C auf 700 °C reduziert die Gefahr der Relaxation. Die Relaxation Δa des SiGe-Films ist aber nicht nur von der Ausheiltemperatur $T_{Ausheil}$, sondern auch vom Germaniumgehalt der SiGe-Verbindung abhängig (Bild 5.20b). Mit steigendem Germaniumanteil x_{Ge} erhöht sich auch die Wahrscheinlichkeit der Relaxation Δa der entsprechenden SiGe-Schicht.

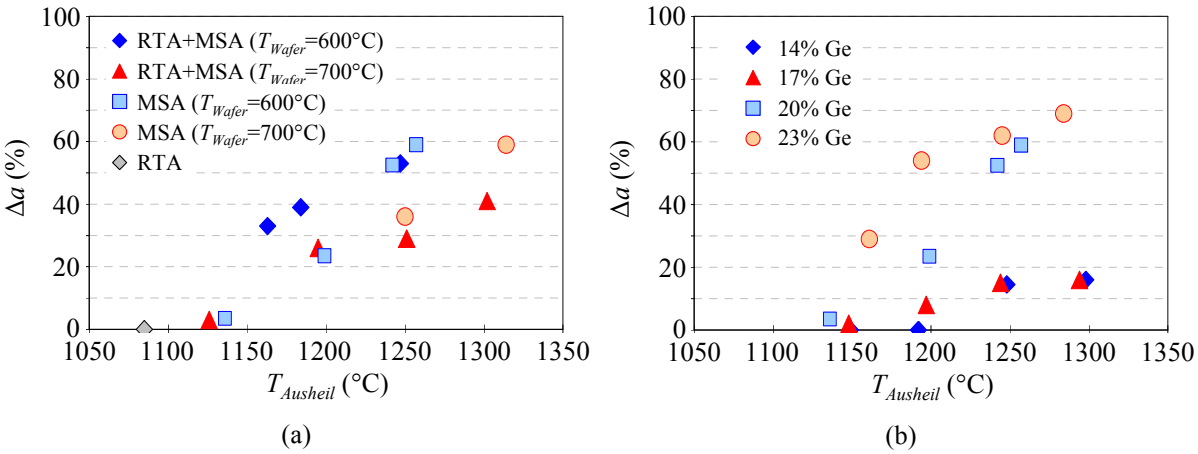


Bild 5.20: Relaxation Δa der SiGe-Schicht in Abhängigkeit von der Ausheiltemperatur $T_{Ausheil}$ für verschiedene (a) Ausheilregime und (b) Germaniumanteile (MSA, 600 °C)

Bild 5.21 zeigt die Waferverbiegung s für eine 40 nm dicke SiGe-Schicht mit einem Germaniumgehalt von 20 % nach verschiedenen Prozessschritten. Nach der SiGe-Abscheidung verursacht die druckverspannte SiGe-Schicht eine konkave Verbiegung des Wafers. Aufgrund der Implantationsschäden, vor allem durch die SD-Implantation, erfolgt eine Relaxation der SiGe-Schicht. Infolge dessen wird die Druckverspannung reduziert, was an der geringeren konkaven Waferverbiegung zu erkennen ist. Die nachfolgende RTA restauriert die entstandenen Implantationsschäden im SiGe, allerdings nur unvollständig, wie man anhand der Waferverbiegung feststellen kann. Die anschließende Kurzzeitausheilung verursacht in Abhängigkeit von Wafer- und Ausheiltemperatur eine signifikante Waferverbiegung, die sich mit zunehmender Ausheiltemperatur $T_{Ausheil}$ von druckverspannt in hoch zugverspannt verändert.

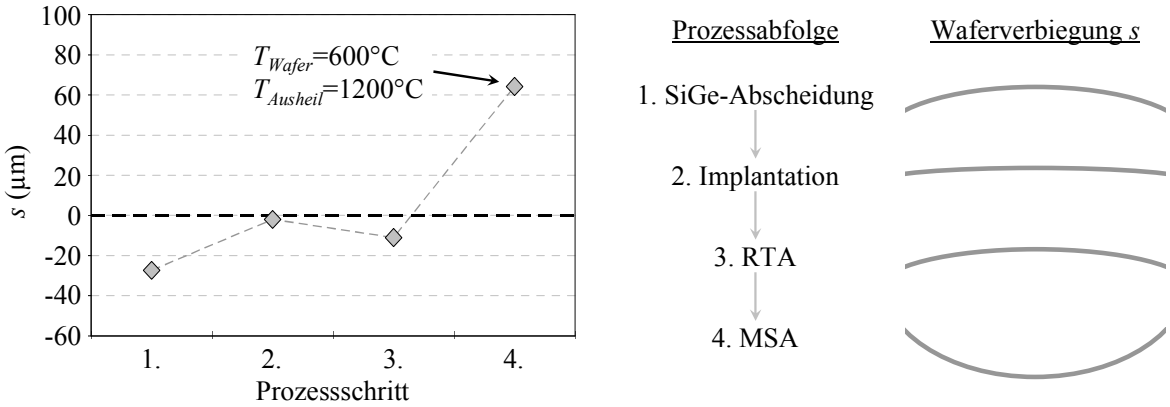


Bild 5.21: Waferverbiegung s nach verschiedenen Prozessschritten

Ursache dafür ist vor allem der durch die Kurzzeitausheilung entstehende Temperaturgradient über dem Wafer (Abschnitt 3.1). Durch die einseitige Erwärmung des Wafers kommt es in Abhängigkeit von der Ausheiltemperatur $T_{Ausheil}$ zu einer Druckverspannung an der Waferoberseite, die der durch die SiGe-Schicht erzeugten Kraft entgegenwirkt. Da die Stärke der durch die Kurzzeitausheilung erzeugten Verspannung sehr stark vom Temperaturgradienten abhängig ist, verringert eine höhere Wafertemperatur T_{Wafer} bei gleicher Ausheiltemperatur $T_{Ausheil}$ die thermische Verspannung und damit die Waferverbiegung.

Weiterhin hat die Stärke und Lage der Implantationsschäden einen wesentlichen Einfluss auf die Relaxation der SiGe-Schicht und folglich auf die Waferverbiegung. Erfolgt eine bis an die Si/SiGe-Grenzfläche heranreichende Zerstörung der SiGe-Struktur, wie es bei der SD- bzw. der Halo-Implantation der Fall ist, wird die Bindungsenergie im Kristall durch die bei der Kurzzeitausheilung erzeugte Verspannung überwunden und die Spannung baut sich über Versetzungen in der Kristallstruktur ab. Es kommt zu einer Relaxation der SiGe-Schicht (Bild 5.22). Dieser Versetzungsfehler pflanzt sich auf sogenannten Versetzungslinien bis in tiefere Regionen des Wafers fort (Bild 5.23). Durch diese Versetzungen kommt es zu der beobachteten hoch zugverspannten Verformung des Wafers, die Verformung wird sozusagen durch die Defekte gespeichert. Bei Implantationen mit einer geringen Implantationstiefe, wie z.B. die SDE-Implantation, erfolgt wie Messungen zeigen weder eine Waferverbiegung noch eine Relaxation der SiGe-Schicht.

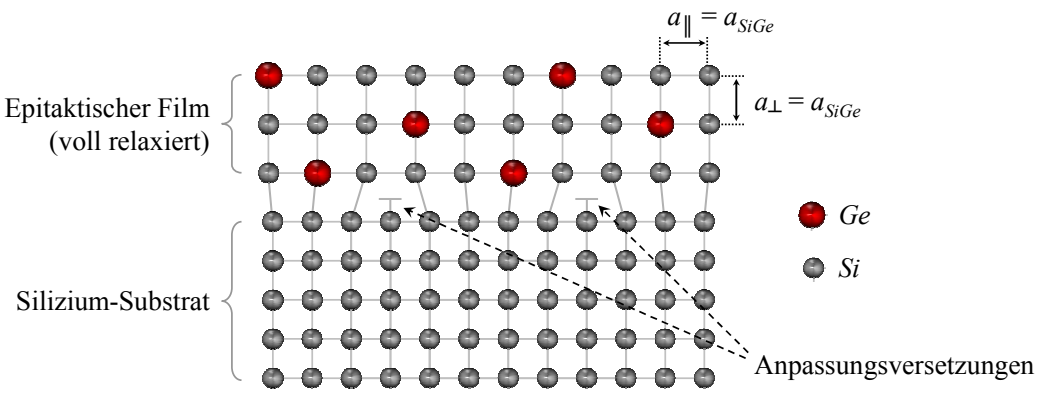


Bild 5.22: Schematische Darstellung einer voll relaxierten SiGe-Schicht ($a_{\parallel} = a_{SiGe}$, $a_{\perp} = a_{SiGe}$) auf einem Silizium-Substrat

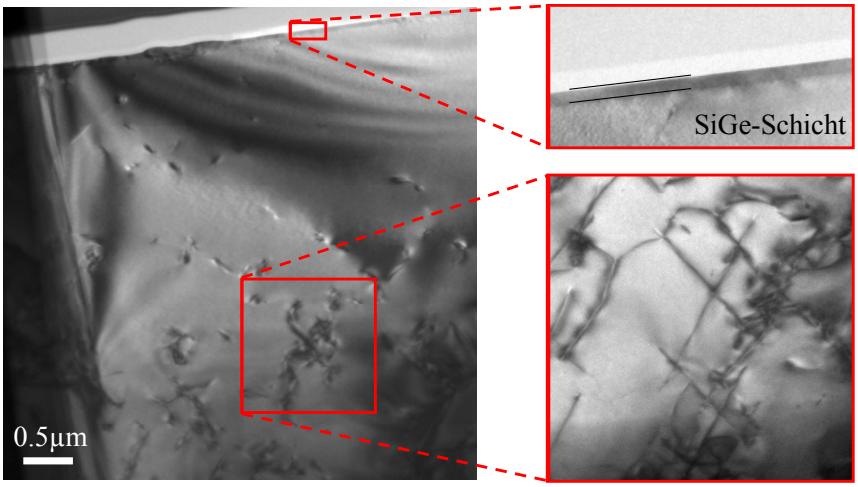


Bild 5.23: TEM-Aufnahme eines mit SiGe beschichteten Wafers nach der Kurzzeitausheilung

Auf SOI-Wafern wird dieses Verhalten interessanterweise nicht beobachtet. Die Messwerte der Waferverbiegung nach den unterschiedlichen Prozessschritten zeigen dort nur geringfügige Veränderungen. Wahrscheinlich dient die BOX zwischen der SOI-Schicht und dem Substrat als Puffer, welcher die Verspannungen aufnimmt und somit keine merklichen Veränderungen in der Waferverbiegung entstehen kann.

Ein weiteres Problem stellt die geringere Schmelztemperatur der SiGe-Verbindung im Vergleich zu Silizium dar. Bild 5.24 zeigt den Schichtwiderstand $R_{Schicht}$ in Abhängigkeit von der Ausheiltemperatur $T_{Ausheil}$ der Kurzzeitausheilung für unterschiedliche Wafertemperaturen T_{Wafer}. Erwartungsgemäß wird der Schichtwiderstand $R_{Schicht}$ mit steigender Ausheiltemperatur $T_{Ausheil}$ kleiner. Für Temperaturen von mehr als 1350 °C findet jedoch ein Aufschmelzen des SiGe-Films statt, wie Bild 5.25 zeigt.

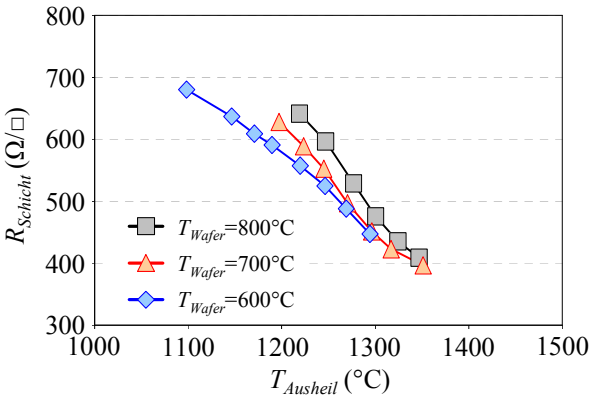


Bild 5.24: Schichtwiderstand $R_{Schicht}$ in Abhängigkeit von der Ausheiltemperatur $T_{Ausheil}$ der Kurzzeitausheilung für unterschiedliche Wafertemperaturen T_{Wafer}

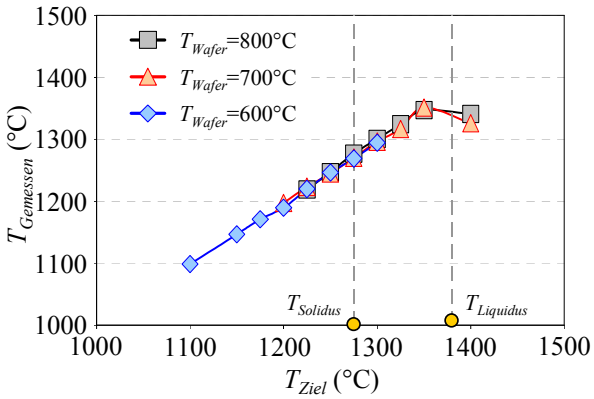


Bild 5.25: Aus der Strahlungsemission des Wafers ermittelte Temperatur $T_{Gemessen}$ in Abhängigkeit von der Zieltemperatur T_{Ziel} für unterschiedliche Wafertemperaturen T_{Wafer} mit Solidus- $T_{Solidus}$ und Liquidus-Temperatur $T_{Liquidus}$ für eine $Si_{0.8}Ge_{0.2}$-Verbindung (nach [119])

Bis zu einer Temperatur von 1350 °C sind die gemessenen Temperaturen $T_{Gemessen}$ im Einklang mit den Zieltemperaturen T_{Ziel}. Der Messpunkt bei 1400 °C fällt jedoch deutlich niedriger aus als erwartet. Betrachtet man das Phasendiagramm von SiGe, so ergibt sich für einen Germaniumanteil von 20 % ein Schmelzpunkt von etwa 1380 °C (Bild 5.26). Da dieser Wert zwischen den letzten beiden Messpunkten liegt, kann man davon ausgehen, dass ein Teil der Energie zum Aufschmelzen der Waferoberfläche verbraucht wurde und der Temperaturanstieg deshalb begrenzt war. Diese Ergebnisse zeigen, dass die geringere Schmelztemperatur der SiGe-Verbindung im Vergleich zu Silizium bei der Integration der Kurzzeitausheilung in einen Prozess mit SiGe mit beachtet werden muss. Wie jedoch Abschnitt 5.1 zeigt, ist bei dem in dieser Arbeit verwendeten Germaniumanteil der SiGe-Verbindung von 20 % vor allem die physikalische Beschädigung des Polysiliziumgates der limitierende Faktor für die maximal mögliche Ausheiltemperatur.

Bild 5.27 stellt den Einfluss der Kurzzeitausheilung auf die Leistungsfähigkeit von p-MOSFETs mit und ohne SiGe dar. Sowohl mit als auch ohne eingebettetes SiGe kann der Sättigungs-Drainstrom $I_{D,sat}$ bei konstantem Sperrstrom $I_{D,off}$ durch den zusätzlichen Ausheilschritt um etwa 3 % verbessert werden. Diese Ergebnisse deuten darauf hin, dass durch die Kurzzeitausheilung weder Relaxationserscheinungen noch Schädigungen der SiGe-Regionen auftreten.

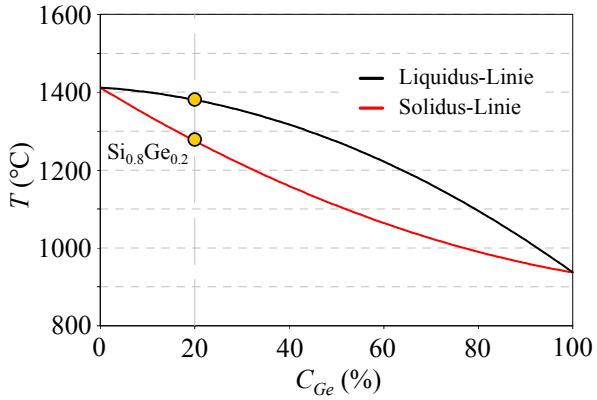


Bild 5.26: Phasendiagramm von $Si_{1-x}Ge_x$ (nach [119])

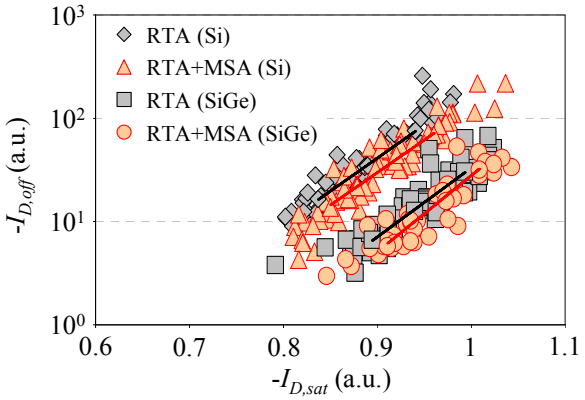


Bild 5.27: Universalkurve von p-MOSFETs mit und ohne SiGe für unterschiedliche Ausheilregime

5.5 Variation der Ausheiltemperatur der konventionellen Ausheilung

Im Folgenden wird der Einfluss der Kurzzeitausheilung auf das Transistorverhalten bei einer Verringerung der Ausheiltemperatur der Spike-RTA untersucht. Zu diesem Zweck wurde die Ausheiltemperatur $T_{Ausheil}$ bei einer konstanten Ausheilzeit $t_{Ausheil}$ von 1.6 s sukzessive von 1085 °C über 1065 °C und 1045 °C auf 1025 °C verringert. Bild 5.28 zeigt den Prozessablauf für die Herstellung der Versuchsproben mit den wesentlichen Kennwerten.

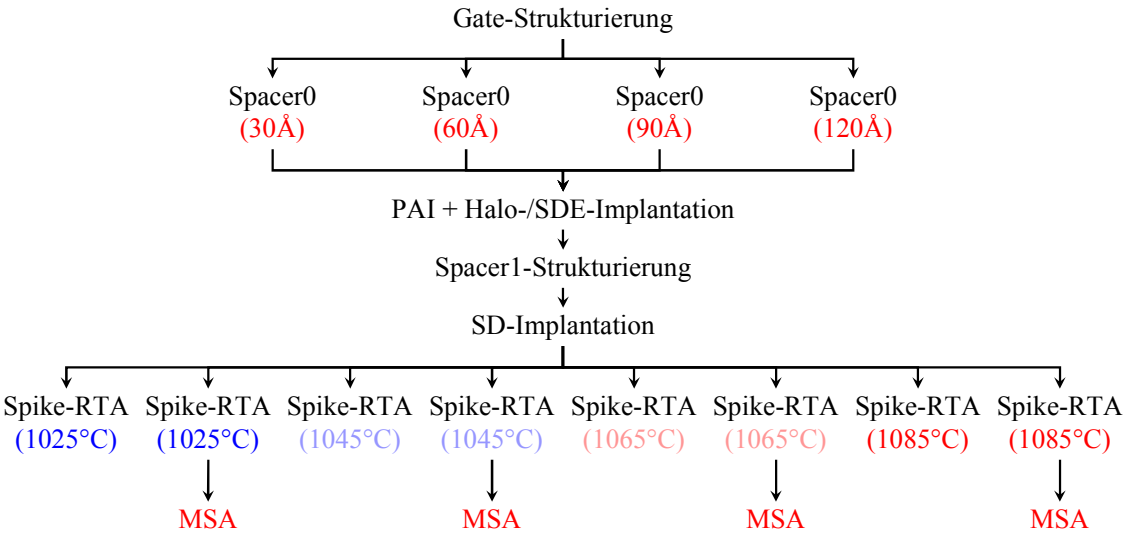


Bild 5.28: Prozessablauf für die Versuchsproben zur Untersuchung des Einflussses der RTA-Ausheiltemperatur bei unterschiedlichen Ausheilregimen

Um eine gleich große Überlappung der SDE-Gebiete unter das Gate zu gewährleisten, ist aufgrund der unterschiedlichen Diffusion eine Anpassung der Weite des Spacer0 erforderlich. Für die Ausheiltemperatur von 1025 °C wird als Abstandshalter für die SDE-Implantationen eine etwa 30 Å dicke Oxidschicht verwendet, die bei der Oxidation des Polysiliziums entsteht.

Diese Oxidschicht dient in der Regel als Schutz des Polysiliziums während der SiGe-Abscheidung, damit kein SiGe auf die Polysiliziumbahnen aufwächst. Die Oxidation des Polysiliziums erfolgt bei allen in dieser Arbeit untersuchten Transistoren unmittelbar nach der Gate-Strukturierung. Um einen einheitlichen Prozessablauf zu gewährleisten, wird dieser Prozessschritt selbst dann durchgeführt, wenn kein SiGe beim p-MOSFET verwendet wird. Für die RTA-Ausheiltemperaturen von 1045 °C, 1065 °C und 1085 °C wird zusätzlich eine SiO_2-Schicht mit einer Stärke von 30 Å, 60 Å bzw. 90 Å nach der Gate-Strukturierung abgeschieden. Demzufolge ergeben sich für die verschiedenen Ausheiltemperaturen Spacer0-Weiten von 30 Å, 60 Å, 90 Å bzw. 120 Å. Vor der Halo- und SDE-Implantation wird die Oxidschicht anisotrop zurückgeätzt. Weiterhin wurde die Dosis der Halo-Implantation angepasst, da es aufgrund der gewinkelten Implantation mit kleineren Spacer0-Weiten bei den betrachteten Gatelängen zur Überlagerung der Halo-Implantationen kommt, was wiederum bei gleichbleibender Dosis der Halo-Implantationen eine erhöhte Schwellspannung zur Folge hat. Aus diesem Grund wurde mit kleinerer Weite des Spacer0 die Dosis der Halo-Implantationen verringert. Alle anderen Prozessschritte (Spacer1-Strukturierung, SD-Implantationen) sind für die verschiedenen Experimente gleich.

Bild 5.29 zeigt den Schichtwiderstand R_{SDE} der SDE-Gebiete für beide Transistortypen für die verschiedenen RTA-Ausheiltemperaturen mit und ohne zusätzlicher Kurzzeitausheilung. Man kann erkennen, dass sich der Widerstand R_{SDE} der SDE-Gebiete mit kleiner werdender RTA-Ausheiltemperatur kontinuierlich erhöht, was sich durch einen flacheren pn-Übergang aufgrund der verringerten Diffusion und eine verminderte Aktivierung erklären lässt. Weiterhin kann man feststellen, dass sich zumindest bei n-MOSFETs durch den zusätzliche Ausheilschritt in Abhängigkeit von der verwendeten RTA-Ausheiltemperatur ein (18…23) % geringerer Widerstand R_{SDE} der SDE-Gebiete realisieren lässt, was entsprechend den Ergebnissen von Abschnitt 3.3.2 als Ursache eine höhere Aktivierung der Dotanden hat. Bei p-MOSFET hingegen ist ein Einfluss der zusätzlichen Kurzzeitausheilung nur bei geringen RTA-Ausheiltemperaturen zu erkennen. In Abhängigkeit von der Ausheiltemperatur verringert sich der Schichtwiderstand R_{SDE} der SDE-Gebiete um etwa (1…6) %.

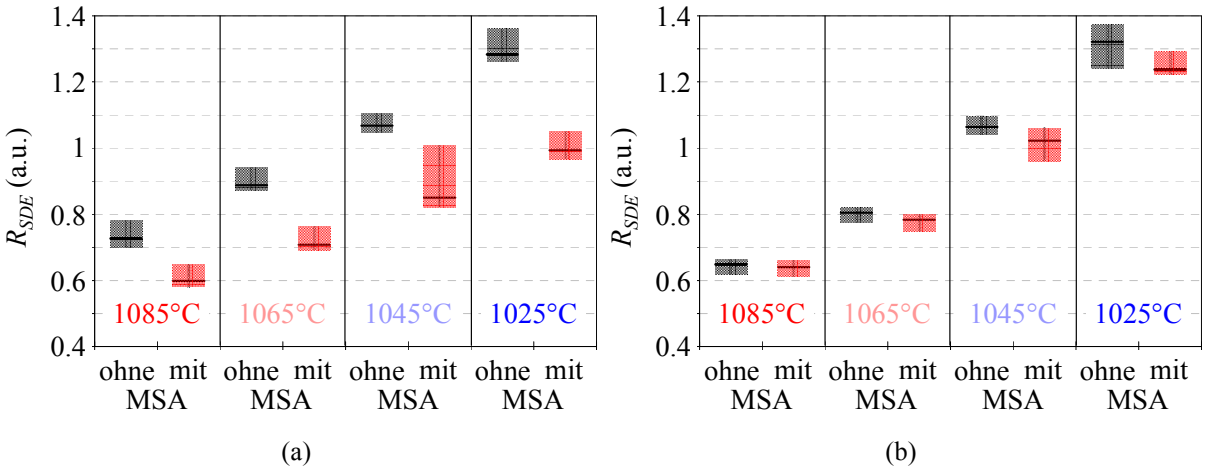


Bild 5.29: Einfluss der Kurzzeitausheilung (MSA) auf den Schichtwiderstand R_{SDE} der SDE-Gebiete für verschiedene RTA-Ausheiltemperaturen für (a) n- und (b) p-MOSFETs

Bild 5.30 zeigt für beide Transistortypen den Einfluss der Kurzzeitausheilung auf die elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$. Bei n-MOSFETs ist die elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$ mit und ohne Kurzzeitausheilung unabhängig von der RTA-Ausheiltemperatur. Durch den zusätzlichen Ausheilschritt verringert sich die elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$ um die bereits beobachteten 0.9 Å.

Ursache dafür ist die höhere Aktivierung der Dotanden im Polysiliziumgate, wie man anhand des geringeren Polysiliziumwiderstands R_{Poly} erkennen kann (Bild 5.31a). Bei p-MOSFETs dagegen zeigt sich ein anderes Bild. Die elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$ ist sowohl mit als auch ohne zusätzlicher Kurzzeitausheilung deutlich von der RTA-Ausheiltemperatur abhängig. Zum Einen kann man feststellen, dass sich die elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$ mit sinkender Ausheiltemperatur der RTA erhöht. Zum Anderen wird der Einfluss der Kurzzeitausheilung auf die elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$ mit kleiner werdender RTA-Ausheiltemperatur bedeutend geringer, obwohl durch den zusätzlichen Ausheilschritt bei kleineren Ausheiltemperaturen ein prozentual höherer Gewinn in der Aktivierung der Dotanden anhand des Polysiliziumwiderstands R_{Poly} erkennbar ist (Bild 5.31b).

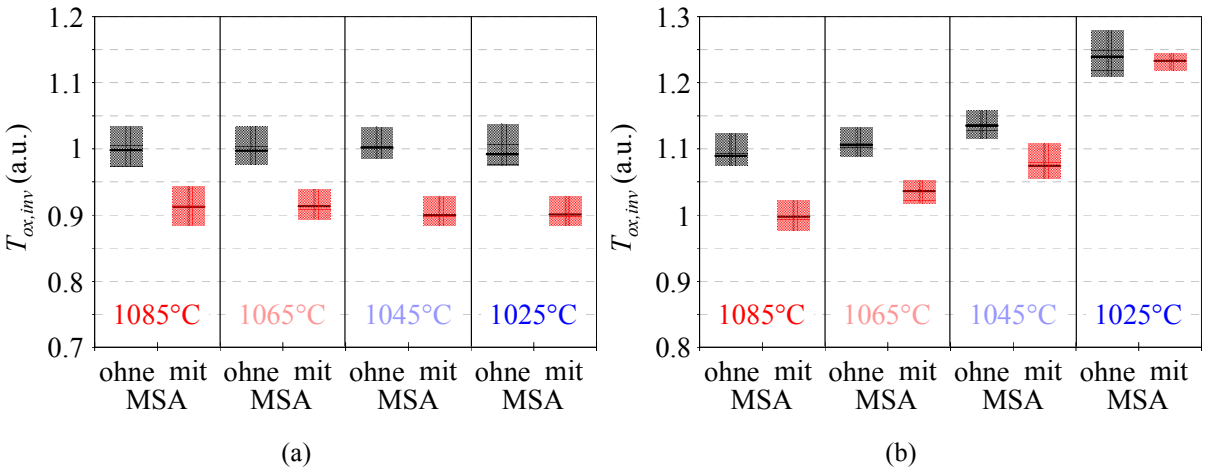


Bild 5.30: Einfluss der Kurzzeitausheilung (MSA) auf die elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$ für verschiedene RTA-Ausheiltemperaturen für (a) n- und (b) p-MOSFETs

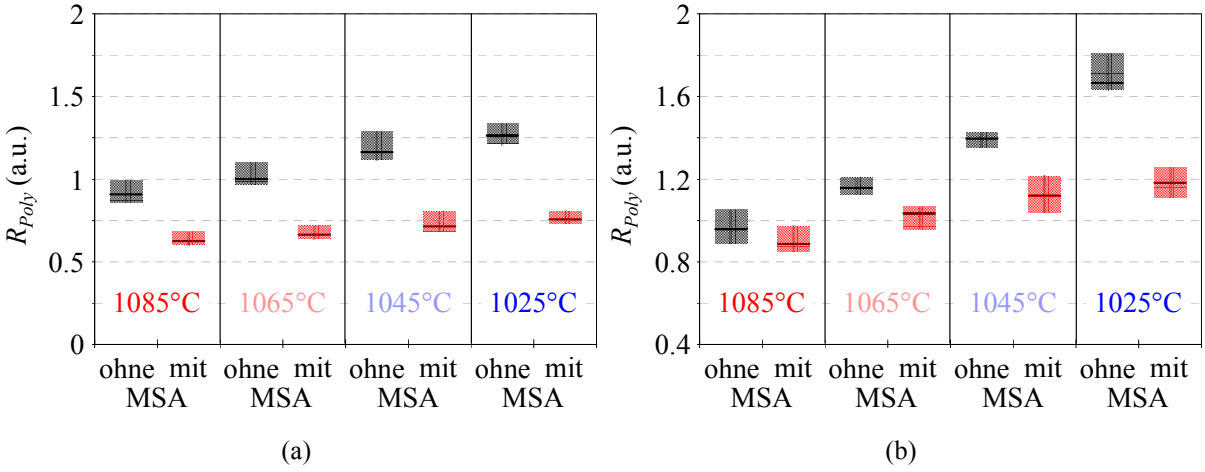


Bild 5.31: Einfluss der Kurzzeitausheilung (MSA) auf den Polysiliziumwiderstand R_{Poly} für verschiedene RTA-Ausheiltemperaturen für (a) n- und (b) p-MOSFETs

Bei n-MOSFETs erfolgt zwischen Polysiliziumabscheidung und Gate-Strukturierung eine Vordotierung der Polysiliziumschicht mit Phosphor, da die Arsen-Implantation der SD-Gebiete keine ausreichend gute n-Dotierung des Polysiliziumgates ermöglicht. Bei p-MOSFETs hingegen wird für die p-leitende Dotierung des Polysiliziumgates die Implantation der SD-Gebiete genutzt.

Mit Verringerung der RTA-Ausheiltemperatur reicht diese Implantation jedoch nicht mehr aus, um das Polysilizium vor allem an der Grenze zum Gateoxid hinreichend p-leitend zu dotieren (Bild 5.32). Dadurch wird auch der Einfluss des zusätzlichen Ausheilschritts mit kleiner werdender RTA-Ausheiltemperatur immer geringer und hat zumindest ab einer RTA-Ausheiltemperatur von 1025 °C unter den gegebenen Prozessbedingungen keine wesentliche Verbesserung in der elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$ mehr zur Folge.

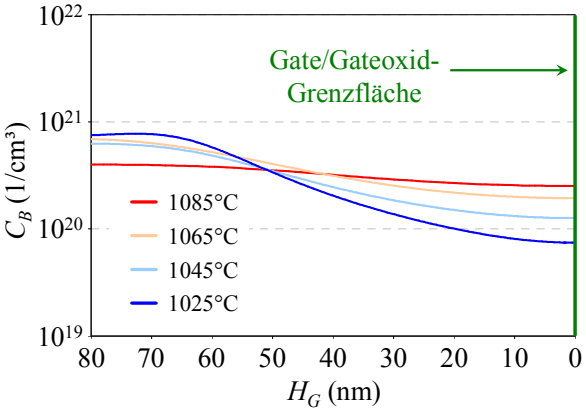


Bild 5.32: Berechnetes Dotierungsprofil C_B von Bor im Polysiliziumgate des p-MOSFETs für verschiedene RTA-Ausheiltemperaturen

Bild 5.33 zeigt die durch die zusätzliche Kurzzeitausheilung verursachte Schwellspannungsverschiebung

$$\Delta U_{th,sat} = |U_{th,sat_RTA+MSA}| - |U_{th,sat_RTA}| \tag{5.15}$$

für Langkanal-n- und -p-MOSFETs für die verschiedenen RTA-Ausheiltemperaturen. Mit abnehmender RTA-Ausheiltemperatur steigt der Einfluss der Kurzzeitausheilung auf die Schwellspannung $U_{th,sat}$. Bei n-MOSFETs sinkt die Sättigungsschwellspannung $U_{th,sat}$ durch den zusätzlichen Ausheilschritt mit kleiner werdender RTA-Ausheiltemperatur um (7…12) mV, wohingegen sie bei p-MOSFETs um (12…29) mV ansteigt. Dieses Verhalten ist jedoch nicht auf eine größere Grenzflächenhaftzentrendichte D_{it} zurückzuführen, da sie für beide Transistortypen mit und ohne zusätzlichen Ausheilschritt unabhängig von der RTA-Ausheiltemperatur ist (Bild 5.34). Im Rückschluss bedeuten diese Ergebnisse, dass die Kurzzeitausheilung mit geringerer Ausheiltemperatur der RTA eine höhere Anzahl an Oxidladungen erzeugt.

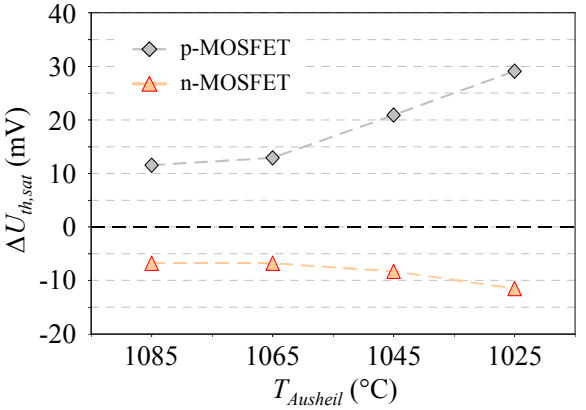


Bild 5.33: Schwellspannungsverschiebung $\Delta U_{th,sat}$ infolge der zusätzlichen Kurzzeitausheilung für Langkanal-n- und -p-MOSFETs in Abhängigkeit von der RTA-Ausheiltemperatur $T_{Ausheil}$

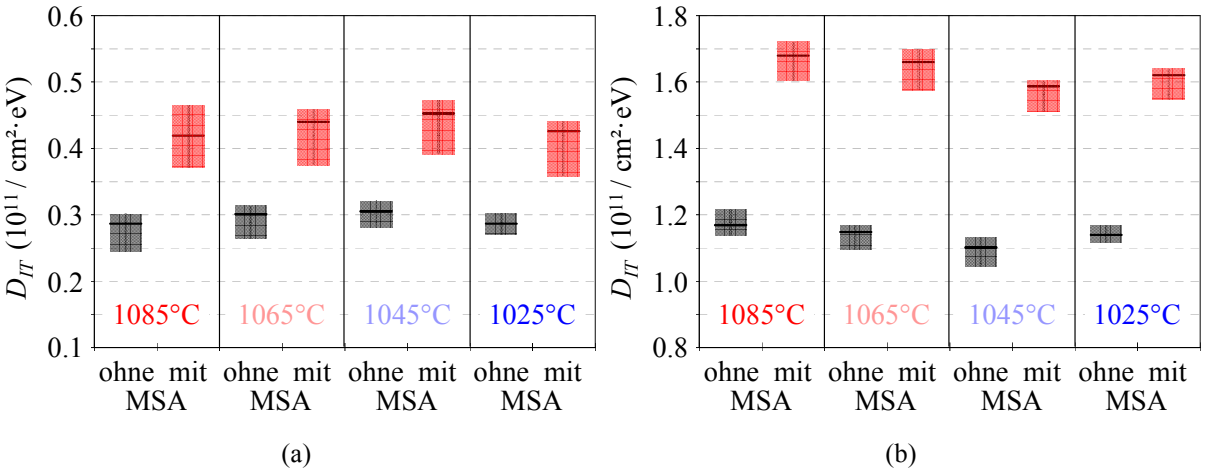


Bild 5.34: Einfluss der Ausheilregimes auf die Grenzflächenhaftzentrendichte D_{it} für verschiedene RTA-Ausheiltemperaturen $T_{Ausheil}$ für (a) n- und (b) p-MOSFETs

In Bild 5.35 ist der Einfluss des zusätzlichen Ausheilschritts bei den verschiedenen RTA-Ausheiltemperaturen auf die Langkanal-Ladungsträgerbeweglichkeit dargestellt. Sowohl die Elektronen- als auch die Löcherbeweglichkeit ist mit und ohne Kurzzeitausheilung unabhängig von der RTA-Ausheiltemperatur, was sich sehr gut mit den Ergebnissen der gemessenen Grenzflächenhaftzentrendichte D_{it} in Bild 5.34 deckt.

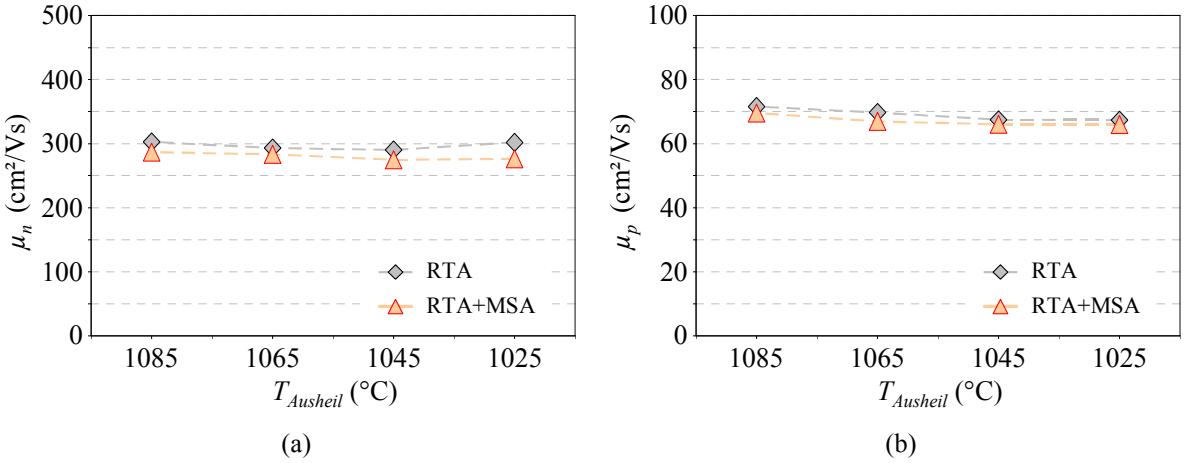


Bild 5.35: Einfluss der Ausheilregimes auf die Langkanal-Ladungsträgerbeweglichkeit μ in Abhängigkeit von der RTA-Ausheiltemperaturen $T_{Ausheil}$ für (a) Elektronen und (b) Löcher

Bild 5.36 zeigt die Universalkurven für beide Transistortypen in Abhängigkeit von der RTA-Ausheiltemperatur ohne zusätzliche Kurzzeitausheilung. Bei n-MOSFETs ist eine kontinuierliche Verringerung des Sättigungs-Drainstroms $I_{D,sat}$ bei konstantem Sperrstrom $I_{D,off}$ mit kleiner werdender RTA-Ausheiltemperatur zu erkennen. Da die elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$ (Bild 5.30a) als auch die Langkanal-Elektronenbeweglichkeit μ (Bild 5.35a) für die betrachteten Ausheiltemperaturen etwa gleich groß ist, kann nur der mit sinkender Temperatur steigende Widerstand R_{SDE} der SDE-Gebiete Ursache für dieses Verhalten sein (Bild 5.29). Die P-MOSFETs zeigen ein etwas anderes Verhalten. Mit kleinerer RTA-Ausheiltemperatur steigt der Sperrstrom $I_{D,off}$, was auf eine ungleichmäßige Einstellung der Schwellspannung U_{th} und der Überlappung der SDE-Gebiete unter das Gate bei der Variation der Ausheiltemperatur zurückgeführt werden kann.

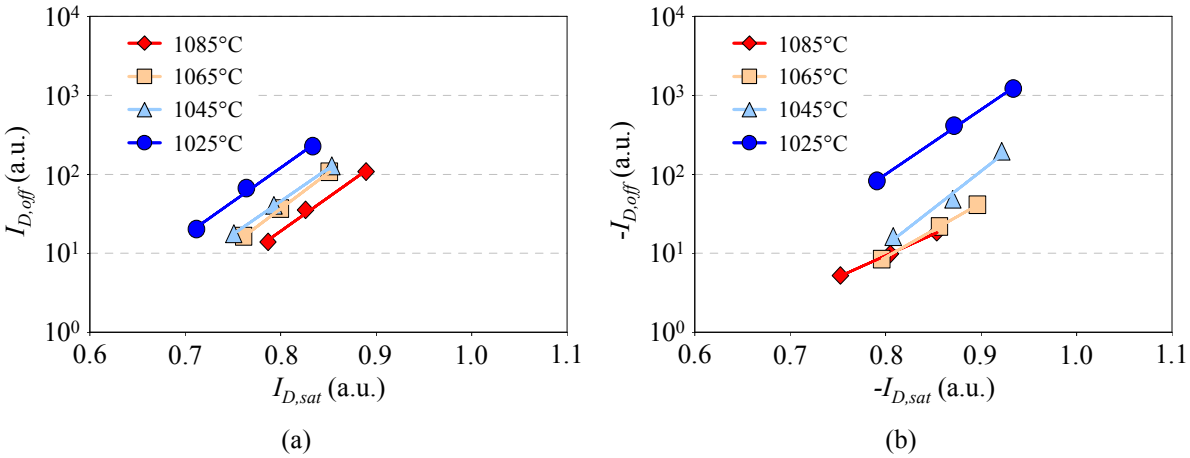


Bild 5.36: Universalkurve für verschiedene RTA-Ausheiltemperaturen ohne Kurzzeitausheilung für (a) n- und (b) p-MOSFETs

In Bild 5.37 ist die Drainstromänderung ΔI_D infolge der zusätzlichen Kurzzeitausheilung im linearen und Sättigungsbereich bei konstantem Sperrstrom $I_{D,off}$ in Abhängigkeit von der RTA-Ausheiltemperatur $T_{Ausheil}$ für beide Transistortypen dargestellt. Bei n-MOSFETs ist eine stetige Zunahme des Drainstromgewinns ΔI_D durch den zusätzlichen Ausheilschritt mit kleiner werdender RTA-Ausheiltemperatur in beiden Arbeitsbereichen zu erkennen, was vor allem auf den verringerten Widerstand R_{SDE} der SDE-Gebiete zurückzuführen ist (Bild 5.29a), da die elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$ unabhängig von der Ausheiltemperatur ist (Bild 5.30a). P-MOSFETs weisen hingegen ein entgegen gesetztes Verhalten auf. Mit Verringerung der Ausheiltemperatur sinkt der Drainstromgewinn ΔI_D durch die Kurzzeitausheilung. Unterhalb einer Ausheiltemperatur von 1065 °C verschlechtert sich sogar die Leistungsfähigkeit der p-MOSFETs durch den zusätzlichen Ausheilschritt. Ursache dafür ist die elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$ (Bild 5.30b). Mit Verringerung der Ausheiltemperatur wird der Einfluss der Kurzzeitausheilung auf die elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$ bedeutend geringer, wodurch die geringere Ladungsträgerbeweglichkeit μ nicht mehr kompensiert wird und zu dem beobachteten Drainstromverhalten führt.

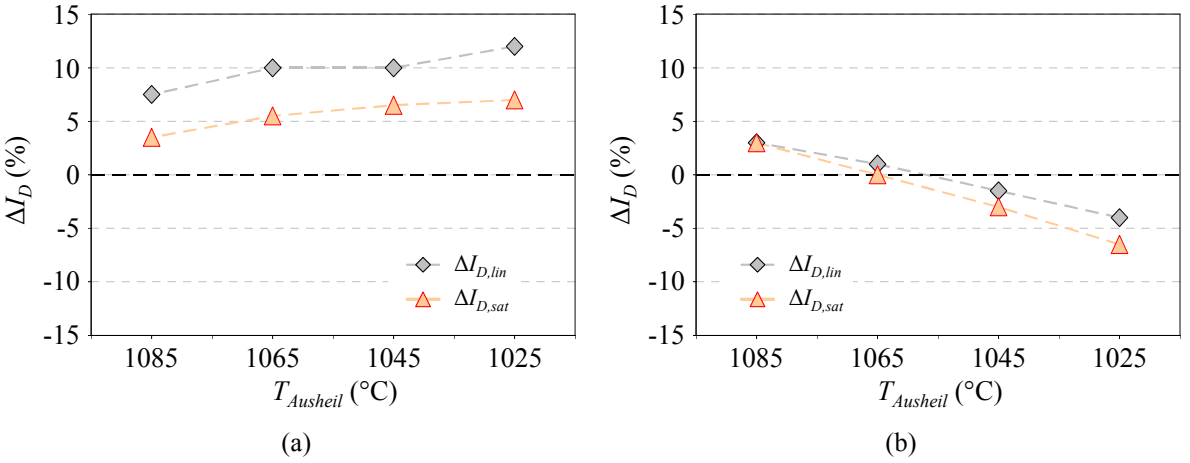


Bild 5.37: Einfluss der Kurzzeitausheilung (MSA) auf die Drainstromänderung ΔI_D im linearen und Sättigungsbereich bei konstantem Sperrstrom $I_{D,off}$ in Abhängigkeit von der RTA-Ausheiltemperatur $T_{Ausheil}$ für (a) n- und (b) p-MOSFETs

Um den Einfluss der elektrischen Gateoxiddicke im Inversionsfall $T_{ox,inv}$ auf die Leistungsfähigkeit des p-MOSFETs zu untersuchen, wurde für die Ausheiltemperatur von 1025 °C Polysiliziumschicht der p-MOSFETs vordotiert. Dabei wurde neben der zusätzlichen Vordotierung gleichzeitig ein Diffusionsschritt (1085 °C, 1.6 s) zwischen der Gate-Strukturierung und der Oxidation des Polysiliziumgates eingefügt.

Bild 5.38a zeigt den Einfluss der zusätzlichen Prozessschritte auf die elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$ für p-MOSFETs. Durch die Vordotierung mit anschließendem Diffusionsschritt verringert sich die elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$ um etwa 1.7 Å, wodurch sich die Leistungsfähigkeit um ca. 3 % im Vergleich zum RTA-Prozess bzw. um nahezu 10 % zum RTA-Prozess mit anschließender Kurzzeitausheilung ohne Vordotierung verbessert (Bild 5.38b).

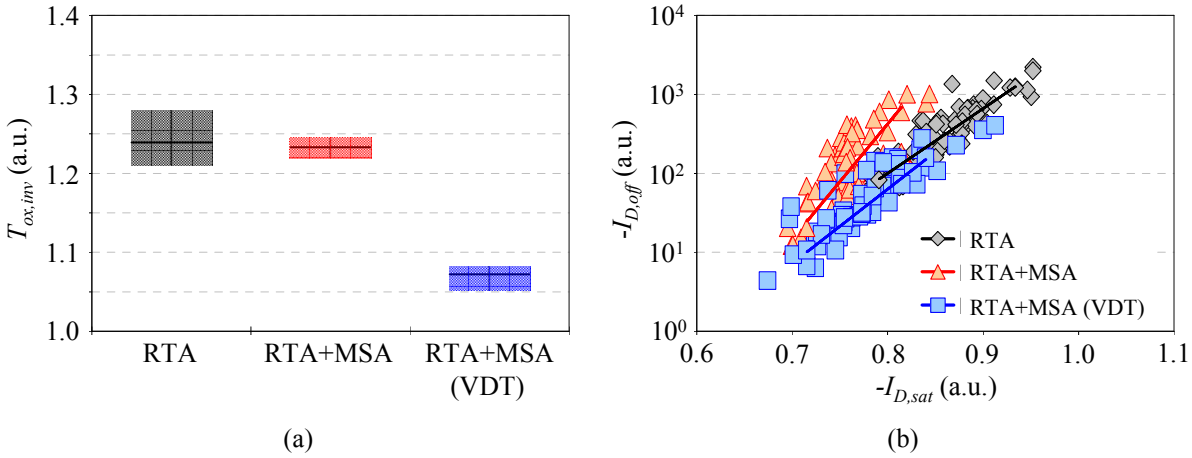


Bild 5.38: Einfluss der Ausheilregime sowie einer Vordotierung (VDT) des Polysiliziums auf (a) die elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$ und (b) die Universalkurve von p-MOSFETs

5.6 Zusammenfassung

In diesem Abschnitt wurde der Einfluss der Kurzzeitausheilung als zusätzlicher Ausheilschritt zur RTA auf verschiedene Transistorparameter untersucht. Der Schichtwiderstand R_{SDE} der SDE-Gebiete bei n-MOSFETs verringert sich bei einer RTA-Ausheiltemperatur von 1085 °C durch die zusätzliche Kurzzeitausheilung um circa 18 %, wohingegen bei p-MOSFETs nur ein geringfügig um (1…2) % kleinerer Schichtwiderstand R_{SDE} der SDE-Gebiete zu verzeichnen ist. Die elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$ verringert sich durch die zusätzliche Kurzzeitausheilung sowohl bei n- als auch bei p-MOSFETs um 0.9 Å. Untersuchungen des Gatedielektrikums ergaben weiterhin, dass durch den zusätzlichen Ausheilschritt Defekte im Gateoxid erzeugt werden, die sowohl eine Verschiebung der Schwellspannung U_{th} als auch eine Verschlechterung der Ladungsträgerbeweglichkeit μ zur Folge haben. Da aber die kleinere elektrisch Gateoxiddicke im Inversionsfall $T_{ox,inv}$ und im Fall der n-MOSFETs der verbesserte Schichtwiderstand R_{SDE} der SDE-Gebiete diesen negativen Einfluss der Kurzzeitausheilung mehr als kompensieren, wird durch den zusätzlichen Ausheilschritt die Leistungsfähigkeit von n- und p-MOSFETs im Sättigungsbereich um 3 % verbessert.

Mit sinkender Ausheiltemperatur der Spike-RTA zeigt sich hinsichtlich der Leistungsfähigkeit mit zusätzlicher Kurzzeitausheilung jedoch ein entgegengesetztes Verhalten von n- und p-MOSFETs. N-MOSFETs weisen durch den zusätzlichen Ausheilschritt mit kleiner werdender RTA-Ausheiltemperatur eine stetige Zunahme des Drainstromgewinns ΔI_D auf, was vor allem auf einen verbesserten Schichtwiderstand R_{SDE} der SDE-Gebiete zurückzuführen ist. Bei p-MOSFETs hingegen sinkt mit Verringerung der Ausheiltemperatur der Drainstromgewinn ΔI_D durch die Kurzzeitausheilung. Unterhalb einer RTA-Ausheiltemperatur von 1065 °C verschlechtert sich sogar die Leistungsfähigkeit von p-MOSFETs durch den zusätzlichen Ausheilschritt, da der Einfluss der Kurzzeitausheilung auf die elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$ unter den gegebenen Prozessbedingungen mit sinkender RTA-Ausheiltemperatur bedeutend geringer wird.

Weitere Untersuchungen zeigten, dass zumindest bis 500 °C durch nachfolgende Prozessschritte noch keine Deaktivierungsvorgänge der durch die Kurzzeitausheilung erzeugten metastabilen Zustände zu befürchten sind.

Da die Implementierung von eingebettetem SiGe in die SD-Gebiete eine sehr effektive Methode zur Verbesserung der Leistungsfähigkeit von p-MOSFETs ist, wurde auch der Einfluss der Kurzzeitausheilung auf verspannte SiGe-Schichten untersucht. Die dabei erzielten Ergebnisse zeigen, dass zumindest auf unstrukturierten Silizium-Wafern mit zunehmender Ausheiltemperatur der Kurzzeitausheilung eine sehr starke Veränderung in der Waferverbiegung zu erkennen ist. Infolge dessen kommt es in Abhängigkeit von der Stärke und der Lage der Implantationsschäden zur Relaxation der SiGe-Schicht. Auf SOI-Wafern wird dieses Verhalten nicht beobachtet. Durch die Kurzzeitausheilung treten weder Relaxationserscheinungen noch Schädigungen der SiGe Regionen auf. Sowohl ohne als auch mit SiGe kann die Leistungsfähigkeit vom p-MOSFETs durch den zusätzlichen Ausheilschritt um etwa 3 % verbessert werden.

6 Diffusionsarmer Transistor

Dieses Kapitel stellt die Herausforderungen bei der Herstellung diffusionsarmer Transistoren, bei denen nur die Kurzzeitausheilung als finaler Ausheilschritt verwendet wird, und die dabei erreichten Ergebnisse dar. Ein Hauptaugenmerk liegt in der Implementierung von in die Source/Drain-Gebiete der Transistoren eingebetteten Verspannungsquellen, um die Erhöhung der Leistungsfähigkeit auch bei diesen Transistortypen zu gewährleisten. Weiterhin werden der Einfluss verschiedener Implantationsparameter sowie zusätzliche Niedrigtemperaturschritte im Prozessablauf auf die Leistungsfähigkeit der Transistoren untersucht. Die vorgestellten Resultate basieren auf Experimenten mit Transistoren der 90 nm-, 65 nm- und 45 nm-Technologiegenerationen.

6.1 Vorbetrachtungen

Als diffusionsarmer Transistor wird ein Transistor bezeichnet, bei dem die Lage der Dotanden hauptsächlich durch die Implantation bestimmt wird. Die Ausheilung der Implantationsschäden soll demzufolge keinen oder nur einen geringen Einfluss auf das letztendlich entstehende Dotierungsprofil haben. Wesentlich ist dabei die Formierung der SDE-Gebiete (siehe Abschnitt 2.3.1), da diese maßgeblich für die mit der Verringerung der Gatelänge einhergehende Verstärkung der Kurzkanaleffekte verantwortlich sind. Prinzipiell existieren zwei verschiedene Ansätze, um dieses Ziel zu erreichen (Bild 6.1).

- Ein erster Ansatz nutzt den aus Kapitel 5 bekannten Herstellungsprozess. Allerdings wird das Polysiliziums für den n- und p-MOSFET vordotiert sowie anschließend mit einer Spike-RTA ausgeheilt, um eine Polysiliziumverarmung und damit eine große elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$ zu verhindern. Nach der Strukturierung des Polysiliziumgates wird durch Oxidation des polykristallinen Gates ein Oxid hergestellt, welches anschließend mit Hilfe der angepassten Spacer0-Ätzung auf den Aktiv-Gebieten anisotrop zurückgeätzt wird. Diese Oxidschicht dient als Abstandshalter für die Implantationen der Halo- und SDE-Gebiete und als Schutz des Gateoxids gegen diese Implantationen. Danach wird als Stoppschicht für die Spacer1-Ätzung eine Oxidschicht abgeschieden. Nach der sich anschließenden Strukturierung des Spacer1, der als Abstandshalter für die Implantationen der tiefen SD-Gebiete dient, wird im Anschluss an die SD-Implantationen nur die Kurzzeitausheilung durchgeführt. Alle weiteren Prozessschritte einschließlich der Silizierung entsprechen dem eines herkömmlichen Prozessablaufs mit RTA.
- In einem zweiten Ansatz werden die SDE- und SD-Gebiete voneinander getrennt ausgeheilt. Nach der Strukturierung des Polysiliziumgates und dessen Oxidation wird ein Hilfs-Spacer abgeschieden und strukturiert. Daran anschließend werden die SD-Implantationen durchgeführt. An dieser Stelle erfolgt im Prozess ein Ausheilschritt mit der Spike-RTA, der die Implantationsschäden beseitigen soll [122]. Nach der Entfernung des Hilfs-Spacers wird die als Stoppschicht für die Spacer-Ätzung dienende Oxidschicht auf den Aktivgebieten anisotrop zurückgeätzt. Erst jetzt werden die Implantationen der Halo- und SDE-Gebiete durchgeführt. Nach der sich anschließenden Spacer1-Strukturierung erfolgt die Kurzzeitausheilung. Alle weiteren Prozessschritte sind identisch zu ersten Variante.

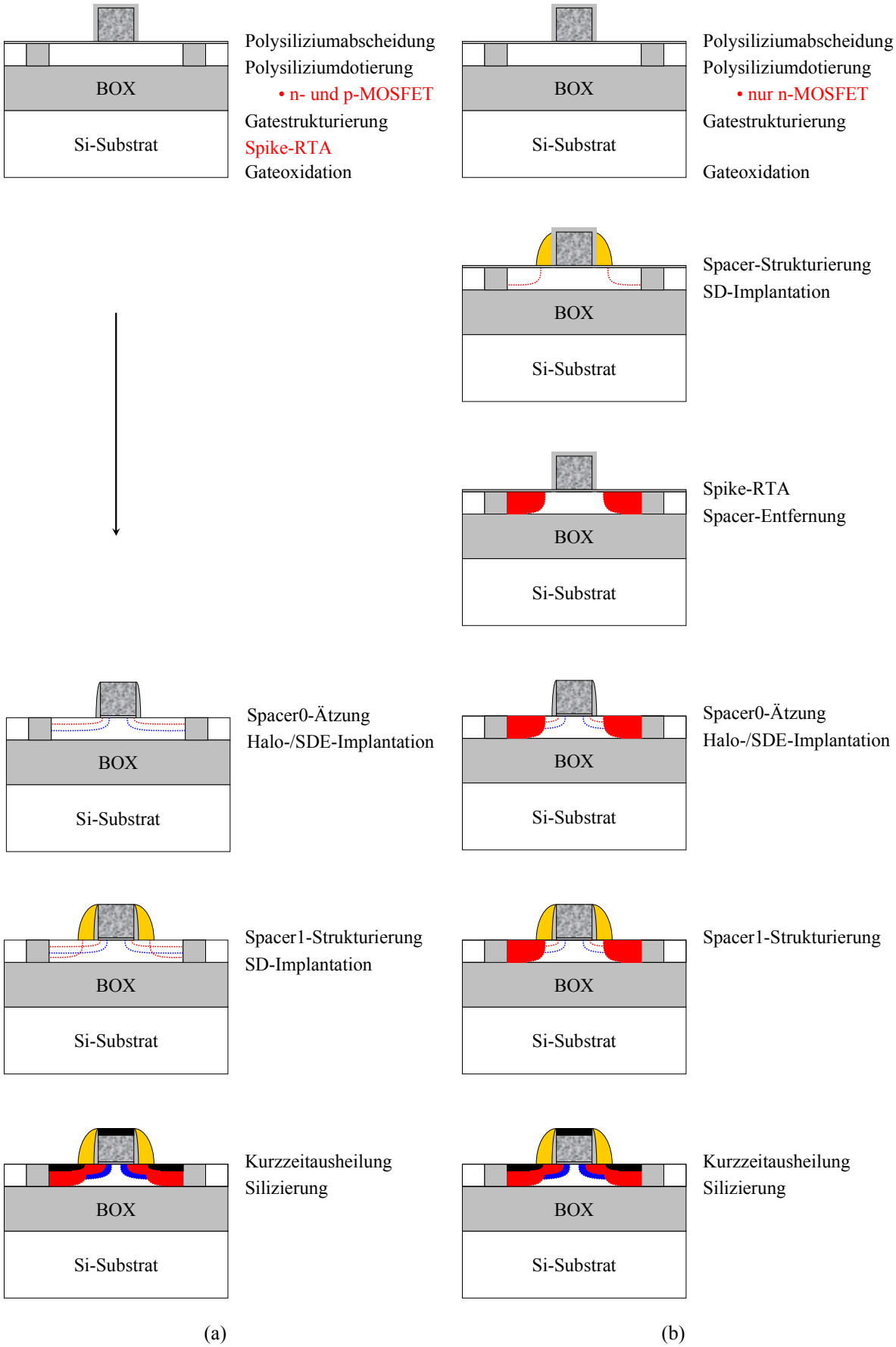


Bild 6.1: Herstellung diffusionsarmer Transistoren nach (a) Variante 1 und (b) Variante 2 (Erläuterung im Text)

Experimente zeigten, dass für p-MOSFETs beide Varianten vergleichbare Transistorergebnisse erzielten, wohingegen bei n-MOSFETs große Unterschiede in der Leistungsfähigkeit beobachtet werden konnten. Durch die Veränderung des Prozessablaufs ändert sich der Abstand der SDE-Implantationen zum Gate, was wiederum eine größere metallurgische Gatelänge zur Folge hat, wie man anhand der um 30 aF geringeren MILLERkapazität für den n-MOSFET nach Variante 2 feststellen kann (Bild 6.2a). Die unzureichende Überlappung der Erweiterungsgebiete unter dem Gate führt zu einer Verschlechterung der Leistungsfähigkeit um 12.5 % (Bild 6.2b). Selbst eine Optimierung der Implantationsparameter brachte keine nennenswerte Verbesserung hinsichtlich der Leistungsfähigkeit des n-MOSFETs. Aufgrund dieser Resultate und des größeren Prozessaufwands von Variante 2, was wiederum mit höheren Herstellungskosten verbunden ist, konzentrierten sich die weiteren Betrachtungen auf den Ansatz mit der gleichzeitigen Ausheilung der SDE- und SD-Gebiete (Variante 1).

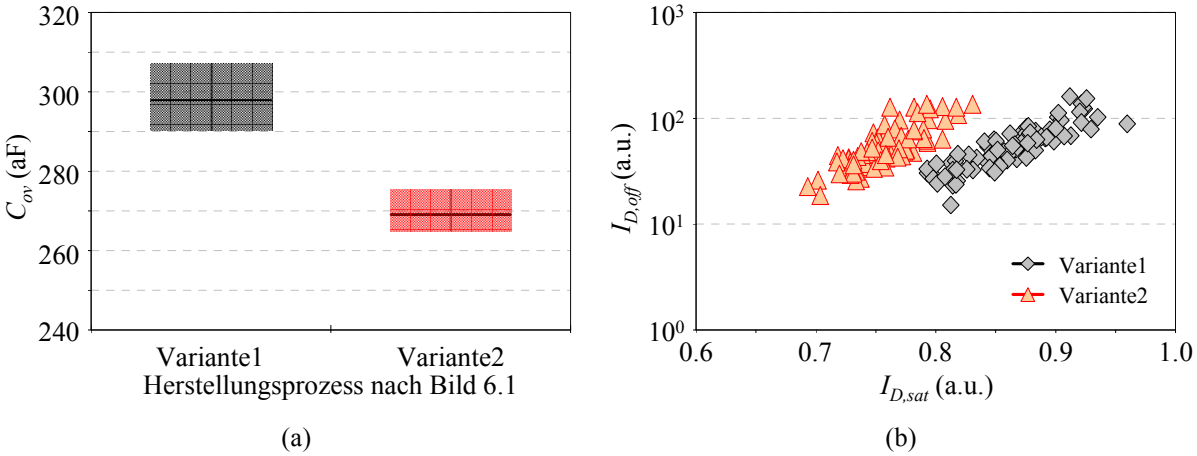


Bild 6.2: Einfluss der Herstellung diffusionsarmer Transistoren auf (a) die MILLERkapazität C_{ov} und (b) die Universalkurve von n-MOSFETs

6.2 Herstellung diffusionsarmer Transistoren

6.2.1 Erste Ergebnisse

Ausgangspunkt für die Experimente zum diffusionsarmen Transistor (nur mit Kurzzeitausheilung) war ein Standardprozess mit einer kombinierten Ausheilung, bestehend aus einer Niedrigtemperatur-RTA und der Kurzzeitausheilung. Im Vergleich zum Standardprozess von Abschnitt 5.2.1/Bild 5.2 bestehen hier folgende Besonderheiten:

- Ein wesentlicher Prozessschritt bei der Verwendung einer Niedrigtemperatur-RTA bzw. nur der Kurzzeitausheilung als finale Ausheilung ist die Vordotierung des Polysiliziums in Verbindung mit einer Hochtemperatur-RTA (1085 °C). Wie man in Abschnitt 5.5 sehen konnte, wird dadurch eine Verarmung des Polysiliziums und damit einhergehend eine größere elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$ vermieden.
- Als Abstandshalter für die SDE-Implantationen dient das Polysiliziumoxid, das durch Oxidation des polykristallinen Gates hergestellt und anschließend auf den Aktivgebieten anisotrop zurückgeätzt wird.

- Nach einer PAI sowie der Implantation der Halo- und SDE-Gebiete wird eine Oxidschicht abgeschieden, welche als Stoppschicht für die Spacer1-Ätzung verwendet wird. Danach wird eine Siliziumnitrid- (SiN-) Schicht als Spacer1 abgeschieden.
- An dieser Stelle im Prozess wird noch vor der Spacer1-Strukturierung ein Diffusionsschritt im Bereich von (600…800) °C für mehrere Minuten durchgeführt, der aber aufgrund seines geringen thermischen Budgets *TB* keinen Einfluss auf die Diffusion der bisher eingebrachten Dotanden hat. Aufgrund dieser Temperaturbehandlung entstehen durch die unterschiedlichen Ausdehnungskoeffizienten der verschiedenen Materialien Verspannungen im Transistor, die auch nach der Strukturierung des Spacer1 noch vorhanden sind. Die Verspannungen werden sozusagen gespeichert und führen zu einer Verbesserung der Elektronenbeweglichkeit und entsprechend zu einer Erhöhung der Leistungsfähigkeit des n-MOSFETs [123].
- Daran anschließend erfolgen die Spacer1-Strukturierung, die Implantation der Dotanden der SD-Gebiete sowie die finale Ausheilung, bestehend aus einer Niedrigtemperatur-RTA und der Kurzzeitausheilung. Der Übergang zum diffusionsarmen Transistor erfolgte unter Vernachlässigung der Niedrigtemperatur-RTA.

Bild 6.3 zeigt für beide Transistortypen den Einfluss der unterschiedlichen Ausheilverfahren auf die elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$ bzw. den Schichtwiderstand R_{SDE} der SDE-Gebiete. Bei n-MOSFETs ist die elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$ unabhängig von der Prozessierung, wohingegen sie bei p-MOSFETs ohne Niedrigtemperatur-RTA trotz Vordotierung des Polysiliziumgates und zusätzlichem Ausheilschritt aufgrund des geringeren thermischen Budgets der finalen Ausheilung um 1.5 Å ansteigt. Experimente zeigen, dass durch eine höhere Implantationsdosis der Vordotierung die elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$ jedoch angepasst werden kann, wodurch sich die Leistungsfähigkeit verbessern kann. Aufgrund der Variation der Gatehöhe zwischen und auch innerhalb von Technologiegenerationen ist eine genaue Bestimmung von Implantationsenergie und -dosis nur schwer möglich, da sie individuell an die jeweilige Gatehöhe angepasst werden müssen. Der Schichtwiderstand R_{SDE} der SDE-Gebiete erhöht sich für den Prozess ohne Niedrigtemperatur-RTA beim n-MOSFET um 28 % und beim p-MOSFET um 52 %. Unter der Annahme einer gleich hohen Aktivierung der Dotanden, infolge der zusätzlichen Kurzzeitausheilung, ist der höhere Schichtwiderstand R_{SDE} durch die verringerte Diffusion und eine damit kleinere Tiefe der pn-Übergänge zu erklären.

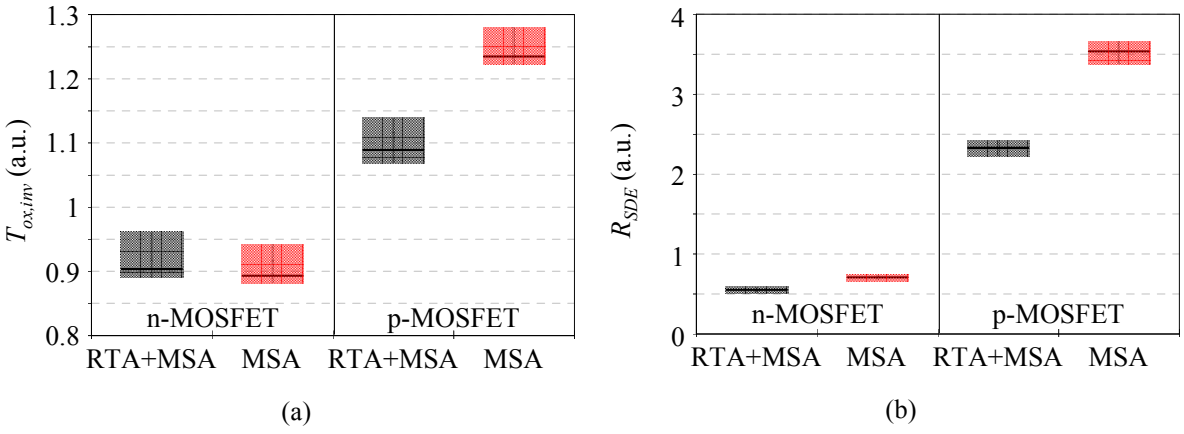


Bild 6.3: Einfluss unterschiedlicher Ausheilbedingungen auf (a) die elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$ und (b) den Schichtwiderstand R_{SDE} der SDE-Gebiete für n- und p-MOSFETs

Bild 6.4 zeigt die Universalkurven für beide Transistortypen für die unterschiedlichen Ausheilbedingungen. Sowohl bei n- als auch bei p-MOSFETs ist eine relative starke Verringerung des Sättigungs-Drainstroms $I_{D,sat}$ zu erkennen. Beim n-MOSFET kann dies auf den erhöhten Bahnwiderstand der SDE- und SD-Gebiete und auf die geringfügig kleinere Überlappung der SDE-Gebiete unter dem Gate zurückgeführt werden. Letzteres lässt sich anhand der etwas geringeren MILLERkapazität C_{ov} für n-MOSFETs nur mit Kurzzeitausheilung erklären (Bild 6.5). Bei p-MOSFETs ist neben dem ebenfalls höheren Widerstand hauptsächlich die deutlich erhöhte elektrische Gateoxiddicke im Inversionsfall $T_{ox,inv}$ sowie die unzureichende Überlappung der SDE-Gebiete unter dem Gate Ursache für die starke Änderung des Sättigungs-Drainstroms $I_{D,sat}$. P-MOSFETs, die nur kurzzeitausheilt wurden, weisen im Vergleich zu p-MOSFETs mit einer kombinierten Ausheilung eine um etwa 50 aF geringere Kapazität C_{ov} auf (Bild 6.5), was aufgrund der gleichen Geometrie der Transistoren auf eine stark verringerte Überlappung zurückzuführen ist.

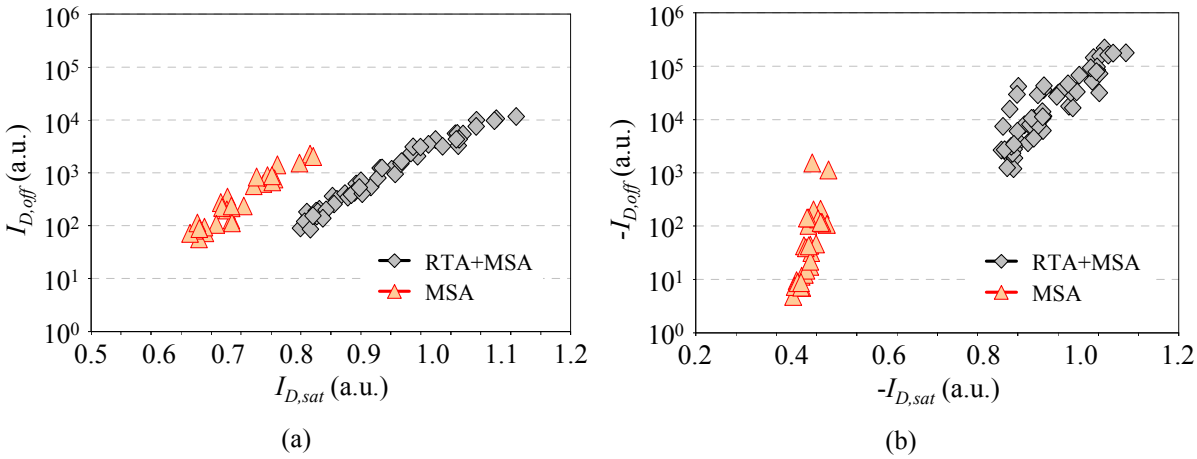


Bild 6.4: Universalkurve für unterschiedliche Ausheilbedingungen für (a) n- und (b) p-MOSFETs

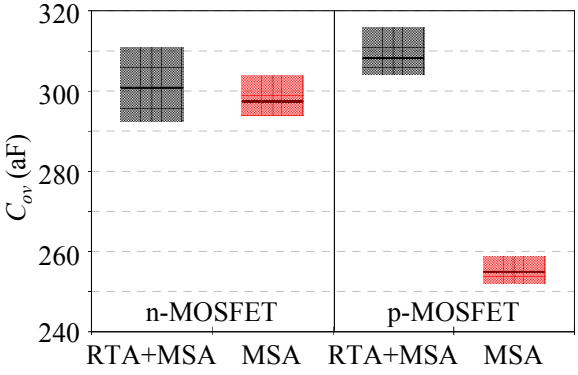


Bild 6.5: MILLERkapazität C_{ov} für unterschiedliche Ausheilbedingungen für n- und p-MOSFETs

Bild 6.6 zeigt die Sättigungsschwellspannung $U_{th,sat}$ in Abhängigkeit von der Gatelänge L_G für n- und p-MOSFETs. Bei n-MOSFETs verbessert sich unter Vernachlässigung der RTA das Kurzkanalverhalten, was jedoch auf die größere metallurgische Gatelänge L_{met} zurückgeführt werden kann. P-MOSFETs nur mit Kurzzeitausheilung zeigen aufgrund der unzureichenden Überlappung der SDE-Gebiete unter dem Gate betragsmäßig eine stark erhöhte Sättigungsschwellspannung $U_{th,sat}$, wodurch eine Aussage über eine etwaige Veränderung des Kurzkanalverhaltens nicht möglich ist.

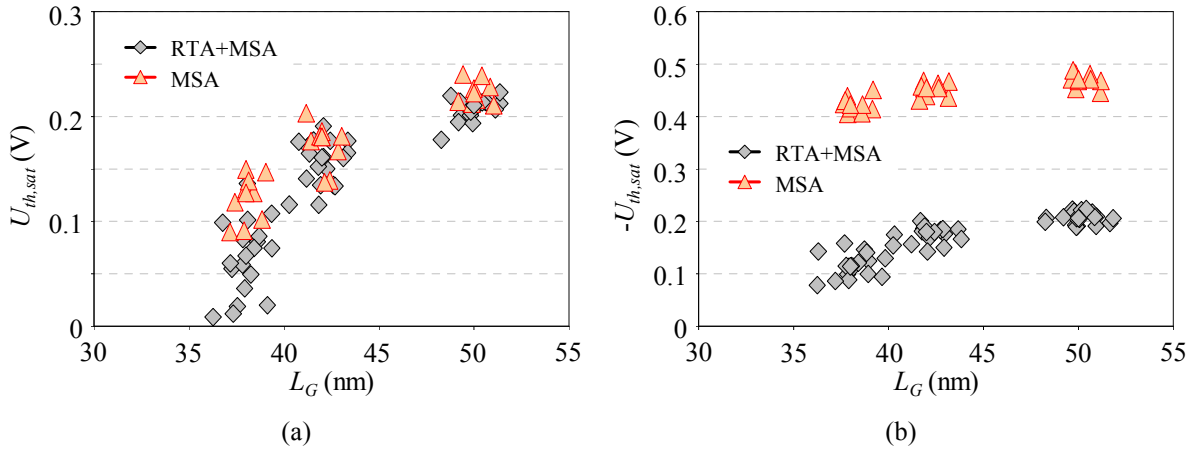


Bild 6.6: Sättigungsschwellspannung $U_{th,sat}$ in Abhängigkeit von der Gatelänge L_G für unterschiedliche Ausheilbedingungen für (a) n- und (b) p-MOSFETs

6.2.2 Einfluss der SDE-Implantation

Um weitere Untersuchungen zum diffusionsarmen Transistor durchführen zu können, war eine gleiche Überlappung der SDE-Gebiete unter dem Gate zumindest beim p-MOSFET angestrebt. Aus diesem Grund wurde in weiterführenden Experimenten die Implantationsenergie und -dosis der SDE-Gebiete entsprechend Tabelle 6.1 variiert. Die Implantationen der Halo-Gebiete wurden nicht verändert.

Tabelle 6.1 Implantationsparameter der SDE-Gebiete zur Untersuchung des Einflusses von Implantationsenergie und -dosis auf die Leistungsfähigkeit von diffusionsarmen n- und p-MOSFETs. Angegeben sind Dotand, Energie und Dosis.

Experiment	n-MOSFET	p-MOSFET
SDE1	As, 3 keV, $2.0 \cdot 10^{15}$ cm^{-2}	BF_2, 3 keV, $0.5 \cdot 10^{15}$ cm^{-2}
SDE2	As, 3 keV, $3.0 \cdot 10^{15}$ cm^{-2}	BF_2, 3 keV, $2.0 \cdot 10^{15}$ cm^{-2}
SDE3	As, 1 keV, $2.0 \cdot 10^{15}$ cm^{-2}	BF_2, 1 keV, $0.5 \cdot 10^{15}$ cm^{-2}
SDE4	As, 1 keV, $3.0 \cdot 10^{15}$ cm^{-2}	BF_2, 1 keV, $2.0 \cdot 10^{15}$ cm^{-2}

Bild 6.7 zeigt den Einfluss der verschiedenen SDE-Implantationen auf den Schichtwiderstand R_{SDE} der SDE-Gebiete. Bei n-MOSFETs wird der Widerstand hauptsächlich durch die Implantationsenergie bestimmt. Durch die verringerte Energie erhöht sich der Schichtwiderstand R_{SDE} aufgrund der kleineren Tiefe der pn-Übergänge um 85.7 %. Eine Erhöhung der Implantationsdosis von $2.0 \cdot 10^{15}$ cm^{-2} auf $3.0 \cdot 10^{15}$ cm^{-2} hat bei beiden Implantationsenergien nur einen sehr geringen Einfluss auf den Schichtwiderstand R_{SDE}. P-MOSFETs zeigen zumindest bei einer Verringerung der Implantationsenergie eine vergleichbare Reaktion. Durch die kleinere Energie verdoppelt (101.4 %) sich der Schichtwiderstand R_{SDE}, was wie beim n-MOSFET vorrangig auf die kleinere Tiefe der pn-Übergänge zurückzuführen ist. Durch die Erhöhung der Implantationsdosis von $0.5 \cdot 10^{15}$ cm^{-2} auf $2.0 \cdot 10^{15}$ cm^{-2} verringert sich der Schichtwiderstand R_{SDE} in Abhängigkeit von der Implantationsenergie um 52.5 % (3 keV) bzw. um 14.2 % (1 keV).

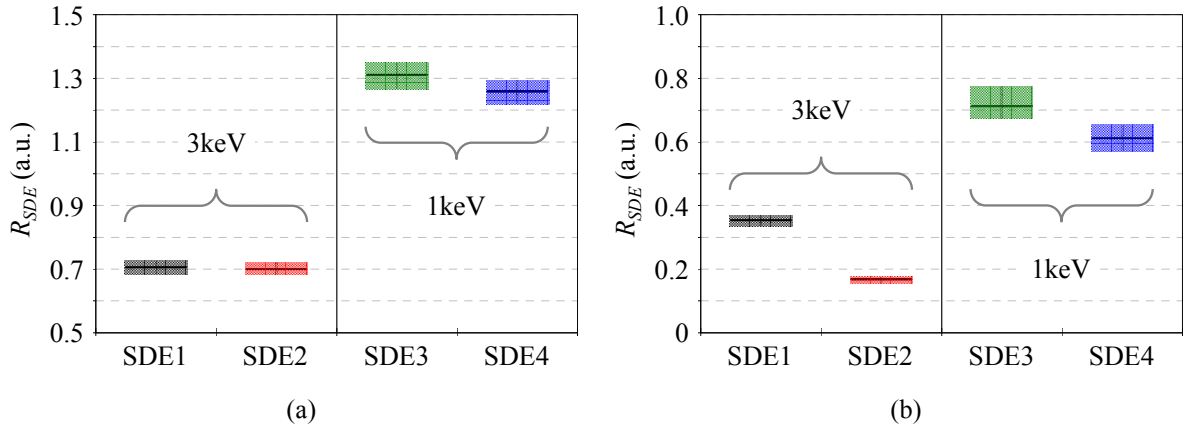


Bild 6.7: Einfluss der SDE-Implantationsenergie und -dosis auf den Schichtwiderstand R_{SDE} der SDE-Gebiete für (a) n- und (b) p-MOSFETs

Bild 6.8 stellt für die betrachteten Implantationsbedingungen die MILLERkapazität für n- und p-MOSFETs dar. Bei n-MOSFETs ist vor allem eine starke Abhängigkeit von der Implantationsenergie zu erkennen. Die geringere Energie hat eine etwa 30 aF kleinere MILLERkapazität C_{ov} zur Folge. Die Implantationsdosis hat dagegen nur einen geringen Einfluss auf die Überlappung der SDE-Gebiete unter dem Gate, wie man anhand der sich nur minimal verändernden MILLERkapazität C_{ov} erkennen kann. P-MOSFETs reagieren nicht so stark auf die Verringerung der Implantationsenergie wie n-MOSFETs, was auf die unzureichende Überlappung zurückgeführt werden kann. Die Erhöhung der Implantationsdosis bei einer Implantationsenergie von 1 keV vergrößert ebenfalls nicht die Überlappung der SDE-Gebiete unter dem Gate. Lediglich eine Implantationsenergie von 3 keV mit einer Dosis von $2.0 \cdot 10^{15}$ cm^{-2} scheint zumindest hinsichtlich der MILLERkapazität C_{ov} eine ausreichend große Überlappung zu gewährleisten.

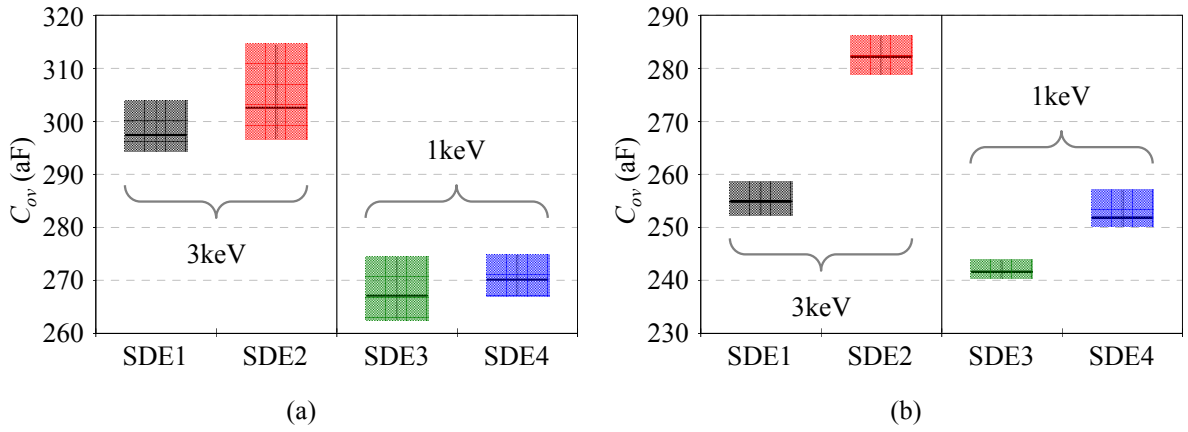


Bild 6.8: Einfluss der SDE-Implantationsenergie und -dosis auf die MILLERkapazität C_{ov} für (a) n- und (b) p-MOSFETs

Bild 6.9 zeigt die Auswirkungen der verschiedenen SDE-Implantationsenergien und -dosen auf die Universalkurve beider Transistortypen. Bei n-MOSFETs lässt sich sowohl eine Abhängigkeit von der Implantationsenergie als auch von der Dosis feststellen. Durch die Erhöhung der Implantationsdosis von $2.0 \cdot 10^{15}$ cm^{-2} auf $3.0 \cdot 10^{15}$ cm^{-2} verbessert sich unabhängig von der Implantationsenergie der Sättigungs-Drainstrom $I_{D,sat}$.

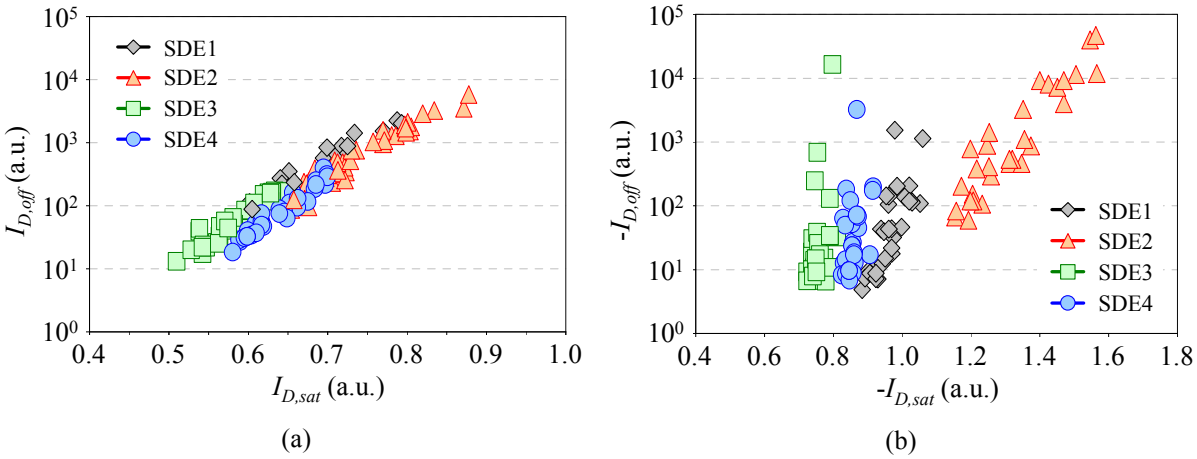


Bild 6.9: Einfluss der SDE-Implantationsenergie und -dosis auf die Universalkurve für (a) n- und (b) p-MOSFETs

Dieses Verhalten lässt sich nur durch einen verringerten Anschlusswiderstand erklären, da sich der Schichtwiderstand R_{SDE} der SDE-Gebiete und die MILLERkapazität C_{ov} nur geringfügig verändern. Durch die Variation der Implantationsenergie erfolgt hauptsächlich eine Veränderung der metallurgischen Gatelänge L_{met}, was an der Verschiebung entlang der Universalkurve zu erkennen ist. Die Universalkurven der p-MOSFETs korrelieren sehr gut mit der MILLERkapazität C_{ov} nach Bild 6.8b. Für sehr geringe MILLERkapazitäten C_{ov}, die durch eine unzureichende Überlappung der SDE-Gebiete unter dem Gate hervorgerufen werden, ist die Leistungsfähigkeit der entsprechenden Transistoren stark verschlechtert. Für eine genügend große Überlappung hingegen lässt sich ein deutlich größerer Sättigungs-Drainstrom $I_{D,sat}$ feststellen (SDE2).

Bild 6.10 stellt für n- und p-MOSFETs die Sättigungsschwellspannung $U_{th,sat}$ in Abhängigkeit von der Gatelänge L_G für die verschiedenen SDE-Implantationen dar. Bei n-MOSFETs ist für die geringere Implantationsenergie eine um etwa 60 mV höhere Schwellspannung $U_{th,sat}$ erkennbar. Ursache dafür ist die mit der geringeren Implantationsenergie einhergehende größere metallurgische Gatelänge L_{met}, wie man anhand der geringeren MILLERkapazität C_{ov} in Bild 6.8a feststellen kann. Der Verlauf der Schwellspannung $U_{th,sat}$ über der Gatelänge L_G ändert sich für die verschiedenen Implantationsenergien nur geringfügig. Bei p-MOSFETs sind, wie bereits die Bilder 6.8b und 6.9b zeigten, geringe Implantationsenergien sowie hohe Implantationsenergien in Verbindung mit geringen Implantationsdosen ungünstig für das Transistorverhalten, so dass diese Varianten im Weiteren nicht mehr betrachtet werden.

Wegen des verbesserten Kurzkanalverhaltens sowie der größeren Leistungsfähigkeit werden als Ausgangspunkt für die weiteren Untersuchungen zum diffusionsarmen Transistor, wenn nicht explizit angegeben, die in Tabelle 6.2 dargestellten Parameter für die SDE-Implantationen verwendet.

Tabelle 6.2 Angepasste Implantationsparameter der SDE-Gebiete für diffusionsarme n- und p-MOSFETs. Angegeben sind Dotand, Energie und Dosis.

Implantation	n-MOSFET	p-MOSFET
SDE	As, 1 keV, $3.0 \cdot 10^{15}$ cm^{-2}	BF_2, 3 keV, $2.0 \cdot 10^{15}$ cm^{-2}

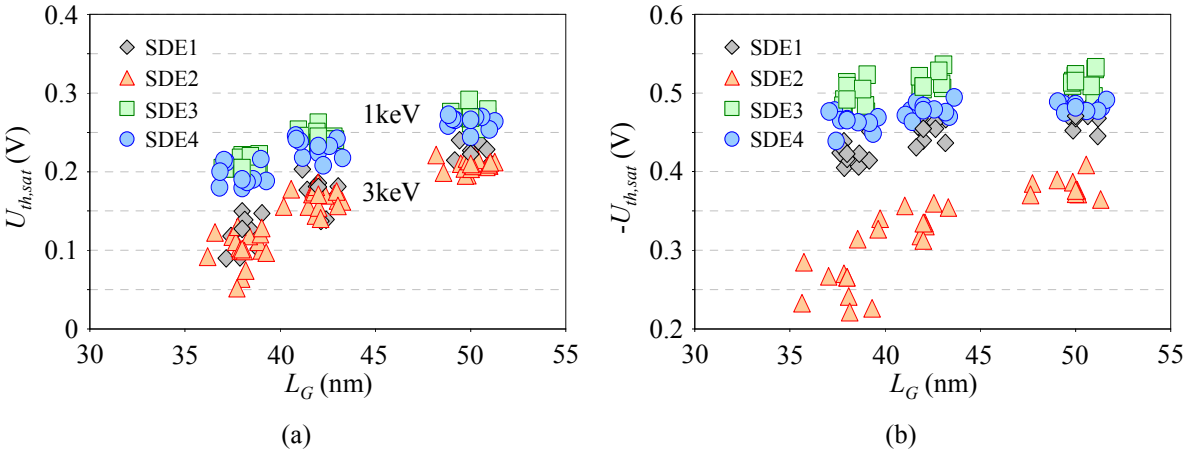


Bild 6.10: Einfluss der SDE-Implantationsenergie und -dosis auf die Sättigungsschwellspannung $U_{th,sat}$ in Abhängigkeit von der Gatelänge L_G für (a) n- und (b) p-MOSFETs

6.2.3 Diskussion

Die Bilder 6.11 bis 6.13 geben einen Überblick auf die wesentlichen elektrischen Parameter für beide Transistortypen nach Anpassung der SDE-Implantationen gemäß Tabelle 6.2 für die unterschiedlichen Ausheilregime. Sowohl bei n- als auch bei p-MOSFETs ist trotz einer Optimierung der SDE-Implantationen eine starke Verringerung in der Leistungsfähigkeit der Transistoren zu erkennen, wenn nur eine Kurzzeitausheilung, aber keine RTA erfolgt (Bild 6.11).

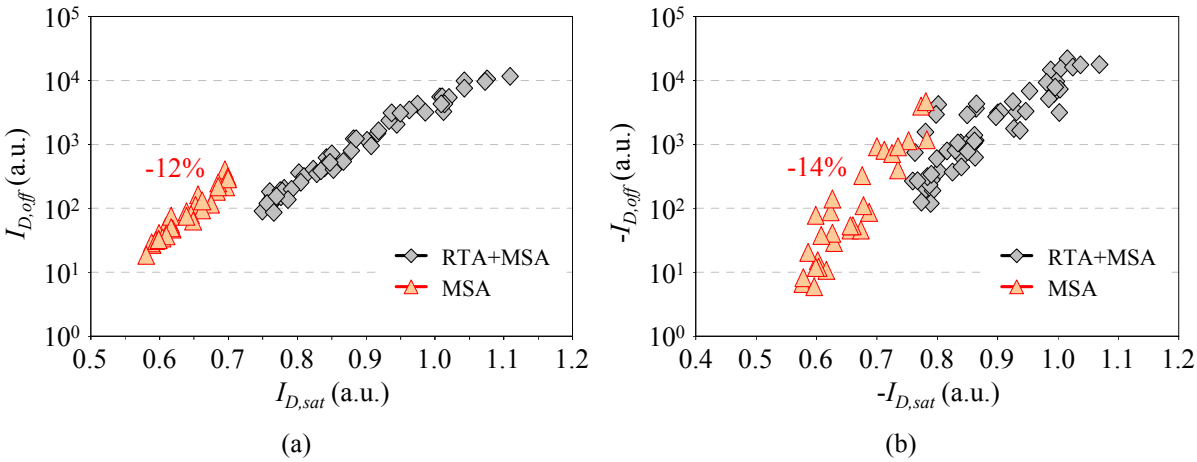


Bild 6.11: Einfluss des Ausheilregimes auf die Universalkurve für (a) n- und (b) p-MOSFETs für die Implantationsparameter von Tabelle 6.2

Wie man aus den Bildern 6.7 und 6.8 von Abschnitt 6.2.2 ableiten kann, ist dies zumindest bei n-MOSFETs nicht nur auf den erhöhten Schichtwiderstand R_{SDE} der SDE-Gebiete oder die verringerte MILLERkapazität C_{ov} zurückzuführen. Vor allem die p-MOSFETs zeigen eine starke Verringerung des Sättigungs-Drainstroms $I_{D,sat}$, obwohl sich der Schichtwiderstand R_{SDE} der SDE-Gebiete ohne RTA nach der Optimierung der SDE-Implantationen sogar verringert (Bild 6.12a). Die kleinere Überlappung der p-MOSFETs nur mit Kurzzeitausheilung (Bild 6.12b) kann zwar durch eine höhere Implantationsenergie bzw. -dosis oder eine gewinkelte Implantation der SDE-Gebiete angepasst werden, wodurch aber das im Vergleich zum RTA-Prozess bereits schlechtere Kurzkanalverhalten weiter beeinträchtigt wird (Bild 6.13b).

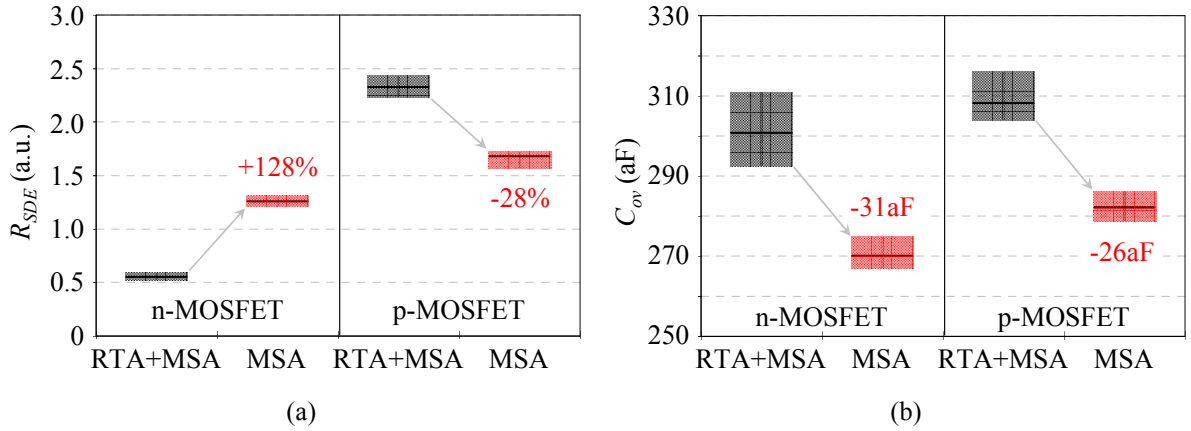


Bild 6.12: Einfluss des Ausheilregimes auf (a) den Schichtwiderstand R_{SDE} der SDE-Gebiete und (b) die MILLERkapazität C_{ov} für n- und p-MOSFETs für die Implantationsparameter von Tabelle 6.2

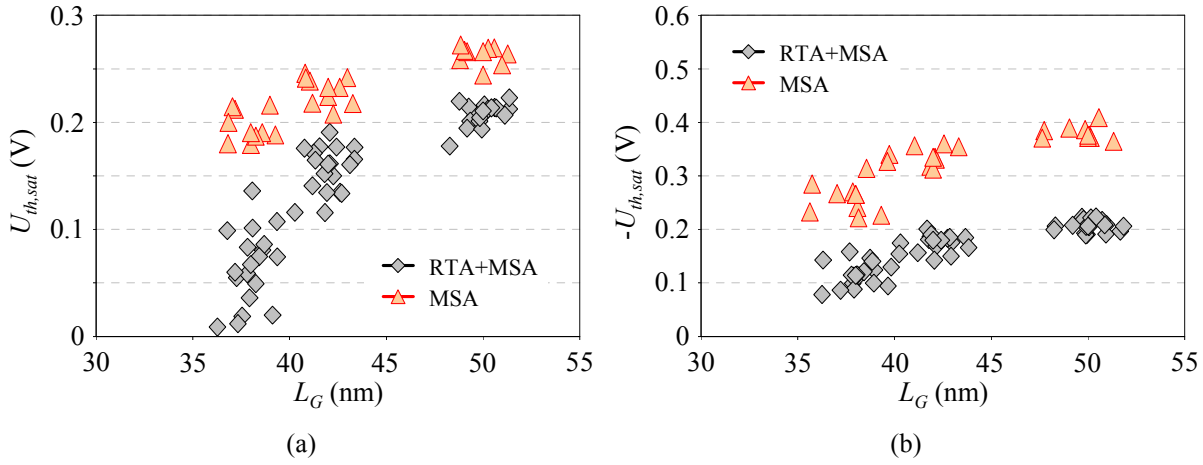


Bild 6.13: Einfluss des Ausheilregimes auf die Sättigungsschwellspannung $U_{th,sat}$ in Abhängigkeit von der Gatelänge L_G für (a) n- und (b) p-MOSFETs für die Implantationsparameter von Tabelle 6.2

Die Verringerung der Leistungsfähigkeit der Kurzkanaltransistoren lässt sich auch nicht mit der geringeren Langkanal-Ladungsträgerbeweglichkeit μ erklären (Bild 6.14), welche auf die erhöhte Grenzflächenhaftzentrendichte D_{it} zurückgeführt werden kann (Bild 6.15). Es müssen demnach noch weitere Effekte auftreten, welche die Leistungsfähigkeit der Transistoren ohne RTA stark verringern.

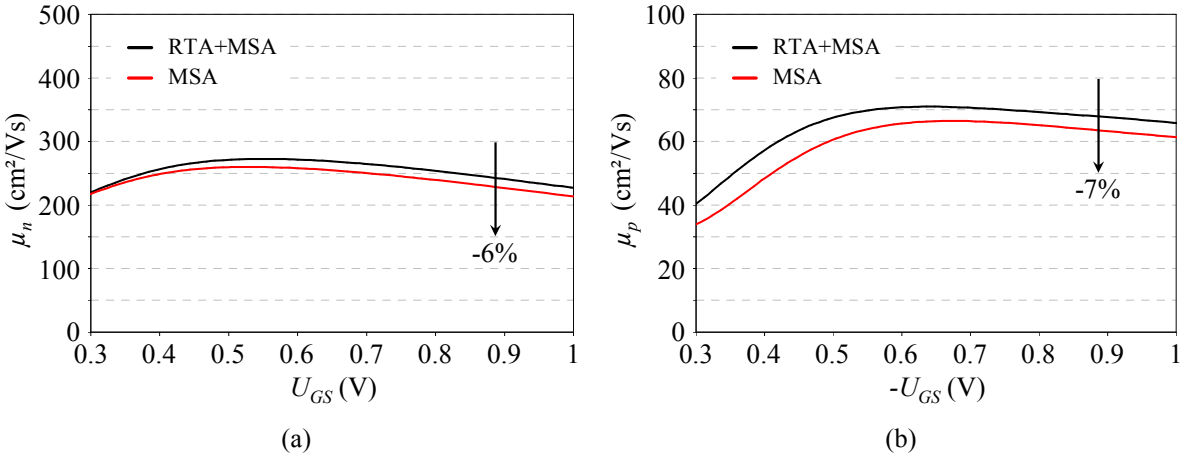


Bild 6.14: Einfluss des Ausheilregimes auf die Langkanal-Ladungsträgerbeweglichkeit μ in Abhängigkeit von der Gate-Source-Spannung U_{GS} für (a) n- und (b) p-MOSFETs für die Implantationsparameter von Tabelle 6.2

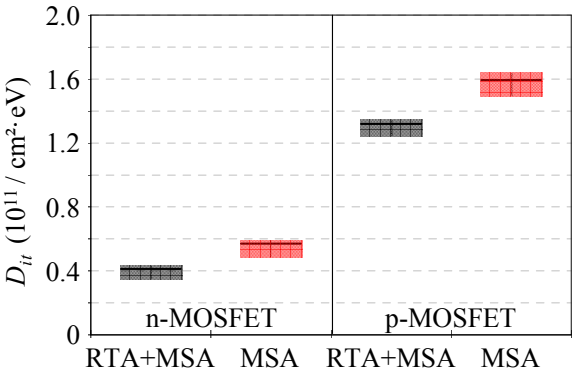


Bild 6.15: Einfluss des Ausheilregimes auf die Grenzflächenhaftzentrendichte D_{it} für n- und p-MOSFETs für die Implantationsparameter von Tabelle 6.2

Bild 6.16 zeigt mögliche Ursachen für die verringerte Leistungsfähigkeit beider Transistortypen:

1. Aufgrund des geringen thermischen Budgets der Kurzzeitausheilung können nicht vollständig ausgeheilte Defekte am Übergang von den Erweiterungsgebieten zum Kanal die Ladungsträgerbeweglichkeit beeinflussen.
2. Durch die verringerte Diffusion und die damit verbundene kleinere Tiefe der pn-Übergänge kann es zu einer Eingrenzung des Stromflusses kommen.
3. Bei den Transistoren mit Kurzzeitausheilung, aber ohne RTA, wird das Dotierungsprofil am Übergang vom Silizium zum Silizid derart verändert, dass sich der Übergangswiderstand erhöht.

Alle drei Möglichkeiten haben einzeln eine Erhöhung des externen Widerstands zur Folge und können somit zu dem beobachteten Verhalten führen.

Bild 6.17 veranschaulicht das Problem der Eingrenzung des Stromflusses. Bild 6.17a zeigt den Stromfluss eines Transistors, der einem RTA-Prozess unterzogen war. Hier treten die Ladungsträger senkrecht zum Source-Kontakt in das Silizium ein und durchlaufen einen Bogen in Richtung zum Drain. Anschließend wird das Gebiet unter dem Spacer, also das SDE-Gebiet annähernd parallel durchquert. Im Überlappgebiet verjüngt sich der Strom trichterförmig bis zum Eintritt in den Kanal auf wenige Nanometer, wo dann schließlich die Ladungsträger durch die Raumladungszone in die Inversionsschicht des inneren Transistors fließen. Im Gegensatz dazu treten beim Transistor ohne RTA die Ladungsträger aufgrund der verringerten Tiefe des pn-Übergangs der SDE-Gebiete direkt aus dem Source-Kontakt in das Gebiet unter dem Spacer ein (Bild 6.17b), wodurch sich der externe Widerstand R_{ext} des Transistors erhöht.

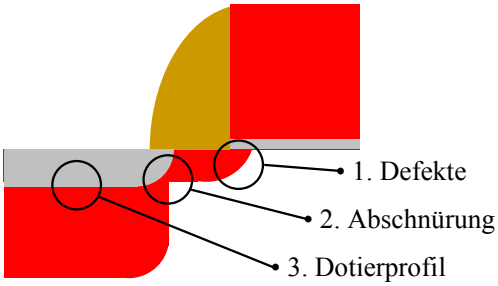


Bild 6.16: Mögliche Ursachen für die verringerte Leistungsfähigkeit von diffusionsarmen Transistoren

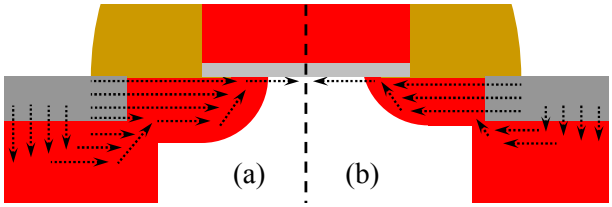


Bild 6.17: Unterschiede im Stromfluss eines Transistors (a) mit einem RTA-Prozess und (b) mit MSA

Ein weiteres Problem ist vor allem das bei p-MOSFETs beobachtete schlechtere Kurzkanalverhalten der Transistoren ohne RTA im Vergleich zum Transistor mit RTA (Bild 6.13). Das Problem dabei ist die bereits sehr viel höhere Sättigungsschwellspannung. Um die Schwellspannung anzupassen, könnte man zwar die Halodotierung verringern oder die Überlappung der SDE-Gebiete vergrößern, was aber in beiden Fällen zu einer weiteren Verschlechterung des Kurzkanalverhaltens führen würde.

Zusammenfassend ist festzustellen, dass unter Verwendung der Kurzzeitausheilung als alleiniger Ausheilschritt erhebliche Beeinträchtigungen in der Leistungsfähigkeit und beim Kurzkanalverhalten zu verzeichnen sind. Eine Ursache hierfür ist das stark veränderte Dotierungsprofil im Vergleich zu einem Transistor mit einem RTA-Schritt.

Bei einem RTA-Prozess wird die finale Lage der eingebrachten Dotanden durch die Implantation sowie die Diffusion während der RTA bestimmt, wohingegen bei einem diffusionsarmen Prozess die Definition des Dotierungsprofils vorwiegend durch die Implantation erfolgt. Aus diesem Grund ist eine Umgestaltung des Dotierungsprofils notwendig, wenn diffusionsarme Transistoren angestrebt werden.

6.3 Variation der Implantationsparameter

Im Folgenden soll durch Experimente und Simulationen, der Einfluss der Parameter der Halo-Implantation, alternativer Spezies für die Implantationen der SDE-Gebiete sowie zusätzlicher SD-Implantationen auf das elektrische Verhalten diffusionsarmer Transistoren untersucht werden.

6.3.1 Halo-Implantation

Wie man anhand des vorherigen Abschnitts sehen konnte, ist das Kurzkanalverhalten unter Verwendung der Kurzzeitausheilung als alleiniger Ausheilschritt zumindest beim n-MOSFET unverändert, beim p-MOSFET verschlechtert sich das Kurzkanalverhalten nur mit Kurzzeitausheilung. Diese Ergebnisse widersprechen damit den Annahmen nach Abschnitt 2.3.1, dass sich mit der Reduzierung der Tiefe der pn-Übergänge auch das Kurzkanalverhalten verbessert. Eine mögliche Ursache für dieses Verhalten kann an der Lage der Halo-Implantationen liegen, da bei den bisherigen Betrachtungen nur die Energie der Implantationen der SDE-Gebiete ohne eine Anpassung der Parameter der Halo-Implantationen verringert wurde. Bild 6.18 zeigt in diesem Zusammenhang die berechnete Sättigungsschwellspannung $U_{th,sat}$ beider Transistortypen in Abhängigkeit von der Gatelänge L_G für unterschiedliche Halo-Implantationsenergien. Dabei ist das deutlich verbesserte Kurzkanalverhalten mit kleineren Implantationsenergien für n- als auch für p-MOSFETs zu erkennen, da sich das Konzentrationsmaximum mit kleiner werdender Implantationsenergie in Richtung Gateoxid verschiebt. Bild 6.19 stellt diesen Sachverhalt beispielhaft für die größte und kleinste Implantationsenergie dar. Weiterhin wird dadurch die Dotierstoffkonzentration im Allgemeinen verringert. Dies ist vor allem im Kanalgebiet von Vorteil, da sich die Anzahl an möglichen Störzentren verringert, was wiederum eine erhöhte Ladungsträgerbeweglichkeit im Kanal zur Folge hat.

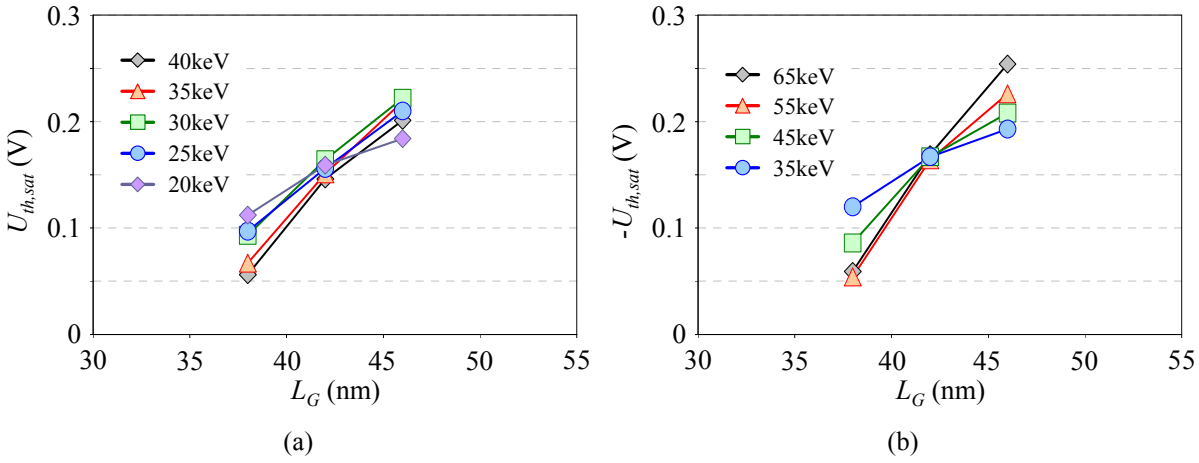


Bild 6.18: Berechnete Sättigungsschwellspannung $U_{th,sat}$ in Abhängigkeit von der Gatelänge L_G (a) für n-MOSFETs für verschiedene BF_2-Halo-Implantationsenergien und (b) für p-MOSFETs für verschiedene Arsen-Halo-Implantationsenergien bei einem Implantationswinkel von 25 °

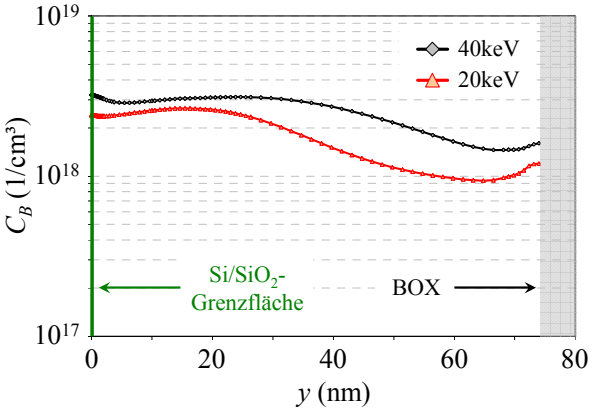


Bild 6.19: Berechnetes Dotierungsprofil für eine BF_2-Halo-Implantation mit einer Energie von 20 keV und 40 keV bei einem Implantationswinkel von 25 ° (vertikaler Schnitt in der Gatemitte, ausgehend von der Si/SiO_2-Grenzfläche in Richtung Substrat)

Der für das verbesserte Kurzkanalverhalten verantwortliche Unterschied zwischen beiden Implantationsenergien ist jedoch am berechneten lateralen Dotierungsprofil zu erkennen (Bild 6.20). Im Gegensatz zur Halo-Implantation mit der höheren Implantationsenergie von 40 keV, die ein relativ homogenes Dotierungsprofil in Kanalrichtung mit einem weniger stark ausgeprägten Maximum in Kanalmitte hat, entstehen durch die 20 keV-Halo-Implantationen zwei Maxima direkt vor den Implantationen der SDE-Gebiete. Dadurch werden die Erweiterungsgebiete in Kanalrichtung besser von der Gegendotierung der Halo-Implantation umschlossen, was wiederum einer Ausbreitung der Raumladungszone der SDE-Gebiete in Kanalrichtung entgegenwirkt. Gleichzeitig wird die Dotierstoffkonzentration im Kanalgebiet im Vergleich zur 40 keV-Halo-Implantation abgesenkt, was ein weiterer Vorteil ist, da dadurch die Ladungsträgerbeweglichkeit im Kanal erhöht wird. Diese Ergebnisse zeigen, dass vor allem die Ausbreitung der Raumladungszone der SDE-Gebiete in Kanalrichtung das Kurzkanalverhalten beeinflussen und weniger in vertikaler Richtung.

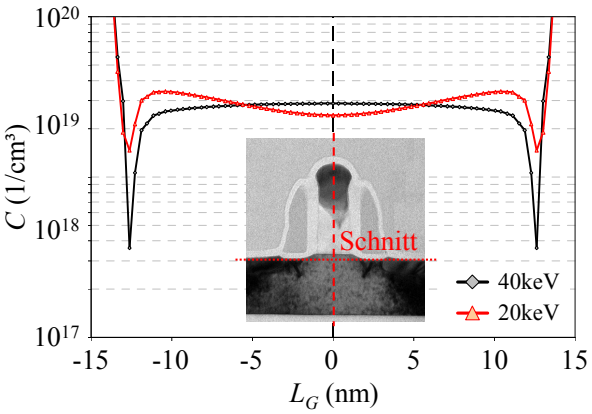


Bild 6.20: Berechnetes Dotierungsprofil für eine BF_2-Halo-Implantation mit einer Energie von 20 keV und 40 keV bei einem Implantationswinkel von 25 ° (horizontaler Schnitt 5 nm unter der Si/SiO_2-Grenzfläche)

Die Bilder 6.21 und 6.22 zeigen den experimentell beobachteten Einfluss der Halo-Implantationsenergie auf verschiedene Parameter der n-MOSFETs. Die Parameter der SDE-Implantationen sowie der Winkel und die Dosis der Halo-Implantationen sind für die betrachteten Energien gleich. Betrachtet man zunächst nur den Verlauf der Sättigungsschwellspannung $U_{th,sat}$ in Abhängigkeit von der Gatelänge L_G für die unterschiedlichen Halo-Implantationsenergien (Bild 6.21a), so ist keine wesentliche Veränderung im Kurzkanalverhalten zu erkennen. Aufgrund der gleichen Halo-Implantationsdosis und der Variation der Implantationsenergie ist nur eine Verschiebung der Schwellspannung $U_{th,sat}$ festzustellen.

Da aber mit sinkender Implantationsenergie die MILLERkapazität C_{ov} größer wird (Bild 6.21b), was wiederum mit einer Verringerung der metallurgischen Gatelänge L_{met} einhergeht, kann man unter Einbeziehung der berechneten Ergebnisse ein verbessertes Kurzkanalverhalten mit sinkender Halo-Implantationsenergie ableiten.

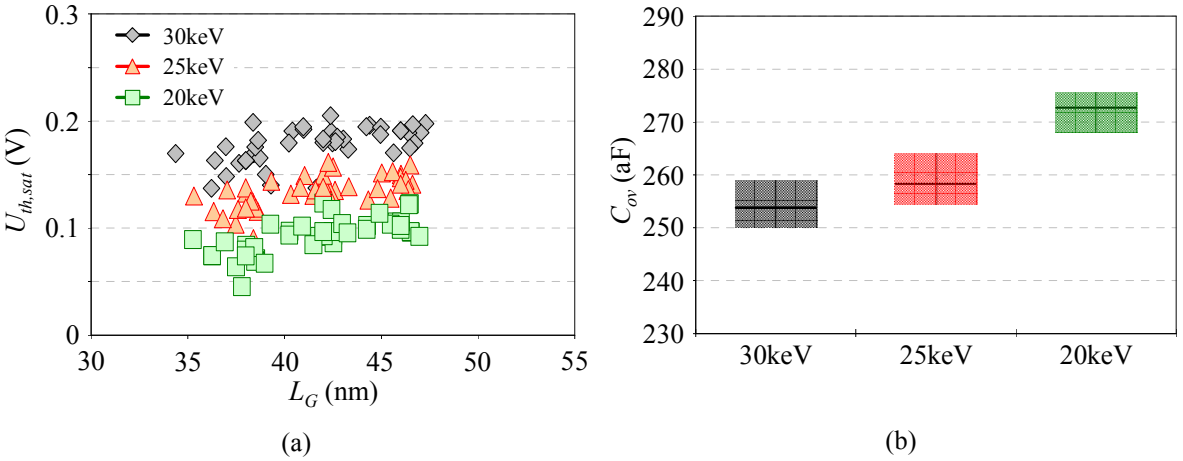


Bild 6.21: Einfluss der Halo-Implantationsenergie bei einem Implantationswinkel von 25 ° auf (a) die Sättigungsschwellspannung $U_{th,sat}$ in Abhängigkeit von der Gatelänge L_G und (b) die MILLERkapazität des n-MOSFETs

Neben der veränderten MILLERkapazität C_{ov} ist ebenfalls eine starke Variation der Sperrschichtkapazität C_j zwischen Source bzw. Drain und Wanne in Abhängigkeit von der verwendeten Halo-Implantationsenergie zu erkennen (Bild 6.22). Infolge der Verschiebung des Konzentrationsmaximums mit sinkender Implantationsenergie in Richtung Waferoberfläche (Bild 6.19) und dem damit einhergehenden Absinken der Dotierstoffkonzentration an der SOI/BOX-Grenzfläche kommt es zur einer stärkeren Ausbreitung der Raumladungszone ausgehend von den SD-Gebieten, was an der mit absinkender Energie kleiner werdender Sperrschichtkapazität erkennbar ist. Ohne entsprechende Gegenmaßnahmen kann es zum Durchgriff der Raumladungszonen kommen, wodurch sich alternative Strompfade ausbilden. Als Lösung für dieses Problem bietet sich z.B. eine Verringerung der Implantationsdosis der SD-Gebiete oder eine Erhöhung der Dosis der Wannen-Implantation an. Ersteres birgt jedoch die Gefahr eines verringerten Übergangswiderstands $R_{Silizid}$ vom Silizium zum Silizid. Weiterhin ist im Bezug auf SOI-Wafer zu beachten, dass die SD-Gebiete auf dem vergrabenen Oxid aufsetzen und so der Vorteil einer geringeren Sperrschichtkapazität C_j von SOI- im Vergleich zu Bulk-Wafern noch gegeben ist.

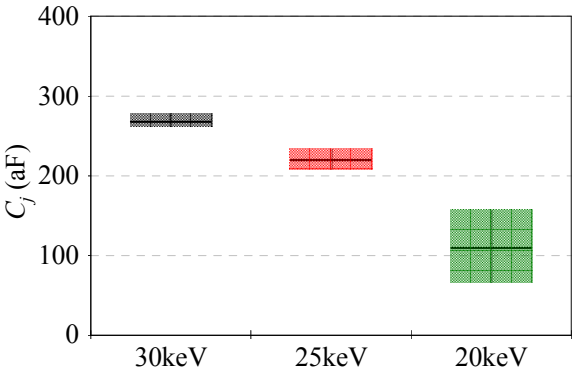


Bild 6.22: Einfluss von unterschiedliche Halo-Implantationsenergien auf die Sperrschichtkapazität C_j zwischen Source bzw. Drain und Wanne von n-MOSFETs bei einem Implantationswinkel von 25 °

6.3.2 SDE-Implantation

Wird die Ausheilung der Transistoren nur über eine Kurzzeitausheilung angestrebt und dabei auf die RTA verzichtet, können prinzipiell auch solche Implantationsspezies verwendet werden, die unter Verwendung der RTA eine zu starke Diffusion aufweisen und somit das Transistorverhalten verschlechtern. Aus diesem Grund werden im Folgenden alternative Implantationsspezies für die SDE-Gebiete von n- und p-MOSFETs untersucht.

a) n-MOSFET

Aufgrund der geringeren Diffusion und der höchsten Festkörperlöslichkeit aller Dotanden für eine n-Dotierung ist im Allgemeinen Arsen der Dotand der Wahl beim n-MOSFET. Unter den veränderten Prozessbedingungen beim Transistor nur mit Kurzzeitausheilung besteht jedoch die Möglichkeit Phosphor als alternative Implantationsspezies für die SDE-Gebiete zu verwenden, obwohl dessen Festkörperlöslichkeit nicht die Werte von Arsen erreicht (Bild 2.20). Der Vorteil von Phosphor ist dessen größere Streuung während des Implantationsprozesses. Um eine genügend große Überlappung der SDE-Gebiete unter das Gate zu gewährleisten, ist bei der Arsen-Implantation eine relativ hohe Implantationsdosis notwendig, die wiederum die Gefahr einer Clusterbildung birgt. Phosphor kann demgegenüber mit einer geringeren Dosis implantiert werden, um eine entsprechende Überlappung sicherzustellen. Der Einfluss einer Phosphordotierung auf die Leistungsfähigkeit von n-MOSFETs wurde anhand von Experimenten untersucht, bei denen die Energie und Dosis der SDE-Implantation gemäß Tabelle 6.3 variiert wurde. Bild 6.23 zeigt dazu die berechneten Profile für die verschiedenen SDE-Implantationen. Alle anderen Prozessparameter blieben unverändert. Als Referenz für diese Untersuchungen dienen Transistoren mit einer Arsen-SDE-Implantation.

Tabelle 6.3 Implantationsparameter der SDE-Gebiete zur Untersuchung des Einflusses von Arsen- bzw. Phosphorimplantationen auf die Leistungsfähigkeit von diffusionsarmen n-MOSFETs. Angegeben sind Dotand, Energie und Dosis.

Experiment	As	P1	P2	P3	P4
SDE-Implantation	As, 1 keV, $3.0\cdot10^{15}$ cm^{-2}	P, 0.5 keV, $1.5\cdot10^{15}$ cm^{-2}	P, 0.5 keV, $2.0\cdot10^{15}$ cm^{-2}	P, 0.7 keV, $1.5\cdot10^{15}$ cm^{-2}	P, 0.7 keV, $2.0\cdot10^{15}$ cm^{-2}

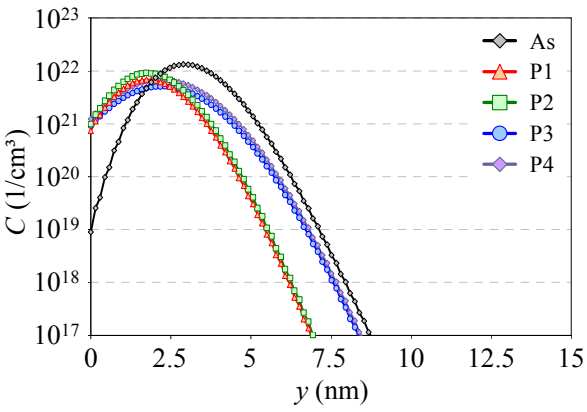


Bild 6.23: Berechnetes Implantationsprofil für die Implantationen von Arsen und Phosphor in amorphes Silizium. P1…P4 sind die Implantationsparameter gemäß Tabelle 6.3

Bild 6.24a zeigt zunächst den Einfluss der SDE-Implantationsparameter auf die MILLERkapazität C_{ov}. Entsprechend den Erwartungen variiert die MILLERkapazität C_{ov} in Abhängigkeit sowohl von der Energie als auch von der Dosis, wodurch die notwendige Überlappung der SDE-Gebiete unter dem Gate durch die entsprechenden Implantationsbedingungen angepasst werden kann.

Um unter den gegebenen Prozessbedingungen eine der Arsenimplantation vergleichbare MILLERkapazität C_{ov} zu erreichen, kann man entsprechend der Ergebnisse entweder:

- eine Phosphorimplantation mit einer Energie von 0.5 keV und einer Dosis von ungefähr $2.23 \cdot 10^{15}$ cm^{-2} oder
- eine Phosphorimplantation mit einer Energie von 0.7 keV und einer Dosis von etwa $1.81 \cdot 10^{15}$ cm^{-2} verwenden.

Der Schichtwiderstand R_{SDE} der SDE-Gebiete verändert sich vergleichbar zur MILLERkapazität C_{ov} abhängig von den verwendeten Parametern der SDE-Implantation (Bild 6.24b). Obwohl beide genannten Phosphorimplantationsregime im Vergleich zur Arsenimplantation eine geringere Tiefe der pn-Übergänge verursachen (Bild 6.27), ergibt sich zumindest für eine Implantationsdosis von $2.0 \cdot 10^{15}$ cm^{-2} ein vergleichbarer bzw. sogar verringerter Schichtwiderstand R_{SDE}. Aufgrund der sehr hohen Arsendotierung, die für eine ausreichende Überlappung der SDE-Gebiete unter dem Gate notwendig ist, kommt es wahrscheinlich zu einer unzureichenden Auflösung von AsV-Clustern bzw. SiAs-Komplexen [73]. Diese wirken als Störstellen und verringern die Ladungsträgerbeweglichkeit innerhalb der Erweiterungsgebiete, wodurch wiederum der Schichtwiderstand R_{SDE} der SDE-Gebiete negativ beeinflusst wird. Aus diesem Grund ist bei einer Phosphorimplantation von 0.5 keV nur eine Dosis von etwa $2.0 \cdot 10^{15}$ cm^{-2} notwendig, um einen zur Arsenimplantation vergleichbaren Schichtwiderstand R_{SDE} zu erreichen. Bei einer weiteren Erhöhung der Phosphorimplantation auf 0.7 keV ist sogar eine noch geringere Dosis notwendig.

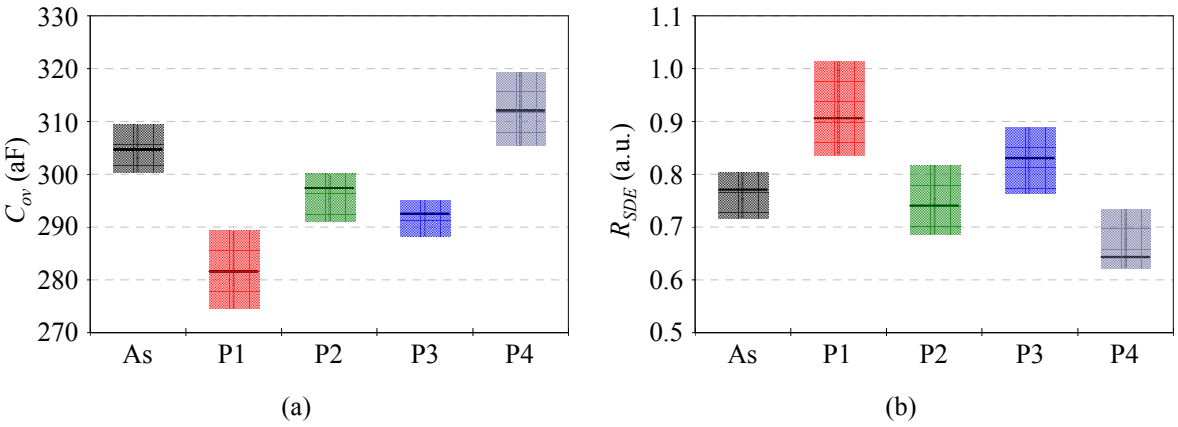


Bild 6.24: Einfluss der SDE-Implantationsparameter gemäß Tabelle 6.3 von Arsen- bzw. Phosphorimplantationen in amorphes Silizium auf (a) die MILLERkapazität C_{ov} und (b) den Schichtwiderstand R_{SDE} der SDE-Gebiete von n-MOSFETs

Bild 6.25 zeigt den Einfluss der SDE-Implantationsparameter auf die Universalkurve von n-MOSFETs. Dabei ist zu erkennen, dass die Leistungsfähigkeit in Abhängigkeit von den verwendeten Implantationsbedingungen sehr stark variiert. Betrachtet man dazu den Sättigungs-Drainstrom $I_{D,sat}$ bei einem konstanten Sperrstrom $I_{D,off}$, so ist eine deutliche Abhängigkeit des Sättigungs-Drainstroms $I_{D,sat}$ von der MILLERkapazität C_{ov} festzustellen (Bild 6.26). Das bedeutet, dass die Leistungsfähigkeit maßgeblich von der Überlappung der SDE-Gebiete unter dem Gate abhängt, wohingegen ein verbesserter Schichtwiderstand R_{SDE} nicht zwingend einen höheren Drainstrom I_D zur Folge haben muss. Als Ergebnis bleibt festzuhalten, dass für eine Phosphorimplantation mit angepassten Implantationsparametern ein zur Arsenimplantation vergleichbares Transistorverhalten erreicht werden kann.

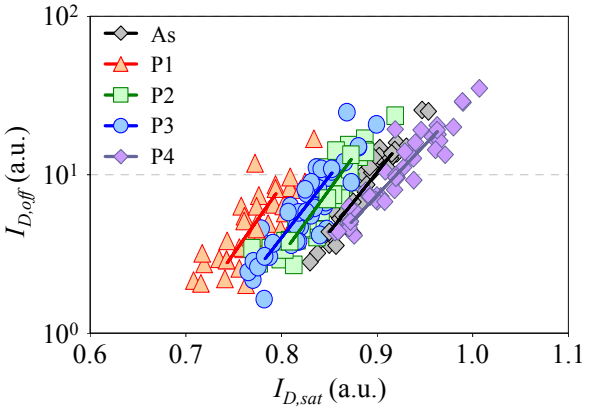


Bild 6.25: Einfluss der SDE-Implantationsparameter gemäß Tabelle 6.3 von Arsen- bzw. Phosphorimplantationen in amorphes Silizium auf die Universalkurve von n-MOSFETs

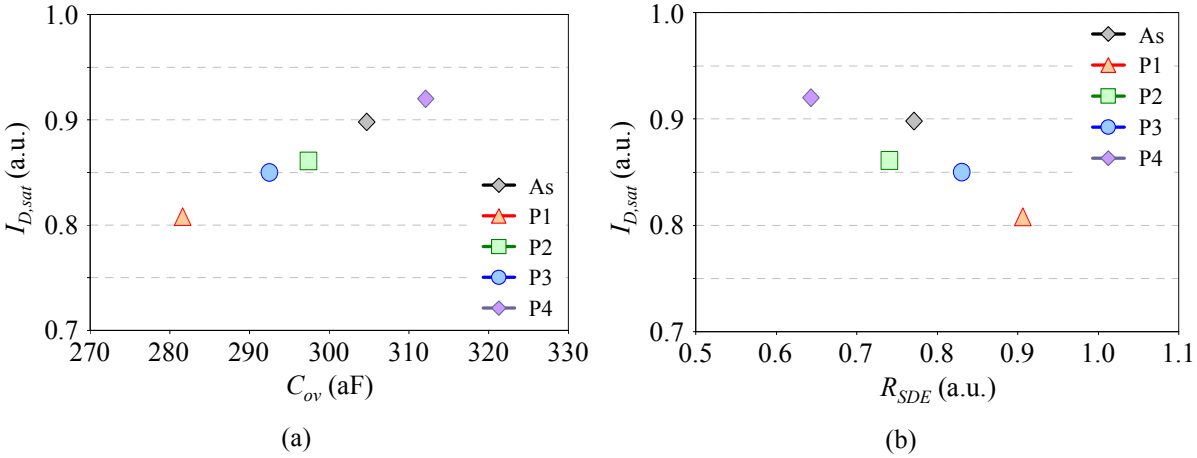


Bild 6.26: Einfluss der SDE-Implantationsparameter gemäß Tabelle 6.3 von Arsen- bzw. Phosphorimplantationen in amorphes Silizium auf den Sättigungs-Drainstrom $I_{D,sat}$ von n-MOSFETs in Abhängigkeit von (a) der MILLERkapazität C_{ov} und (b) dem Schichtwiderstand R_{SDE} der SDE-Gebiete bei $I_{D,off}$ = 100 nA/µm = konst.

b) p-MOSFET

Bor ist die erste Wahl als Dotand für eine p-Dotierung. Dabei kann Bor als Dotand in Verbindung mit Fluor als BF_2-Molekül implantiert werden. Weiterhin besteht die Möglichkeit, diesen Dotanden über eine Monomer-Borimplantation in den Halbleiter einzubringen. Beim herkömmlichen RTA-Prozess wird vorwiegend das BF_2-Molekül verwendet, da das gleichzeitig implantierte Fluor die Diffusion von Bor hemmt [120]. Unter den veränderten Prozessbedingungen bei p-MOSFETs ohne RTA besteht jedoch die Möglichkeit, Monomer-Bor als Implantationsspezies für die SDE-Gebiete zu verwenden. Um den Einfluss dieser Implantation auf die Leistungsfähigkeit des p-MOSFETs zu untersuchen, wurden Experimente mit unterschiedlichen Energien und Dosen der SDE-Implantation durchgeführt. Bild 6.27 zeigt dazu die berechneten Profile für die verschiedenen SDE-Implantationen entsprechend der Implantationsparameter nach Tabelle 6.4. Alle anderen Prozessparameter blieben unverändert. Als Referenz für diese Untersuchungen dienen Transistoren mit einer BF_2-Implantation.

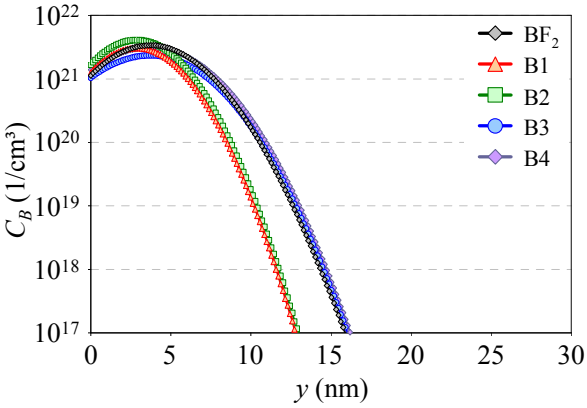


Bild 6.27: Berechnetes Implantationsprofil für die Implantationen von BF_2 und Bor in amorphes Silizium. B1…B4 sind die Implantationsparameter gemäß Tabelle 6.4

Tabelle 6.4 Implantationsparameter der SDE-Gebiete zur Untersuchung des Einflusses von BF_2- bzw. Borimplantationen auf die Leistungsfähigkeit von diffusionsarmen p-MOSFETs. Angegeben sind Dotand, Energie und Dosis.

Experiment	BF2	B1	B2	B3	B4
SDE-Implantation	BF_2, 3 keV, $2.0\cdot10^{15}$ cm^{-2}	B, 0.5 keV, $1.5\cdot10^{15}$ cm^{-2}	B, 0.5 keV, $2.0\cdot10^{15}$ cm^{-2}	B, 0.7 keV, $1.5\cdot10^{15}$ cm^{-2}	B, 0.7 keV, $2.0\cdot10^{15}$ cm^{-2}

Bild 6.28a zeigt die MILLERkapazität C_{ov} für die SDE-Implantationsparameter nach Tabelle 6.4. Um unter den gegebenen Prozessbedingungen eine der BF_2-Implantation vergleichbare MILLERkapazität C_{ov} zu erreichen, kann nur die Monomer-Borimplantation mit einer Energie von 0.5 keV verwendet werden. Die Dosis hat dabei nur einen geringen Einfluss auf die MILLERkapazität C_{ov} des p-MOSFETs. Der Schichtwiderstand R_{SDE} der SDE-Gebiete verändert sich vergleichbar zur MILLERkapazität C_{ov} abhängig von den verwendeten SDE-Implantationsparametern (Bild 6.28b). Obwohl die Monomer-Borimplantation mit einer Energie von 0.5 keV eine geringere Tiefe der pn-Übergänge im Vergleich zur BF_2-Implantation verursacht (Bild 6.27), ergibt sich für diese Implantationsenergie sogar mit der geringeren Dosis ein kleinerer Schichtwiderstand R_{SDE}. Die Ursache für den höheren Widerstand der BF_2-Implantation kann auf den Einfluss von Fluor auf die elektrische Aktivierung von Bor zurückgeführt werden. Durch das überlagerte Fluorprofil werden Boratome gebunden, was eine verringerte Aktivierung verursacht, aber auch flachere pn-Übergänge ermöglicht [121].

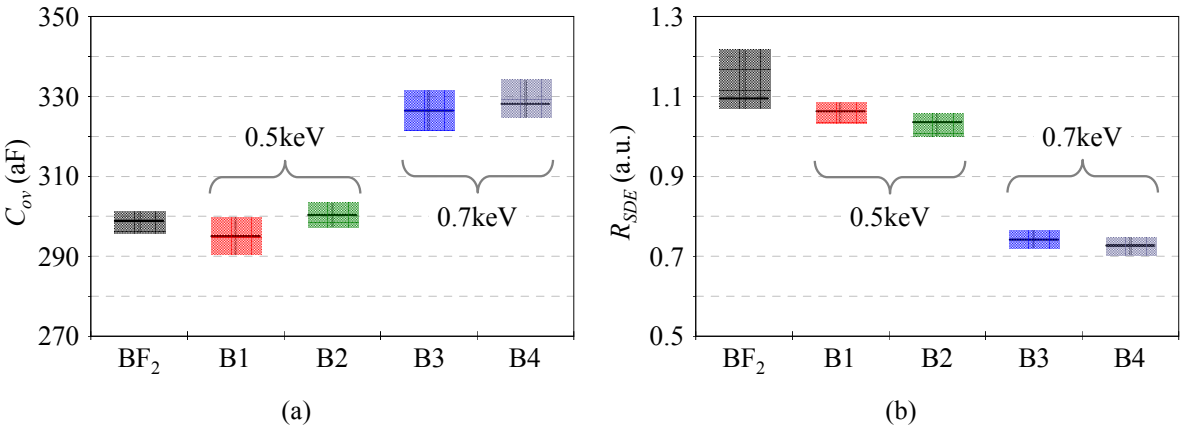


Bild 6.28: Einfluss der SDE-Implantationsparameter gemäß Tabelle 6.4 von BF_2- bzw. Borimplantationen in amorphes Silizium auf (a) die MILLERkapazität C_{ov} und (b) den Schichtwiderstand R_{SDE} der SDE-Gebiete von p-MOSFETs

Bild 6.29 zeigt den Einfluss der verschiedenen SDE-Implantationen auf die Universalkurve des p-MOSFETs. Dabei ist zu erkennen, dass die Leistungsfähigkeit des p-MOSFETs mit einer Monomer-Borimplantation sehr stark von den verwendeten Implantationsbedingungen abhängt. Eine zur BF_2-Implantation vergleichbare MILLERkapazität C_{ov} hat eine entsprechende Leistungsfähigkeit des Transistors zur Folge. Demgegenüber weist die 0.7 keV Monomer-Borimplantation aufgrund der stark erhöhten MILLERkapazität C_{ov} und der damit einhergehend verkleinerten metallurgischen Gatelänge L_{met} einen stark erhöhten Sperrstrom $I_{D,off}$ auf. Daraus lässt sich ableiten, dass mit einer Monomer-Borimplantation bei Anpassung der Implantationsparameter im Vergleich zur BF_2-Implantation analoge Transistorergebnisse erzielt werden können.

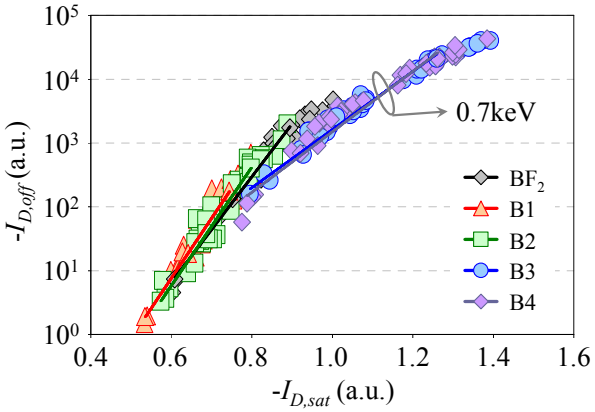


Bild 6.29: Einfluss der SDE-Implantationsparameter gemäß Tabelle 6.4 von BF_2- bzw. Borimplantationen in amorphes Silizium auf die Universalkurve von p-MOSFETs

6.3.3 SD-Implantation

Wie in Abschnitt 0 bereits angesprochen, kann es mit Hilfe der Kurzzeitausheilung, aber ohne RTA, durch die verringerte Diffusion und der damit verbundenen kleineren Tiefe der pn-Übergänge im Vergleich zu einem Prozess mit einer RTA zu einer Eingrenzung des Stromflusses kommen. Dadurch ist es möglich, dass bei Transistoren ohne RTA das Dotierungsprofil am Übergang vom Silizium zum Silizid derart verändert wird, dass sich der Übergangswiderstand $R_{Silizid}$ erhöht (Bild 6.17). Beides lässt sich durch eine Anpassung des Dotierungsprofils der SD-Implantationen erreichen. Dazu muss jedoch:

1. das Konzentrationsmaximum an den Silizium/Silizid-Übergang platziert werden und
2. die Konzentration der SD-Implantation an der Grenzfläche Silizium/BOX ausreichend sein, um ein Aufsetzen der SD-Gebiete auf das vergrabene Oxid zu gewährleisten.

Mit Hilfe einer zweiten SD-Implantation lassen sich beide Vorgaben vorteilhaft voneinander entkoppeln. Man kann z.B. eine tiefere Implantation mit einer geringeren Dosis unter einem Implantationswinkel von 0 ° verwenden, um die SD-Gebiete auf das BOX aufzusetzen und eine zweite flache Implantation mit einer höheren Dosis dazu benutzen, den Übergangswiderstand $R_{Silizid}$ zwischen Silizium und Silizid zu verringern. Um außerdem das Problem der Eingrenzung des Stromflusses zu vermeiden, kann man die zweite zusätzliche SD-Implantation unter einem Winkel in den Halbleiter einbringen (Bild 6.30).

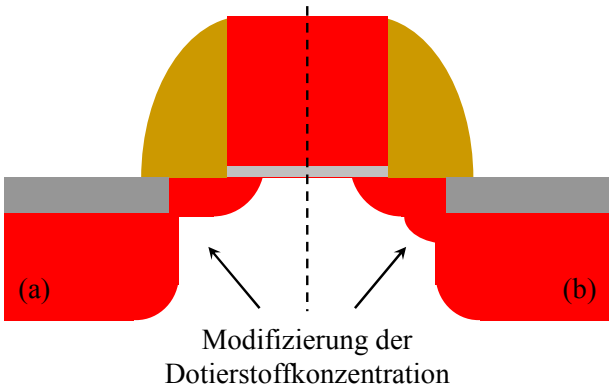


Bild 6.30: Vergleich der Dotierungsprofile (a) ohne und (b) mit zusätzlicher SD-Implantation

Im Folgenden wird der Einfluss einer gewinkelten SD-Implantation auf das Transistorverhalten betrachtet. Bei den durchgeführten Untersuchungen wurden die zusätzlichen SD-Implantationen mit unterschiedlicher Energie bei konstanter Dosis und mit unterschiedlicher Dosis bei konstanter Energie eingebracht (Tabelle 6.5). Als Referenz dient der Transistor mit einer einfachen SD-Implantation. Auf die Parameter dieser Implantation wird an dieser Stelle nicht näher eingegangen, da diese wie im Fall der Vordotierung des Polysiliziumgates aufgrund der Variation der Gatehöhe zwischen und auch innerhalb von Technologiegenerationen individuell an die jeweilige Gatehöhe angepasst werden müssen, um eine Durchimplantation durch das Gateoxid zu vermeiden.

Tabelle 6.5 Implantationsparameter zur Untersuchung des Einflusses zusätzlicher SD-Implantationen auf die Leistungsfähigkeit von diffusionsarmen n- und p-MOSFETs. Angegeben sind Dotand, Energie, Dosis und Winkel.

Experiment		n-MOSFET	p-MOSFET
Energie1	bei einer konstanten Dosis von $1.0 \cdot 10^{15}$ cm^{-2}	P, 2 keV, 30 °	B, 0.7 keV, 30 °
Energie2		P, 4 keV, 30 °	B, 1.0 keV, 30 °
Energie3		P, 6 keV, 30 °	B, 1.3 keV, 30 °
Dosis1	bei einer konstanten Energie von 4 keV (n-MOSFET) von 1 keV (p-MOSFET)	P, $1.0 \cdot 10^{15}$ cm^{-2}, 30 °	B, $1.0 \cdot 10^{15}$ cm^{-2}, 30 °
Dosis2		P, $2.0 \cdot 10^{15}$ cm^{-2}, 30 °	B, $2.0 \cdot 10^{15}$ cm^{-2}, 30 °
Dosis3		P, $4.0 \cdot 10^{15}$ cm^{-2}, 30 °	B, $4.0 \cdot 10^{15}$ cm^{-2}, 30 °

a) Einfluss der Implantationsenergie

Bild 6.31 zeigt zunächst den Einfluss der zusätzlichen SD-Implantationen auf die Universalkurve für beide Transistortypen. Sowohl bei n- als auch bei p-MOSFETs verbessert sich der Sättigungs-Drainstrom $I_{D,sat}$ infolge der zusätzlichen Implantation. Bei n-MOSFETs ist eine kontinuierliche Verbesserung mit zunehmender Implantationsenergie zu erkennen, wohingegen p-MOSFETs lediglich auf die höchste Implantationsenergie mit einem geringfügig verbesserten Transistorverhalten reagieren. Betrachtet man dazu die berechnete Konzentrationsänderung 15 nm unter der Siliziumgrenzfläche am Übergang vom Silizium zum Silizid (Bild 6.32), so ist in Abhängigkeit von der Implantationsenergie vor allem bei n-MOSFETs eine fast 30 % höhere Dotierstoffkonzentration festzustellen, wodurch die beobachtete Drainstromerhöhung um etwa 3 % zu erklären ist. Die Konzentrationsänderung ist bei p-MOSFETs nicht so ausgeprägt wie bei n-MOSFETs, was sich auch in der geringeren Drainstromerhöhung von etwa 2 % widerspiegelt.

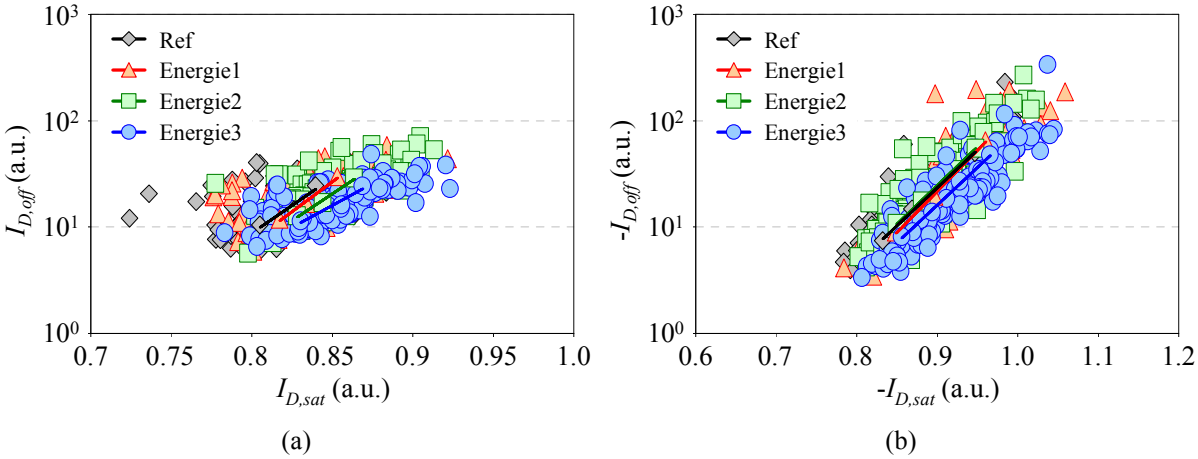


Bild 6.31: Einfluss der Implantationsenergie bei konstanter Dosis der zusätzlichen SD-Implantationen auf die Universalkurve für (a) n- und (b) p-MOSFETs

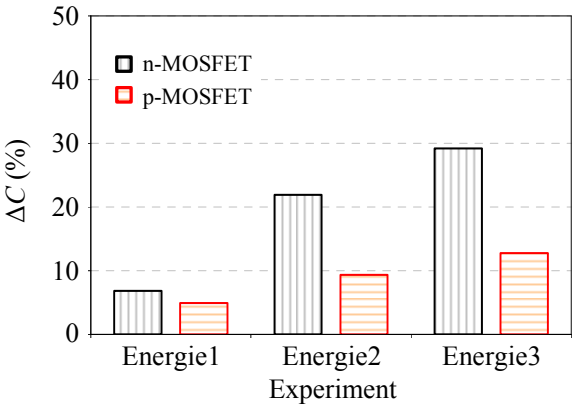


Bild 6.32: Berechnete Konzentrationsänderung ΔC 15 nm unter der Siliziumgrenzfläche am Übergang vom Silizium zum Silizid für verschiedene Implantationsenergien bei konstanter Dosis der zusätzlichen SD-Implantationen für n- und p-MOSFET

b) Einfluss der Implantationsdosis

Bild 6.33 zeigt die berechnete Änderung der Dotierstoffkonzentration am Silizium-Silizid-Übergang. Bei n-MOSFETs erhöht sich in Abhängigkeit von der Implantationsdosis die Konzentration C an Dotanden um bis zu 70%, was sich auch in einer nennenswerten Erhöhung des Sättigungs-Drainstroms $I_{D,sat}$ um bis zu 9 % ausdrückt (Bild 6.34a).

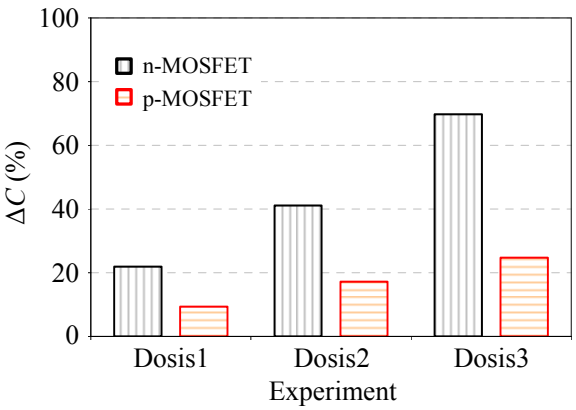


Bild 6.33: Berechnete Konzentrationsänderung ΔC 15 nm unter der Siliziumgrenzfläche am Übergang vom Silizium zum Silizid für verschiedene Implantationsdosen bei konstanter Energie der zusätzlichen SD-Implantationen für n- und p-MOSFET

P-MOSFETs reagieren leider nicht so deutlich auf eine Erhöhung der Implantationsdosis. Bei der größten Implantationsdosis ist eine etwa 23 % höhere Dotierstoffkonzentration erkennbar, wodurch sich der Sättigungs-Drainstrom $I_{D,sat}$ nur um etwa 3 % erhöht (Bild 6.34b). Die Ursache für die nicht so stark ausgeprägte Erhöhung des Drainstroms I_D wie bei n-MOSFETs kann in der nicht ganz optimalen Wahl der Implantationsenergie liegen, da die richtige Tendenz mit höheren Implantationsdosen, also eine Steigerung des Drainstroms I_D erkennbar ist. Diese Experimente zeigen aber auch das Potenzial dieses Ansatzes, sowohl für n- und p-MOSFET mit relativ einfachen Mitteln die Leistungsfähigkeit eines Transistors ohne RTA weiter zu verbessern. Durch eine weitere Optimierung der Implantationsparameter ist vor allem bei p-MOSFETs eine weitere Verbesserung in der Leistungsfähigkeit möglich.

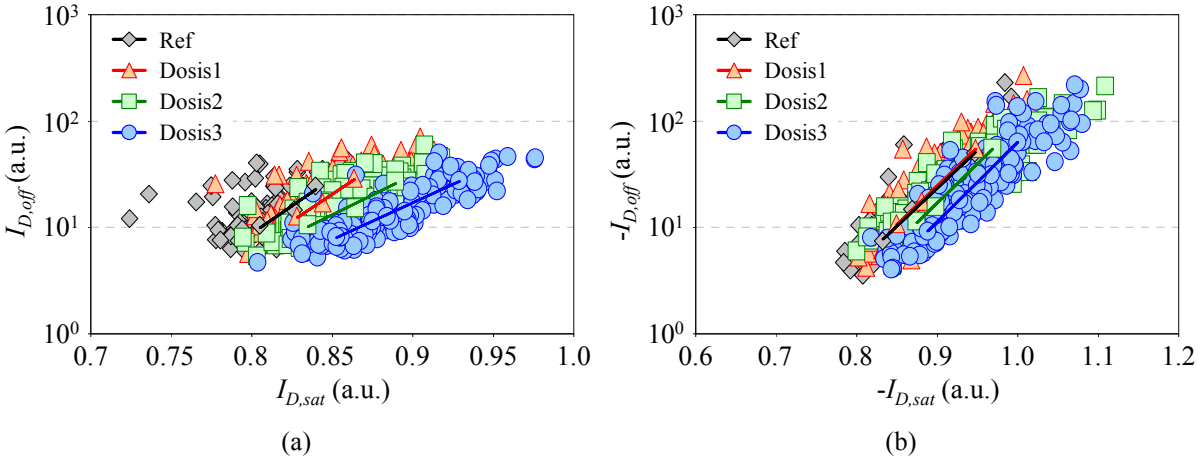


Bild 6.34: Einfluss der Implantationsdosis bei konstanter Energie der zusätzlichen SD-Implantationen auf die Universalkurve für (a) n- und (b) p-MOSFETs

6.3.4 Zusammenfassung

Aus den Ergebnissen der Untersuchungen zum Einfluss der Implantationsparameter der Halo-, SDE- und SD-Implantationen kann man für Transistoren, die eine Kurzzeitausheilung, aber keine RTA erfahren, folgende Schlussfolgerungen ziehen:

- Das Kurzkanalverhalten lässt sich nur durch eine kombinierte Optimierung des Dotierungsprofils der Wannen-, SDE-, SD- und Halo-Gebiete erreichen und nicht durch eine separate Halo-Implantation verbessern.
- Bei Verzicht auf die RTA lassen sich Monomer-Bor und Phosphor als alternative Implantationsspezies für eine p- bzw. n-Dotierung verwenden, ohne die Leistungsfähigkeit der Transistoren negativ zu beeinflussen.
- Erfolgt eine zusätzliche Implantation in die SD-Gebiete des Transistors unter einem Winkel verringert sich der externe Widerstand der n- als auch des p-MOSFETs und erhöht dadurch die Leistungsfähigkeit beider Transistortypen erheblich.

6.4 Einfluss zusätzlicher Ausheilprozesse

Wie in Abschnitt 6.2.1 beschrieben, wird beim herkömmlichen RTA-Prozess vor der Spacer1-Strukturierung ein zusätzlicher Diffusionsschritt im Temperaturbereich von (600…800) °C für mehrere Minuten durchgeführt, der aber aufgrund seines geringen thermischen Budgets *TB* keinen Einfluss auf die Diffusion der bis zu diesem Prozessschritt eingebrachten Dotanden hat. Aufgrund dieser Temperaturbehandlung entstehen im Zusammenhang mit einer PAI, die für diese Verspannungstechnik notwendig ist, Verspannungen im Transistor, welche zu einer Verbesserung der Elektronenbeweglichkeit im Kanal und in Folge dessen zu einer Erhöhung der Leistungsfähigkeit von n-MOSFETs führen [123]. Im Folgenden soll der Einfluss dieser Verspannungstechnik auf Transistoren nur mit Kurzzeitausheilung untersucht werden. Dazu wurden Transistoren mit und ohne zusätzlichem Temperaturschritt (Niedrigtemperatur-RTA, NT-RTA) prozessiert. Alle weiteren Prozessparameter blieben unverändert.

Bild 6.35 zeigt den Einfluss eines zusätzlichen Niedrigtemperaturschritts auf die Universalkurve für beide Transistortypen. Im Gegensatz zur höheren Leistungsfähigkeit von n-MOSFETs in [123] wird durch die zusätzliche Temperaturbehandlung der Sättigungs-Drainstrom $I_{D,sat}$ um 6 % verringert. Die Leistungsfähigkeit von p-MOSFETs ist ebenfalls um 9 % reduziert. Die Verbesserung der Leistungsfähigkeit von n-MOSFETs beim herkömmlichen Hochtemperatur-RTA-Prozess werden demnach bei einem Transistor nur mit Kurzzeitausheilung durch andere Effekte „kaputt gemacht“.

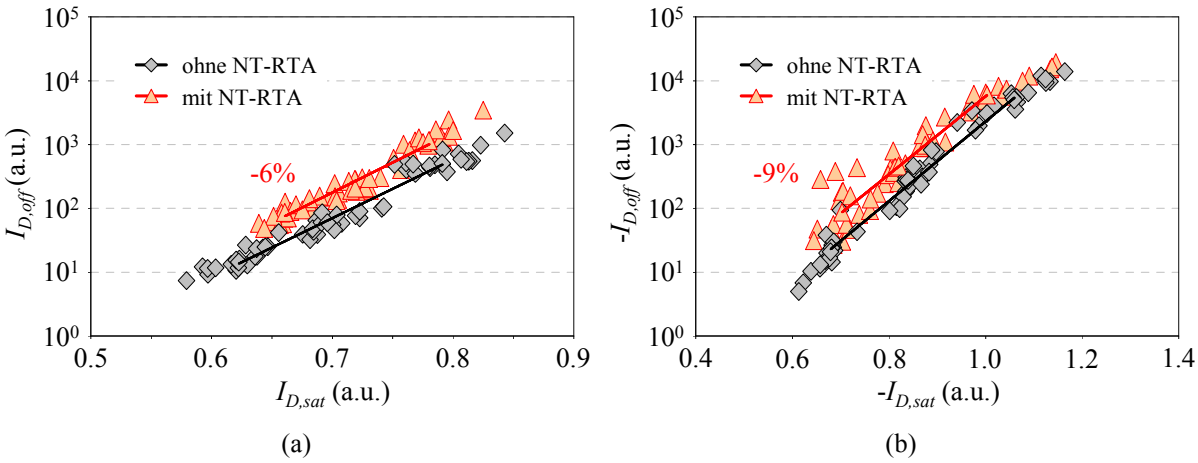


Bild 6.35: Einfluss eines zusätzlichen Niedrigtemperaturschritts (NT-RTA) auf die Universalkurve für (a) n- und (b) p-MOSFETs

Bild 6.36 zeigt den Einfluss des zusätzlichen Prozessschritts auf den Schichtwiderstand R_{SDE} der SDE-Gebiete beider Transistortypen. Vor allem bei n-MOSFETs ist durch die zusätzliche Temperaturbehandlung eine sehr starke Erhöhung des Schichtwiderstands um 169 % zu erkennen. Der Schichtwiderstand R_{SDE} der SDE-Gebiete bei p-MOSFETs erhöht sich zwar nicht so stark wie bei n-MOSFETs, aber immer noch um 48 %.

Neben dem größeren Schichtwiderstand R_{SDE} der SDE-Gebiete, was als Hauptgrund für die verringerte Leistungsfähigkeit der Transistoren angenommen werden kann, ist ebenfalls eine verstärkte Dotandendiffusion festzustellen. Dieser Sachverhalt wird anhand der größeren MILLERkapazität C_{ov} sowie der kleineren Sperrschichtkapazität C_j der pn-Übergänge zwischen Source bzw. Drain und Wanne deutlich (Bild 6.37).

Die erhöhte MILLERkapazität C_{ov} ist ein Zeichen für eine größere Überlappung der SDE-Gebiete unter dem Gate, was gleichbedeutend mit einer verstärkten lateralen Diffusion der SDE-Gebiete in Richtung Kanal ist. Die geringere Sperrschichtkapazität C_j ist auf eine erhöhte Dotandendiffusion der tiefen SD-Gebiete in Richtung Wannengebiet zurückzuführen.

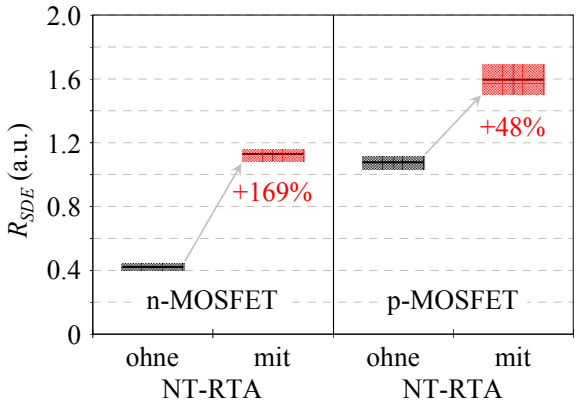


Bild 6.36: Einfluss eines zusätzlichen Niedrigtemperaturschritts (NT-RTA) auf den Schichtwiderstand R_{SDE} der SDE-Gebiete für n- und p-MOSFETs

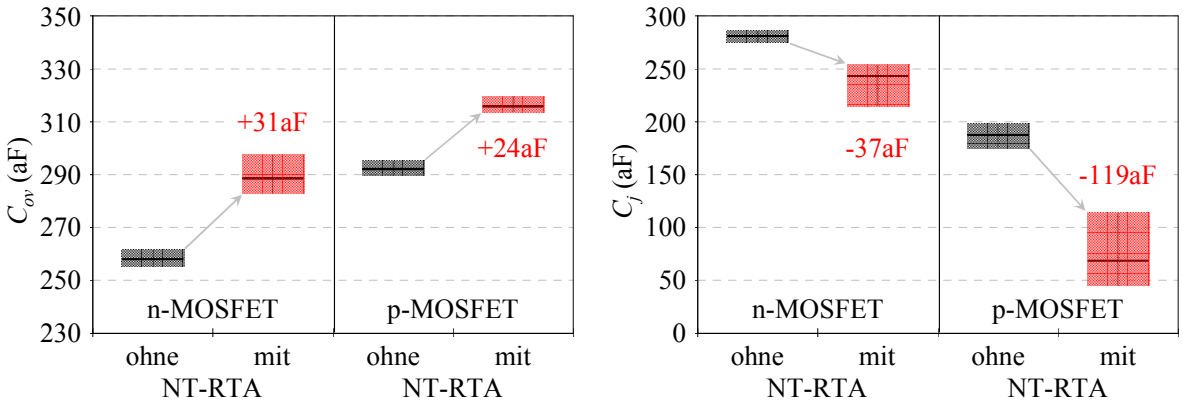


Bild 6.37: Einfluss eines zusätzlichen Niedrigtemperaturschritts (NT-RTA) auf (a) die MILLERkapazität C_{ov} und (b) die Sperrschichtkapazität C_j zwischen Source bzw. Drain und Wanne für n- und p-MOSFETs

Bild 6.38 zeigt schematisch den Einfluss des zusätzlichen Temperaturschritts bei der Herstellung von diffusionsarmen Transistoren. Aufgrund der vor den Implantationen der Halo- und SDE-Gebiete durchgeführten PAI entsteht eine etwa 40 nm dicke, vollständig amorphisierte Schicht (Bild 6.38a). Infolge der Niedrigtemperaturbehandlung erfolgt die Rekristallisation der amorphen Schicht durch die Festphasenepitaxie, die eine Aktivierung der SDE-Gebiete, aber auch eine große Anzahl von EOR-Defekten an der originalen Grenzfläche zwischen kristallinem und amorphisiertem Silizium verursacht. Diese Defekte existieren noch nach der Spacer-Strukturierung und den Implantationen der SD-Gebiete (Bild 6.38c). Durch die sich an die Implantation anschließende Kurzzeitausheilung werden die EOR-Defekte aufgelöst. Die dabei freigesetzten Silizium-Zwischengitteratome führen zu einer verstärkten Diffusion der Dotanden der SDE- und SD-Gebiete in Richtung Kanal bzw. Wanne (Bild 6.38d), wodurch sich wiederum die größere MILLERkapazität C_{ov} in Bild 6.37a und die kleinere Sperrschichtkapazität C_j in Bild 6.37b erklären lassen.

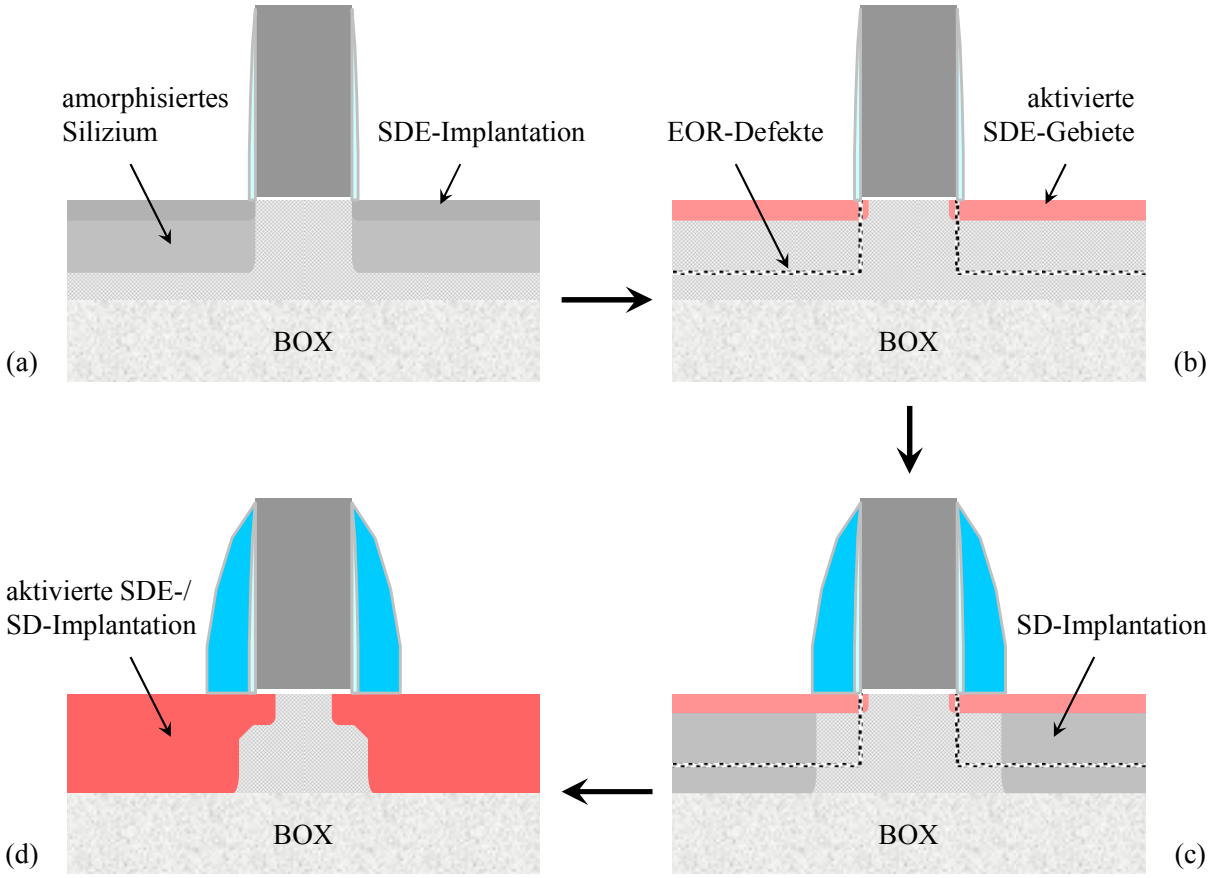


Bild 6.38: Schematische Darstellung des Einflusses einer Niedrigtemperaturbehandlung bei diffusionsarmen Transistoren
(a) nach PAI sowie Halo- und SDE-Implantation, (b) nach Niedrigtemperaturbehandlung,
(c) nach SD-Implantation, (d) nach Kurzzeitausheilung.

Bild 6.39 zeigt einen Vergleich der sich theoretisch aus den elektrischen Parametern ergebenden Dotierungsprofile mit und ohne Niedrigtemperaturbehandlung. Die zusätzliche Niedrigtemperaturbehandlung hat eine erhöhte Dotandendiffusion zur Folge, was anhand der gestichelten Linien in Bild 6.39b dargestellt wird. Weiterhin erfolgt eine drastische Erhöhung des Schichtwiderstands R_{SDE} der SDE-Gebiete, was als Hauptgrund für die verringerte Leistungsfähigkeit der Transistoren angenommen werden kann. Diese Ergebnisse zeigen, dass bei einem diffusionsarmen Transistor das thermische Budget jedes einzelnen Prozessschritts während der Herstellung beachtet werden muss, um eine unnötige Diffusion oder Deaktivierung der eingebrachten Dotanden zu vermeiden.

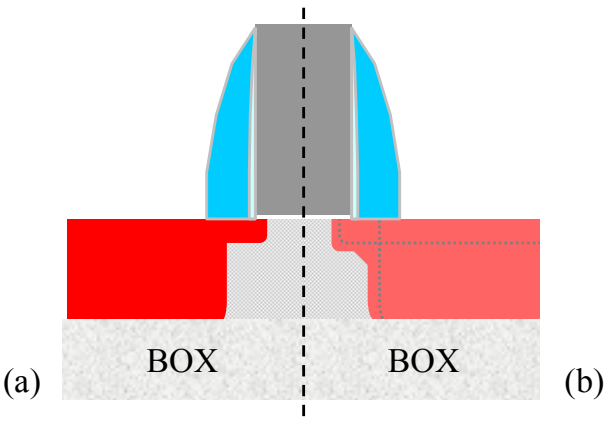


Bild 6.39: Schematische Darstellung der Dotierungsprofile (a) ohne und (b) mit zusätzlichem Niedrigtemperaturschritt. Die gestrichelten Linien in (b) stellen dabei das Dotierungsprofil des Transistors ohne zusätzlichen Temperaturschritt nach (a) dar.

6.5 Einbau von verspannten Silizium-Germanium-Schichten

6.5.1 Vorbetrachtungen

Wie man in Abschnitt 4.4.2 sehen konnte, kann durch SiGe in den SD-Gebieten die Leistungsfähigkeit des p-MOSFETs verbessert werden. Dabei treten zumindest auf SOI-Wafern weder Relaxationserscheinungen noch Schädigungen der SiGe-Regionen durch die Kurzzeitausheilung auf. Das SiGe-Modul (Bild 4.25) wird bei diffusionsarmen p-MOSFETs nach der Gate-Strukturierung und vor der Spacer0-Ätzung eingefügt (Bild 6.40). Wenn im Folgenden nicht anders angegeben, ist die Tiefe der Mulden 500 Å und der Germaniumgehalt im SiGe beträgt 22 %.

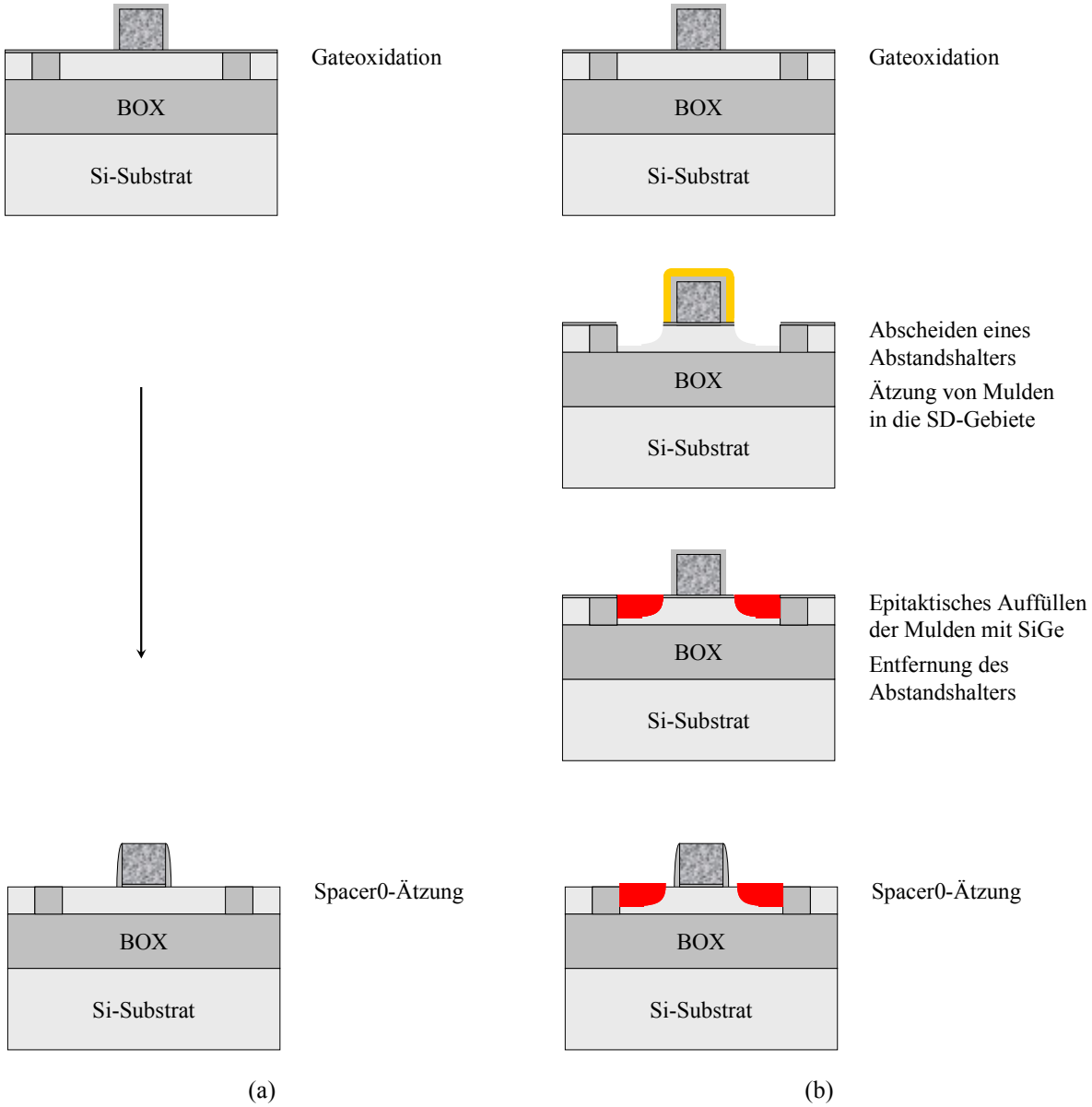


Bild 6.40: Vergleich der Prozessabfolge für diffusionsarme p-MOSFETs (a) ohne und (b) mit SiGe (vgl. Gesamtprozess in Bild 6.1)

6.5.2 Einfluss auf die Leistungsfähigkeit von p-MOSFETs

Um den Einfluss von SiGe auf das Transistorverhalten des diffusionsarmen p-MOSFETs zu untersuchen, wurde der Abstand *DDS* des SiGe zum Gate von 250 Å über 180 Å, 130 Å auf 80 Å verkleinert. Alle weiteren Prozessparameter blieben unverändert. Als Referenz (Ref) dient ein Transistor ohne SiGe.

Bild 6.41 zeigt den Einfluss des Abstands *DDS* des SiGe zum Gate auf die Universalkurve für p-MOSFETs. Durch das SiGe erhöht sich der Sättigungs-Drainstrom $I_{D,sat}$ bei einem SiGe-Gate-Abstand von 250 Å um etwa 10 %. Dieser Gewinn im Drainstrom erhöht sich linear um circa 3 % pro 50 Å kleinerem SiGe-Gate-Abstand. Dies führt bei einem SiGe-Gate-Abstand von 80 Å zu einem Drainstromgewinn von nahezu 21 %. Im Vergleich zum RTA-Prozess für verschiedene Technologiegenerationen sowie berechneten Werten zeigen die beobachteten Ergebnisse eine recht gute Übereinstimmung (Bild 6.42). Wie die durchgeführten Berechnungen zeigen, ist die verbesserte Leistungsfähigkeit der p-MOSFETs auf eine erhöhte Druckverspannung in Kanalrichtung zurückzuführen (Bild 6.43). Pro 50 Å kleinerem SiGe-Gate-Abstand erhöht sich die Druckverspannung des Kanals in Stromrichtung um etwa 84 MPa, was nach Abschnitt 4.3.3 zu einer steigenden Löcherbeweglichkeit führt und somit den höheren Sättigungs-Drainstrom zur Folge hat.

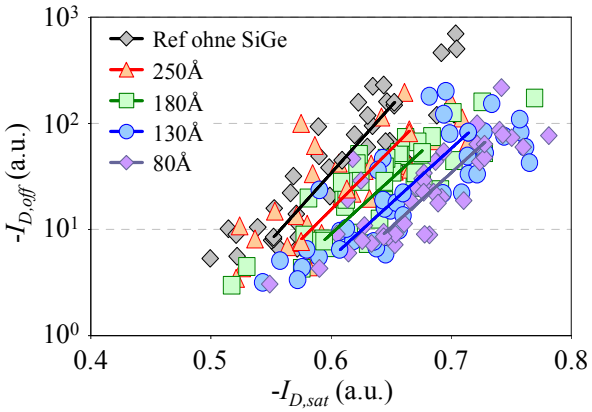


Bild 6.41: Einfluss des SiGe-Gate-Abstands *DDS* auf die Universalkurve für p-MOSFETs

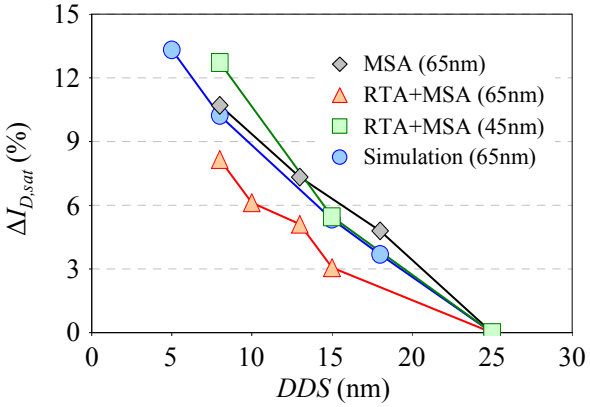


Bild 6.42: Einfluss der Ausheilsequenz bzw. Technologiegeneration auf die Sättigungs-Drainstromänderung $\Delta I_{D,sat}$ für p-MOSFETs mit SiGe in Abhängigkeit vom SiGe-Gate-Abstand *DDS*

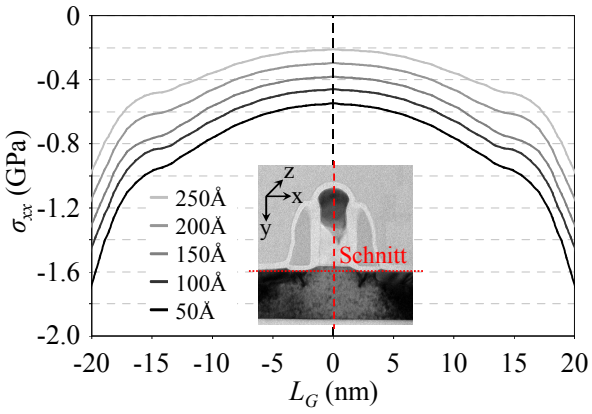


Bild 6.43: Einfluss des SiGe-Gate-Abstands *DDS* auf die berechnete Verspannungsverteilung σ_{xx} in Kanalrichtung 3 nm unter der Si/SiO_2-Grenzfläche nach der Silizierung

Die im Nanometerbereich lokal aufgelöste Verspannungsverteilung innerhalb eines Transistors ist mit den derzeitigen Charakterisierungsmethoden nur sehr schwer möglich. Über die Ladungsträgerbeweglichkeit, die nach der in Abschnitt 2.2.2 beschriebenen Methode bestimmt werden kann, lassen sich jedoch Rückschlüsse über mögliche Verspannungen innerhalb des Kanalgebiets machen.

Bild 6.44 zeigt den Einfluss des SiGe-Gate-Abstands *DDS* auf den Transistorwiderstand R_{Total} für p-MOSFETs in Abhängigkeit von der Gatelänge L_G. Beim Vergleich der Transistoren mit und ohne SiGe ist ein klarer Unterschied im Anstieg *m* der Kurven zu erkennen. Durch den Einbau von SiGe erhöht sich die Löcherbeweglichkeit im Kanal, was an dem flacheren Anstieg *m* der Kurven der Transistoren mit SiGe deutlich wird. Mit kleiner werdendem SiGe-Gate-Abstand *DDS* verringert sich der Anstieg der Kurven weiter, was auf eine steigende Löcherbeweglichkeit schließen lässt. Im Rückschluss deuten diese Resultate auf eine höhere Verspannung mit geringerem SiGe-Gate-Abstand *DDS* hin, was wiederum sehr gut mit den Ergebnissen der Simulation übereinstimmt (Bild 6.43).

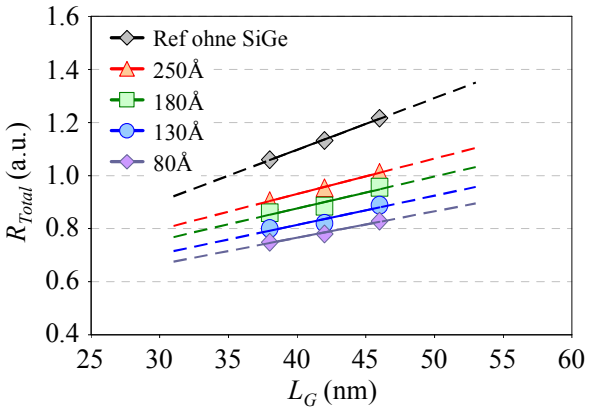


Bild 6.44: Einfluss des SiGe-Gate-Abstands *DDS* auf den Transistorwiderstand R_{Total} in Abhängigkeit von der Gatelänge L_G für p-MOSFETs. Die eingezeichneten Linien sind lineare Regressionsgeraden.

Tabelle 6.6 gibt eine Übersicht auf die Anstiege der einzelnen Kurven aus Bild 6.44 sowie die nach Gleichung (2.6) daraus ermittelten Beweglichkeitsänderungen im Vergleich zu Silizium. Durch die Verringerung des SiGe-Gate-Abstands *DDS* von 250 Å auf 80 Å erhöht sich die Ladungsträgerbeweglichkeit im Vergleich zum Referenztransistor ohne SiGe von 46 % auf 96 %.

Tabelle 6.6 Anstiege *m* der in Bild 6.44a dargestellten Kurven sowie die daraus ermittelten Beweglichkeitsänderungen μ_{Ref}/μ_{DDS} der untersuchten Transistoren mit SiGe im Vergleich zum Referenztransistor ohne SiGe

Experiment	Ref	250 Å	180 Å	130 Å	80 Å
m (Ω/nm)	4.87	3.34	3.01	2.75	2.49
μ_{Ref}/μ_{DDS} (%)	-	+46	+62	+77	+96

Neben der Verringerung des Kanalwiderstands infolge der erhöhten Löcherbeweglichkeit ergibt sich ebenfalls durch den Einbau von SiGe in die SD-Gebiete des Transistors ein kleinerer externer Widerstand R_{ext} aufgrund

- eines geringeren Übergangswiderstands $R_{Silizid}$ vom SiGe zum Silizid durch die geringere Barrierenhöhe,
- eines geringeren Schichtwiderstands R_{SDE} der SDE-Gebiete aufgrund der höheren Löslichkeit von Bor in SiGe (Bild 6.45).

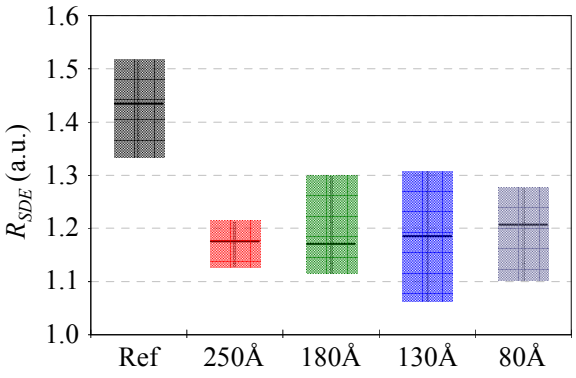


Bild 6.45: Einfluss des SiGe-Gate-Abstands *DDS* auf den Schichtwiderstand R_{SDE} der SDE-Gebiete für p-MOSFETs

Die Verringerung des Transistorwiderstands R_{Total} in Abhängigkeit vom SiGe-Gate-Abstand *DDS* um (16…31) % setzt sich somit aus einem geringeren Kanalwiderstand R_{Kanal} sowie einem kleineren externen Widerstand R_{ext} zusammen. Dennoch bleibt die durch das eingebettete SiGe hervorgerufene Kanalverspannung die Hauptursache für die Steigerung der Leistungsfähigkeit des p-MOSFETs.

6.5.3 Einfluss der Implantationsspezies

a) Vergleich zwischen BF_2 und Bor

Wie in Abschnitt 6.5.1 beschrieben, erfolgt die Implementierung des SiGe-Moduls in den Herstellungsprozess zwischen der Gate-Strukturierung und der Spacer0-Ätzung. Somit werden alle Implantationen nach der SiGe-Abscheidung durchgeführt. Aus diesem Grund wird in Verbindung mit eingebettetem SiGe wie beim herkömmlichen RTA-Prozess keine PAI verwendet, um eine zu starke Zerstörung und somit Relaxation der SiGe-Struktur zu vermeiden. Da aber eine BF_2-Implantation in Abhängigkeit von der verwendeten Implantationsenergie und -dosis ebenfalls eine amorphisierende Wirkung besitzt, ist eine Relaxation der SiGe-Schicht und die damit einhergehende Verringerung der Leistungsfähigkeit des Transistors möglich. Aus diesem Grund wurde der Einfluss einer B-SDE-Implantation auf das elektrische Verhalten von p-MOSFETs mit SiGe untersucht. Der Abstand des SiGe zum Gate beträgt dabei 180 Å. Parallel erfolgte die Untersuchung der beiden Implantationsspezies an einem p-MOSFET ohne SiGe.

Die Parameter der Halo-Implantationen wurden für die untersuchten Transistoren identisch gewählt, um das Dotierungsprofil im Kanal gleich zu gestalten. Weiterhin wurde die Implantationsenergie für die SDE-Gebiete so gewählt, dass das Maximum der Bordotierung für beide Implantationen an derselben Stelle liegt (Bild 6.46).

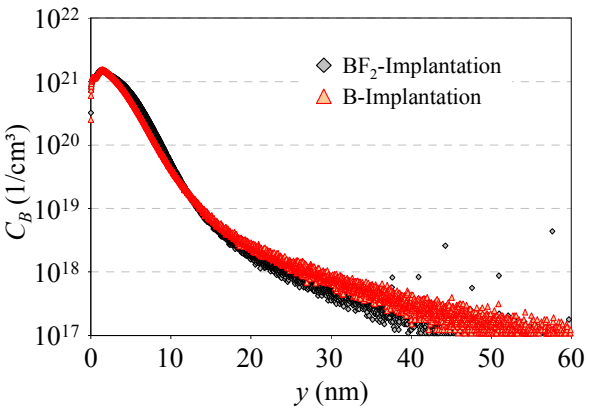


Bild 6.46: Implantationsprofil der Bor- und der BF_2-Implantation in kristallinem SiGe

Die BF_2- und die B-Implantationen wurden mit einer Energie von 2.3 keV und einer Dosis von $1.2 \cdot 10^{15}$ cm^{-2} bzw. mit einer Energie von 0.5 keV und einer Dosis von $1.0 \cdot 10^{15}$ cm^{-2} in das kristalline Silizium bzw. SiGe eingebracht. Die Wahl der Implantationsparameter erfolgte hinsichtlich identischer Schwellspannungen und gleicher metallurgischer Gatelängen der Transistoren. Letzteres wurde anhand von identischen MILLERkapazitäten bewiesen. Damit war ein sinnvoller Vergleich zwischen beiden Implantationsspezies möglich. Bild 6.47 zeigt die TEM-Aufnahmen von p-MOSFETs mit BF_2-SDE-Implantation bzw. mit B-SDE-Implantation. Helle Bereiche kennzeichnen dabei amorphe Gebiete. Durch die BF_2-Implantation wird eine etwa 10 nm dicke amorphe Schicht erzeugt (Bild 6.47a). Die teilweise amorphisierten Bereiche in den SD-Gebieten sind die Folge der Halo-Implantation. Demgegenüber hat die B-SDE-Implantation keine amorphisierende Wirkung (Bild 6.47b).

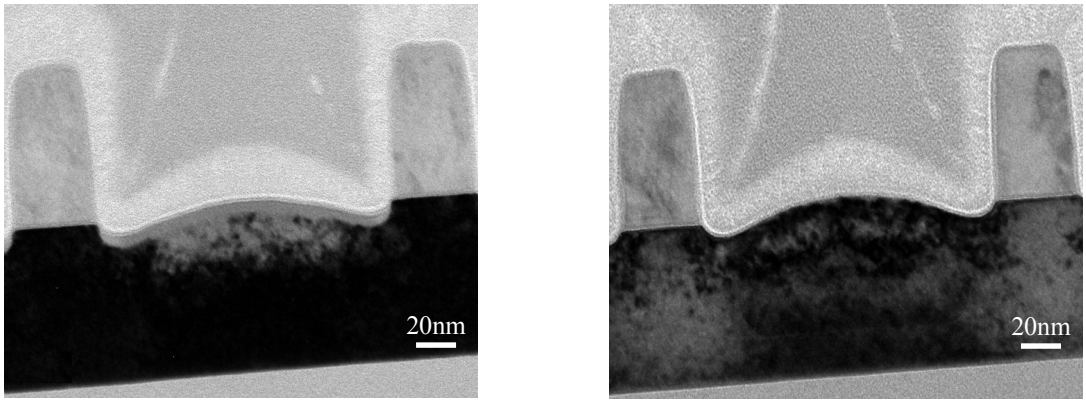


Bild 6.47: TEM-Aufnahmen von p-MOSFETs mit SiGe mit (a) BF_2-SDE-Implantation und (b) B-SDE-Implantation

Bild 6.48a zeigt die Auswirkungen der beiden SDE-Implantationen auf die Leistungsfähigkeit von p-MOSFETs mit und ohne SiGe. Im Gegensatz zu p-MOSFETs ohne SiGe verbessert sich durch die B-SDE-Implantation die Leistungsfähigkeit von p-MOSFETs mit SiGe um etwa 10 %. In Bild 6.48b ist der Transistorwiderstand R_{Total} in Abhängigkeit von der Gatelänge L_G für die verschiedenen Transistoren abgebildet. Tabelle 6.7 gibt dazu eine Übersicht auf die Anstiege der einzelnen Kurven aus Bild 6.48b sowie die nach Gleichung (2.6) daraus ermittelten Beweglichkeitsänderungen.

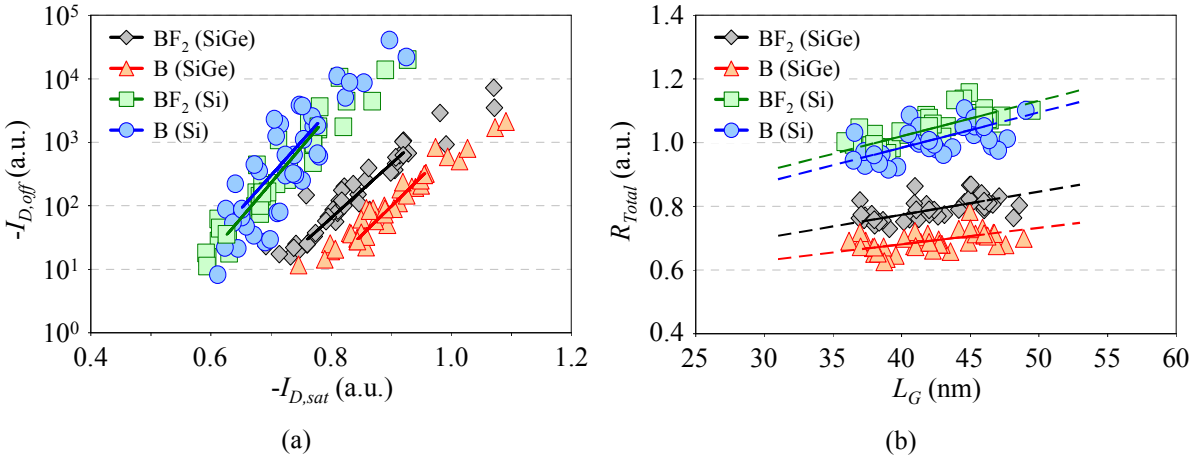


Bild 6.48: Einfluss der BF_2- bzw. der B-SDE-Implantation auf (a) die Universalkurve und (b) den Transistorwiderstand R_{Total} in Abhängigkeit von der Gatelänge L_G für p-MOSFET mit und ohne SiGe

Tabelle 6.7 Anstiege m der in Bild 6.48b dargestellten Kurven sowie die daraus ermittelten Beweglichkeitsänderungen μ_{Si}/μ_{SiGe} und μ_{BF2}/μ_B zum Einen im Vergleich zu Silizium und zum Anderen im Vergleich zu BF_2-SDE-Implantationen

Experiment	BF_2 (Si)	BF_2 (SiGe)	Bor (Si)	Bor (SiGe)
m (Ω/nm)	2.77	1.83	2.77	1.29
μ_{Si}/μ_{SiGe} (%)	-	51.4	-	114.7
μ_{BF2}/μ_B (%)	-	-	0	41.9

Zwei Schlussfolgerungen lassen sich aus Bild 6.48b und Tabelle 6.7 ziehen:

1. Die Ladungsträgerbeweglichkeit μ und somit der Kanalwiderstand R_{Kanal} sind im Fall der Transistoren ohne SiGe für beide Implantationsspezies gleich (gleich große Anstiege m der R_{Total}-L_G-Kurven). Die Kurve der BF_2-SDE-Implantation ist um etwa 25 Ω zu einem größeren Transistorwiderstand R_{Total} hin versetzt, was aufgrund des gleichen Kanalwiderstands R_{Kanal} nur durch einen höheren externen Widerstand R_{ext} zu erklären ist.
2. Im Fall der p-MOSFETs mit SiGe verringert sich durch die B-SDE-Implantationen der Anstieg der R_{Total}-L_G-Kurve von 1.83 Ω/nm auf 1.29 Ω/nm, was nach Gleichung (2.6) einer um 41.9 % höheren Ladungsträgerbeweglichkeit μ im Kanal entspricht.

Der erhöhte externe Widerstand R_{ext} mit BF_2-SDE-Implantationen lässt sich auf den höheren Schichtwiderstand R_{SDE} der SDE-Gebiete im Vergleich zu B-SDE-Implantationen zurückführen (Bild 6.49). Der Hauptgrund für die vergrößerte Ladungsträgerbeweglichkeit mit B-SDE-Implantationen ist in der bereits angesprochenen amorphisierenden Wirkung der BF_2-SDE-Implantation zu finden.

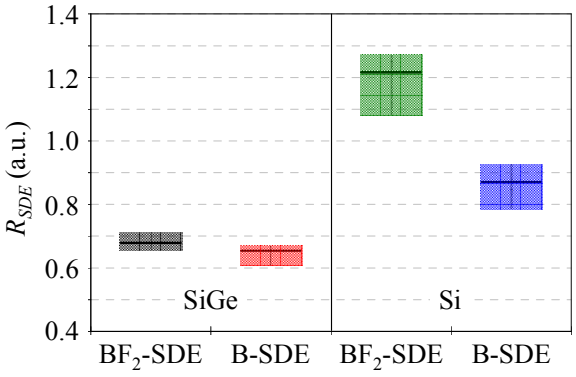


Bild 6.49: Einfluss der BF_2- bzw. der B-SDE-Implantationen auf den Schichtwiderstand R_{SDE} der SDE-Gebiete für p-MOSFETs mit und ohne SiGe

Die verringerte Ladungsträgerbeweglichkeit durch die BF_2-SDE-Implantation kann zwei Ursachen haben:

- Einerseits ist es möglich, dass die Ausheilung der Implantationsschäden eine teilweise Relaxation der zuvor amorphisierten SiGe-Schicht verursacht.
- Andererseits können die durch die Ausheilung entstandenen und nicht vollständig aufgelösten EOR-Defekte die Übertragung der Verspannung vom SiGe zum Kanal sowie die allgemeine Ladungsträgerbeweglichkeit beeinflussen.

Bild 6.50 zeigt dazu den Einfluss beider Implantationsspezies auf den Sperrstrom I_S der pn-Übergänge zwischen Source bzw. Drain und Wannengebiet. Unter Verwendung der BF_2-SDE-Implantationen ist mit und ohne SiGe eine beträchtliche Zunahme des Sperrstroms I_S zu erkennen. Da die Kanal-, Wannen- und SD-Dotierung aufgrund der verwendeten Implantationen bei den untersuchten Transistoren gleich ist, kann der höhere Sperrstrom I_S nur durch eine größere Anzahl an Defekten im pn-Übergang verursacht worden sein, welche einen erhöhten Generations-Rekombinations-Strom zur Folge haben.

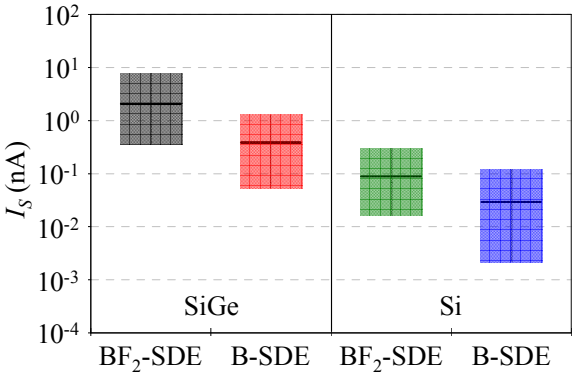


Bild 6.50: Einfluss der BF_2- bzw. der B-SDE-Implantationen auf den Sperrstrom I_S der pn-Übergänge zwischen Source bzw. Drain und Wannengebiet bei $U_{SW} = U_{DW} = -1.2$ V für p-MOSFETs mit und ohne eSiGe

Die experimentellen Ergebnisse sollten durch Simulationen erklärt werden. Bild 6.51 zeigt in diesem Zusammenhang den berechneten Transistorwiderstand R_{Total} in Abhängigkeit von der Gatelänge L_G und dem Germaniumgehalt. Wie erwartet, erhöht sich mit steigendem Germaniumgehalt die Verspannung im Kanal (Bild 6.52), was wiederum zu einer erhöhten Löcherbeweglichkeit und somit zu einer stetigen Verringerung des Anstiegs der R_{Total}-L_G-Kurve führt. Ein Vergleich der berechneten R_{Total}-L_G-Kurven mit der experimentell bestimmten Kurve für die BF_2-SDE-Implantation in SiGe mit einem Germaniumgehalt von 22 % ergibt, dass die berechnete R_{Total}-L_G-Kurve für eine B-SDE-Implantation in SiGe mit einem Germaniumgehalt von 14 % sehr gut mit den experimentellen Daten übereinstimmt (Bild 6.53). Das bedeutet, dass die Verringerung der Ladungsträgerbeweglichkeit infolge der BF_2-SDE-Implantationen einer Relaxation der SiGe-Schicht von etwa 36 % entspricht.

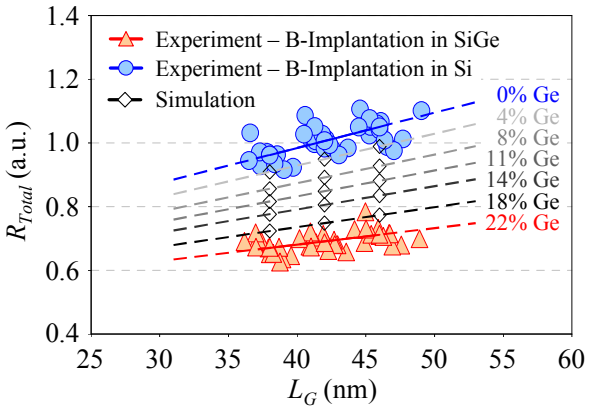


Bild 6.51: Berechneter Transistorwiderstand R_{Total} in Abhängigkeit von der Gatelänge L_G und dem Germaniumgehalt des SiGe

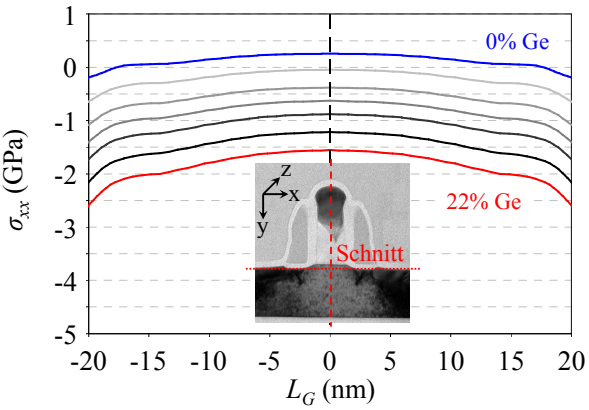


Bild 6.52: Berechnete Verspannungsverteilung σ_{xx} in Kanalrichtung 3 nm unter der Si/SiO_2-Grenzfläche nach der Silizierung in Abhängigkeit vom Germaniumgehalt im SiGe

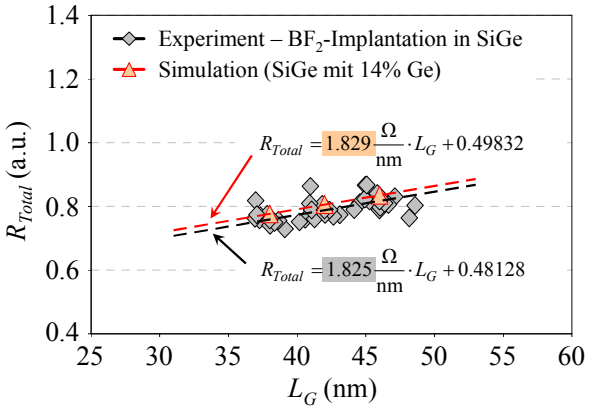


Bild 6.53: Vergleich des berechneten mit dem experimentell bestimmten Transistorwiderstand R_{Total} in Abhängigkeit von der Gatelänge L_G. Die Gleichungen geben die Regressionsgeraden für Simulation und Experiment an.

b) Vergleich zwischen B und $B_{18}H_{22}$

Wie man aus den Untersuchungen der Abschnitte 6.2.2, 6.3.2 und 6.5.3 a erkennen kann, werden für die Implantationen der SDE-Gebiete typischerweise Implantationsenergien von weniger als 3 keV für Arsen bzw. 1 keV für Bor und Phosphor sowie Implantationsdosen im Bereich von 10^{15} cm^{-2} verwendet. Um einen hohen Durchsatz des Implantationsprozesses zu gewährleisten, ist es notwendig, die Dotanden mit höheren Energien aus der Ionenquelle zu extrahieren. Diese werden mit Hilfe eines elektrischen Feldes kurz vor der Waferoberfläche auf die gewünschte Implantationsenergie abgebremst. Aufgrund der Interaktion einzelner Ionen mit Gasmolekülen werden dabei jedoch Neutrale erzeugt, die nicht durch das elektrische Feld abgebremst werden können und demzufolge den Wafer mit der Extraktionsenergie erreichen. Dieser Effekt wird als Energiekontamination bezeichnet und führt zu tieferen als den gewünschten Implantationsprofilen [124], [125].

Um dieses Problem zu umgehen, können anstatt eines einzelnen mehrere Dotandenatome pro Ion gleichzeitig implantiert werden. Durch diese sogenannten Cluster-Implantationen ist es möglich, höhere Extraktionsenergien zu verwenden. Beim Auftreffen dieser Ionenspezies auf der Waferoberfläche zerfällt das Molekül. Dabei verhält sich jedes einzelne Atom entsprechend seinem Massenanteil an der Totalenergie des implantierten Cluster-Ions wie ein herkömmlich implantiertes Atom. Anzumerken ist, dass alle Atome genau denselben Anteil an der Energie des implantierten Ions erhalten, wodurch sich im Vergleich zur Implantation mit einem Dotierstoffatom pro Ion keine Schwankung im Implantationsprozess ergibt. Auf diese Weise ist es möglich, Dotanden mit sehr geringen Energien bei gleichzeitig hohem Durchsatz zu implantieren, was für die Herstellung von pn-Übergängen mit sehr geringer Tiefe von großem Interesse ist.

Ein Beispiel für ein solches Cluster-Ion ist Octadecaboran $B_{18}H_{22}$ [126]. Dieses Cluster-Molekül enthält 18 Dotierstoffatome, d.h. in diesem Fall 18 Boratome. Unter Verwendung dieser Cluster-Implantation kann eine Borimplantation von 0.5 keV mit einer Extraktionsenergie von 10 keV durchgeführt werden. Jedoch werden durch die Implantation von Cluster-Molekülen erheblich größere Schäden am Siliziumkristall hervorgerufen als mit der herkömmlichen Borimplantation. Aufgrund der Größe des Cluster entsteht durch die Implantation selbst bei Verwendung relativ geringer Dosen ein amorphes Gebiet (Bild 6.54). Der Vorteil dieser Amorphisierung ist, dass dadurch das Channeling ohne eine zusätzliche Amorphisierungsimplantation verhindert wird.

Bild 6.55 zeigt die Implantationsprofile einer Bor- und einer $B_{18}H_{22}$-Implantation mit und ohne PAI entsprechend den Implantationsbedingungen. Wie man erkennen kann, wird infolge der Amorphisierung durch die Cluster-Implantation der Channeling-Anteil gegenüber der Monomer-Borimplantation sehr stark vermindert. Im Vergleich zu einer Borimplantation mit einer vorherigen PAI ist aber immer noch ein geringer Channeling-Effekt zu beobachten.

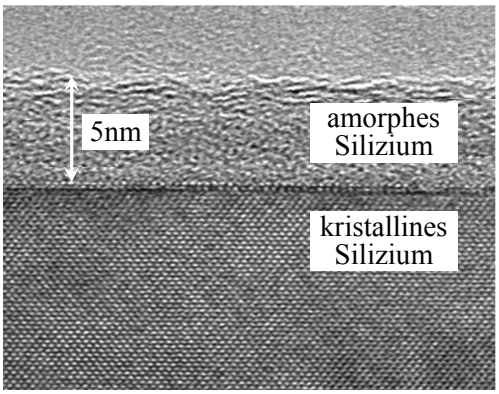


Bild 6.54: TEM-Aufnahme einer $B_{18}H_{22}$-Implantation mit einer 0.25 keV-äquivalenten B-Implantationsenergie (nach [127])

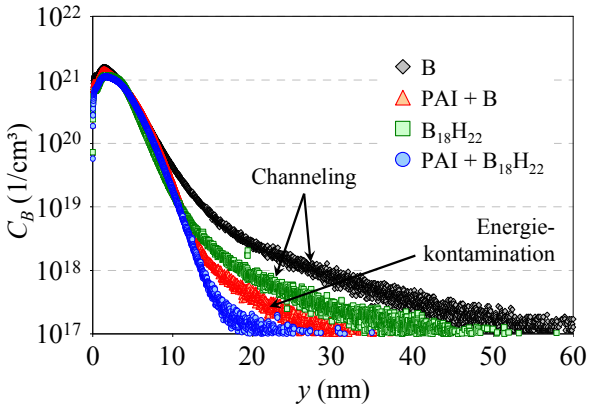


Bild 6.55: Implantationsprofile von B- und $B_{18}H_{22}$-Implantationen in Silizium (mit und ohne vorherige PAI)

Ursache dafür ist, dass während der Implantation erst ein amorphisiertes Gebiet entsteht und dadurch das Channeln der Dotanden nicht vollständig beseitigt werden kann. Weiterhin ist deutlich der Effekt der Energiekontamination beim Vergleich der B- und $B_{18}H_{22}$-Implantationen mit vorheriger PAI zu erkennen. Durch eine PAI wird zwar das Channeln der Dotierstoffatome verhindert, nicht aber das tiefere Eindringen der Dotanden aufgrund einer höheren Implantationsenergie. In [128] wurde gezeigt, dass mit Hilfe dieser neuen Implantationstechnologie bei p-MOSFETs ohne SiGe analoge Transistorergebnisse im Vergleich zu einer Borimplantation erreicht werden, aber mit dem Vorteil eines höheren Waferdurchsatzes. Wie man aber am Beispiel der BF_2-SDE-Implantation sehen kann, sollte die mit der Implantation einhergehende Amorphisierung insbesondere beim Einsatz von SiGe vermieden werden. Hinsichtlich einer weiteren Verringerung der Implantationstiefe und der damit verbundenen Gefahr der Energiekontamination ist es notwendig, das Zusammenspiel von Cluster-Implantationen und SiGe zu überprüfen. Aus diesem Grund wurde der Einfluss einer $B_{18}H_{22}$-SDE-Implantation auf das elektrische Verhalten von p-MOSFETs mit SiGe untersucht. Die Implantationsparameter der beiden unterschiedlichen SDE-Implantationen wurden erneut so gewählt, dass sich identische Schwellspannungen sowie gleicher metallurgischer Gatelängen zwischen den Transistoren einstellen. Letzteres wurde wiederum anhand der MILLERkapazitäten geprüft. Tabelle 6.8 gibt eine Übersicht auf die Implantationsbedingungen der SDE-Gebiete.

Tabelle 6.8 Implantationsparameter der SDE-Implantationen zur Untersuchung des Einflusses von Cluster-Implantationen auf die Leistungsfähigkeit von diffusionsarmen p-MOSFETs mit SiGe. Angegeben sind Dotand, Energie und Dosis.

Experiment	B-SDE	B_{18}-SDE
SDE-Implantation	B, 0.5 keV, $1.5 \cdot 10^{15}$ cm^{-2}	$B_{18}H_{22}$, 10 keV, $2.0 \cdot 10^{15}$ cm^{-2}

Bild 6.56a zeigt den Einfluss der beiden SDE-Implantationen auf die Leistungsfähigkeit der p-MOSFETs mit SiGe. Trotz des aufgrund der höheren Implantationsdosis um 13.9 % geringeren Schichtwiderstands R_{SDE} der SDE-Gebiete (Bild 6.56b) verringert sich unter Verwendung der Cluster-B-SDE-Implantation der Sättigungs-Drainstrom $I_{D,sat}$ bei konstantem Sperrstrom $I_{D,off}$ im Vergleich zur B-SDE-Implantation um etwa 5 %.

Der Hauptgrund für dieses Transistorverhalten ist wie bei der BF_2-SDE-Implantation die amorphisierenden Wirkung der Cluster-Implantation. Bild 6.57a stellt dazu den Transistorwiderstand R_{Total} in Abhängigkeit von der Gatelänge L_G dar. Wie im Fall der BF_2-Implantation verringert sich die Ladungsträgerbeweglichkeit μ im Kanal, da wie bereits angesprochen nicht vollständig aufgelöste EOR-Defekte die Übertragung der Verspannung vom SiGe zum Kanal sowie die allgemeine Ladungsträgerbeweglichkeit beeinflussen. Das Defekte infolge der Cluster-Implantation vorhanden sind, erkennt man am höheren Sperrstrom I_S der pn-Übergänge zwischen Source bzw. Drain und Wannengebiet (Bild 6.57b). Anzumerken sei aber noch, dass das um 5 % schlechtere Transistorverhalten mit einer Cluster-B-SDE-Implantation weniger stark ausgeprägt ist als bei einer BF_2-SDE-Implantation (-10 %). Dies kann im Vergleich zur BF_2-Implantation am kleineren Schichtwiderstand R_{SDE} der SDE-Gebiete oder an der Qualität der Grenzfläche zwischen amorphem und kristallinem Silizium bzw. SiGe liegen. Durch die Cluster-Implantation wird im Vergleich zu einer BF_2-Implantation ein wesentlich gleichmäßigerer, glatter Übergang geschaffen, wodurch bei der anschließenden Kurzzeitausheilung weniger EOR-Defekte entstehen [126].

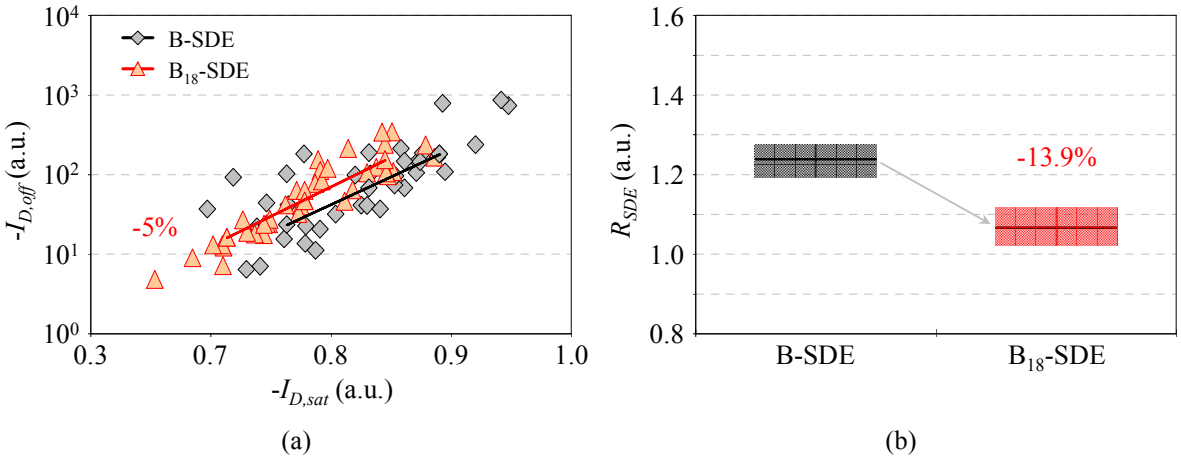


Bild 6.56: Einfluss der Cluster-B-SDE-Implantation auf (a) die Universalkurve und (b) den Schichtwiderstand R_{SDE} der SDE-Gebiete für p-MOSFETs mit SiGe.

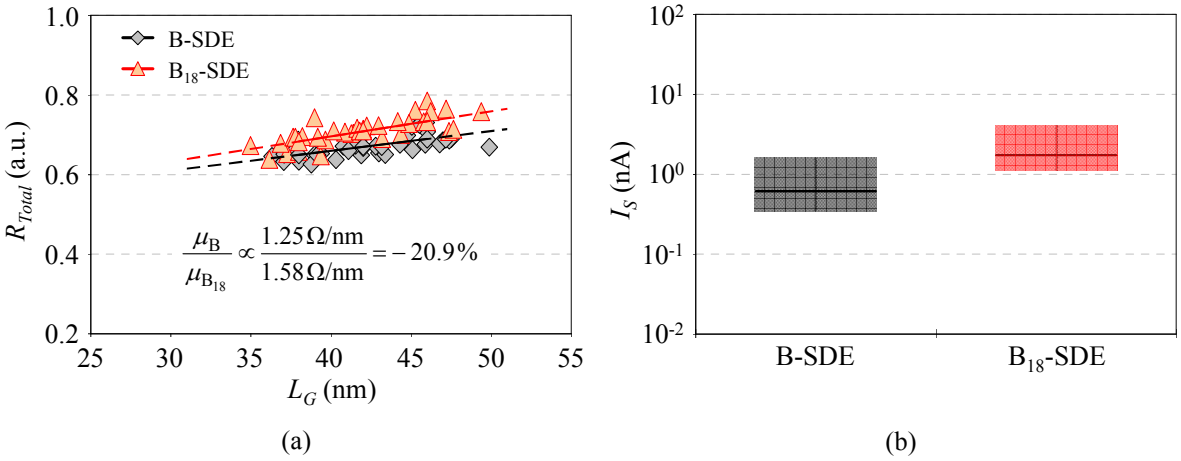


Bild 6.57: Einfluss der Cluster-B-SDE-Implantation auf (a) den Transistorwiderstand R_{Total} in Abhängigkeit von der Gatelänge L_G und (b) den Sperrstrom I_S der pn-Übergänge zwischen Source bzw. Drain und Wannengebiet bei $U_{SW} = U_{DW} = -1.2$ V für p-MOSFET mit SiGe

6.5.4 Einfluss der Spacer0-Ätzung

Im Zuge der Herstellung des diffusionsarmen Transistors dient ein 23 Å dickes Polysiliziumoxid, das durch Oxidation des polykristallinen Gates hergestellt und anschließend auf den Aktivgebieten anisotrop zurückgeätzt wird, als Abstandshalter für die Halo- und SDE-Implantationen. Bei dieser sogenannten Spacer0-Ätzung wird aufgrund der dabei ablaufenden chemischen Reaktionen ebenfalls ein Teil des Aktivgebiets mit entfernt, was bei einem unverspannten Transistor von geringer Bedeutung ist. Werden jedoch Verspannungsquellen in die SD-Gebiete eingebettet, spielt dieser Prozessschritt eine größere Rolle, da die Wirkung der Verspannungstechnik vermindert werden kann. Aus diesem Grund wird im Folgenden der Einfluss der Spacer0-Ätzung auf das Verhalten von p-MOSFETs mit SiGe untersucht. Als Vergleich dient ein Transistor ohne Spacer0-Ätzung. Da aber die Oxidationsschicht in der Regel als Schutz des Polysiliziums während der SiGe-Abscheidung dient, damit kein SiGe auf die Polysiliziumbahnen aufwächst, ist eine Oxidation des Polysiliziumgates zwingend notwendig. Aus diesem Grund wird beim Vergleichstransistor das Polysilizium nur so lange thermisch behandelt, bis sich eine etwa 10 Å dicke Oxidschicht gebildet hat. Diese Oxidschicht hat wiederum einen positiven Nebeneffekt, da sie gleichzeitig als Streuoxid für die sich anschließenden SDE-Implantationen dient, wodurch das Channeln auf natürliche Weise unterdrückt wird. Bild 6.58 zeigt TEM-Aufnahmen eines Transistors mit Spacer0-Ätzung und des Vergleichstransistors ohne Spacer0-Ätzung.

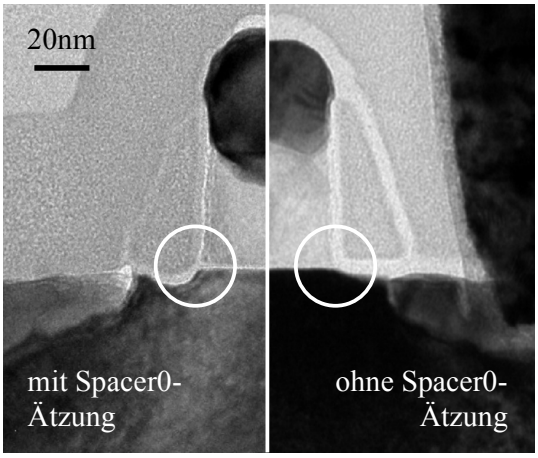


Bild 6.58: TEM-Aufnahme von diffusionsarmen p-MOSFETs mit und ohne Spacer0-Ätzung

Bild 6.59a zeigt den Einfluss der Spacer0-Ätzung auf die Universalkurve von p-MOSFETs mit SiGe. Bei den p-MOSFETs ohne Spacer0-Ätzung erhöht sich der Sättigungs-Drainstrom $I_{D,sat}$ bei einem konstanten Sperrstrom $I_{D,off}$ um etwa 11 %. Ursache dafür ist die verbesserte Verspannungseinkopplung vom SiGe zum Kanal. Bild 6.59b zeigt dazu den Transistorwiderstand R_{Total} in Abhängigkeit von der Gatelänge L_G mit und ohne Spacer0-Ätzung. Aufgrund der unterschiedlichen Prozessierung erhöht sich beim p-MOSFET ohne Spacer0-Ätzung die Ladungsträgerbeweglichkeit μ im Kanal um annähernd 32 %, was sich wiederum in der verbesserten Leistungsfähigkeit widerspiegelt. Simulationen zeigen, dass ohne Spacer0-Ätzung die für p-MOSFETs vorteilhafte Druckverspannung in Kanalrichtung um 63 % von maximal 459 MPa auf 751 MPa ansteigt (Bild 6.60), was den experimentell beobachteten Trend des Transistorwiderstands R_{Total} in Abhängigkeit von der Gatelänge L_G bestätigt.

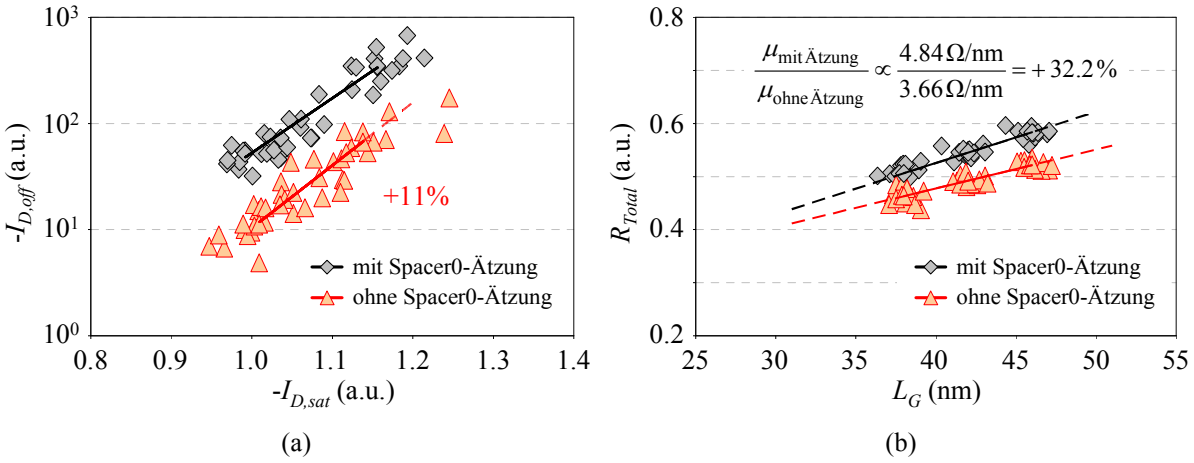


Bild 6.59: Einfluss der Spacer0-Ätzung auf (a) die Universalkurve und (b) den Transistorwiderstand R_{Total} in Abhängigkeit von der Gatelänge L_G für p-MOSFETs mit SiGe.

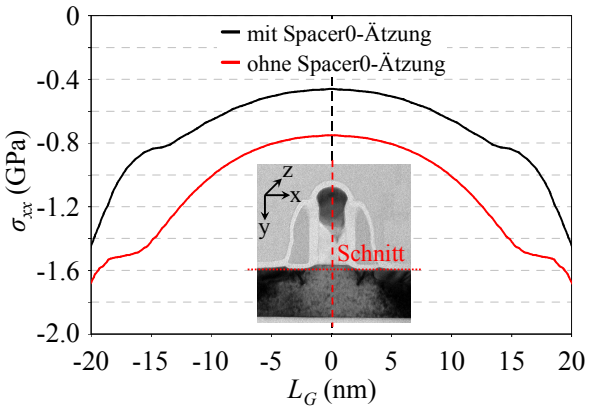


Bild 6.60: Berechnete Verspannungsverteilung σ_{xx} in Kanalrichtung 3 nm unter der Si/SiO_2-Grenzfläche nach der Silizierung mit und ohne Spacer0-Ätzung für p-MOSFETs mit SiGe

Neben dem veränderten Drainstromverhalten werden ebenfalls eine Reihe weiterer Transistorparameter beeinflusst. So wird durch das Auslassen der Spacer0-Ätzung die Sättigungsschwellspannung $U_{th,sat}$ betragsmäßig um 50.8 mV erhöht (Bild 6.61), was ebenfalls an der Verschiebung der Universalkurve in Richtung kleinerer Sperrströme $I_{D,off}$ zu erkennen ist (Bild 6.59a). Dies lässt durch das veränderte Dotierungsprofil und damit die Lage der Halo-Implantationen erklären (Bild 6.62a). Aufgrund der fehlenden Spacer0-Ätzung erhöht sich die Donatorenkonzentration im Kanalgebiet (Bild 6.62b), was wiederum eine betragsmäßig höhere Schwellspannung $U_{th,sat}$ zur Folge hat. Des Weiteren erhöht sich die MILLERkapazität C_{ov} um 12.3 aF (Bild 6.63), was seinen Ursprung vorwiegend in der verringerten Polysiliziumoxidation hat.

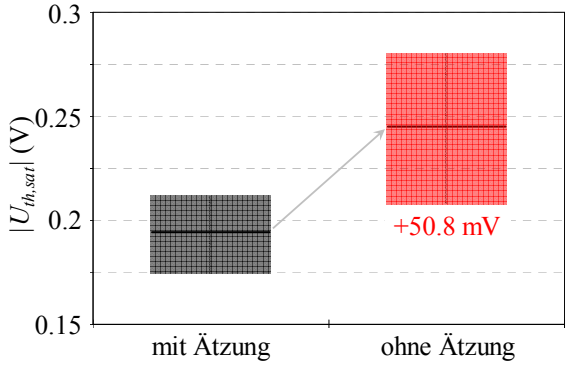


Bild 6.61: Einfluss der Spacer0-Ätzung auf die Sättigungsschwellspannung $U_{th,sat}$ für p-MOSFETs mit SiGe

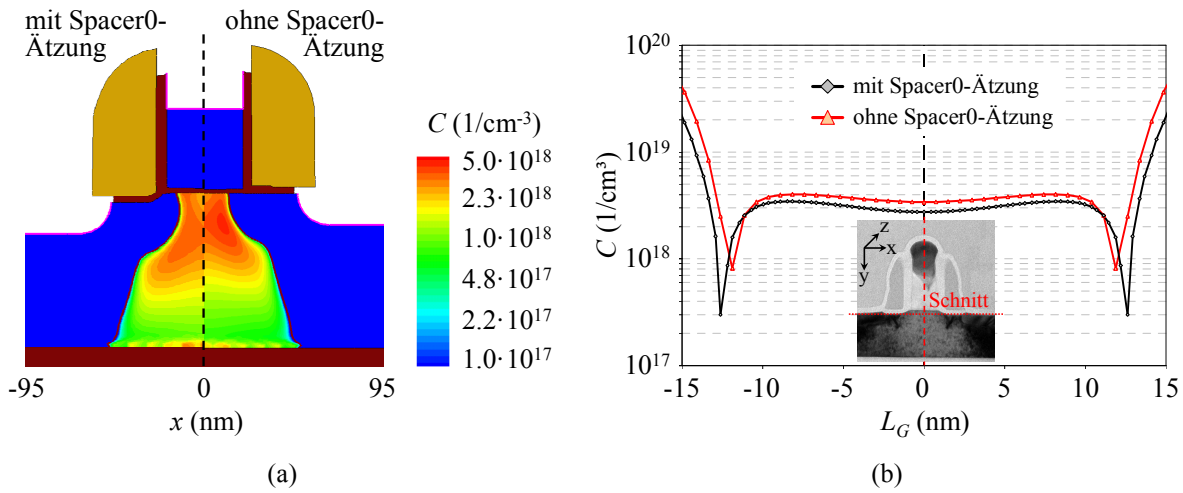


Bild 6.62: Berechnetes Dotierungsprofil mit und ohne Spacer0-Ätzung gemäß Bild 6.58; (a) Querschnitt; (b) 5 nm unter der Si/SiO_2-Grenzfläche

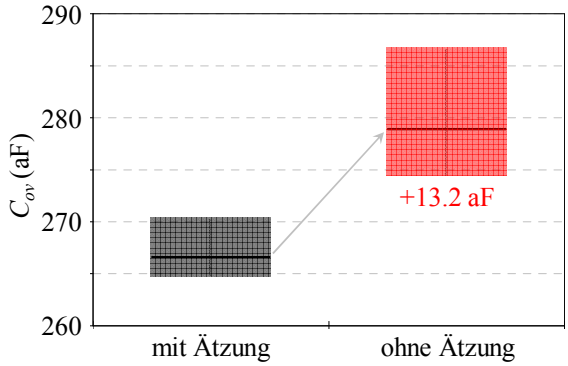


Bild 6.63: Einfluss der Spacer0-Ätzung auf die MILLERkapazität C_{ov} für p-MOSFETs mit SiGe

Bild 6.64 zeigt dazu die berechnete Verteilung des elektrischen Felds in Höhe der Überlappungsgebiete an der Si/SiO_2/Poly-Grenzfläche mit und ohne Spacer0-Ätzung. Aufgrund der verringerten Polysiliziumoxidation wird weniger Polysilizium an der Gatekante verbraucht. Dies hat zur Folge, dass sich ohne Spacer0-Ätzung eine kleinere physikalische Gateoxiddicke in Höhe der Überlappungsgebiete ausbildet, wodurch die direkte Überlappungskapazität größer wird (Gebiet 1 in Bild 6.64b). Weiterhin erhöht sich der Anteil der äußeren, umsäumenden Kapazität aufgrund der veränderten Geometrie (Gebiet 2 in Bild 6.64b). Die infolge des veränderten Dotierungsprofils etwas größere Überlappung der SDE-Gebiete unter das Gate hat ebenfalls einen Anteil an der höheren MILLERkapazität C_{ov} (Gebiet 3 in Bild 6.64b). Den größten Einfluss hat jedoch die verringerte physikalische Gateoxiddicke im Bereich der Überlappungsgebiete.

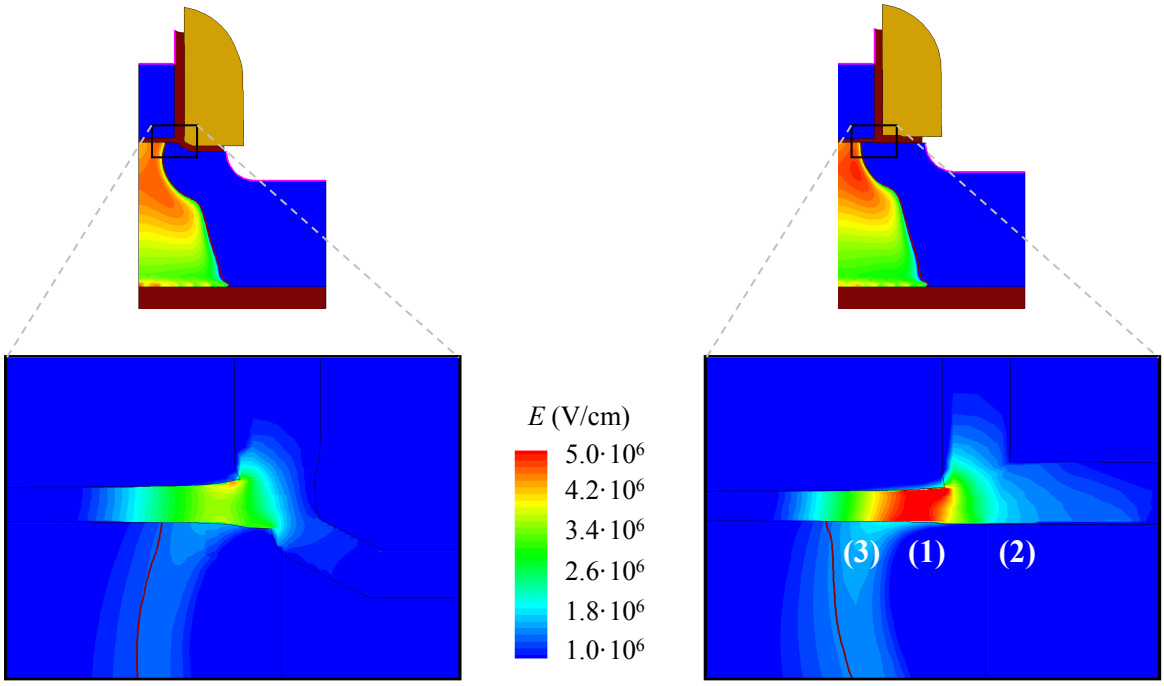


Bild 6.64: Berechnetes elektrisches Feld in Höhe der Überlappungsgebiete an der Si/SiO_2/Poly-Grenzfläche (a) mit und (b) ohne Spacer0-Ätzung

6.5.5 *In situ* bordotiertes Silizium-Germanium

Aufgrund der verringerten Diffusion beim Transistor ohne RTA bietet sich die Möglichkeit, *in situ* bordotiertes (ISBD) SiGe in die SD-Gebiete von p-MOSFETs einzubetten, was den Vorteil eines verringerten Anschlusswiderstands mit sich bringt und die Gefahr der Relaxation durch die beim herkömmlichen SiGe notwendige SD-Implantation verhindert.

Bild 6.65 zeigt das mit SIMS bestimmte Borprofil der epitaktisch abgeschiedenen ISBD-SiGe-Schicht mit einer Dotierstoffkonzentration von etwa 10^{20} Boratomen/cm³. Durch die Abscheidung lässt sich im Gegensatz zur Implantation mit seinem vorwiegend gaußförmigen Profil, ein sehr homogenes Dotierungsprofil über fast die gesamte Tiefe von circa 50 nm erzeugen.

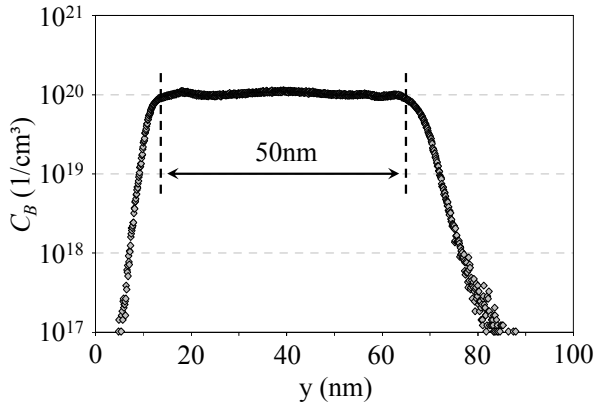


Bild 6.65: SIMS-Profil einer 50 nm dicken, epitaktisch abgeschiedenen ISBD-SiGe-Schicht

Für die folgenden Untersuchungen wurde ISBD-SiGe mit einem Germaniumgehalt von 20 % und einer Borkonzentration C_B von 10^{20} cm^{-3} verwendet. Als Referenz dient ein Transistor mit undotiertem, in die SD-Gebiete eingebettetem SiGe mit einem Germaniumgehalt von 22 % und einem SiGe-Gate-Abstand *DDS* von 100 Å. Die Mulden sind für alle untersuchten Transistoren 500 Å tief. Bild 6.66 zeigt die sich daraus ergebende Transistorstruktur. Alle weiteren Prozessparameter blieben unverändert und sind für die betrachteten Transistoren gleich.

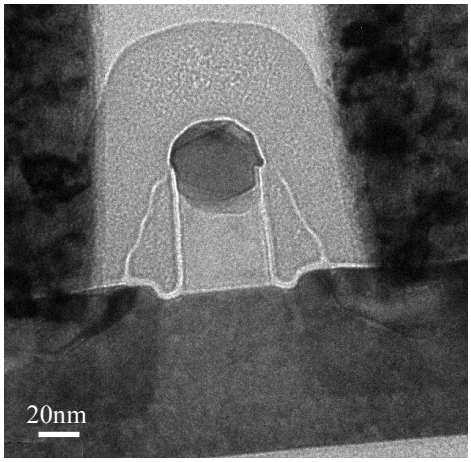


Bild 6.66: TEM-Aufnahme von p-MOSFETs mit ISBD-SiGe

a) Möglichkeiten der Implementierung

Im Gegensatz zur Integration von undotierten SiGe lässt sich ISBD-SiGe in zwei Varianten in den Herstellungsprozess einbinden (Tabelle 6.9). So kann das eSiGe-Modul nach den Halo- und SDE-Implantationen (Variante 1) oder wie bei undotiertem SiGe vor den Implantationen der Halo- und SDE-Gebiete in den Prozess implementiert werden (Variante 2). In beiden Varianten lassen sich auch die Halo- und SDE-Implantation entkoppeln. So kann z.B. die Halo-Implantation vor und die SDE-Implantation nach der SiGe-Abscheidung in den Herstellungsprozess eingebaut werden, was aber in Hinsicht auf eine erhöhte Parametervariation vermieden werden sollte.

Tabelle 6.9 Möglichkeiten der Integration von ISBD-SiGe in den Herstellungsprozess von diffusionsarmen p-MOSFETs

Variante 1	Variante 2
- Gateoxidation (10 Å) - Halo-/SDE-Implantation - eSiGe-Modul - Spacer1-Strukturierung	- Gateoxidation (10 Å) - eSiGe-Modul - Halo-/SDE-Implantation - Spacer1-Strukturierung

Bild 6.67a zeigt den Einfluss der Implantationsreihenfolge auf die MILLERkapazität C_{ov} für p-MOSFETs mit ISBD-SiGe. Der Abstand des ISBD-SiGe zum Gate beträgt 250 Å. Erfolgen die Halo- und SDE-Implantationen vor dem SiGe-Modul (Variante 1), ist eine deutlich erhöhte MILLERkapazität C_{ov} zu erkennen, was auf eine größere Überlappung der SDE-Gebiete unter dem Gate zurückzuführen ist. Da dies wiederum mit einer kleineren metallurgischen Gatelänge L_{met} einhergeht, verringert sich dementsprechend auch die Sättigungsschwellspannung $U_{th,sat}$ um etwa 400 mV (Bild 6.67b). Demgegenüber zeigt Variante 2 in Bezug auf die MILLERkapazität C_{ov} und die Sättigungsschwellspannung $U_{th,sat}$ ein unverändertes Transistorverhalten. Aus diesem Grund wurden alle weiteren Versuche mit ISBD-SiGe entsprechend Variante 2 prozessiert.

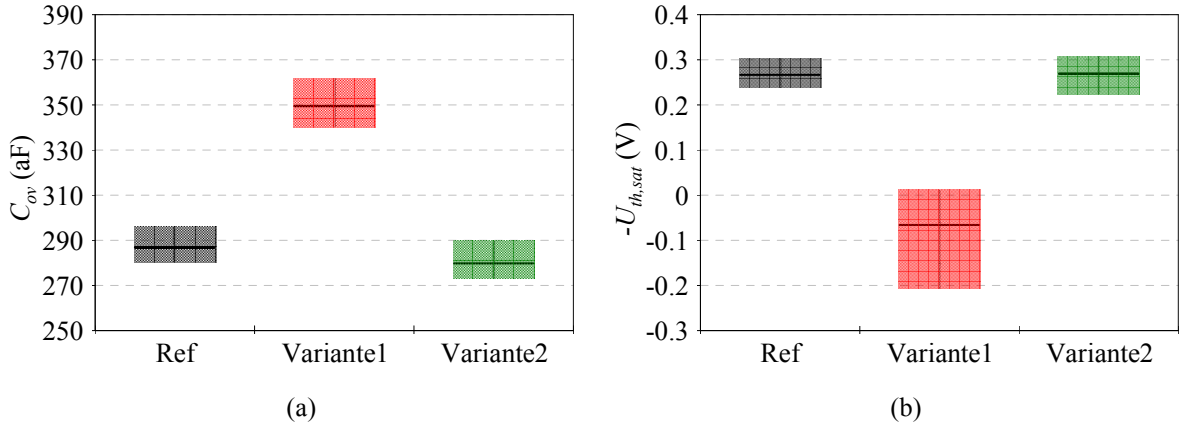


Bild 6.67: Einfluss der Implantationsreihenfolge auf (a) die MILLERkapazität C_{ov} und (b) die Sättigungsschwellspannung $U_{th,sat}$ für p-MOSFETs mit ISBD-SiGe

b) Variation des SiGe-Gate-Abstands

Im Folgenden wird der Einfluss des SiGe-Gate-Abstands *DDS* auf das Transistorverhalten von diffusionsarmen p-MOSFETs untersucht. Bei den dazu durchgeführten Experimenten wurde der Abstand des eingebetteten ISBD-SiGe zum Gate von 250 Å über 180 Å und 130 Å auf 100 Å verkleinert.

Bild 6.68a zeigt die Universalkurve für p-MOSFETs mit ISBD-SiGe in Abhängigkeit vom SiGe-Gate-Abstand *DDS*. Dabei kann man feststellen, dass p-MOSFETs mit einem SiGe-Gate-Abstand *DDS* von 250 Å ein vergleichbares Transistorverhalten wie der Referenztransistor mit einem SiGe-Gate-Abstand *DDS* von 100 Å aufweisen, obwohl die Ladungsträgerbeweglichkeit im Kanal entsprechend Bild 6.68b und Tabelle 6.10 geringer ist. Ursache für den gleichen Drainstrom I_D ist der verringerte externe Widerstand R_{ext} des Transistors aufgrund des bordotierten SiGe. Dieses Ergebnis bedeutet wiederum, dass bei den verwendeten Transistorabmessungen ein kleinerer externer Widerstand R_{ext} einen größeren Einfluss auf das Transistorverhalten hat als ein verringerter Kanalwiderstand R_{Kanal}. Weiterhin ist mit kleiner werdendem Abstand des SiGe zum Gate eine Erhöhung des Sättigungs-Drainstroms $I_{D,sat}$ sowie des Sperrstroms $I_{D,off}$ zu erkennen.

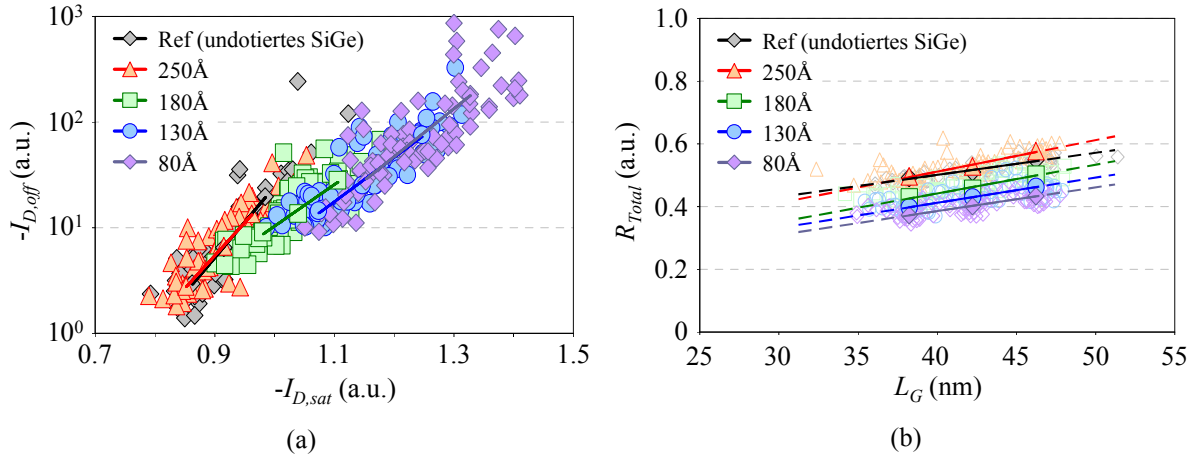


Bild 6.68: Einfluss des SiGe-Gate-Abstands *DDS* auf (a) die Universalkurve und (b) den Transistorwiderstand R_{Total} in Abhängigkeit von der Gatelänge L_G für p-MOSFETs mit ISBD- und undotiertem SiGe.

Tabelle 6.10 Anstiege *m* der in Bild 6.68b dargestellten Kurven sowie die daraus ermittelten Beweglichkeitsänderungen μ_{Ref}/μ_{DDS} der p-MOSFETs mit ISBD-SiGe im Vergleich zum Referenztransistor

Experiment	Ref (100 Å)	250 Å	180 Å	130 Å	100 Å
m (Ω/nm)	3.53	5.02	4.58	4.02	3.79
μ_{Ref}/μ_{DDS} (%)	-	-30	-23	-12	-7

Bild 6.69a zeigt dazu die Sättigungs-Drainstromänderung $\Delta I_{D,sat}$ in Abhängigkeit vom SiGe-Gate-Abstand *DDS* für p-MOSFETs mit ISBD- und undotiertem SiGe. Im Vergleich zum SiGe-Gate-Abstand *DDS* von 250 Å erhöht sich unter Verwendung von ISBD-SiGe der Sättigungs-Drainstrom $I_{D,sat}$ mit kleiner werdendem SiGe-Gate-Abstand *DDS* nahezu linear um etwa 12 %. Dies führt bei einem SiGe-Gate-Abstand *DDS* von 100 Å zu einem Drainstromgewinn von circa 38 %. Demgegenüber erhöht sich der Sättigungs-Drainstrom $I_{D,sat}$ bei undotiertem SiGe mit einem SiGe-Gate-Abstand *DDS* von 80 Å nur um etwa 10 % (Bild 6.41). Die Ursache für die unterschiedliche Drainstromänderung von undotiertem und ISBD-SiGe lässt sich im veränderten Bahnwiderstand der SDE- und SD-Gebiete zu finden. Aufgrund der sehr hohen und nahezu homogenen Dotierung der ISBD-SiGe-Gebiete verringert sich der externe Widerstand R_{ext} mit kleinerem SiGe-Gate-Abstand *DDS* der der Transistoren mit ISBD-SiGe.

Bild 6.69b zeigt den Sperrstrom $I_{D,off}$ in Abhängigkeit vom SiGe-Gate-Abstand *DDS* für p-MOSFETs mit ISBD- und undotiertem SiGe. Im Gegensatz zu undotiertem SiGe ist bei ISBD-SiGe eine stetige Erhöhung des Sperrstroms $I_{D,off}$ mit kleiner werdendem SiGe-Gate-Abstand zu erkennen, was sich durch den höheren *DIBL* erklären lässt (Bild 6.70a). Dieser Effekt lässt sich anhand der MILLERkapazität C_{ov} erklären, die eine ähnliche Abhängigkeit vom SiGe-Gate-Abstand *DDS* zeigt wie der *DIBL* (Bild 6.70b).

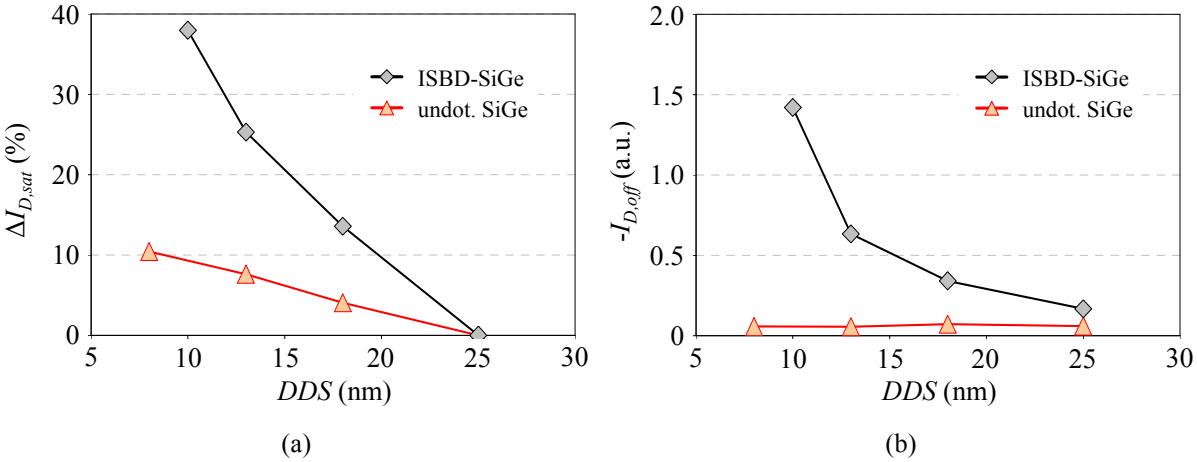


Bild 6.69: Einfluss des SiGe-Gate-Abstands *DDS* auf (a) die Sättigungs-Drainstromänderung $\Delta I_{D,sat}$ und (b) den Sperrstrom $I_{D,off}$ für p-MOSFETs mit ISBD- und undotiertem SiGe

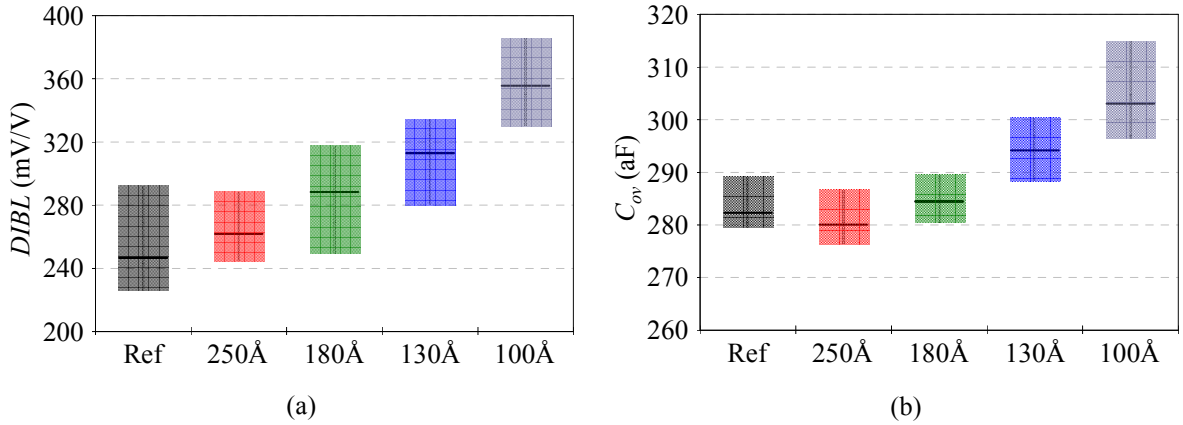


Bild 6.70: Einfluss des SiGe-Gate-Abstands *DDS* auf (a) den *DIBL* und (b) die MILLERkapazität C_{ov} für p-MOSFETs mit ISBD- und undotiertem SiGe (Ref)

Bild 6.71 zeigt den Zusammenhang zwischen berechnetem *DIBL*, berechneter MILLERkapazität C_{ov} und berechneter metallurgischer Gatelänge L_{met} für p-MOSFETs mit ISBD-SiGe entsprechend der Transistorstruktur von Bild 6.66. Analog zu den experimentellen Ergebnissen erhöht sich sowohl der *DIBL* als auch die MILLERkapazität C_{ov} mit sinkendem SiGe-Gate-Abstand *DDS* (Bild 6.71a). Betrachtet man die MILLERkapazität C_{ov} und die metallurgische Gatelänge L_{met} in Abhängigkeit vom SiGe-Gate-Abstand *DDS* (Bild 6.71b), kann man feststellen, dass sich die Überlappung mit kleinerem SiGe-Gate-Abstand *DDS* sogar verringert, was an der größeren metallurgischen Gatelänge L_{met} zu erkennen ist. Die geringere Überlappung mit kleinerem SiGe-Gate-Abstand wird durch die unterschiedlichen Diffusionseigenschaften von Bor in Si und SiGe verursacht. Abhängig vom SiGe-Gate-Abstand liegen die Implantationen der SDE-Gebiete mehr oder weniger im SiGe bzw. im Silizium. Da sich die Diffusionseigenschaften zwischen SiGe und Si sprunghaft ändern, ergibt sich eine starke Reaktion auf veränderte Abstände des SiGe zum Kanal, da die Gesamtmenge der Dotanden zu verschiedenen Anteilen im SiGe oder Si liegt. Mit kleinerem SiGe-Gate-Abstand *DDS* befinden sich weniger Dotanden im Silizium als im Vergleich zu einem größeren SiGe-Gate-Abstand *DDS*, wodurch sich die kleinere Überlappung und damit größere metallurgische Gatelänge ergibt. Dieses Ergebnis bedeutet aber auch, dass die MILLERkapazität C_{ov} durch andere Effekte beeinflusst wird.

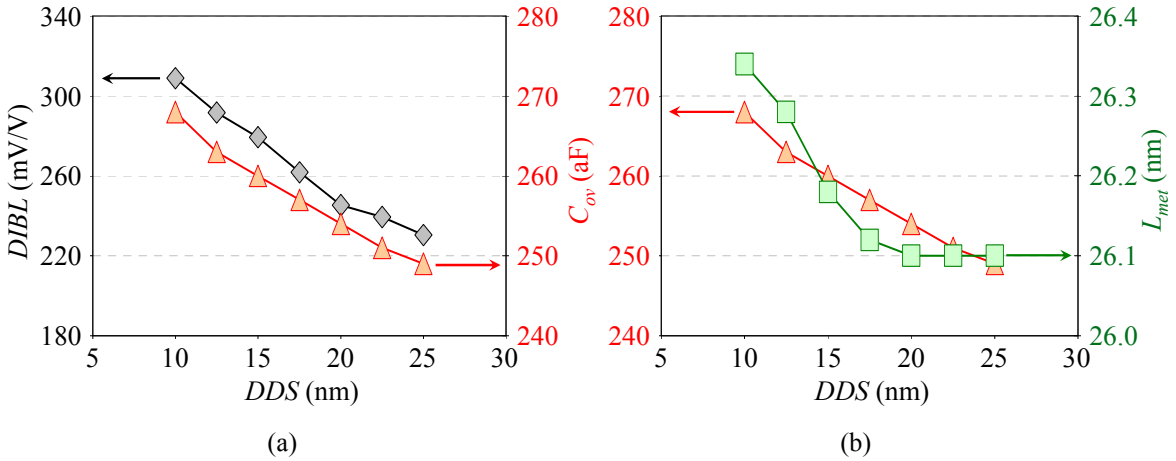


Bild 6.71: Zusammenhang zwischen SiGe-Gate-Abstand *DDS*, berechnetem *DIBL*, berechneter MILLERkapazität C_{ov} und berechneter metallurgischer Gatelänge L_{met} für p-MOSFETs mit ISBD-SiGe

Bild 6.72a zeigt die berechnete MILLERkapazität C_{ov} für p-MOSFETs mit ISBD- und undotiertem SiGe sowie für verschiedene Füllhöhen des ISBD-SiGe. Mit kleiner werdendem SiGe-Gate-Abstand *DDS* erhöht sich die MILLERkapazität C_{ov} bei größeren Füllhöhen unabhängig von der Dotierung. Erfolgt jedoch keine Überfüllung mit SiGe, bleibt die MILLERkapazität C_{ov} nahezu konstant. Diese Ergebnisse zeigen, dass die mit sinkendem SiGe-Gate-Abstand *DDS* steigende MILLERkapazität C_{ov} ihren Ursprung somit in der veränderten geometrischen Struktur hat, da sich der Anteil der äußeren, umsäumenden Kapazität an der MILLERkapazität C_{ov} erhöht.

Bild 6.72b stellt den *DIBL* für p-MOSFETs mit ISBD- und mit undotiertem SiGe sowie für verschiedene Füllhöhen des ISBD-SiGe dar. Im Gegensatz zur MILLERkapazität zeigt der *DIBL* eine Abhängigkeit von der Dotierung, aber keine von der Füllhöhe. Aufgrund der sehr hohen, aber vor allem tiefen Bordotierung der näher am Gate liegenden SiGe-Gebiete erreichen die Feldlinien der drainseitigen Raumladungszone bei $U_{DS} = 1.0$ V die sourceseitige Raumladungszone. Damit vermindert sich die Potenzialbarriere φ_{SD} zwischen Source und Drain im Kanalgebiet, wodurch die Kanalladung und der Drainstrom I_D nicht mehr vom Gate allein, sondern maßgeblich mit von der Drain-Source-Spannung U_{DS} beeinflusst wird. Durch eine Anpassung der Halo-Implantationen kann die Potenzialbarriere φ_{SD} wieder erhöht werden, wodurch der Sperrstrom $I_{D,off}$ verringert wird. Gleichzeitig wird damit aber ein steilerer pn-Übergang erzeugt, der höhere Tunnelströme verursacht und somit wiederum den Sperrstrom $I_{D,off}$ anhebt.

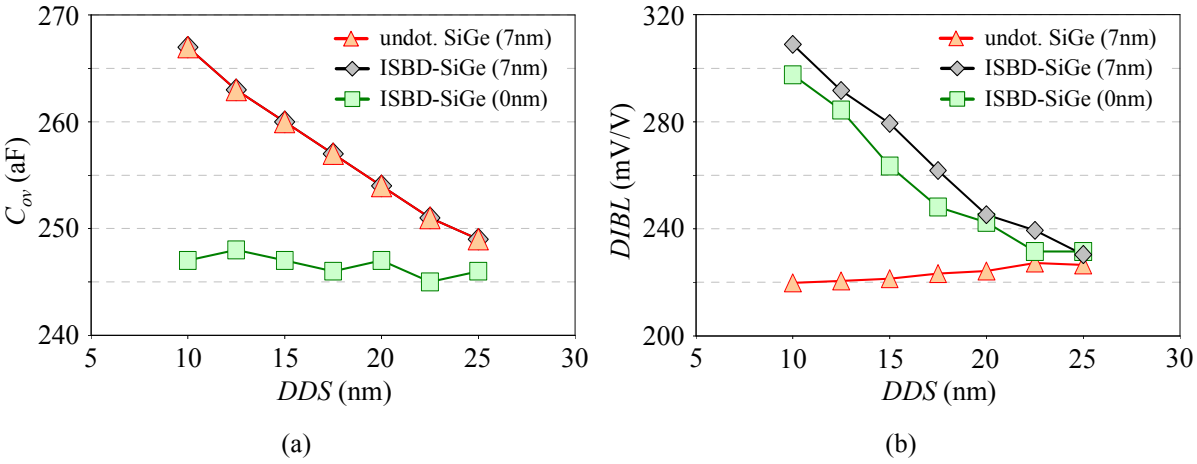


Bild 6.72: Einfluss des SiGe-Gate-Abstands *DDS* und der ISBD-Füllhöhe auf (a) die berechnete MILLERkapazität C_{ov} und (b) den berechneten *DIBL* für p-MOSFETs mit ISBD- und mit undotiertem SiGe

c) Einfluss einer zusätzlichen SD-Implantation

Im Gegensatz zu p-MOSFETs mit undotiertem SiGe ist es bei p-MOSFETs mit ISBD-SiGe möglich, den Einfluss der SD-Implantation auf mögliche Relaxationen der SiGe-Schicht zu untersuchen. Bild 6.73a zeigt dafür die Universalkurve mit und ohne SD-Implantation. Durch die SD-Implantation verringert sich der Sättigungs-Drainstrom $I_{D,sat}$ bei einem konstanten Sperrstrom $I_{D,off}$ um etwa 8 %. Ursache dafür ist die verringerte Verspannungseinkopplung vom SiGe zum Kanal. Bild 6.73b zeigt dazu den Transistorwiderstand R_{Total} in Abhängigkeit von der Gatelänge L_G mit und ohne SD-Implantation. Aufgrund der unterschiedlichen Prozessierung der Transistoren verringert sich mit der SD-Implantation die Ladungsträgerbeweglichkeit im Kanal um annähernd 20 %, was sich wiederum in der geringeren Leistungsfähigkeit widerspiegelt.

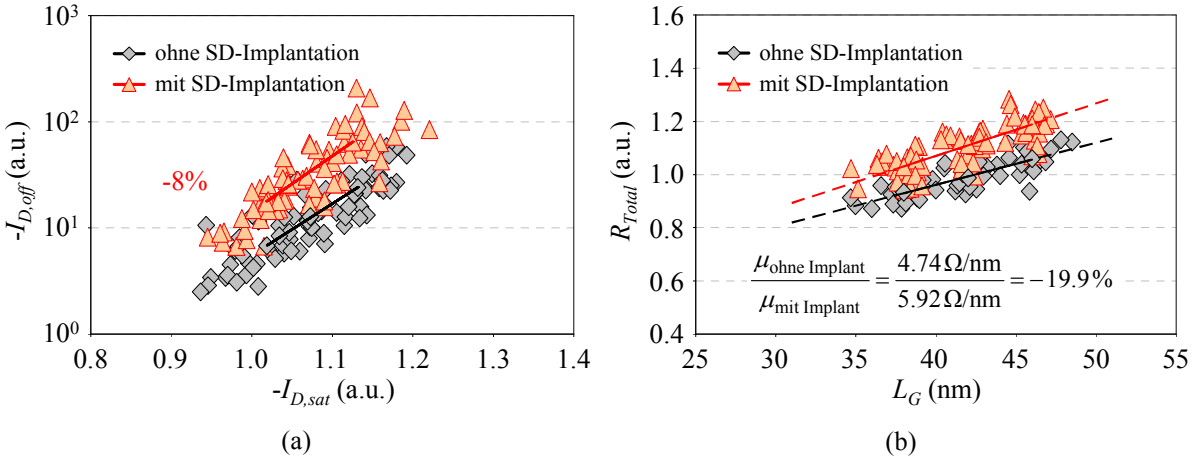


Bild 6.73: Einfluss einer SD-Implantation auf (a) die Universalkurve und (b) den Transistorwiderstand R_{Total} in Abhängigkeit von der Gatelänge L_G für p-MOSFETs mit ISBD-SiGe.

6.5.6 Zusammenfassung

Aus den Ergebnissen dieses Abschnitts kann man für p-MOSFETs mit SiGe, die nur eine Kurzzeitausheilung, aber keine RTA erfahren haben, folgende Schlussfolgerungen ziehen:

- Das Einbetten von SiGe in die SD-Gebiete von diffusionsarmen p-MOSFETs ist wie beim RTA-Prozess eine Möglichkeit, die Leistungsfähigkeit des Transistors in Abhängigkeit vom SiGe-Gate-Abstand *DDS* weiter zu verbessern.
- In Bezug auf die Leistungsfähigkeit von p-MOSFETs mit SiGe ist die Wahl der richtigen Implantationsspezies von entscheidender Bedeutung.
- Bei p-MOSFETs ohne Spacer0-Ätzung kann aufgrund der verbesserten Verspannungseinkopplung ein höherer Sättigungs-Drainstrom $I_{D,sat}$ erreicht werden. Die damit einhergehende größere MILLERkapazität C_{ov} verringert jedoch etwas den Gewinn in der Leistungsfähigkeit im dynamischen Betrieb.
- Durch die Verwendung von ISBD-SiGe kann die Leistungsfähigkeit von diffusionsarmen p-MOSFETs weiter verbessert werden. Nachteilig auf das Transistorverhalten wirkt sich jedoch der mit sinkendem SiGe-Gate-Abstand *DDS* stark erhöhte Sperrstrom $I_{D,off}$ aus.
- Die SD-Implantation verursacht eine teilweise Relaxation der SiGe-Schicht und verringert dadurch die Leistungsfähigkeit.

6.6 Einbau von verspannten Silizium-Kohlenstoff-Schichten

6.6.1 Vorbetrachtungen

So wie SiGe bei p-MOSFETs durch die uniaxiale Druckverspannung in Kanalrichtung die Löcherbeweglichkeit μ_p erhöht, lässt sich theoretisch das Gleiche mit SiC bei n-MOSFETs erreichen. In die SD-Gebiete eingebettetes SiC induziert eine uniaxiale Zugverspannung in Kanalrichtung, wodurch sich nach Abschnitt 4.3.3 die Elektronenbeweglichkeit μ_n im Kanal erhöht. Die Implementierung dieser Verspannungsquelle ist jedoch aufgrund der starken thermischen Instabilität dieses Verbindungshalbleiters eine große Herausforderung.

Die Festkörperlöslichkeit von substitutionell in die Kristallstruktur von Silizium eingebautem Kohlenstoff liegt im Temperaturbereich von (1200…1400) °C bei etwa (10^{-3}…10^{-4}) % [129], was eine unzureichende Kohlenstoffkonzentration C_C darstellt, um den Transistorkanal zur Erhöhung der Ladungsträgerbeweglichkeit ausreichend genug zu verspannen. Abhilfe schafft hier die epitaktische Erzeugung von SiC-Schichten, wodurch die Festkörperlöslichkeit von substitutionell gelöstem Kohlenstoff um mehr als das 10^4 fache erhöht werden kann [130]. Mit Hilfe des während dieser Arbeit zur Verfügung stehenden Epitaxiesystems war es möglich, verspannte SiC-Schichten mit 1.8 % C auf Gitterplätzen herzustellen, was nach Gleichung (4.8) einer äquivalenten Gitterverformung von etwa 14.5 % Germanium im SiGe entspricht. Dieser Wert wurde auf unstrukturierten Silizium-Wafern indirekt über Röntgenbeugungsmessungen bestimmt, die eine Deformation der Gitterkonstanten bezüglich der unverspannten Siliziumzelle nachweist.

Im Herstellungsprozess eines Transistors wird der epitaktisch erzeugte SiC-Film verschiedenen thermischen Ausheilprozessen ausgesetzt. Um diese relativ hohe Konzentration an metastabil auf substitutionellen Gitterplätzen eingebautem Kohlenstoff beizubehalten, muss das thermische Budget aller nachfolgenden Prozessschritte, vor allem der finalen Ausheilung, minimiert werden, da dieses metastabil kohlenstoffdotierte Mischkristall eine starke thermische Instabilität aufweist [131]. Ein großes thermisches Budget kann die Verlagerung von substitutionell eingebautem Kohlenstoff auf Zwischengitterplätze zur Folge haben, wodurch wiederum die Verformung sowie die Qualität des Kristalls negativ beeinflusst werden.

Bild 6.74 zeigt in diesem Zusammenhang die Abhängigkeit der Kohlenstoffkonzentration C_C vom thermischen Budget des verwendeten Ausheilverfahrens. Ausgangspunkt bzw. Referenz ist die Kohlenstoffkonzentration C_C von 1.8 % direkt nach der Abscheidung der SiC-Schicht auf unstrukturierten Silizium-Wafern. Die Beurteilung der Ausheilung auf die Verspannung des SiC-Films erfolgte durch eine Röntgenbeugungsmessung. Unter Verwendung der herkömmlichen RTA mit einer Ausheiltemperatur von 1070 °C (RTA1) erfolgt eine massive Verringerung der Konzentration an substitutionellem Kohlenstoff auf 0.93 %. Selbst eine Reduzierung der Ausheiltemperatur auf 1000 °C (RTA2) verringert die Kohlenstoffkonzentration C_C auf 1.65 %. Erst ab einer Ausheiltemperatur von 950 °C (RTA3) wird die Konzentration an substitutionellem Kohlenstoff im SiC-Film nicht durch das thermische Budget der Ausheilung beeinträchtigt. Hervorzuheben ist jedoch der Vorteil der Kurzzeitausheilung als alleiniger Ausheilschritt in Verbindung mit der SiC-Epitaxie. Durch die Kurzzeitausheilung erhöht sich sogar die Konzentration an Kohlenstoff um 0.12 % auf 1.92 %, da auf Zwischengitterplätzen gelagerter Kohlenstoff zu substitutionell auf Gitterplätzen eingebautem Kohlenstoff umgewandelt wird [132], [133].

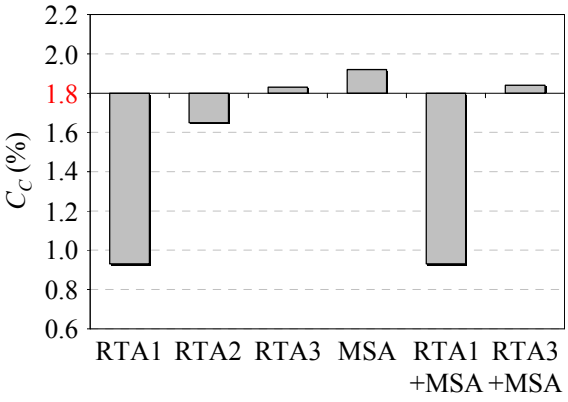


Bild 6.74: Gemessene Kohlenstoffkonzentration C_C in Abhängigkeit von der Ausheilung (RTA1 = 1070 °C, RTA2 = 1000 °C, RTA3 = 950 °C)

Wird die Kurzzeitausheilung jedoch als zusätzlicher Ausheilschritt nach der RTA durchgeführt (RTA1+MSA, RTA3+MSA), hat dies im Vergleich zur Ausheilung nur mit der RTA (RTA1, RTA3) unabhängig von deren Ausheiltemperatur keinen bzw. nur einen sehr geringen Einfluss auf die Konzentration von Kohlenstoff auf Gitterplätzen. Diese Ergebnisse zeigen, dass die Kurzzeitausheilung ohne RTA vorteilhaft für diese Verspannungstechnologie ist. Weiterhin hat die Stärke der Implantationsschäden wie bei der SiGe-Schicht einen wesentlichen Einfluss auf die Relaxation Δa des SiC-Films. Bild 6.75 stellt dazu den Relaxationsgrad Δa der SiC-Schicht für verschiedene Implantationen mit anschließender Kurzzeitausheilung dar. Referenz ist die SiC-Epitaxie ohne Implantation mit anschließender Kurzzeitausheilung. Sowohl für die SDE- als auch für die Halo-Implantation ist nur eine geringe Relaxation von 2.1 % bzw. 7.3 % festzustellen. Demgegenüber hat eine tiefe SD-Implantation eine fast 35 %ige Relaxation Δa des SiC-Films zur Folge.

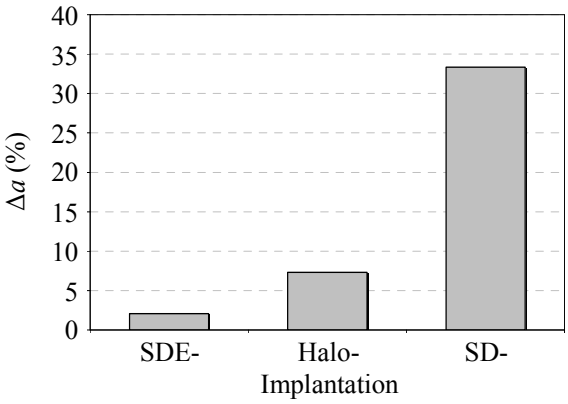


Bild 6.75: Relaxation Δa der SiC-Schicht in Abhängigkeit von der Implantation mit nachfolgender Kurzzeitausheilung

Da aber das zur Verfügung stehende Epitaxiesystem in der Lage ist, *in situ* phosphordotiertes (ISPD-) SiC abzuscheiden, kann auf die ansonsten notwendige SD-Implantation verzichtet werden. Durch die Abscheidung lässt sich wie im Fall des ISBD-SiGe eine sehr homogene Phosphordotierung über die entsprechende Muldentiefe erzeugen. Für die folgenden Untersuchungen wurde ISPD-SiC mit einem Kohlenstoffgehalt von 1.8 % und einer Dotierstoffkonzentration von etwa $4 \cdot 10^{20}$ Phosphoratomen/cm³ verwendet. Der SiC-Gate-Abstand *DDS* ist 130 Å und die Tiefe der Mulden beträgt 500 Å. Alle anderen Prozessparameter blieben unverändert und sind für die betrachteten Transistoren gleich.

6.6.2 Experimentelle Ergebnisse

Wie im Fall des ISBD-SiGe bei p-MOSFETs bestehen analog zu Tabelle 6.9 zwei Möglichkeiten der Einbindung der ISPD-SiC-Epitaxie in den Herstellungsprozess von n-MOSFETs, entweder nach den Halo- und SDE-Implantationen (Variante 1) oder vor den Implantationen der Halo- und SDE-Gebiete (Variante 2).

Bild 6.76a zeigt den Einfluss der Implantationsreihenfolge auf die MILLERkapazität C_{ov} für n-MOSFETs mit ISPD-SiC. Die Halo- und SDE-Implantationen nach Variante 1 rufen eine deutlich erhöhte MILLERkapazität C_{ov} hervor, was auf eine größere Überlappung der SDE-Gebiete unter dem Gate zurückzuführen ist. Da dies wiederum mit einer kleineren metallurgischen Gatelänge L_{met} einhergeht, verringert sich dementsprechend auch die Sättigungsschwellspannung $U_{th,sat}$ um etwa 200 mV (Bild 6.76b). Demgegenüber zeigt die Implementierung des SiC-Moduls nach Variante 2 eine zum Referenztransistor ohne SiC leicht verringerte MILLERkapazität C_{ov}, wodurch sich auch die erhöhte Sättigungsschwellspannung $U_{th,sat}$ erklären lässt.

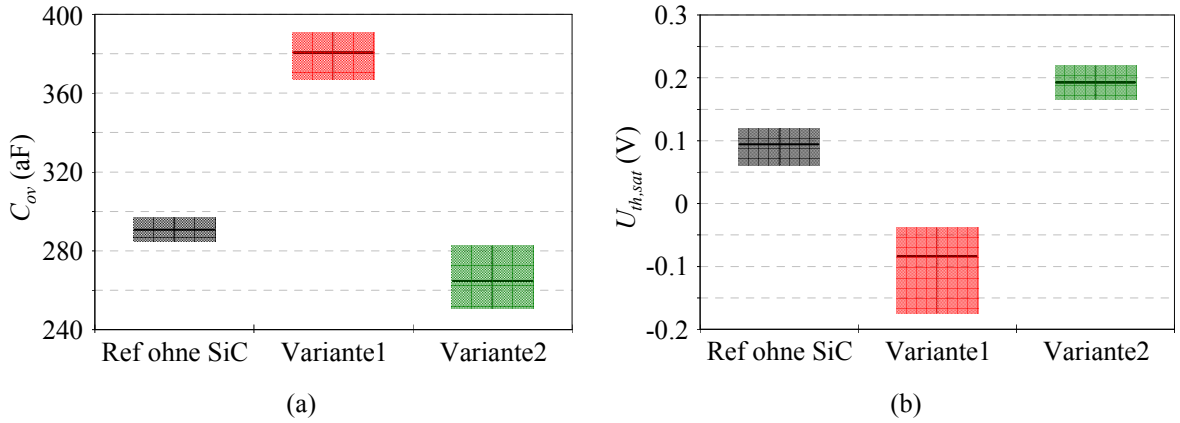


Bild 6.76: Einfluss der Implantationsreihenfolge auf (a) die MILLERkapazität C_{ov} und (b) die Sättigungsschwellspannung $U_{th,sat}$ für n-MOSFETs mit ISPD-SiC

Bild 6.77 zeigt für die beiden unterschiedlichen Implementierungsmöglichkeiten das Universalkurvenverhalten im Vergleich zum Referenztransistor ohne SiC. Aufgrund der unterschiedlichen Schwellspannungen $U_{th,sat}$ ist ein direkter Vergleich zwischen den Transistoren mit und ohne SiC nicht möglich. Betrachtet man jedoch die Regressionskurve (Strichlinie) zwischen den unterschiedlichen SiC-Implementierungsmöglichkeiten, würde eine Anpassung der Sättigungsschwellspannung $U_{th,sat}$ an die des Referenztransistors eine Verringerung des Sättigungs-Drainstroms $I_{D,sat}$ bei konstantem Sperrstrom $I_{D,off}$ von etwa 4% bedeuten. Diese Ergebnisse bedeuten jedoch auch, dass entweder der Transistorkanal nicht durch den in die SD-Gebiete eingebetteten Mischkristall verspannt oder die Verspannung des Kanals durch andere Effekte überlagert wird.

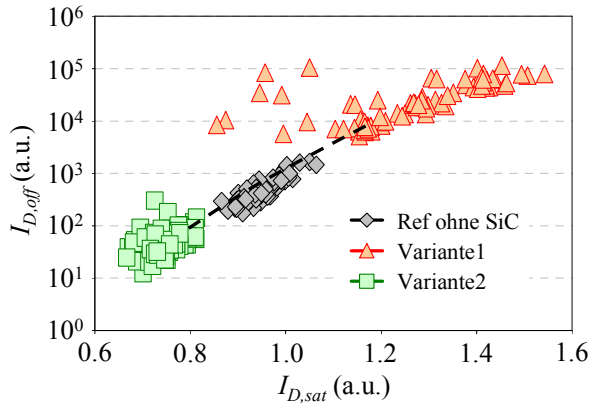


Bild 6.77: Einfluss der Implantationsreihenfolge auf die Universalkurve von n-MOSFETs mit ISPD-SiC

Zur Untersuchung dieser Problematik wurden weitere Experimente mit in den SD-Gebieten des Transistors eingebettetem SiC nach Variante 2 durchgeführt. Dabei wurde der Einfluss einer tiefen PAI sowie einer zusätzlich unter einem Winkel von 30 ° durchgeführten Phosphor-SD-Implantation untersucht. Obwohl aus den Ergebnissen der Relaxationsexperimente auf unstrukturierten Wafern eine deutliche Relaxation der SiC-Schicht mit einer zusätzlichen SD-Implantation erkennbar ist (Bild 6.75), zeigen die Universalkurven mit und ohne PAI keine Unterschiede im Transistorverhalten (Bild 6.78a). Demgegenüber wird der Sättigungs-Drainstrom $I_{D,sat}$ durch die zusätzliche Phosphor-SD-Implantation sogar geringfügig erhöht. Diese Ergebnisse weisen darauf hin, dass der Transistorkanal durch das SiC nicht verspannt wird.

Die Darstellung des Transistorwiderstands R_{Total} über der Gatelänge L_G zeigt ebenfalls keinen Hinweis auf eine veränderte Elektronenbeweglichkeit im Kanal und damit ebenfalls nicht auf eine mögliche Kanaldeformation, da alle Kurven der Transistoren mit SiC im Vergleich zum Referenztransistor ohne SiC den gleichen Anstieg m aufweisen (Bild 6.78b). Die Ursache für dieses Verhalten ist vermutlich auf Kristall-defekte an der vertikalen Grenzfläche zwischen Si und SiC und den damit verbundenen Problemen beim SiC-Wachstum zurückzuführen. Gleichzeitig erhöht sich der parasitäre SD-Widerstand der Transistoren mit SiC um etwa 25 Ω (Variante 1, Variante 2, PAI), wodurch auch die verringerte Leistungsfähigkeit in Bild 6.77 begründet ist. Dies ist erstaunlich, da aufgrund der sehr hohen und zudem über die Tiefe der SiC-SD-Gebiete homogenen ISPD die Schichtwiderstände R_{SDE} der Aktivgebiete der Transistoren mit SiC im Vergleich zu den Transistoren ohne SiC nur 1/6-mal so groß sind (Bild 6.79).

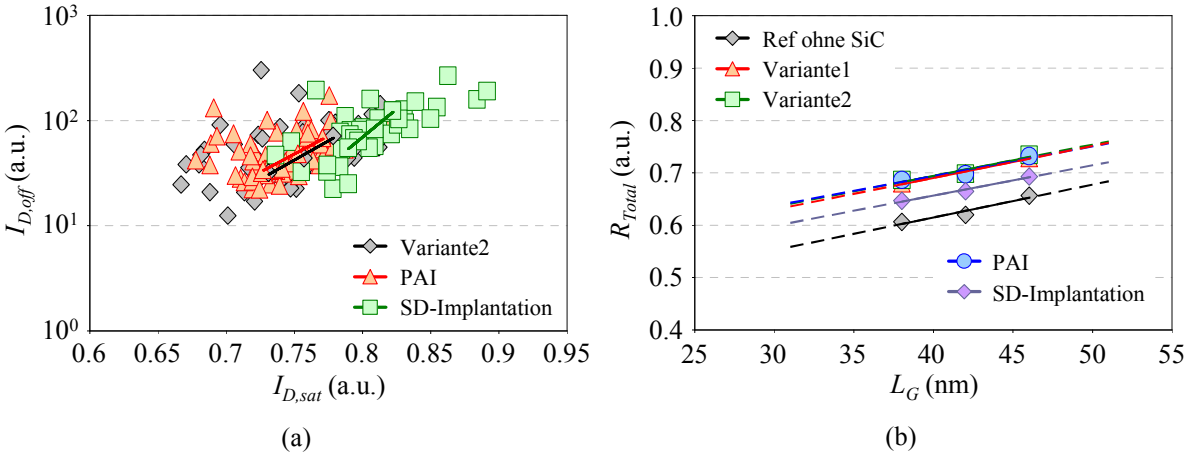


Bild 6.78: Einfluss einer zusätzlichen PAI bzw. SD-Implantation auf (a) die Universalkurve und (b) den Transistorwiderstand R_{Total} in Abhängigkeit von der Gatelänge L_G für n-MOSFETs mit ISPD-SiC.

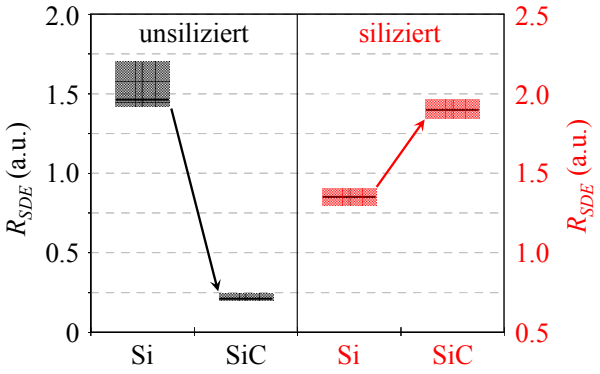


Bild 6.79: Unsilizierter und silizierter Schichtwiderstand R_{SDE} der SDE-Gebiete von n MOSFETs mit und ohne ISPD-SiC

Der silizierte Schichtwiderstand R_{NiSi} der Aktivgebiete mit SiC ist dagegen etwa 40 % höher (Bild 6.79), was auf Probleme bei der Silizierung des kohlenstoffdotiertem Siliziums hindeutet, obwohl SiC im Vergleich zu Silizium eine geringere Barrierenhöhe $\Phi_{B,n}$ aufweist [134], was wiederum nach Gleichung (2.14) einen kleineren Übergangswiderstand $R_{Silizid}$ zur Folge haben sollte. Dieses Verhalten lässt sich nur durch eine veränderte Silizierung mit schlechteren elektrischen Eigenschaften erklären, was schließlich den Übergangswiderstand $R_{Silizid}$ verändert und somit den Bahn- und folglich den externen Widerstand R_{ext} erhöht. Die zusätzliche unter einem Winkel durchgeführte Phosphor-implantation verringert den Übergangswiderstand $R_{Silizid}$ zum Silizid (Abschnitt 6.3.3), wodurch auch der geringfügig höhere Sättigungs-Drainstrom $I_{D,sat}$ in Bild 6.78a zu erklären ist.

Aus diesen Ergebnissen ist zu erkennen, dass anders als beim SiGe bei p-MOSFETs durch das Einbetten von SiC in die SD-Gebiete von n-MOSFETs keine Verbesserung in der Leistungsfähigkeit des Transistor erreicht wird. Obwohl im Vergleich zu einem Germaniumgehalt von 22 % im SiGe durch einen Kohlenstoffgehalt von 1.8 % im SiC nur 66 % der im Kanal entstehenden Gitterverformung erreicht werden kann und sich der Bahnwiderstand durch das Einbetten des SiC erhöht, müsste dennoch ein Unterschied in der Verspannung des Kanals anhand der dR/dL-Methode zu erkennen sein. Wahrscheinlichste Ursache für dieses Verhalten ist, dass bereits während der Epitaxie eine Relaxation des SiC in den SD-Gebieten erfolgt, was wiederum den ausbleibenden Effekt der PAI erklärt. Die Implementierung dieser Verspannungsquelle ist und bleibt somit eine große Herausforderung.

6.7 Zusammenfassung

In diesem Kapitel wurde der Einfluss der Kurzzeitausheilung als alleiniger Ausheilschritt auf verschiedene Transistorparameter von diffusionsarmen n- und p-MOSFETs untersucht. Bei ersten Experimenten war festzustellen, dass der Verzicht auf die RTA erhebliche Beeinträchtigungen in der Leistungsfähigkeit und beim Kurzkanalverhalten bei beiden Transistortypen erbrachte. Eine Ursache hierfür ist das stark veränderte Dotierungsprofil im Vergleich zu einem Transistor mit einem RTA-Schritt. Aus diesem Grund ist eine Umgestaltung des Dotierungsprofils beim Übergang zum diffusionsarmen Transistor notwendig. Die dazu durchgeführten Untersuchungen zeigten, dass zur Verbesserung des Kurzkanalverhaltens vor allem bei SOI-Transistoren eine kombinierte Optimierung des Dotierungsprofils der Wannen-, SDE-, SD- und Halo-Gebiete erfordert und nicht nur durch eine einfache Anpassung der Halo-Implantationen erreicht werden kann. Außerdem zeigten Experimente und Simulationen, dass eine zusätzliche Implantation in die SD-Gebiete des Transistors unter einem Winkel von 30 ° den externen Widerstand R_{ext} sowohl von n- als auch von p-MOSFETs verringert und dadurch die Leistungsfähigkeit beider Transistortypen erheblich erhöht. Weiterhin war zu beobachten, dass im Gegensatz zum herkömmlichen RTA-Prozess alternative Implantationsspezies für eine BF_2- und Arsen-Dotierung verwendet werden können, ohne die Leistungsfähigkeit der diffusionsarmen Transistoren negativ zu beeinflussen.

Neben der Umgestaltung des Dotierungsprofils muss bei einem diffusionsarmen Transistor das thermische Budget jedes einzelnen Prozessschritts während der Herstellung beachtet werden, um eine unnötige Diffusion oder Deaktivierung der eingebrachten Dotanden zu vermeiden. Selbst ein Diffusionsschritt mit einer Ausheiltemperatur von 600 °C und einer Ausheilzeit von 240 s, der bei einem herkömmlichen RTA-Prozess aufgrund der Erzeugung einer mechanischen Verspannung im Transistorkanal zu einer Erhöhung der Leistungsfähigkeit des n-MOSFETs führt, hat einen negativen Einfluss auf das elektrische Verhalten eines diffusionsarmen Transistors.

Bei einem Herstellungsprozess nur mit einer Kurzzeitausheilung als finalem Ausheilschritt führt diese zusätzliche Temperaturbehandlung sowohl bei n- als auch bei p-MOSFETs zu einem verringerten Sättigungs-Drainstrom $I_{D,sat}$ bei gleichzeitig stark erhöhter MILLERkapazität C_{ov}.

Eine weitere Herausforderung bei der Herstellung von diffusionsarmen Transistoren ist die Implementierung von eingebetteten Verspannungsquellen, um die Leistungsfähigkeit auch bei diesem Transistortyp zu erhöhen. Aus den Ergebnissen zum p-MOSFET mit SiGe kann man ableiten, dass diese Verspannungstechnik wie beim RTA-Prozess eine Möglichkeit ist, die Leistungsfähigkeit des Transistors in Abhängigkeit vom SiGe-Gate-Abstand *DDS* weiter zu verbessern. Bei diffusionsarmen p-MOSFETs mit eingebettetem SiGe ist jedoch im Gegensatz zum RTA-Prozess die Wahl der richtigen Implantationsspezies von entscheidender Bedeutung.

Eine zu starke Zerstörung der Kristallstruktur des SiGe-Gebiets z.B. infolge einer amorphisierenden BF_2- oder $B_{18}H_{22}$-Implantation führt im Vergleich zu einer Standard-Borimplantation zu einer verringerten Ladungsträgerbeweglichkeit im Kanalgebiet des Transistors und somit zu einem kleineren Drainstrom I_D. Weiterhin war festzustellen, dass unter Vernachlässigung von Ätzschritten ein höherer Drainstrom aufgrund einer verbesserten Verspannungseinkopplung erreicht werden kann. Eine weitere Steigerung der Leistungsfähigkeit lässt sich durch den Einsatz von *in situ* bordotiertem SiGe erreichen, da neben der höheren Ladungsträgerbeweglichkeit im Kanal auch der Bahn- und damit der externe Widerstand R_{ext} durch die *in Situ* Bordotierung der eingebetteten SiGe-Gebiete verringert wird.

So wie SiGe bei p-MOSFETs durch die uniaxiale Druckverspannung in Kanalrichtung die Löcherbeweglichkeit μ_p erhöht, lässt sich das Gleiche mit SiC bei n-MOSFETs erreichen. Im Gegensatz zum SiGe weist der kohlenstoffdotierte Mischkristall jedoch eine hohe thermische Instabilität auf, was die Implementierung dieser Verspannungsquelle in den herkömmlichen Herstellungsprozess stark einschränkt. Ergebnisse auf unstrukturierten Wafern mit epitaktisch erzeugten, *in situ* phosphordotierten SiC-Schichten zeigen, dass eine RTA-Ausheiltemperatur von mehr als 950 °C den Anteil an substitutionell eingebauten Kohlenstoff stark verringert. Demgegenüber kann beim Verzicht der RTA nur unter alleiniger Verwendung der Kurzzeitausheilung dieser Anteil sogar erhöht werden, damit der Ansatz der Transistorherstellung nur mit Kurzzeitausheilung für diese Verspannungstechnologie von Vorteil ist. Das Einbetten von ISPD-SiC in die SD-Gebiete von n-MOSFETs erweist sich jedoch als eine große Herausforderung, so dass im Vergleich zum diffusionsarmen Referenztransistor ohne SiC noch keine verbesserten Transistorergebnisse erzielt werden konnten.

Zusammenfassend gibt Tabelle 6.11 noch einmal einen Überblick über verschiedene Optimierungsmöglichkeiten für diffusionsarme n- und p-MOSFETs.

Tabelle 6.11 Optimierungsmöglichkeiten für diffusionsarme n- und p-MOSFETs

Transistortyp	n-MOSFETs	p-MOSFETs
Kurzkanalverhalten	- kombinierte Optimierung der Wannen-, SDE-, SD- und Halo-Implantation	
Leistungsfähigkeit	- zusätzliche, gewinkelte SD-Implantation	- zusätzliche, gewinkelte SD-Implantation - SiGe oder ISBD-SiGe
Implantationsspezies (Majoritäten)	- As, P	- B, BF_2, $B_{18}H_{22}$ (Si) - B (SiGe)

Die Bilder 6.80 und 6.81 zeigen abschließend dazu die Universalkurven bzw. die Sättigungsschwellspannung $U_{th,sat}$ in Abhängigkeit von der Gatelänge L_G für optimierte diffusionsarme n- und p-MOSFETs im Vergleich zu einem Transistor der 45 nm-Technologie. Letzterer basiert auf einem Prozess mit SiGe beim p-MOSFET sowie einer kombinierten Ausheilung von RTA und Kurzzeitausheilung. Die Ausheiltemperatur der RTA beträgt 1070 °C. Im Fall von diffusionsarmen n-MOSFETs konnte durch eine Optimierung der Implantationsparameter und Spacer-Strukturierung eine zum RTA-Prozess vergleichbare Leistungsfähigkeit erreicht werden. Bei diffusionsarmen p-MOSFETs konnte durch die Implementierung von ISBD-SiGe sowie der Anpassung des Prozessablaufs und der Implantationsparameter die Leistungsfähigkeit des Transistors soweit gesteigert werden, dass dessen Sättigungs-Drainstrom $I_{D,sat}$ im Vergleich zum RTA-Transistor nur um etwa 7 % verringert ist.

Das Kurzkanalverhalten ist für beide diffusionsarmen Transistortypen im Vergleich zum jeweiligen RTA-Transistor stark verbessert (Bild 6.81), wodurch die physikalische Gatelänge L_G verkleinert werden kann. Diese Ergebnisse zeigen, dass eine weitere Skalierung der planaren Transistorstruktur mit Hilfe diffusionsarmer Transistoren, bei gleicher Leistungsfähigkeit im Fall von n-MOSFETs und nur geringfügig kleinerer Leistungsfähigkeit im Fall von p-MOSFETs, möglich ist.

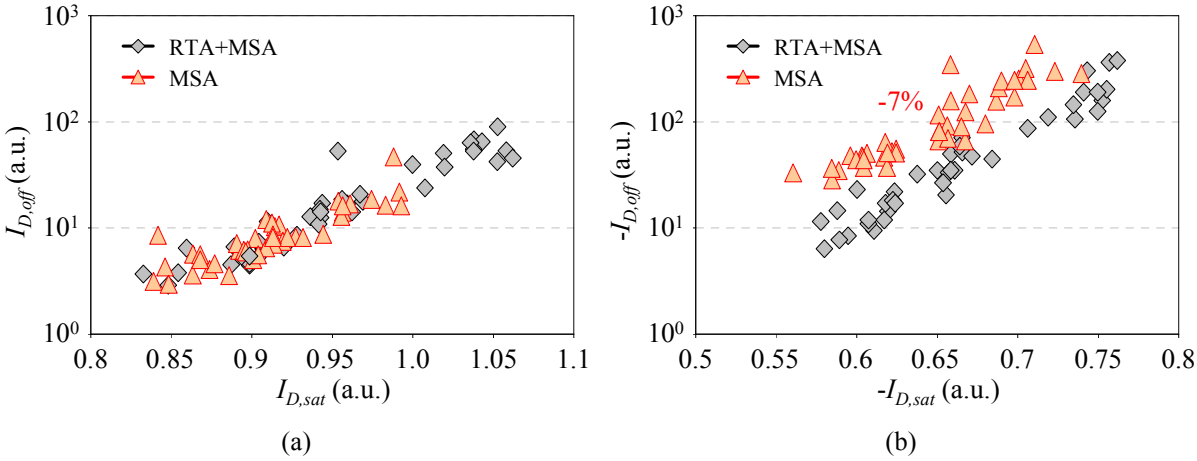


Bild 6.80: Einfluss unterschiedlicher Ausheilregime mit RTA und MSA auf die Universalkurve von (a) n- und (b) p-MOSFETs

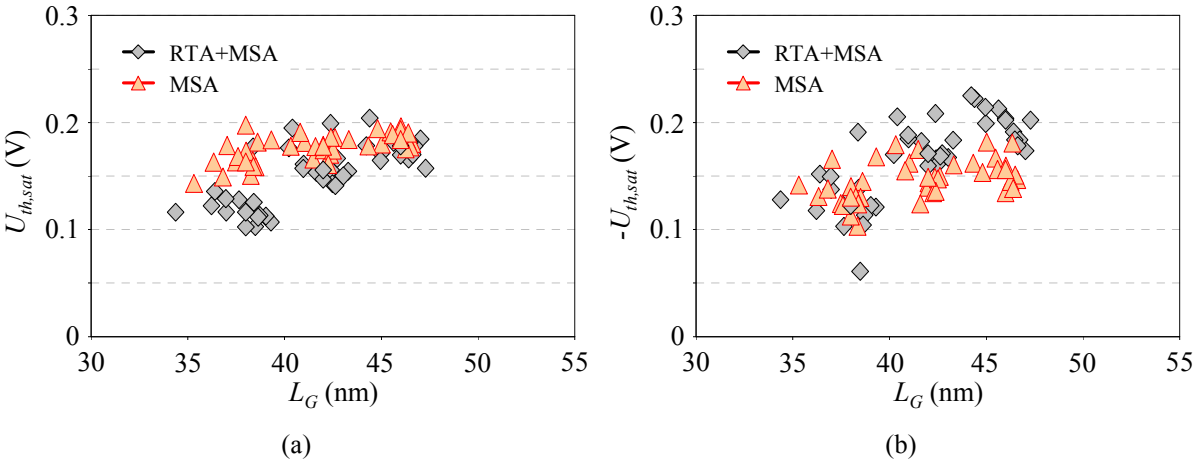


Bild 6.81: Einfluss unterschiedlicher Ausheilregime mit RTA und MSA auf die Sättigungsschwellspannung $U_{th,sat}$ in Abhängigkeit von der Gatelänge L_G von (a) n- und (b) p-MOSFETs

7 Zusammenfassung und Ausblick

7.1 Zusammenfassung

Der bisherige Erfolg der Mikroelektronik während der letzten Jahrzehnte beruht auf der stetigen Strukturverkleinerung von MOSFETs, da dadurch die Eigenschaften dieser Bauelemente erheblich verbessert werden und auch der Preis pro Schaltfunktion sinkt. Im Allgemeinen erfolgt bei dieser Skalierung der Bauelementeabmessungen eine Anpassung des Ausheilungsprozesses durch eine Erhöhung der Ausheiltemperatur bei gleichzeitiger Verringerung der Ausheilzeit. Das Ziel ist dabei, möglichst flache, abrupte pn-Übergänge mit maximaler elektrischer Aktivierung zu erhalten, um sowohl die im Zuge der Miniaturisierung immer stärker auftretenden Kurzkanaleffekte als auch die parasitären Widerstände zu verringern. Das derzeitig angewandte Ausheilverfahren, die RTA, stößt dabei auf einen Zielkonflikt zwischen der Diffusion der Dotanden und ihrer elektrischen Aktivierung, da mit diesem Ausheilverfahren nur Ausheilzeiten von minimal 1 s ereicht werden können.

Diese Einschränkungen hinsichtlich der Ausheilzeit haben die Forschung nach neuen Verfahren angeregt, wodurch Ausheilzeiten im Millisekundenbereich bei Temperaturen nahe dem Schmelzpunkt von Silizium realisiert werden können. Bei diesen neuartigen Ausheilverfahren wird ausgenutzt, dass nur die oberste Schicht des Wafers durch einen Energieimpuls erhitzt wird, wodurch Aufheiz- und Abkühlraten im Bereich von 10^6 K/s und damit Pulslängen von weniger als 10 ms erreicht werden können. Aufgrund der großen Temperaturunterschiede zwischen Ober- und Unterseite des Wafers kommt es durch die thermische Ausdehnung der wärmeren Schichten zur Ausbildung einer thermischen Verspannung des Wafers. Durch diesen Temperaturunterschied können wiederum Defekte am Kristall entstehen, was zu einer Verformung des Wafers und im schlimmsten Fall zum Waferbruch führen kann. Die Stärke der Verspannung ist sehr stark vom Temperaturgradienten über dem Wafer abhängig. Eine höhere Wafertemperatur bei gleicher Ausheiltemperatur verringert die thermische Verspannung des Wafers, wohingegen eine höhere Ausheiltemperatur bei gleicher Wafertemperatur die thermische Verspannung verstärkt.

Für die Untersuchungen zum Einfluss der Kurzzeitausheilung auf das Transistorverhalten wurden während dieser Arbeit zwei unterschiedlichen Technologien der Kurzzeitausheilung, die Blitzlampen-Ausheilung und die Laser-Ausheilung, verwendet (Tabelle 7.1). Die Blitzlampen-Anlage arbeitet mit Xenon-Blitzlichtlampen im sichtbaren Wellenlängenbereich von λ = (0.3…0.6) µm, wodurch die emittierte Strahlung im Silizium fast zu 100 % durch Interband-Übergänge absorbiert wird.

Tabelle 7.1 Vergleich der wesentlichsten Technologieparameter der Blitzlampen- und Laser-Ausheilung

Technologieparameter	Blitzlampen-Ausheilung	Laser-Ausheilung
Energiequelle	Xenon-Blitzlichtlampen	CO_2-Laser
Wellenlängenbereich	(0.3…0.6) µm	(9.8…10.4) µm
Absorptionsmechanismus	Interband-Übergänge	Intraband-Übergänge
Belichtungsfläche	ganzer Wafer	100 µm x 10 mm

Demgegenüber nutzt die Anlage für die Laser-Ausheilung einen ungepulsten CO_2-Laser mit einer Wellenlänge im mittleren infraroten Bereich von $\lambda = (9.8…10.4)$ µm. Im Gegensatz zu den Blitzlampen-Anlagen, bei denen der gesamte Wafer auf einmal belichtet wird, erfolgt bei der Laser-Ausheilung durch die zeilenweise Belichtung des Wafers nur eine lokale Erwärmung der Waferoberfläche, wodurch neben dem vertikalen auch ein lateraler Temperaturgradient entsteht. Aufgrund der bei der Laser-Ausheilung verwendeten Wellenlänge erfolgt die Absorption der einfallenden Strahlung durch Intraband-Übergänge mit Hilfe der Freien-Ladungsträger-Absorption.

Anhand von Experimenten auf unstrukturierten Wafern wurde der Einfluss der Kurzzeitausheilung am Beispiel der Laser-Ausheilung auf die Diffusion und Aktivierung der Dotierstoffe für eine p- und n-Dotierung untersucht. In Verbindung mit der RTA hat die Kurzzeitausheilung sowohl bei Bor als auch bei Arsen als Dotierstoff für eine p- bzw. n-Dotierung effektiv keine weitere Diffusion der Dotanden mehr zur Folge. Durch die zusätzliche Ausheilung im Anschluss an die RTA verringert sich der Schichtwiderstand der Arsendotierung, wohingegen sich der Schichtwiderstand für eine Bordotierung nicht verändert. Die Verwendung der Kurzzeitausheilung als alleiniger Ausheilschritt führt im Vergleich zur RTA vor allem für eine p-Dotierung zu erheblich kürzeren Diffusionslängen bei einem nur geringfügig höheren Schichtwiderstand. Der Einfluss der Kurzzeitausheilung auf die Diffusionslänge und den Schichtwiderstand für eine n-Dotierung ist demgegenüber nicht so stark ausgeprägt. Die Erhöhung der Ausheiltemperatur der Kurzzeitausheilung als alleiniger Ausheilschritt hat aufgrund einer verbesserten Kristallqualität für beide Dotierungsarten eine kontinuierliche Verringerung des Schichtwiderstands bei einer gleichzeitig nur geringfügig größeren Diffusionslänge zur Folge. Tabelle 7.2 fasst den Einfluss der Kurzzeitausheilung auf die Diffusionslänge und den Schichtwiderstand für eine p- und n-Dotierung im Vergleich zur RTA zusammen.

Tabelle 7.2 Einfluss der Kurzzeitausheilung (MSA) auf die Diffusionslänge L_D und den Schichtwiderstand $R_{Schicht}$ für eine p- und n-Dotierung im Vergleich zur RTA. Plus (+) bedeutet eine geringere Diffusionslänge L_D bzw. einen kleineren Schichtwiderstand $R_{Schicht}$, Minus (-) demgegenüber eine größere Diffusionslänge L_D bzw. einen höheren Schichtwiderstand $R_{Schicht}$.

Dotierung	Parameter	RTA+MSA	nur MSA
p-Dotierung (Bor)	L_D	0	++
	$R_{Schicht}$	0	0
n-Dotierung (Arsen)	L_D	0	+
	$R_{Schicht}$	+	–

Im Fall von Ausheilungstechniken, die auf Lichteinstrahlung basieren, was die Blitzlampen- und Laser-Ausheilung mit einschließt, können die unterschiedlichen optischen Eigenschaften von Bereichen mit verschiedenen Schichtstapeln erhebliche lokale Temperaturunterschiede über dem Mikroprozessor verursachen. Da die beiden Ausheilverfahren unterschiedliche Wellenlängen und damit Absorptionsmechanismen nutzen, kann es auf strukturierten Wafern zu Unterschieden in der Ausheiltemperatur zwischen der Blitzlampen- und Laser-Ausheilung kommen. Mit Hilfe eines speziellen Maskensatzes, wodurch statistisch relevante Rückschlüsse auf Parametervariationen von Transistoren im Mikroprozessor möglich sind, wurde der Einfluss der beiden Ausheilverfahren innerhalb eines Mikroprozessors anhand von elektrischen Parametern experimentell untersucht. Die durchgeführten Analysen zeigen, dass in Verbindung mit einer RTA keine Unterschiede zwischen beiden Ausheilverfahren erkennbar sind, wohingegen die Laser-Ausheilung ohne eine RTA aufgrund der verwendeten Wellenlänge eine geringere Interaktion mit der vorhandenen Waferoberfläche und damit Vorteile gegenüber der Blitzlampen-Ausheilung aufweist.

Ein weiterer Schwerpunkt der vorliegenden Arbeit war die experimentelle Untersuchung von Möglichkeiten zur Integration der Kurzzeitausheilung in den Herstellungsprozess von planaren Hochleistungs-SOI-CMOSFETs. Dabei wurden zwei verschiedene Ansätze untersucht (Bild 7.1):

- Kurzzeitausheilung als zusätzlicher Ausheilschritt in Kombination mit der RTA,
- Kurzzeitausheilung als alleiniger Ausheilschritt ohne RTA.

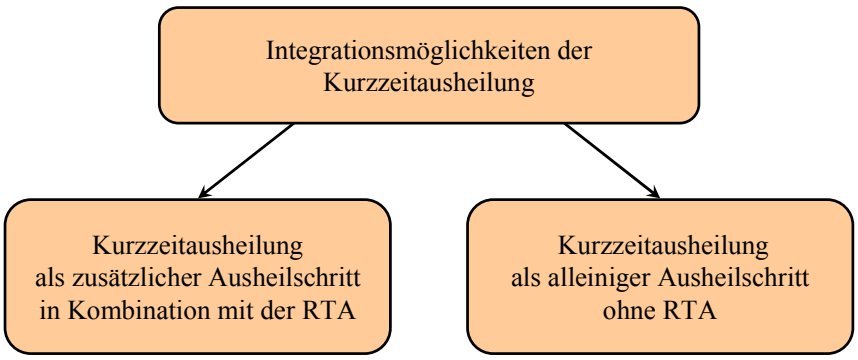


Bild 7.1: Möglichkeiten zur Integration der Kurzzeitausheilung in den Herstellungsprozess von planaren Hochleistungs-SOI-CMOSFETs

Kurzzeitausheilung als zusätzlicher Ausheilschritt mit der RTA

Zunächst stand die Implementierung der Kurzzeitausheilung als zusätzlicher Ausheilschritt in Kombination mit der zurzeit verwendeten RTA im Vordergrund. Im Zuge dieser Untersuchungen wurde die Kurzzeitausheilung im Anschluss an die RTA in den Herstellungsprozess von CMOSFETs integriert und:

1. der Einfluss der Kurzzeitausheilung auf die Leistungsfähigkeit von n- und p-MOSFETs anhand von verschiedenen Transistorparametern bewertet.
2. der Einfluss von Prozessschritten analysiert, die nach der finalen Ausheilung durchgeführt werden und dadurch eine Deaktivierung der eingebrachten Dotanden verursachen können.
3. die Wechselwirkung der Kurzzeitausheilung mit verspannten SiGe-Schichten betrachtet, da die Implementierung von eingebettetem SiGe in die SD-Gebiete von p-MOSFETs eine sehr effektive Methode ist, die Leistungsfähigkeit dieses Transistortyps zu verbessern.

Einfluss der Kurzzeitausheilung auf die Leistungsfähigkeit von n- und p-MOSFETs

Die Ergebnisse dieser Untersuchungen zeigen, dass bei hohen RTA-Ausheiltemperaturen im Bereich von 1085 °C durch die zusätzliche Kurzzeitausheilung die elektrische Gateoxiddicke im Inversionsfall und der Schichtwiderstand, letzteres zumindest bei n-MOSFETs, merklich verringert werden können. Untersuchungen des Gatedielektrikums ergaben allerdings, dass durch die Kurzzeitausheilung zusätzliche Defekte im Gateoxid sowie an der Grenzfläche zwischen Gateoxid und Kanalgebiet erzeugt werden, was eine Verschiebung der Schwellspannung und eine Verringerung der Ladungsträgerbeweglichkeit zur Folge hat. Diese negativen Auswirkungen der Kurzzeitausheilung werden aber durch die kleinere elektrische Gateoxiddicke im Inversionsfall und im Fall von n-MOSFETs durch den verringerten Schichtwiderstand mehr als kompensiert, wodurch letztendlich die Leistungsfähigkeit von n- als auch von p-MOSFETs durch den zusätzlichen Ausheilschritt erhöht werden kann. Der Einfluss der Kurzzeitausheilung als zusätzlicher Ausheilschritt auf die Leistungsfähigkeit beider Transistortypen verändert sich jedoch mit sinkender Ausheiltemperatur der RTA. Bei n-MOSFETs ist mit kleiner werdender RTA-Ausheiltemperatur durch den zusätzlichen Ausheilschritt eine stetige Zunahme des Drainstromgewinns zu beobachten. Bei p-MOSFETs hingegen sinkt mit Verringerung der Ausheiltemperatur der Drainstromgewinn durch die Kurzzeitausheilung. Unterhalb einer RTA-Ausheiltemperatur von 1065 °C verschlechtert sich sogar die Leistungsfähigkeit von p-MOSFETs durch den zusätzlichen Ausheilschritt. Mit Verringerung der Ausheiltemperatur wird der Einfluss der Kurzzeitausheilung auf die elektrisch Gateoxiddicke im Inversionsfall bedeutend kleiner, wodurch die geringere Ladungsträgerbeweglichkeit nicht mehr kompensiert wird und zu dem beobachteten Drainstromverhalten führt.

Einfluss nachfolgender Prozessschritte

Eine weitere Herausforderung bei der Integration der Kurzzeitausheilung in den Herstellungsprozess ist die Empfindlichkeit der aktivierten Dotanden gegenüber nachfolgenden Temperaturschritten. Nach der Ausheilung der Implantationsschäden müssen noch verschiedene thermische Prozesse einschließlich der Silizidbildung durchgeführt werden, was aufgrund der durch die Kurzzeitausheilung erzeugten metastabilen Zustände eine Deaktivierung der Dotanden zur Folge haben kann. Maßgeblich ist dabei der Prozess mit dem größten thermischen Budget. Untersuchungen zu diesem Problem ergaben, dass zumindest für Prozesse unter 500 °C noch keine Deaktivierungsvorgänge der durch die Kurzzeitausheilung erzeugten metastabilen Zustände durch nachfolgende Prozessschritte zu befürchten sind.

Wechselwirkung der Kurzzeitausheilung mit verspannten SiGe-Schichten

Die Ergebnisse dieser Untersuchungen zeigen, dass zumindest auf unstrukturierten Wafern sowohl die Waferverbiegung als auch der Relaxationsgrad der SiGe-Schicht mit zunehmender Ausheiltemperatur der Kurzzeitausheilung negativ beeinflusst werden. Ursachen für dieses Verhalten sind der durch die Kurzzeitausheilung entstehende Temperaturgradient zwischen Ober- und Unterseite des Wafers und die Stärke und Lage der Implantationsschäden. Die Untersuchungen auf strukturierten SOI-Wafern deuten jedoch darauf hin, dass weder Relaxationserscheinungen noch Schädigungen der SiGe-Gebiete durch die Kurzzeitausheilung auftreten, weil wahrscheinlich die BOX zwischen der SOI-Schicht und dem Substrat als Puffer dient, der die Verspannungen aufnimmt. Sowohl mit als auch ohne eingebettetes SiGe erhöht sich durch den zusätzlichen Ausheilschritt der Sättigungs-Drainstrom, was auf eine Kompatibilität von verspannten SiGe-Gebieten und der Kurzzeitausheilung hinweist.

Kurzzeitausheilung als alleiniger Ausheilschritt ohne RTA

Das Hauptaugenmerk dieser Arbeit lag in der Entwicklung, Herstellung und Charakterisierung von diffusionsarmen Transistoren, bei denen die Lage der Dotanden hauptsächlich durch die Implantation bestimmt und durch die Ausheilung nicht mehr signifikant verändert wird. Das lässt sich dadurch erreichen, wenn nur eine Kurzzeitausheilung verwendet und auf die RTA vollständig verzichtet wird. Im Zuge dieser Untersuchungen wurde:

1. die Leistungsfähigkeit von diffusionsarmen n- und p-MOSFETs anhand von verschiedenen Transistorparametern bewertet.
2. eine Optimierung der Implantationsparameter der Halo-, SDE- und SD-Gebiete durchgeführt.
3. der Einfluss zusätzlicher Niedrigtemperaturschritte im Prozessablauf auf die Leistungsfähigkeit diffusionsarmer Transistoren untersucht.
4. die Integration von verspannten SiGe- und SiC-Schichten in die SD-Gebiete von diffusionsarmen p- bzw. n-MOSFETs untersucht.

Bewertung der Leistungsfähigkeit von diffusionsarmen n- und p-MOSFETs

Die Ergebnisse dieser Untersuchungen zeigen, dass bei beiden Transistortypen durch die Kurzzeitausheilung als alleiniger Ausheilschritt ohne weitere Anpassungen im Herstellungsprozess erhebliche Beeinträchtigungen in der Leistungsfähigkeit und beim Kurzkanalverhalten zu verzeichnen sind. Eine Ursache hierfür ist das stark veränderte Dotierungsprofil im Vergleich zu einem Transistor mit einem RTA-Schritt. Bei einem RTA-Prozess wird die finale Lage der eingebrachten Dotanden durch die Implantation sowie die Diffusion während der RTA bestimmt, wohingegen bei einem diffusionsarmen Prozess die Definition des Dotierungsprofils vorwiegend durch die Implantation erfolgt. Aus diesem Grund ist eine Umgestaltung des Dotierungsprofils notwendig, wenn diffusionsarme Transistoren angestrebt werden.

Optimierung der Implantationsparameter der Halo- SDE- und SD-Gebiete

Untersuchungen und parallel zu diesem Thema durchgeführte Simulationen zeigen, dass sich das Kurzkanalverhalten nur durch eine kombinierte Optimierung des Dotierungsprofils der Wannen-, SDE-, SD- und Halo-Gebiete und nicht nur durch eine separate Halo-Implantation verbessert werden kann. Weiterhin war zu beobachten, dass alternative Implantationsspezies wie Bor und Phosphor für eine BF_2- bzw. Arsen-Dotierung für die SDE-Gebiete verwendet werden können, ohne im Gegensatz zum herkömmlichen RTA-Prozess die Leistungsfähigkeit der Transistoren negativ zu beeinflussen. Ein weiteres Ergebnis dieser Untersuchungen ist, dass eine zusätzliche unter einem Winkel in die SD-Gebiete des Transistors durchgeführte Implantation den externen Widerstand bei beiden Transistortypen verringert und dadurch die Leistungsfähigkeit erheblich erhöht.

Einfluss zusätzlicher Niedrigtemperaturschritte

Die in diesem Zusammenhang durchgeführten Untersuchungen zeigen, dass selbst Prozesse mit sehr geringen Temperaturen von etwa (600…800) °C im Vorfeld der Kurzzeitausheilung zum Einen das Kurzkanalverhalten infolge einer erhöhten Dotandendiffusion verschlechtern und zum Anderen die Leistungsfähigkeit von diffusionsarmen n- und p-MOSFETs aufgrund einer drastischen Erhöhung der Bahnwiderstände verringern. Aus diesem Grund muss bei diffusionsarmen Transistoren ebenfalls das thermische Budget jedes einzelnen Prozessschritts während der Herstellung beachtet werden, um eine unnötige Diffusion oder Deaktivierung der eingebrachten Dotanden zu vermeiden.

Integration von verspannten SiGe- und SiC-Schichten

Die Ergebnisse dieser Untersuchungen zeigen, dass bei diffusionsarmen p-MOSFETs mit SiGe die Leistungsfähigkeit der Transistoren wie beim RTA-Prozess stark verbessert wird. Im Gegensatz zum RTA-Prozess ist jedoch die Wahl der richtigen Implantationsspezies von entscheidender Bedeutung, da eine zu starke Schädigung der SiGe-Schicht durch die Implantation die Leistungsfähigkeit der diffusionsarmen p-MOSFETs mit SiGe sehr stark verringert. Weiterhin war festzustellen, dass infolge einer Optimierung des Herstellungsprozesses eine verbesserte Topographie der Aktivgebiete erreicht werden kann. Dies hat einen höheren Drainstrom aufgrund einer verbesserten Verspannungseinkopplung zur Folge. Eine weitere Steigerung der Leistungsfähigkeit lässt sich durch den Einsatz von *in situ* bordotiertem SiGe erreichen, da neben der höheren Ladungsträgerbeweglichkeit im Kanal auch der Bahn- und damit der externe Widerstand durch die *in situ* Bordotierung der eingebetteten SiGe-Gebiete verringert wird.

So wie SiGe bei p-MOSFETs durch die uniaxiale Druckverspannung in Kanalrichtung die Löcherbeweglichkeit erhöht, lässt sich theoretisch das Gleiche mit SiC bei n-MOSFETs erreichen. Die dazu durchgeführten Untersuchungen zeigen, dass die Kurzzeitausheilung als alleiniger Ausheilschritt zumindest auf unstrukturierten Wafern die Konzentration an substitutionell auf Gitterplätzen eingebautem Kohlenstoff zur Erzeugung von SiC-Schichten erhöht. Auf strukturierten Wafern wird demgegenüber anders als beim SiGe bei p-MOSFETs durch das Einbetten von *in situ* phosphordotiertem SiC in die SD-Gebiete von diffusionsarmen n-MOSFETs keine Verbesserung in der Leistungsfähigkeit der Transistoren beobachtet, was vermutlich auf Kristalldefekte an der vertikalen Grenzfläche zwischen Si und SiC und den damit verbundenen Problemen beim SiC-Wachstum zurückzuführt werden kann.

Abschließende Untersuchungen zu den diffusionsarmen Transistoren zeigen, dass im Fall von n-MOSFETs durch die Optimierung der Implantationsparameter und Spacer-Strukturierung eine zum RTA-Prozess vergleichbare Leistungsfähigkeit erreicht werden kann. Bei diffusionsarmen p-MOSFETs ist es durch die Implementierung von ISBD-SiGe sowie der Optimierung des Prozessablaufs und der Implantationsparameter möglich, die Leistungsfähigkeit dieser Transistoren soweit zu steigern, dass deren Sättigungs-Drainstrom im Vergleich zum RTA-Transistor nur geringfügig kleiner ist. Das Kurzkanalverhalten wird für beide diffusionsarmen Transistortypen im Vergleich zum jeweiligen RTA-Transistor stark verbessert. Anhand der Untersuchungen dieser Arbeit konnte damit gezeigt werden, dass im Gegensatz zur herkömmlichen RTA-Ausheilung eine weitere Miniaturisierung der planaren Transistorstruktur mit Hilfe der Kurzzeitausheilung möglich ist.

7.2 Ausblick

Der Nachteil der Kurzzeitausheilung als alleiniger Ausheilschritt ohne RTA ist, dass einerseits bei einer sehr starken Schädigung der Kristallstruktur infolge einer amorphisierenden Implantation das thermische Budget dieser Ausheiltechnologie nicht ausreicht, um alle Implantationsdefekte vollständig bzw. in energetisch stabilere Konfigurationen aufzulösen, was wiederum höhere Leckströme der pn-Übergänge zur Folge haben kann. Andererseits kommt es wie bereits erwähnt aufgrund des großen Temperaturunterschieds zwischen Ober- und Unterseite des Wafers zur Ausbildung einer thermischen Verspannung des Wafers, was zu einer Verformung des Wafers und im schlimmsten Fall zum Waferbruch führen kann. Eine mögliche Lösung für diese verschiedenen Probleme kann an dieser Stelle ein leicht verändertes Temperaturprofil der Kurzzeitausheilung sein.

So geht die Entwicklung der Anlagenhersteller immer mehr in Richtung eines etwas höheren thermischen Budgets, ohne die Vorteile der Kurzzeitausheilung, also das die Lage der Dotanden nicht mehr signifikant durch die Ausheilung verändert wird, zu verlieren. In [135] und [136] wurden dazu für die Blitzlampen- bzw. die Laser-Ausheilung ein Konzept vorgestellt, bei dem der Wafer etwa 10 ms lang auf eine Temperatur von 1000 °C vorgewärmt wird, bevor die eigentliche Kurzzeitausheilung bei einer Ausheiltemperatur von 1300 °C und einer Ausheilzeit von 2 ms stattfindet. Dadurch besteht zum Einen die Möglichkeit den Temperaturgradienten und damit die Verspannung des Wafers während der Ausheilung drastisch zu verringern und zum Anderen Implantationsdefekte aufgrund des höheren thermischen Budgets dieses Ausheilkonzepts besser auszuheilen.

Zukünftige Untersuchungen müssen deshalb zeigen, ob durch dieses Ausheilkonzept bessere Transistorergebnisse erreicht werden können und inwieweit eine Kompatibilität zu Gateoxidmaterialien mit hoher Permittivität gegeben ist, da diese Gateoxidmaterialien für eine weitere Skalierung der Bauelementeabmessungen ebenfalls unentbehrlich sind.

Literaturverzeichnis

[1] J.S. Kilby: *Invention of the integrated circuit*, Institute of Electrical and Electronics Engineers (IEEE) Transactions on Electron Devices, Vol. 23, No. 7, pp. 648-654, 1976.

[2] G.E. Moore: *Cramming more components onto integrated circuits*, Electronics, Vol. 38, No. 8, p. 114-117, 1965.

[3] G.E. Moore: *Progress in digital integrated circuit*, International Electron Devices Meeting (IEDM) - Conference Proceedings, p. 11, 1975.

[4] *Process Integration, Devices, and Structures*, ITRS 2006, www.itrs.net.

[5] T. Ito, K. Suguro, M. Tamura, T. Taniguchi, Y. Ushiku et al.: *14 nm-depth low resistance boron doped extension by optimized flash lamp annealing*, Solid State Devices and Materials (SSDM) - Conference Proceedings, p. 19, 2002.

[6] S. McCoy, E.A. Arevalo, J. Gelpey, D.F. Downey: *Advanced doping and millisecond annealing for ultra-shallow junctions for 65 nm and beyond*, Advanced Thermal Processing of Semiconductors (RTP) - Conference Proceedings, pp. 99-108, 2004.

[7] D. Jennings, A. Mayur, V. Parihar, L. Haifan, R. Mcintosh, et al.: *Dynamic surface anneal: activation without diffusion*, RTP - Conference Proceedings, pp. 47-51, 2004.

[8] S. Talwar, D. Markle, M. Thompson: *Junction scaling using lasers for thermal annealing*, Solid State Technology, Vol. 46, No. 7, p. 83, 2003.

[9] T. Ito, K. Suguro, T. Itani, K. Nishinohara, K. Matsuo, et al.: *Improvement of threshold voltage roll-off by ultra-shallow junction formed by flash lamp annealing*, Symposium on Very-Large-Scale Integration (VLSI) Technology - Conference Proceedings, pp. 53-54, 2003.

[10] C.-C. Chen, C.F. Nieh, D.W. Lin, K.C. Ku, J.C. Sheu, et al.: *Channel stress modulation and pattern loading effect minimization of milli-second super anneal for Sub-65nm high performance SiGe CMOS*, Symposium on VLSI Technology - Conference Proceedings, pp. 174-175, 2006.

[11] A. Pouydebasque, B. Dumont, S. Denorme, F. Wacquant, M. Bidaud, et al.: *High density and high speed SRAM bit-cells and ring oscillators due to laser annealing for 45nm Bulk CMOS*, IEDM - Conference Proceedings, pp. 679-682, 2005.

[12] S.K.H. Fung, H.T. Huang, S.M. Cheng, K.L. Cheng,, S.W. Wang, et al.: *65nm CMOS high speed, general purpose and low power transistor technology for high volume foundry application*, Symposium on VLSI Technology - Conference Proceedings, pp. 92-93, 2004.

[13] T. Yamamoto, T. Kubo, T. Sukegawa, Y. Wang, L. Feng, et al.: *Performance enhancement in aggressively scaled CMOS devices by higher carrier activation with laser spike annealing*, International Workshop on Junction Technology (IWJT) - Conference Proceedings, pp. 15-18, 2005.

[14] T. Yamashita, P.A. Fisher, O. Gluschenkov, H. Kimura, R. Ramachandran, et al.: *High performance 65nm SOI transistors using laser spike annealing*, European Solid-State Device Research Conference (ESSDERC) - Conference Proceedings, pp. 347-350, 2006.

[15] T. Sanuki, T. Iwamoto, K. Ota, T. Komoda, H. Yamazaki, et al.: *High-performance 45nm node CMOS transistors featuring flash lamp annealing (FLA)*, IEDM - Conference Proceedings, pp. 281-284, 2007.

[16] T. Hoffmann, T. Noda, S. Felch, S. Severi, V. Parihar, et al.: *Laser annealed junctions: process integration sequence optimization for advanced CMOS technologies*, IWJT - Conference Proceedings, pp. 137-140, 2007.

[17] S. Endo, Y. Maruyama, Y. Kawasaki, T. Yamashita, H. Oda, et al.: *Novel junction engineering scheme using combination of LSA and Spike-RTA*, IWJT - Conference Proceedings, pp. 135-136, 2007.

[18] T. Yamamoto, T. Kubo, T. Sukegawa, E. Takii, Y. Shimamune, et al.: *Junction profile engineering with a novel multiple laser spike annealing scheme for 45-nm node high performance and low leakage CMOS technology*, IEDM - Conference Proceedings, pp. 143-146, 2007.

[19] G.H. Buh, G.-H. Yon, T.-S. Park, J.-W.Lee, J. Kim, et L:: *Integration of sub-melt laser annealing on metal gate CMOS devices for Sub 50 nm Node DRAM*, IEDM - Conference Proceedings, pp. 863-866, 2006.

[20] T. Feudel: *Advanced annealing schemes for high-performance SOI logic technologies*, Materials Science Forum - Rapid Thermal Processing and beyond: Applications in Semiconductor Processing, Vol. 573-574, pp. 387-400, 2008.

[21] A. Shima, Y. Wang, S. Talwar, A. Hiraiwa: *Ultra-shallow junction formation by non-melt laser spike annealing for 50-nm gate CMOS*, Symposium on VLSI Technology - Conference Proceedings, pp. 174-175, 2004.

[22] A. Pouydebasquel, B. Dumont, F. Wacquant, A. Halimaoui, C. Laviron, et al.: *NMOS-junction integration study with ultra-high temperature non-diffusive laser annealing for the 45nm node and below*, IWJT - Conference Proceedings, pp. 11-14, 2005

[23] M. Narihiro, T. Iwamoto, T. Yamamoto, T. Ikezawa, K. Yako, et al.: *Sub-30nm MOSFET fabrication technology incorporating precise dopant profile design using diffusion-less high-activation laser annealing*, RTP - Conference Proceedings, pp. 147-151, 2006.

[24] S. Severi, E. Augendre, S. Thirupapuliyur, K. Ahmed, S. Felch: *Optimization of sub-melt laser anneal: performance and reliability*, IEDM - Conference Proceedings, pp. 1-4, 2006.

[25] K. Yako, K. Uejima, T. Yamamoto, A. Mineji, T. Nagumo, et al.: *Aggressive design of millisecond annealing junctions for near-scaling-limit bulk CMOS using raised source/drain Extensions*, IEDM - Conference Proceedings, pp. 1-4, 2008.

[26] K. Rim, S. Narasimha, M. Longstreet, A. Mocuta, J. Cai: *Low field mobility characteristics of sub-100nm unstrained and strained Si MOSFETs*, IEDM - Conference Proceedings, pp. 43-46, 2002.

[27] Y. Taur, T.H. Ning: *Fundamentals of Modern VLSI Devices*, Cambridge University Press, 1998.

[28] L.D. Yau: *A simple theory to predict the threshold voltage of short-channel IGFET's*, Solid State Electronics, Vol. 17, No. 10, pp. 1059-1063, 1974.

[29] S.E. Thompson, P.A. Packan, M.T. Bohr: *Linear versus saturated drive current: tradeoffs in super steep retrograde well engineering*, Symposium on VLSI Technology - Conference Proceedings, pp. 154-155, 1996.

[30] J.B. Jacobs, D. Antoniadis: *Channel profile engineering for MOSFET's with 100 nm channel lengths*, IEEE Transactions on Electron Devices, Vol. 42, No. 5, pp. 870-875, 1995.

[31] C.F. Codella, S. Ogura: *Halo doping effects in submicron DI-LDD device design*, IEDM - Conference Proceedings, pp. 230-233, 1985.

[32] T. Hori: *A 0.1 μm CMOS technology with tilt-implanted punchthrough stopper (TIPS)*, IEDM - Conference Proceedings, pp. 75-78, 1994.

[33] Y. Taur, E.J. Nowak: *CMOS devices below 0.1 μm: how high will performance go?*, IEDM - Conference Proceedings, pp. 215-218, 1997.

[34] D.M. Caughey, R.E. Thomas: *Carrier mobilities in silicon empirically related to doping and field*, Proceedings of the IEEE, Vol. 55, No. 12, pp. 2192-2193, 1967.

[35] S. Flachowsky, A. Wei, T. Herrmann, R. Illgen, M. Horstmann, et al.: *Gate length scaling trends of drive current enhancement in CMOSFETs with dual stress overlayers and embedded-SiGe*, Materials Science and Engineering: B, Vol. 154-155, pp. 98-101, 2008.

[36] S.E. Thompson: *Source-drain series resistance: the real limiter to MOSFET scaling*, Advanced Short-Time Thermal Processing for Si-Based CMOS Devices II, pp. 412-419, 2004.

[37] A.M. Noori, M. Balseanu, P. Boelen, A. Cockburn, S. Demuynck: *Manufacturable processes for ≤ 32-nm-node CMOS enhancement by synchronous optimization of strain-engineered channel and external parasitic resistances*, IEEE Transactions on Electron Devices, Vol. 55, No. 5, pp. 1259-1264, 2008.

[38] S.-D. Kim, C.-M. Park, J.S. Woo: *Advanced model and analysis of series resistance for CMOS scaling into nanometer regime—Part I: theoretical derivation*, IEEE Transactions on Electron Devices, Vol. 49, No. 3, pp. 457-466, 2002.

[39] L.J. Lin, K.L. Prey, W.K. Choi, E.A. Fitzgerald, D.A. Antoniadis, D.Z. Chi: *Effects of platinum on NiPtSiGe/n-SiGe and NiPtSi/n-Si schottky contacts*, Advanced Materials for Micro- and Nano-Systems, 2005.

[40] S.M. Sze: *Physics of Semiconductor Devices*, John Wiley & Sons, New York, 1981.

[41] L. Bergmann, C. Schaefer: *Lehrbuch der Experimentalphysik – Band 6*, Walter de Gruyter, Berlin - New York, 1992.

[42] J.D. Plummer, M.D. Deal, P.B. Griffin: *Silicon VLSI Technology – Fundamentals, Practice and Modeling*, Prentice Hall, 2000.

[43] J.M. Poate and J.S. Williams: *Ion Implantation and Beam Processing*, Academic Press, 1984.

[44] L. Csepregi, E.F. Kennedy, J.W. Mayer, T.W. Sigmon: *Substrate-orientation dependence of the epitaxial regrowth rate from Si-implanted amorphous Si*, Journal of Applied Physics, Vol. 49, No. 7, pp. 3906-3911, 1978.

[45] L. Csepregi, E.F. Kennedy, T.J. Gallagher, J.W. Mayer, T.W. Sigmon: *Reordering of amorphous layers of Si implanted with ^{31}P, ^{75}As, and ^{11}B ions*, Journal of Applied Physics, Vol. 48, No. 10, pp. 4234-4240, 1977.

[46] E.J.H. Collart, A.J. Murrell, M.A. Foad, J.A. van den Berg, S. Zhang, et al.: *Cluster formation during annealing of ultra-low-energy boron-implanted silicon*, Journal of Vacuum Science & Technology B: Microelectronics and Nanometer Structures, Vol. 18, pp. 435-439, 2000.

[47] A.D. Lilak, V. Krishnamoorthy, D. Vieira, M.E. Law, K.S. Jones: *A study of boron clustering transients and mechanisms in doped silicon*, Materials Research Society (MRS) - Conference Proceedings, Vol. 610, B 5.4.1, 2000.

[48] P.A. Stolk, H.-J. Gossmann, D.J. Eaglesham, D.C. Jacobson, C.S. Rafferty, et al.: *Physical mechanisms of transient enhanced dopant diffusion in ion-implanted silicon*, Journal of Applied Physics, Vol. 81, No. 9, pp. 6031-6050, 1997.

[49] D.W. Lawther, U. Myler, P.J. Simpson, P.M. Rousseau, P.B. Griffin, et al.: *Vacancy generation resulting from electrical deactivation of arsenic*, Applied Physics Letters, Vol. 67, No. 24, pp. 3575-3577, 1995.

[50] K. Saarinen, J. Nissilä, H. Kauppinen, M. Hakala, M.J. Puska, et al.: *Identification of vacancy-impurity complexes in highly n-type Si*, Physical Review Letters, Vol. 82, No. 9, pp. 1883-1886, 1999.

[51] R. Subrahmanyan, M. Orlowski, A.R. Sitaram: *Confirmation of a vacancy role in the electrical activation and deactivation of arsenic using a point-defect injection experiment*, ECS - Conference Proceedings - Process Physics and Modeling in Semiconductor Technology, Vol. 93, No. 6, pp. 120-129, 1993.

[52] F.A. Trumbore: *Solid Solubility of Impurity Elements in Germanium and Silicon*, Bell System Technical Journal, Vol. 39, pp. 205-233, 1960.

[53] A. Fick: *Über Diffusion*, Annalen der Physik und Chemie von Poggendorff, Vol. 94, pp. 59-86, 1855.

[54] R.F. Scholz: *Untersuchungen zur Diffusion von Eigenpunktdefekten in Silizium und Galliumarsenid*, Dissertation, Universität Halle, 1999.

[55] J. Crank: *The Mathematics of Diffusion*, Second Edition, Oxford University Press, 1975.

[56] P.M. Fahey, P.B. Griffin, J.D. Plummer: *Point defects and dopant diffusion in silicon*, Reviews of Modern Physics, Vol. 61, No. 2, pp. 289-388, 1989.

[57] A. Ural, P.B. Griffin, J.D. Plummer: *Fractional contributions of microscopic diffusion mechanisms for common dopants and self-diffusion in silicon*, Journal of Applied Physics, Vol. 85, No. 9, pp. 6440-6446, 1999.

[58] U. Gösele, T.Y. Tan: *The nature of point defects and their influence on diffusion processes in silicon at high temperatures*, MRS - Conference Proceedings, Vol. 14, pp. 45-59, 1983.

[59] W. Frank, U. Gösele, H. Mehrer, A. Seeger: *Diffusion in Silicon and Germanium*, Diffusion in Crystalline Solids, S. 63-142, London Academic Press, 1984.

[60] P. Pichler: *Intrinsic Point defects, Impurities, and Their Diffusion in Silicon*, Springer-Verlag Wien, 2004.

[61] F.J. Morin, J.P. Maita: *Electrical properties of silicon containing arsenic and boron*, Physical Review, Vol. 96, No. 1, pp. 28-35, 1954.

[62] M.J. Aziz, Y. Zhao, H.-J. Gossmann, S. Mitha, S.P. Smith, et al.: *Pressure and stress effects on the diffusion of B and Sb in Si and Si-Ge alloys*, Physical Review B, Vol. 73, No. 5, pp. 054101/01-054101/20, 2006.

[63] R.B. Fair: *Challenges in manufacturing submicron, ultra-large scale integrated circuits*, Proceedings of the IEEE, Vol. 78, No. 11, pp. 1687-1705, 1990.

[64] R.B. MacKnight, P.J. Timans, S.-P. Tay, Z. Nenyei: *RTP applications and technology options for the sub-45 nm nodes*, RTP - Conference Proceedings, pp. 3-36, 2004.

[65] A. Mokhberi, P.B. Griffin, J.D. Plummer, E. Paton, S. McCoy, et al.: *A comparative study of dopant activation in boron, BF_2, arsenic, and phosphorus implanted silicon*, IEEE Transactions on Electron Devices, Vol. 49, No. 7, pp. 1183-1191, 2002.

[66] A.T. Fiory, K.K. Bourdelle: *Electrical activation kinetics for shallow boron implants in silicon*, Applied Physics Letters, Vol. 74, No. 18, pp. 2658-2660, 1999.

[67] P.J. Timans: *Rapid Thermal Processing*, Handbook of Semiconductor Manufacturing Technology (Second Edition), CRC Press, Taylor & Francis Group, 2008.

[68] P.J. Timans, N. Acharya: *Fundamental issues in millisecond annealing*, Advanced Short-Time Thermal Processing for Si-Based CMOS Devices II, pp. 11-18, 2004.

[69] A. Colin, P. Morin, F. Cacho, H. Bono, R. Beneyton, et al.: *Simulation of the sub-melt laser anneal process in 45 CMOS technology - application to the thermal pattern effects*, Materials Science and Engineering: B, Vol. 154-155, pp. 31-34, 2008.

[70] T. Ito, K. Suguro, M. Tamura, T. Taniguchi, Y. Ushiku et al.: *Low-resistance ultrashallow extension formed by optimized flash lamp annealing*, IEEE Transactions on Semiconductor Manufacturing, Vol. 16, No. 3, pp. 417-422, 2003.

[71] W. Skorupa, T. Gebel, R.A. Yankov, S. Paul, W. Lerch, et al.: *Advanced thermal processing of ultrashallow implanted junctions using flash lamp annealing*, Journal of The Electrochemical Society, Vol. 152, No. 6, pp. G436-G440, 2005.

[72] W. Lerch, S. Paul, J. Niess, S. McCoy, J. Gelpey, et al.: *Advanced activation and deactivation of arsenic-implanted ultra-shallow junctions using flash and spike + flash annealing*, RTP - Conference Proceedings, pp. 191-196, 2007.

[73] W. Lerch, S. Paul, J. Niess, J. Chan, S. McCoy, et al.: *Experimental and theoretical results of dopant activation by a combination of spike and flash annealing*, IWJT - Conference Proceedings, pp. 129-134, 2007.

[74] W. Lerch, S. Paul, J. Niess, S. McCoy, T. Selinger, et al.: *Advanced activation of ultra-shallow junctions using flash-assisted RTP*, Materials Science and Engineering: B, Vol. 124-125, pp. 24-31, 2005.

[75] D.K. Schroder, R.N. Thomas, J.C. Swartz: *Free carrier absorption in silicon*, IEEE Journal of Solid-State Circuits, Vol. SC-13, No. 1, pp. 180-187, 1978.

[76] R. Beneyton, A. Colin, H. Bono, F. Cacho, M. Bidaud, et al.: *Origin of local temperature variation during spike anneal and millisecond anneal*, RTP - Conference Proceedings, pp. 183-193, 2008.

[77] T. Kubo, T. Sukegawa, E. Takii, T. Yamamoto, S. Satoh, et al.: *First quantitative observation of local temperature fluctuation in millisecond annealing*, RTP - Conference Proceedings, pp. 321-326, 2007.

[78] T. Kubo, T. Sukegawa, M. Kase: *Total temperature tluctuation of a patterned wafer in the millisecond annealing*, RTP - Conference Proceedings, pp. 195-199, 2008.

[79] L.M. Feng, Y. Wang, D.A. Markle: *Minimizing pattern dependency in millisecond annealing*, IWJT - Conference Proceedings, pp. 25-30, 2006.

[80] Sentaurus TCAD, Release Z-2007.03 ed., Synopsys, Mountain View, CA, USA, 2007.

[81] A.F. Tasch, H. Shin, C. Park, J. Alvis, S. Novak: *An improved approach to accurately model shallow B and BF_2 implants in silicon*, Journal of the Electrochemical Society, Vol. 136, No. 3, pp. 810-814, 1989.

[82] Z. Ren, G. Pei, J. Li, B.(F). Yang, R. Takalkar, et al.: *On implementation of embedded phosphorus-doped SiC stressors in SOI nMOSFETs*, Symposium on VLSI Technology - Conference Proceedings, pp. 172-173, 2008.

[83] M.C. Öztürk, N. Pesovic, I. Kang, J. Liu, H. Mo, S. Gannavaram: *Ultra-shallow source/drain junctions for nanoscale CMOS usingselective silicon-germanium technology*, IWJT - Conference Proceedings, pp. 77-82, 2001.

[84] R. Stenzel, C. Pigorsch, W. Klix, R. Schüffny: *Simulation von Nanobauelementen*, Proc. Dresdner Arbeitstagung „Schaltungs- und Systementwurf“, Fraunhofer Institut für Integrierte Schaltungen.

[85] S.E. Laux, A.C. Warren: *Self-consistent calculation of electron states in narrow channels*, IEDM - Conference Proceedings, pp. 567-570, 1986.

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

[87] C. Moglestue: *Monte Carlo Simulation of Semiconductor Devices*, Chapman & Hall, 1993.

[88] A. Abou-Elnour: *Modeling and simulation of heterostructure-FET*, Verlag Shaker, 1995.

[89] K. Tomizawa: *Numerical simulation of submicron semiconductor devices*, Artech House, Inc., 1993.

[90] S. Selberherr: *Analysis and Simulation of Semiconductor Devices*, Springer Verlag, 1984.

[91] J. Höntschel: *Simulation und Optimierung komplexer nanoelektronischer Strukturen*, Dissertation, Technische Universität Dresden, 2004.

[92] H.R. Lazar: *Mobility Degradation of Advanced CMOS Gatestacks*, Dissertation, North Carolina State University, 2005.

[93] A. Schenk: *Halbleiterbauelemente – Physikalische Grundlagen und Simulation*, Skript Eidgenössische Technische Hochschule Zürich, Integrated Systems Laboratory, 2003.

[94] N.D. Arora, J.R. Hauser, D.J. Roulston: *Electron and hole mobilities in silicon as a function of concentration and temperature*, IEEE Transactions on Electron Devices, Vol. ED-29, No. 2, pp. 292-295, 1982.

[95] C. Lombardi, S. Manzini, A. Saporito, M. Vanzi: *A physically based mobility model for numerical simulation of nonplanar devices*, IEEE Transactions on Computer-Aided Design, Vol. 7, No. 11, pp. 1164-1171, 1988.

[96] C. Canali, G. Majni, R. Minder, G. Ottaviani: *Electron and hole drift velocity measurements in silicon and their empirical relation to electric field and temperature*, IEEE Transactions on Electron Devices, Vol. ED-22, No. 11, pp. 1045-1047, 1975.

[97] J.G. Fossum: *Computer-aided numerical analysis of silicon solar cells*, Solid-State Electronics, Vol. 19, No. 4, pp. 269-277, 1976.

[98] R. Häcker, A. Hangleiter: *Intrinsic upper limits of the carrier lifetime in silicon*, Journal of Applied Physics, Vol. 75, No. 11, pp. 7570-7572, 1994.

[99] M. Reisch: *Halbleiter-Bauelemente*, Springer-Verlag Berlin Heidelberg, 2005.

[100] A. Schütz, S. Selberherr, H.W. Pötzl: *A two-dimensional model of the avalanche effect in MOS transistors*, Solid-State Electronics, Vol. 25, No. 3, pp. 177-183, 1982.

[101] G.A.M. Hurkx, D.B.M. Klaassen, M.P.G. Knuvers: *A new recombination model for device simulation including tunneling*, IEEE Transactions on Electron Devices, Vol. 39, No. 2, pp. 331-338, 1992.

[102] E. Ungersböck: *Advanced Modeling of Strained CMOS Technology*, Dissertation, TU Wien, 2007.

[103] M. Levinshtein, S. Rumyantsev, M. Shur: *Handbook Series on Semiconductor Parameters Volume 1*, World Scientific Publishing, 1996.

[104] N. Mohta, S.E. Thompson: *Mobility enhancement - the next vector to extend Moores law*, IEEE Circuits & Devices Magazine, pp. 18-23, 2005.

[105] K. Uchida, T. Krishnamohan, K.C. Saraswat, Y. Nishi: *Physical mechanisms of electron mobility enhancement in uniaxial stressed MOSFETs and impact of uniaxial stress engineering in ballistic regime*, IEDM - Conference Proceedings, pp. 129-132, 2005.

[106] I. Goroff, L. Kleinmann: *Deformation potentials in silicon III - effects of a general strain on conduction and valence levels*, Physical Review, Vol. 132, No. 3, pp. 1080-1084, 1964.

[107] G.L. Bir, G.E. Pikus: *Symmetry and Strain-Induced Effects in Semiconductors*, John Wiley & Sons, New York, 1974.

[108] J.L. Egley, D. Chidambarrao: *Strain effects on device characteristics: implementation in drift-diffusion simulators*, Solid-State Electronics, Vol. 36, No. 12, pp. 1653-1664, 1993.

[109] B. Obradovic, P. Matagne, L. Shifren, E. Wang, M. Stettler, et al.: *A physically-based analytic model for stress-induced hole mobility enhancement*, Journal of Computational Electronics, Vol. 3, No. 3-4, pp. 161-164, 2004.

[110] Y. Sun, S.E. Thompson, T. Nishida: *Physics of strain effects in semiconductors and metal-oxide-semiconductor field effect transistors*, Journal of Applied Physics, Vol. 101, No. 10, pp. 104503/1-104503/22, 2007.

[111] R.F. Lever, J.M. Bonar, A.F.W. Willoughby: *Boron diffusion across silicon - silicon germanium boundaries*, Journal of Applied Physics, Vol. 83, No. 4, pp. 1988-1994, 1998.

[112] R.T.-P. Lee, K.-M. Tan, A.E.-J. Lim, T.-Y. Liow, G.S. Samudra, et al.: *P-channel tri-gate FinFETs featuring NiPtSiGe sourcedrain contacts for enhanced drive current performance*, IEEE Electron Device Letters, Vol. 29, No. 5, pp. 438-441, 2008.

[113] B.E. Deal: *Standardized terminology for oxide charges associated with thermally oxidized silicon*, IEEE Transactions on Electron Devices, Vol. 27, No. 3, pp. 606-608, 1980.

[114] M. Trentzsch: *Integration von SiON Gatedielektrika im Bereich sub-1,5nm für die Anwendung in 90-65nm High-Performance CMOS Technologien*, Dissertation, TU Chemnitz, 2008.

[115] G. Groeseneken, H.E. Maes, N. Beltran, R.F. De Keersmaecker: *A reliable approach to charge-pumping measurements in MOS transistors*, IEEE Transactions on Electron Devices, Vol. 31, No. 1, pp. 42-53, 1984.

[116] T. Hayashi, T. Yamashita, K. Shiga, K. Hayashi, H. Oda, et al.: *Suppression of boron penetration from source/drain-extension to improve gate leakage characteristics and gate-oxide reliability for 65-nm node CMOS and beyond*, Japanese Journal of Applied Physics, Vol. 44, No. 4B, pp. 2157-2160, 2005.

[117] J.C. Gelpey, S. McCoy, W. Lerch, S. Paul, J. Niess, et al: *Advanced activation and stability of ultra-shallow junctions using flash-assisted RTP*, RTP - Conference Proceedings, pp. 73-83, 2005.

[118] S. Flachowsky, R. Illgen, T. Herrmann, A. Wei, J. Höntschel, et al.: *Strain engineering for performance enhancement in advanced nano scaled SOI-MOSFETs*, Nanofair - Conference Proceedings, 2009.

[119] R.W. Olesinski, G.J. Abbaschian: *The Ge-Si system*, Bulletin of Alloy Phase Diagrams, Vol. 5, No. 2, pp. 180-183, 1984.

[120] D.F. Downey, J.W. Chow, E. Ishida, K.S. Jones: *Effect of fluorine on the diffusion of boron in ion implanted Si*, Applied Physics Letters, Vol. 73, No. 9, pp. 1263-1265, 1998.

[121] A. Mokhberi, R. Kasnavi, P.B. Griffin, J.D. Plummer: *Fluorine interaction with point defects, boron, and arsenic in ion implanted silicon*, Applied Physics Letters, Vol. 80, No. 19, pp. 3530-3532, 2002.

[122] A. Oishi, T. Komoda, Y. Morimasa, T. Sanuki, H. Yamasaki, et al.: *High performance CMOSFET technology for 45nm generation*, Symposium on VLSI Technology - Conference Proceedings, pp. 166-167, 2004.

[123] A. Wei, M. Wiatr, A. Mowry, A. Gehring, R. Boschke, et al.: *Multiple stress memorization in advanced SOI CMOS technologies*, Symposium on VLSI Technology - Conference Proceedings, pp. 216-217, 2007.

[124] Z. Wan, J. Chen, D. Tang, L. Chen: *Energy contamination control during ion beam deceleration for low energy ion implantation*, Ion Implantation Technology (IIT) - Conference Proceedings, pp. 221-224, 2002.

[125] M. Herden, D. Gehre, T. Feudel, L. Herrmann: *Influence of energy contamination at S/D-extension dopand implantation using ultra fast annealing*, IIT - Conference Proceedings, pp. 13-16, 2006.

[126] W. A. Krull: *Ion cluster technology*, INSIGHT in Semiconductor Device Fabrication, Metrology and Modeling - Conference Proceedings, 2007.

[127] S. Heo, S. Oh, M. Hasan, H.T. Cho, W.A. Krull, et al.: *Characteristics of ultrashallow p^+/n junction prepared cluster boron ($B_{18}H_{22}$) ion implantation and excimer laser annealing*, IWJT - Conference Proceedings, pp. 48-49, 2006.

[128] Y. Kawasaki, T. Kuroi, T. Yamashita, K. Horita, T. Hayashi, et al.: *Ultra-shallow junction formation by $B_{18}H_{22}$ ion implantation*, Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms, Vol. 237, pp. 25-29, December 2005.

[129] R.W. Olesinski, G.J. Abbaschian: *The C-Si system*, Bulletin of Alloy Phase Diagrams, Vol. 5, No. 5, pp. 486-489, 1984.

[130] J. Tersoff: *Enhanced solubility of impurities and enhanced diffusion near crystal surfaces*, Physical Review Letters, Vol. 74, No. 25, pp. 5080-5083, 1995.

[131] J.W. Strane, H.J. Stein, S.R. Lee, S.T. Picraux, J.K. Watanabe, et al.: *Precipitation and relaxation in strained $Si_{1-y}C_y/Si$ heterostructures*, Journal of Applied Physics, Vol. 76, No. 6, pp. 3656-3668, 1994.

[132] Y. Wang, S. Chen, M. Shen, X. Wang, S. Zhou, et al.: *Laser spike annealing and its application to lading edge logic devices*, RTP - Conference Proceedings, pp. 57-63, 2008.

[133] P. Grudowski, V. Dhandapani, S. Zollner, D. Goedeke, K. Loiko, et al.: *An embedded silicon-carbon S/D stressor CMOS integration on SOI with enhanced carbon incorporation by laser spike annealing*, SOI Conference - Conference Proceedings, pp. 17-18, 2007.

[134] M. Sinha: *Schottky barrier height tuning of silicide on $Si_{1-x}C_x$*, Applied Physics Letters, Vol. 91, No. 24, pp. 242108-242110, 2007.

[135] T. Onizawa, S. Kato, T. Aoyama, K. Ikeda, Y. Ohji: *The fabrication of low leakage junction with ultra shallow profile by the combination annealing of 10-ms low power and 2-ms high power FLA*, Symposium on VLSI Technology - Conference Proceedings, pp. 162-163, 2009.

[136] J. Hebb, Y. Wang, S.Chen, M. Shen, S. Zhou, X. Wang, D.Owen: *Expanded application space for laser spike annealing of CMOS devices*, RTP - Conference Proceedings, pp. 1-7, 2009.

Nachwort

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter an der Fakultät Elektrotechnik der Hochschule für Technik und Wirtschaft Dresden in Zusammenarbeit mit der Abteilung Technologieentwicklung (Technology Development, TD) von GLOBALFOUNDRIES in Dresden im Zeitraum von November 2005 bis Dezember 2010.

Ich möchte mich an dieser Stelle für jede erhaltene Unterstützung sowohl seitens der Hochschule als auch von Seiten GLOBALFOUNDRIES bedanken.

Mein Dank gilt in erster Linie Herrn Prof. Dr.-Ing. habil. Gerald Gerlach für die wohlwollende Förderung meiner wissenschaftlichen Tätigkeit, sein entgegengebrachtes Vertrauen und die Übernahme des Erstgutachtens.

Besonders bedanken möchte ich mich weiterhin bei Herrn Prof. Dr.-Ing. habil. Roland Stenzel, der mir die Möglichkeit zur Durchführung dieser Arbeit gegeben hat, mir stets mit seiner Diskussionsbereitschaft zur Seite stand, mich durch seine Begeisterungsfähigkeit motiviert hat und durch seine Unterstützung in jeder Hinsicht entscheidend zum Gelingen dieser Arbeit beitrug.

Eine besondere Freundschaft verbindet mich mit Herrn Dr.-Ing. Stefan Flachowsky, der mich während meines Studiums und meiner bisherigen beruflichen Laufbahn hilfreich und motivierend unterstützt hat und mir durch seine unschätzbare Fachkompetenz bei wissenschaftlichen Problemen stets neue Perspektiven eröffnen konnte.

Einen besonderen Dank möchte ich an dieser Stelle weiterhin Herrn Dr.-Ing. Tom Herrmann für die sehr gute Zusammenarbeit und die stete Bereitschaft zu vielfältigen und intensiven Diskussionen zukommen lassen, die wesentlich zum Erfolg dieser Arbeit beigetragen haben.

Ein weiterer Dank gilt allen Studenten, Praktikanten, Diplomanden und Doktoranden, die im Laufe meiner wissenschaftlichen Tätigkeit ein Zimmer bzw. einen Cube mit mir geteilt haben. Namentlich erwähnen möchte ich an dieser Stelle: Ina Ostermay, Andreas Naumann, Torben Kelwing und Tim Baldauf.

Des Weiteren bin ich den Kolleginnen und Kollegen der Fakultät Elektrotechnik der Hochschule für Technik und Wirtschaft Dresden für das freundliche und harmonische Arbeitsklima und für die vielfältige Unterstützung zu Dank verpflichtet.

Auch möchte ich mich sehr herzlich bei den Mitarbeitern der Abteilung Technologieentwicklung von GLOBALFOUNDRIES bedanken. In zahlreichen Diskussionen und freundschaftlichen Gesprächen haben sie mir weit mehr auf meinen weiteren Weg mitgegeben als ihr Fachwissen.

Ganz besonders weiß ich die Geduld, die Rücksichtnahme und die Motivation für diese Arbeit seitens meiner Frau Diana zu schätzen, bei ihr habe ich stets Verständnis und Rückhalt auch in schwierigen Situationen erhalten, um diese Aufgabe zu bewältigen.

Lebenslauf

Persönliche Daten

Name:	Ralf Illgen
Geburtsdatum:	24.11.1972 in Riesa

Studium

09/2001 – 09/2005	Studium der Kommunikationstechnik an der *Hochschule für Wirtschaft und Technik Dresden*, Fachbereich Elektrotechnik
09/2003 – 03/2004	Praktikumssemester bei *Vodafone D2 GmbH*, Radebeul, Abteilung Netzplanung und Optimierung
09/2006	7. Dresdner Sommerschule Mikroelektronik

Beruflicher Werdegang

04/2006 – 02/2010	Wissenschaftlicher Mitarbeiter am *Zentrum für angewandte Forschung und Technologie e.V.* an der *Hochschule für Wirtschaft und Technik Dresden*
seit 03/2010	Entwicklungsingenieur bei *GLOBALFOUNDRIES LLC & Co. KG* Dresden