

CMOS auf hochohmigem Silizium für integrierte Mikrowellenschaltungen (MMIC)

Von der Fakultät Elektrotechnik und Informationstechnik
der Universität Stuttgart zur Erlangung der Würde eines
Doktor-Ingenieurs (Dr.-Ing.) genehmigte Abhandlung

Vorgelegt von
Dietmar Beck
geboren in Holzminden

Hauptberichter:	Prof. Dr. phil. E. Kasper
Mitberichter:	Prof. Dr.-Ing. M. Berroth
Tag der Einreichung:	1. Oktober 1998
Tag der mündlichen Prüfung:	19. Oktober 1999

Institut für Halbleitertechnik
Universität Stuttgart
1999

Inhaltsverzeichnis

Einleitung	9
1 Verwendete CMOS-Prozeßtechnik	13
1.1 Halbleitertechnologie	13
1.1.1 Reinigungs- und Ätzlösungen	13
1.1.2 Ofenprozesse	15
1.2 CMOS-Prozeß-Varianten für SIMMWIC	17
1.2.1 Standard-CMOS	17
1.2.2 CMOS auf hochohmigem Silizium	18
1.3 Prozeßmeßtechnik unter besonderer Berücksichtigung der niedrigen Dotierung	22
1.3.1 Vierpunkt-Messung und Schrägschliff	22
1.3.2 Spreading-Resistance Verfahren	25
1.3.3 SIMS	30
1.3.4 CV-Profilng	31
2 Bauelemente auf hochohmigem Siliziumsubstrat	33
2.1 MOSFETs	33
2.1.1 NMOS-FET der Variante A	34
2.1.2 NMOS-FET der Varianten B und C	42
2.1.3 PMOS-FET	46
2.1.4 Temperung in H_2N_2	49
2.2 MOS-Varaktoren	51
2.2.1 Dielektrische Integrität	51
2.2.2 Inversionskanäle	52
2.2.3 CV-Messungen bei Tieftemperatur	58
2.3 Source-Wannen-Dioden	60
2.4 pin-Dioden für Schalter	82
3 Hybridtechnik mit Multi-Chip-Modul	87
3.1 Unterschiede Hybridtechnik - monolithische Integration	87
3.2 Hybride CMOS-SIMMWIC-Rectenna	88

4 Vorarbeiten zur monolithischen Integration	91
4.1 Vorteile einer monolithischen Integration	91
4.2 Anforderungen an einen Integrationsprozeß	91
4.2.1 Untersuchungen mittels MOS-Varaktoren	93
4.2.2 Auswirkungen auf MOSFETs	100
4.3 Integrationsprozeß	102
4.4 Ansteuerschaltung für einen Hochfrequenz-pin-Schalter	104
4.5 Beeinflussung der HF-Dämpfung durch CMOS	108
5 Wertung und Ausblick	115
Anhang	
A 1 Entwurfsdaten aktiver und passiver Bauelemente	119
A 1.1 Geometrische Kenngrößen von MOS-Transistoren	119
A 1.2 Geometrische Kenngrößen passiver Bauelemente	125
A 1.3 Isolationen und Mindestabstände	129
A 2 Modellierung der MOS-Transistoren in SPICE	130
A 3 CV-Charakterisierung der Grenzfläche Si-SiO ₂	132
A 3.1 Kapazitäts-Spannungs-Kennlinien des MOS-Varaktors	132
A 3.2 Grenzflächenzustandsdichte	135
A 3.3 Problematik großer Serienwiderstände	141
A 4 Lebensdauerbestimmung an pn- und pin-Dioden	143
A 5 CMOS-Schaltungstechnik	148
A 5.1 Analoge Verstärkermeßtechnik	149
A 5.2 Analoge Verstärkerschaltungen	152
A 5.3 Ansteuerschaltung für pin-Schalter	163
Literaturverzeichnis	167

Formelzeichen

A	Querschnittsfläche
A_v	Spannungsverstärkung
a, b	Wellenamplitude
b	Breite
C	Kapazität
C_{inv}	Kapazität der Inversionsschicht
C_{it}	Kapazität durch Grenzflächenzustände
C_o	Oberflächenkonzentration
C'_{ox}	spezifische Oxidkapazität
C_{si}	Halbleiterkapazität
$C(V)$	spannungsabhängige Kapazität
D	Diffusionskonstante eines Elements in Silizium
D_a	ambipolare Diffusionskonstante
D_{it}	Grenzflächenzustandsdichte
D_p, D_n	Diffusionskonstante von Löchern und Elektronen
d	Abstand, Durchmesser
d_n, d_p	Länge der Bahngebiete
E	Energie
E_A	Aktivierungsenergie
E_f	Fermienergie
E_{gap}	Bandlücke
E_i	intrinsische Energie
E_{ms}	Austrittsenergie zwischen Metall und Halbleiter
dE	Energieänderung
e	Elektronenanzahl, Elementarladung
f	Frequenz
f_{3dB}	3dB-Frequenz
f_T	Transitfrequenz
F	Fläche
G	Leitwert
g_m	Stromsteilheit
g_o	Ausgangsleitwert
h	Substrathöhe
I	Strom
I_d	Kanalstrom
$I_{d,sat}$	Drainsättigungsstrom
I_s	Sättigungsstrom eines pn-Übergangs
I_{sub}	Substratstrom
k_p	geometrieunabhängiger Steilheitsparameter
L	Kanallänge des MOS
L_p, L_n	Diffusionslänge von Löchern und Elektronen
ΔL	Kanallängenänderung
M	Lawinenmultiplikationsfaktor
m_i	zentrales Moment i-ter Ordnung
N	Volumendichte
N_A	Akzeptorenkonzentration
N_C, N_V	effektive Zustandsdichten des Leitungs- und des Valenzbands
N_D	Donatorenkonzentration
N_{ox}	Dichte der Oxidladungen
$N(x)$	Ortsprofil der Dotierkonzentration
N_{\square}	Flächendosis

n	Elektronendichte, Atomkernanzahl
n_i	intrinsische Trägerdichte
n_p	Minoritätsträgerdichte von Elektronen im p-Gebiet
P	Punkt im Raum
p	Druck
p	Löcherdichte
p_n	Minoritätsträgerdichte von Löchern im n-Gebiet
Q_B	Sperrschichtladung
Q_f	feste Oxidladung
Q_i	Ladung der Inversionsschicht
Q_{it}	Grenzflächenladung
Q_m	mobile Ladung
Q_o	Gesamtdosis bei Implantationen
Q_{ot}	Trapladungen
Q_{ox}	Oxidladung
δQ	Ladungsänderung
R_b	Substratserienwiderstand
R_k	Kontaktwiderstand
R_m, R_{sr}	Ausbreitungswiderstand
R_p	mittlere Reichweite implantierter Ionen
ΔR_p	Standardabweichung der mittleren Reichweite implantierter Ionen
R_s	Schichtwiderstand, Serienwiderstand eines pn-Übergangs
ΔR_t	Transversale Aufweitung der Verteilung implantierter Ionen
r	Abstand
S	Gatespannungsintervall zur zehnfachen Drainstromerhöhung
S	Strahlungsdichte elektromagnetischer Wellen
S_e, S_n	Bremsquerschnitte von Elektronen und Atomkernen
$S_{11}, S_{12}, S_{21}, S_{22}$	Streuparameter
s	Abstand
ds	Weglänge
T	Temperatur
t	Zeit
t_{Al}	Aluminiumschichtdicke
t_{Au}	Goldschichtdicke
t_{ox}	Oxiddicke
U	Spannung, Umfang
U_d	Diffusionsspannung
U_t	Temperaturspannung
V_{bs}	Bulk-Source-Spannung
V_{ds}	Drain-Source-Spannung
$V_{ds,db}$	Durchbruchspannung
V_g	Gatespannung
V_{pt}	Durchgriffspannung
V_{th}	Schwellspannung
V_{FB}	Flachbandspannung
W	Kanalbreite des MOS
ΔW	Kanalbreitenänderung
w	Dicke einer leitfähigen Schicht oder eines Wafers
w	Sperrschichtweite
x	Ortskoordinate
x_j	Sperrschichttiefe
Y	komplexe Admittanz
z	Tiefenkoordinate
Z	komplexe Impedanz
Z_o	charakteristischer Wellenwiderstand

α	Winkel, Ionisationsrate
$\alpha(f)$	frequenzabhängige Dämpfungskonstante
β	Steilheitsparameter, Kurtosis einer Verteilung
γ	Substratkonstante, Schiefe einer Verteilung
δ	Eindringtiefe in metallische Leiter (skin depth)
Δ	Änderung, Fehlerbeitrag
ϵ_r	relative Dielektrizitätskonstante
η	Idealitätsfaktor
λ	Debye-Länge, Wellenlänge
μ	Beweglichkeit
μ_r	relative Permeabilitätskonstante
ϕ	Phase, Potential
ϕ_B	Barrierenhöhe
ϕ_s	Oberflächenpotential
ψ	Bandverbiegung
ρ	Spezifischer Widerstand
σ	Standardabweichung
σ	spezifische elektrische Leitfähigkeit
τ	Lebensdauer
ω	Kreisfrequenz

Einleitung

Im Zuge weiter fortschreitender Automatisierungsbemühungen in Bereichen des Verkehrswesens oder auch der industriellen Produktion kommt intelligenten und kostengünstigen Kommunikationssystemen steigende Bedeutung zu. Gerade nicht-ortsgebundene Objekte, wie etwa Transportschlitten in komplexen Produktionsanlagen, erfordern dabei einen effizienten Datenaustausch mittels elektromagnetischer Wellen anstelle leitungsgebundener Kommunikation. Auch der Einsatz in Großserienprodukten bedarf sehr kostengünstiger Lösungen, die sich mit den zur Zeit eingesetzten Technologien nur schwer erreichen lassen. Nach heutigem Stand bieten allein die Methoden der modernen Halbleitertechnik die Aussicht, diese Systeme unter den gesetzten Bedingungen in absehbarer Zukunft zur Verfügung stellen zu können.

Erst die Entwicklung elektronischer Schaltungen hin zu immer höheren Arbeitsfrequenzen erlaubt eine zunehmend schnellere Signalverarbeitungsgeschwindigkeit, die zur Realisierung komplexer Funktionen erforderlich ist. Bezogen auf Hochfrequenzschaltungen lassen sich mit steigender Frequenz außerdem vergrößerte Signalbandbreiten bei Kommunikationssystemen, verbesserte Auflösungen bei Sensoren und zunehmende Gewinne bei Antennen mit gleichbleibenden geometrischen Dimensionen erreichen. Gerade bei elektromagnetischen Signalen im Millimeterwellenbereich, d.h. bei Signalfrequenzen oberhalb 30 GHz, werden durch die geringe Wellenlänge sehr kompakte Antennenstrukturen möglich. So beträgt die Vakuumwellenlänge bei $f=100$ GHz nur noch $\lambda_0=3$ mm.

Die Integration von aktiven Bauelementen, wie Transistoren und Dioden, mit planaren Antennen und Wellenleiterstrukturen führt zum Konzept der MMICs (Millimeter-Wave Monolithic Integrated Circuits). Für diese Art von Technik spricht die bei entsprechend hoher Stückzahl erzielbare Verringerung der Produktionskosten, sowie eine gegenüber diskret aufgebauten Schaltungen erhöhte Zuverlässigkeit bei geringerer Größe und reduziertem Gewicht.

Die Hauptanwendungsgebiete von Millimeterwellen werden in Zukunft verstärkt im Bereich der Nahfeldkommunikation mit hohen Übertragungsraten und bei der Automobilsensorik, wie etwa den Antikollisionsradars, gesehen. Die Verbreitung solch hochfrequenter Schaltungen in nicht-militärischen Anwendungsgebieten setzt allerdings eine deutliche Kostensenkung der MMICs voraus. Die zur Zeit vorherrschende Technik auf Basis von III-V-Verbindungshalbleitern, wie z.B. Gallium-Arsenid (GaAs), hat diesbezüglich mit dem Nachteil hoher Substrat- und Prozeßkosten zu kämpfen. Während diese Technologie aufgrund der im Vergleich zu Silizium deutlich höheren Elektronenbeweglichkeit als Höchsthochfrequenztechnik einige Vorzüge bietet, existieren viele Anwendungen, in denen Bauelemente, die mittels einer entsprechend modifizierten Siliziumtechnik hergestellt werden, potentiell kostengünstigere Alternativen bieten können.

Für siliziumbasierte MMICs, sogenannte SIMMWICs (Silicon Monolithic Millimeter-Wave Integrated Circuits), stehen zahlreiche Silizium- und Silizium-Germaniumbauelemente mit Arbeitsfrequenzen bis etwa 100 GHz zur Verfügung. Als Beispiele gelten hier die IMPATT-Diode (IMPACT Avalanche Transit-Time) zur Erzeugung hochfrequenter Schwingungen, die Schottky-Diode zur Detektion von Millimeterwellensignalen [0.1] und der Heterobipolartransistor (HBT) als aktives Bauelement [0.2]. Diese Bauelemente werden vornehmlich unter Ausnutzung der Möglichkeiten der Molekularstrahlepitaxie (MBE) hergestellt. Mit Hilfe dieser Bauelemente sind sowohl Ein-Chip-Empfänger und Sender als auch Mischerschaltungen realisiert worden [0.3]. Pin-Dioden kommen als HF-Schalter und zur Modulation von Hochfrequenzsignalen in entsprechenden Schaltungen zum Einsatz.

Bei einem Vergleich der Substrateigenschaften [0.4] zwischen semi-isolierendem GaAs mit spezifischen Widerständen von $\rho > 10^6 \Omega\text{cm}$ und hochohmigem Silizium mit $\rho > 10^4 \Omega\text{cm}$ zeigt sich, daß trotz des geringfügig höheren Verlustfaktors bei Silizium keine signifikant unterschiedliche Dämpfung in den Wellenleiterstrukturen auftritt. Dies liegt in der Dominanz der metallischen Leiterverluste gegenüber den Substratverlusten begründet. Die in etwa gleich große Dielektrizitätskonstante, $\epsilon_{r,\text{GaAs}} = 12.9$ bei Gallium-Arsenid und $\epsilon_{r,\text{Si}} = 11.7$ bei Silizium, führt zu einer deutlichen Reduktion der Substratwellenlänge verglichen mit der Vakuumwellenlänge und folglich zu einer Verkleinerung der erforderlichen Strukturen. Die thermische Leitfähigkeit des Siliziums ist etwa dreimal so hoch wie die von GaAs und vereinfacht die Wärmeableitung von Leistungsbaulementen beträchtlich. Dies ist insbesondere bei Sendeoszillatoren von Vorteil. Mittels thermischer Oxidation gebildete Isolations-schichten stellen nur einen der prozeßtechnischen Vorteile des Siliziums gegenüber GaAs dar, die in ihrer Summe zur beherrschenden Stellung des Siliziums auf dem Halbleitermarkt geführt haben.

Diese Dominanz des Siliziums als Basismaterial der modernen Halbleitertechnik zeigt sich in der hohen Verbreitung von Bauelementen, die heute zum weitaus größten Teil in CMOS- (Complementary Metal Oxide Semiconductor), n-MOS [0.5] oder Bipolartechnik gefertigt werden. Gerade aus der teilweisen Übernahme von etablierten Prozeßmodulen dieser Technologien bezieht die SIMMWIC-Technik ihre besondere Attraktivität.

Die CMOS-Technologie wiederum stellt heute die am weitesten verbreitete Methode zur Herstellung integrierter Schaltungen auf Silizium dar. Dies erklärt sich insbesondere aufgrund ihrer Eignung für rein digitale Schaltungen, so werden z.B. Mikroprozessoren und Speicherchips in dieser Technik produziert. Seit einigen Jahren haben aber auch analoge Schaltungsfunktionen bei gemischt analog-digitalen Applikationen zusätzliche Bedeutung für die CMOS-Technik erhalten. Die Gründe der großen Verbreitung von CMOS liegen sowohl in der vergleichsweise einfachen Integrationsfähigkeit der MOS-Bauelemente als auch in ihrer guten technologischen Reproduzierbarkeit. Durch den Einsatz komplementärer Bauelemente, d.h. p- und n-Kanal-MOS-Transistoren, lassen sich auch komplexe Schaltungen mit recht geringer Leistungsaufnahme herstellen. Ein hoher Signal-Rauschspannungsabstand und große Ausgangsspannungshübe vervollständigen die positiven Eigenschaften dieser Technik. Mittels der spannungsgesteuerten Eingangskopplung von MOS-Fets ist ein gegenüber der n-MOS- oder Bipolar-technik vereinfachter schaltungstechnischer Aufbau möglich.

Zu den Nachteilen der CMOS-Technik gehören eine im Vergleich zur Bipolar-technik verringerte maximale Schaltgeschwindigkeit und eine Beschränkung auf geringere Ströme und Spannungen. Einschränkend ist jedoch zu sagen, daß aufgrund der Schrittmacherfunktion der CMOS-Technik, insbesondere bei den hohen Integrationsdichten der ULSI-Technologie (Ultra Large Scale Integration) [0.6], die Geschwindigkeitsdifferenz zur konventionellen Silizium-Bipolar-technik abnimmt. Erweiterungen der reinen CMOS-Technik führen zur Inkorporation von Bipolartransistoren für schnelle Schaltfunktionen sowie zu höheren Leistungen und bilden die BiCMOS-Technik (Bipolar CMOS). Spezielle Bauformen, wie etwa der VMOS, erweitern die Einsatzgebiete von MOS-Transistoren durch ihre Verfügbarkeit bei höheren Spannungen.

Wie man beim Übergang von diskret aufgebauten Hochfrequenzschaltungen zu MMICs und bei den Erweiterungen der CMOS- zur BiCMOS-Technik gesehen hat, ist die Inkorporation zusätzlicher Funktionalität ein Haupttrend in der modernen Elektronik. Auf die SIMMWIC-Technik bezogen eröffnet nun eine Integration dieser Technik mit CMOS neue Perspektiven für komplette Ein-Chip-Systeme. Diese Systeme könnten dann neben der gesamten Hochfrequenzelektronik in SIMMWIC-Technik ebenfalls signalverarbeitende Funktionen in CMOS-Technik ausführen. Da eine Signalverarbeitung sowohl auf der Sender- als auch auf der Empfängerseite meist auf einer niedrigeren Zwischenfrequenz stattfindet, bietet sich hierfür eine monolithische Integration der SIMMWIC-Front-Ends mit komplexen Ansteuer- und Auswerteschaltungen in CMOS-Technologie an.

Durch die monolithische Integration von SIMMWIC- und CMOS-Schaltungen ist neben ökonomischen Vorteilen, die sich bei der Produktion von großen Stückzahlen ergeben, auch eine erhöhte Zuverlässigkeit der Gesamtschaltung zu erwarten. Bei einer geschickten Ausnutzung der Freiräume zwischen den Wellenleiterstrukturen lassen sich auch komplexe analoge und digitale Funktionen ohne nennenswerte Vergrößerung der Chipfläche einbeziehen.

Zu den Nachteilen einer Integration gehören neben einer vergrößerten Prozeßkomplexität im Vergleich zu den beiden Einzeltechnologien auch eine daraus resultierende verringerte Ausbeute. Allerdings sind diese Nachteile insofern zu relativieren, da sich in der Vergangenheit diese Schwierigkeiten durch Fortschritte in der Prozeßtechnik stets so gut beherrschen ließen, daß die Vorteile der weitergehenden Integration überwogen.

Die Untersuchung der prinzipiellen Machbarkeit und eine Abschätzung der zu erwartenden Schwierigkeiten bei einer Integration von CMOS- und SIMMWIC-Technik stehen daher im Vordergrund dieser Arbeit.

Bei der Integration sind einige grundlegende Aspekte zu berücksichtigen: Das sehr hochohmige Substrat der SIMMWIC-Schaltungen, üblicherweise mit spezifischen Widerständen im Bereich einiger 1000 Ωcm , ist für das Erreichen niedriger Signaldämpfungen auf den Wellenleiterstrukturen erforderlich. Im Gegensatz dazu erfordern MOS-Transistoren zur Optimierung ihrer Schwellspannungen und Spannungsfestigkeiten mittlere Dotierungen (um etwa 10^{16} cm^{-3}), die zu vergleichsweise niederohmigen Substraten führen.

Ein mögliches Integrationskonzept bedingt daher zuerst die Anpassung eines CMOS-Prozesses auf das sehr niedrig dotierte Substrat der SIMMWIC-Technologie. Das durch den Prozeß des Zonenziehen (engl. float zone) hergestellte Substrat für SIMMWIC-Schaltungen weist in Bezug auf seine innere Struktur außerdem Abweichungen zu den üblichen Substraten für CMOS auf.

Um im Bereich der Wellenleiterstrukturen die geringe Leitfähigkeit zu konservieren, kommt den Maskierungen gegen eine unerwünschte Dotierung der entsprechenden Gebiete besondere Bedeutung zu. Sowohl die durch CMOS-Prozeßschritte auf den Wafer aufgetragenen Dotierstoffe als auch die noch in den Prozeßöfen befindlichen Dotierstoffreste (sog. Autodoping) können zu einer Beeinträchtigung der geringen Leitfähigkeit des Substrats führen. Die Beschreibung des als Basis eingesetzten institutseigenen CMOS-Prozesses und die im Rahmen dieser Arbeit untersuchten Lösungsstrategien für die Umsetzung dieses vorhandenen Prozesses auf das hochohmige Substrat sind Thema des ersten Kapitels. Um die Vergleichbarkeit der zu entwickelnden Prozeßvarianten mit dem Ausgangsprozess sicherzustellen, sollte diese Anpassung mit möglichst geringen Modifikationen erfolgen.

Beim Vergleich der auf hochohmigem Substrat erhaltenen Transistoren mit Bauelementen, die auf dem moderat dotierten n-Substrat des als Vergleichsstandard dienenden Instituts-Prozesses realisiert worden sind, konnten Unterschiede im elektrischen Verhalten beobachtet werden. Diese Abweichungen werden im zweiten Kapitel anhand verschiedener Modellansätze erklärt. Zur genaueren Charakterisierung des Transistorverhaltens wird den Sperrschichtübergängen der Source- und Draingebiete besondere Aufmerksamkeit geschenkt. Neben einer umfassenden Gleich- und Wechselstromcharakterisierung der Transistoren erfordert die Grenzfläche zwischen Silizium und Siliziumoxid besondere Beachtung. Hierzu ist die Anwendbarkeit von Kapazität-Spannungsmessungen (CV) auf MOS-Varaktoren mit sehr großen Serienwiderständen erweitert worden. Diese Messungen bilden die Basis für die Bewertung der durch die Prozessschritte zur Technologie-Integration verursachten Auswirkungen auf das Gateoxid.

Grundlage der Arbeiten zur Integration ist die Demonstration der Funktionsfähigkeit von Bauelementen auf hochohmigem p⁻-Substrat anhand gering komplexer CMOS-Schaltungen. Diese im Anhang beschriebenen Schaltkreise erlauben den Vergleich mit entsprechenden Schaltungen auf konventionellem Substratmaterial. Sie wurden im Rahmen der Integrationsversuche als Verstärker und Ansteuerschaltungen eingesetzt.

Die hybride Kombination von SIMMWIC- und CMOS-Schaltungen als sinnvoller Zwischenschritt zur monolithischen Integration ist das Thema des dritten Kapitels. Insbesondere bei kleineren Produktionsstückzahlen ist dieser Ansatz oft auch die kostengünstigere Alternative. Eine SIMMWIC-Empfangsschaltung mit nachgeschaltetem CMOS-Verstärker bildet hierbei einen Empfänger im Gigahertz-Bereich mit beachtlicher Empfindlichkeit. Diese und alle weiteren im Rahmen dieser Arbeit eingesetzten SIMMWIC-Komponenten wurden von der Abteilung F2H der Daimler-Benz-Forschung, Ulm, entwickelt und zur Verfügung gestellt.

Die Voruntersuchungen zur grundsätzlichen Integrationsfähigkeit der entwickelten CMOS-Prozesse auf hochohmigem Silizium mit einem SIMMWIC-Prozeß wurden anhand von MOS-Varaktoren und FETs durchgeführt. Die Konzeption eines Integrationsprozesses unter Berücksichtigung der unterschiedlichen Temperaturbudgets, Schichtfolgen und Metallisierungen der beiden Prozesse steht im Mittelpunkt des vierten Kapitels. Erste Ergebnisse der Anwendung dieser Prozeßfolge werden anhand eines integrierten pin-Dioden-Hochfrequenzschalters erläutert. Neben der Problematik des oben erwähnten Autodoping ist auch durch die Präsenz der CMOS-Schaltungsteile in unmittelbarer Nähe der Wellenleiter eine Beeinträchtigung der Hochfrequenzdämpfung zu erwarten. Zur vergleichenden Analyse des HF-Dämpfungsverhaltens und der gegenseitigen Beeinflussung beider Technologien wurden Versuche mit Aluminium- und Gold-Mikrostreifenstrukturen auf lokal dotierten Silizium-Wafern durchgeführt. Die Bestimmung des Mindestabstands zwischen den beiden Schaltungsteilen mit Hilfe spezieller Teststrukturen und Messungen im Gigahertz-Bereich ist daher ebenso Thema dieses Kapitels.

Während der im Rahmen dieser Arbeit erfolgten Untersuchungen konnte die grundsätzliche Machbarkeit einer monolithischen Integration gezeigt werden. Eine Bewertung der verfolgten Ansätze sowie Empfehlungen für zukünftige Arbeiten sind im fünften Kapitel aufgeführt.

In der nun folgenden Darstellung werden die Vorgehensweisen bei diesen Untersuchungen beschrieben und die erzielten Ergebnisse vorgestellt.

Verwendete CMOS-Prozeßtechnik

In diesem Kapitel steht die Konzeption der CMOS-Prozesse auf hochohmigem p⁻-Substrat im Vordergrund. Im Rahmen der Beschreibung einzelner Technologie-Module (siehe Abschnitt 1.1) wird auch die Frage der ausreichenden Maskierwirkung herkömmlicher Oxiddicken für hochohmiges Siliziumsubstrat behandelt (Abs. 1.1.2). In Abschnitt 1.2 wird die Prozeßfolge des als Vergleichsbasis dienenden CMOS-Prozesses auf n-Substrat dargestellt und den im Rahmen dieser Arbeit entwickelten Prozeßvarianten auf hochohmigem Silizium gegenübergestellt.

Die zur Charakterisierung der Dotierungsprofile eingesetzten Meßverfahren dienen zur Überprüfung der mittels Prozeßsimulation vorausgesagten Ergebnisse und sind in Abschnitt 1.3 aufgeführt.

1.1 Halbleitertechnologie

Aufgrund der zahlreichen Einflußgrößen auf die Prozeßergebnisse bei der Fabrikation von Halbleiterbauelementen bedient man sich gemeinhin fester Module, wie Reinigungs-, Ätz-, Lithographie-, Temperatur-, Implantations- und Metallisierungsschritte, deren zentrale Parameter unter Beibehaltung der weiteren Größen angemessen modifiziert werden. Auch in dieser Arbeit ist ein solcher Weg beschritten worden, so daß weitgehend auf Prozeßmodule des Instituts-CMOS-Prozesses auf n-Silizium zurückgegriffen werden konnte. Durch diese Kontinuität konnte eine bessere Vergleichbarkeit der entwickelten CMOS-Prozesse auf hochohmigem p⁻-Substrat mit dem vorhandenen Prozeß erreicht werden.

1.1.1 Reinigungs- und Ätzlösungen

Heute in der Halbleiterfertigung eingesetzte Reinigungsverfahren haben die Entfernung unerwünschter Metalle und sonstiger Verunreinigen in Form von Partikeln und Filmen von der Waferoberfläche zum Ziel. Man unterscheidet hierbei zwischen

(i) organischen Verunreinigungen, wie sie etwa beim Einsatz organischer Lösungsmittel (z. B. Aceton bei der Entfernung von Positiv-Photolacken) auf der Oberfläche zurückbleiben. Aus der Luft aufgenommene Partikel, Rückstände durch manuelles Wafer-Handling oder Ölrreste bei diversen Vakuum-Prozeßschritten gehören ebenfalls zu dieser Gruppe.

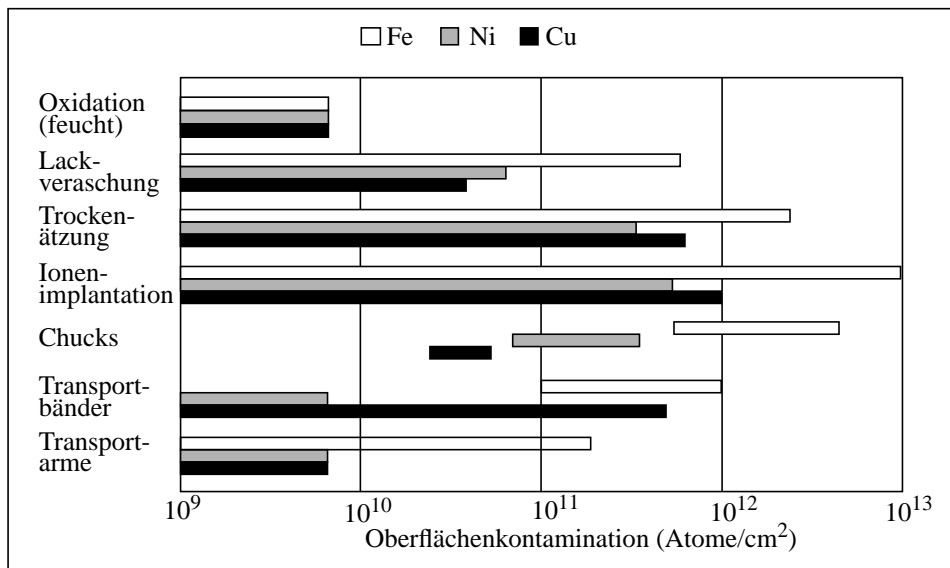


Bild 1.1: Darstellung metallischer Kontaminationsquellen und typischer Konzentrationen in der Halbleiterfertigung [1.1].

(ii) Metalle gelangen bei nahezu allen Prozeßschritten in geringer Konzentration auf den Si-Wafer (vgl. Bild 1.1) und gehören zur zweiten wichtigen Gruppe, den nicht-organischen Verunreinigungen. Diese können in Form von Partikeln und Filmen an der Oberfläche vorliegen oder bei Bildung des sog. natürlichen Oxides darin eingebaut werden. Die Entfernung dieses sich in natürlicher Atmosphäre sehr schnell bildenden Oxides - anfangs $t_{\text{ox},n} = 15$ bis 20 \AA stark und bis auf $t_{\text{ox},n} = 40 \text{ \AA}$ anwachsend [1.2] - besteht daher als eine Hauptforderung an eine effiziente Reinigung. Da sich neben Ionen auch ungeladene Metallatome auf der Oberfläche anlagern, ist eine Oxidation dieser Substanzen zur besseren Löslichkeit vorteilhaft. Aufgrund dieser und weiterer Anforderungen hat sich eine von Kern [1.3] eingeführte und in der Folge leicht modifizierte Reinigungssequenz in der Halbleiterfabrikation durchgesetzt. Die ursprüngliche Form dieser als Standard-Reinigung oder „RCA-clean“ bezeichneten Sequenz besteht aus den beiden folgenden Schritten:

(i) Eine alkalische Lösung auf Basis von Ammoniaklauge (NH_4OH 29 %ig) im Mischungsverhältnis von 1:1:5 mit Wasserstoffperoxid (H_2O_2 30 %ig) und de-ionisiertem Wasser (DI-Wasser H_2O) dient zur Entfernung organischer Filme und metallischer Verunreinigungen. Die Komplexbildung der Lauge mit verschiedenen Metallen erhöht deren Löslichkeit und erlaubt die Entfernung von Gold, Silber, Kupfer, Nickel, Kadmium, Zink, Kobalt und Chrom. Die Wirkung des Wasserstoffperoxides besteht in einer Reoxidation der durch die Lauge entfernten Oxidschicht (Ätzrate je nach Oxidqualität 0.1 ... 0.4 nm/min bei $T = 80 \text{ }^\circ\text{C}$ [1.4]), so daß eine fortwährende langsame Abtragung der Siliziumoberfläche erfolgt. Obwohl eine über $T = 80 \text{ }^\circ\text{C}$ steigende Temperatur während der Reinigung die Reaktionsgeschwindigkeit und die -effizienz erhöhen sollte, kann aufgrund des dann beschleunigten H_2O_2 -Zerfalls die Temperatur nicht über diesen Wert erhöht werden.

(ii) Die anschließende saure Reinigungslösung basierend auf Salzsäure (HCl 37 %ig) im Mischungsverhältnis 1:1:6 mit Wasserstoffperoxid (H_2O_2 30 %ig) und de-ionisiertem Wasser (H_2O) dient zur Entfernung von alkalischen Ionen, z.B. Kalium und Natrium, sowie zur weiteren Reduktion der Konzentration der metallischen Verunreinigungen, wie z.B. Aluminium, Eisen und Mangan. Auch hier verhindert die Komplexbildung eine Wiederablagerung der gelösten Verunreinigungen auf dem Wafer. Die Oxidation des Siliziums hinterläßt eine passivierte Waferoberfläche wie bei der alkalischen Lösung.

Für beide Schritte wird eine Temperatur von $T = 80 \text{ }^\circ\text{C}$ und eine Prozeßdauer von $t = 10 \text{ min}$ empfohlen.

Bei starken organischen Verunreinigungen, wie nach einem Lithographieschritt durch Photolackreste oder nach Implantationsprozessen, ist als vorausgehender Reinigungsschritt eine Schwefelsäurereinigung (H_2SO_4 96 %ig und H_2O_2 30 %ig im Verhältnis 2:1) empfehlenswert, die in der Literatur auch als „Caro'sche“ Säure oder „Piranha-etch“ bekannt ist. Bei Temperaturen oberhalb von $T = 80 \text{ }^\circ\text{C}$ und einer Prozeßzeit von $t = 30 \text{ min}$ erfolgt die Reinigung unter Bildung eines Oxides an der Si-Oberfläche. Wird vor Anschluß der Standardreinigung eine Entfernung dieses Oxides in 1 %iger Flußsäurelösung (HF) vorgenommen, so lassen sich die metallischen Verunreinigungen deutlich reduzieren [1.5]. Voraussetzung dafür ist jedoch eine sehr hohe Qualität des zum Spülen eingesetzten DI-Wassers, da die hydrophobe Oberfläche des Siliziums nur vorübergehend durch Fluoratomate abgesättigt ist und somit eine starke Anlagerung von im Wasser befindlichen Verunreinigungen an unbesättigten

und damit bindungsfreudigen Si-Atomen erfolgen kann. Diese Spülung von am Wafer befindlichen gelösten Ionenkomplexen, die nach jedem Reinigungsschritt durchgeführt wird, ist bei schon passivierten Oberflächen relativ unkritisch, da die Oxidoberfläche deutlich weniger reaktionsfreudig ist.

Eine nach Durchführung der gesamten Reinigungssequenz (vgl. Bild 1.2) erfolgende Oxidentfernung mittels HF-Lösung ist bei Vorbelegungsschritten (siehe Abschnitt 1.1.2) und vor der Metallisierung unbedingt notwendig, da sich andernfalls durch das Oxid Diffusionsbarrieren bzw. Probleme bei der Bildung ohmscher Kontakte ergeben können [1.6].

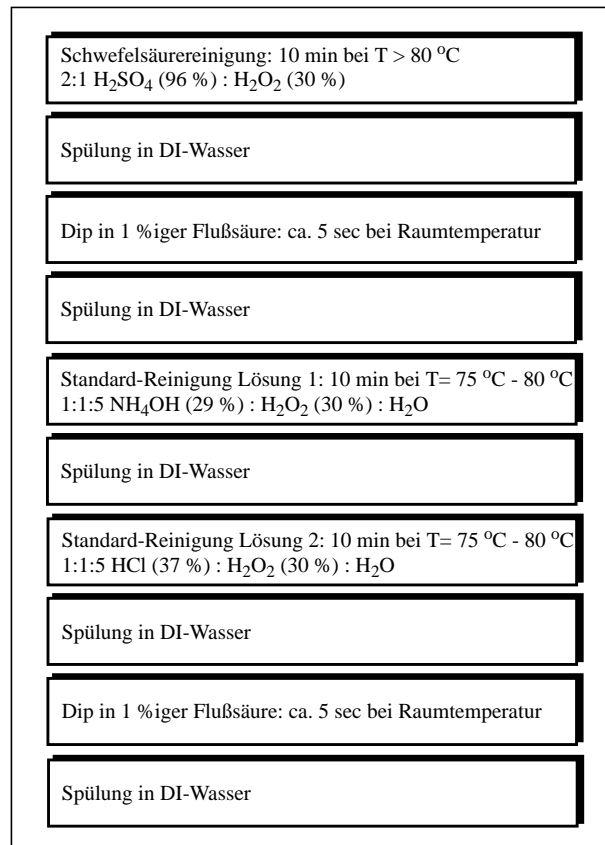


Bild 1.2: Komplette Reinigungssequenz, wie für die CMOS-Prozessierung in dieser Arbeit eingesetzt.

Bei der CMOS-Prozessierung in dieser Arbeit sind ausnahmslos feuchte Ätzprozesse eingesetzt worden. Für die Definition von Öffnungen in der Oxidmaske ist eine Ätzung in gepufferter Flußsäure BHF ($\text{HF}/\text{NH}_4\text{F}/\text{H}_2\text{O}$) bei Raumtemperatur für dünne und bei $T = 55\text{ °C}$ für dickere thermische Oxide eingesetzt worden. Die Strukturierung des Aluminiums mittels einer die gewünschten Strukturen abdeckenden Lackmaske erfolgte mit Phosphorsäure (H_3PO_4 85 %ig) bei $T = 55\text{ °C}$.

1.1.2 Ofenprozesse

Im CMOS-Prozß auftretende Temperaturschritte beinhalten

- (i) Oxidationen in trockener und feuchter Atmosphäre, die Funktionsschichten, Isolierungen und Oberflächenpassivierungen erzeugen,
 - (ii) Vorbelegungsschritte zur Einbringung hochdotierter P^+ - und N^+ -Gebiete mittels gasförmigen Diboran (B_2H_6) und Phosphoroxichlorid- (POCl_3) Quellen,
 - (iii) Diffusionsschritte zur Erreichung der gewünschten Sperrschichttiefen, sog. Drive-In-Schritte nach Vorbelegungen und Implantationen, sowie
 - (iv) Temperschritte zur Bildung ohmscher Kontakte zwischen Aluminiumleiterbahnen und hochdotierten Bauelementanschlüssen in reiner (N_2) oder wasserstoffhaltiger (N_2H_2) Stickstoffatmosphäre.
- Oxidationsschritte zur Erzeugung geringer Oxiddicken (bis ca. $t_{\text{ox}} = 100\text{ nm}$) erfolgen in sog. trockener Atmo-

sphäre, d.h. einer O₂-N₂-Atmosphäre, und führen zu qualitativ hochwertigen thermischen Oxiden. Diese Oxide dienen zum einen als Streuoxide mit einer Dicke t_{ox} = 60 nm bei Implantationen oder als Gateoxid mit t_{ox} = 95 nm. Um bei letzterem eine möglichst geringe Konzentration an Verunreinigungen während des Wachstumsprozesses zu erreichen, d.h. insbesondere den Gehalt an mobilen Ladungen Q_m (vgl. Anhang 3.1) zu minimieren, ist der Gateoxidation eine Chlorgetterung des Quarzrohres vorgeschaltet [1.7]. Beim Wachstum von Feldoxiden mit Oxiddicken von t_{ox} = 1.0 - 1.3 µm geht der deutlich rascher ablaufenden sog. feuchten Oxidation eine 5 minütige trockene Oxidation voraus, die insbesondere während der thermischen Stabilisierung der Wafer bei Aufheißvorgängen ein stark unterschiedliches Oxidwachstum verhindern soll. Bei der feuchten Oxidation wird der Sauerstoff vor Eintritt in das Reaktionsrohr in einem sog. Bubbler-Gefäß in ca. 95 °C heißem de-ionisierten Wasser mit Wasserdampf gesättigt. Die höhere Diffusivität von Wassermolekülen (H₂O) verglichen mit Sauerstoffmolekülen (O₂) durch die Oxidschicht (SiO₂) an die Grenzfläche zum Silizium (Si) wird für die schnellere Oxidation in feuchter Atmosphäre verantwortlich gemacht [1.8]. Die geforderte Vergleichbarkeit der CMOS-Prozesse auf hochohmigem p⁻-Substrat mit dem als Referenz dienenden Prozeß auf n-Substrat bedingt die gleiche Dicke des Gateoxides und ähnliche Dicken der Feld- und Wannenoxyde.

Aufgrund der geringen Dotierung der p- und n-Wannen können auch unterhalb des Feldoxides parasitäre Feldefekttransistoren entstehen. Anstelle einer technologisch problematischen zusätzlichen Erhöhung der Oxiddicke ist in dieser Arbeit der Weg der selektiven Einbringung hochdotierter Isolationsbahnen verfolgt worden (vgl. Anhang 1.3), der insbesondere zur Vermeidung von Feldeffekten im Bereich der nicht-dotierten p⁻-Oberfläche unumgänglich ist. Entscheidend ist auf hochohmigem Substrat ferner die sichere Maskierfähigkeit der Oxide, da schon geringe Konzentrationen durch das Oxid diffundierender Dotierstoffe die niedrigdotierte Oberfläche des p⁻-Substrats beeinträchtigen können. Zwar existieren für wichtige Elemente, wie Bor und Phosphor, Angaben über das Diffusionsvermögen in SiO₂, jedoch beträgt die Bandbreite dieser Werte bis zu drei Größenordnungen [1.9], da das Diffusionsverhalten sehr stark von den jeweiligen Eigenschaften des Oxides abhängt. Im Rahmen einer Studienarbeit [1.10] wurden daher die in der Literatur angegebenen notwendigen Maskierdicken [1.11] für hochohmigem p⁻-Substrat überprüft und bestätigt. Den dort angegebenen Oxiddicken t_{ox,min} liegt das parabolische Diffusionsgesetz

$$\frac{t_{ox,min}^2}{t_d} = D_o \cdot \exp\left(-\frac{E_A}{kT_d}\right) \tag{1.1}$$

zugrunde; hierbei ist t_d die Diffusionszeit bei der Temperatur T_d. Empirische Werte für die Aktivierungsenergie E_A und die Diffusionskonstante D_o sind für Bor mit E_{A,B} = 2.8 eV und D_{o,B} = 4.9*10⁻⁵ cm²/sec sowie für Phosphor mit E_{A,P} = 1.46 eV und D_{o,P} = 1.7*10⁻⁷ cm²/sec angegeben.

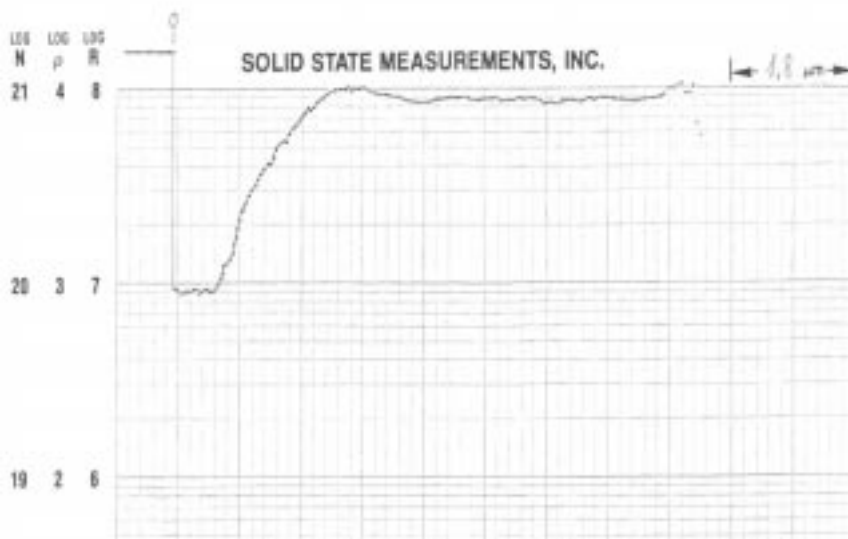


Bild 1.3: Der Spreading-Resistance Meßplot der p⁻-Substratoberfläche nach kompletter CMOS-Prozessierung zeigt den Erhalt des hochohmigen Verhaltens.

Die gute Maskierwirkung verdeutlicht der in Bild 1.3 dargestellte Spreading-Resistance-Meßplot, der die Konservierung der hochohmigen Charakteristik des p⁻-Substrats unterhalb des Feldoxides nach der kompletten CMOS-Prozessierung zeigt. Nur sehr oberflächennah, im Bereich bis ca. 1 µm von der Grenzfläche Si-SiO₂ entfernt, zeigt sich eine geringe Reduktion des spezifischen Widerstands, die wahrscheinlich von Inversionseffekten der Oberfläche, bedingt durch die Präsenz von Oxidladungen, herrührt (vgl. Abs. 2.2.2).

Eine beobachtete erhöhte Bruchempfindlichkeit der mittels des float-zone Verfahrens (FZ) hergestellten p⁻-Wafer im Vergleich zu mittels des Czochralski-Verfahrens (CZ) erzeugten Wafern liegt vermutlich in den fehlenden Oxidausscheidungen der FZ-Wafer begründet. Im Volumen, d.h. unterhalb der sog. denuded zone, befindlicher Sauerstoff in CZ-Wafern sammelt sich während des ersten Temperaturschritts zu Sauerstoff-Ausscheidungen, sog. O₂-clustern [1.12]. Diese ortsfesten Kristalldefekte können bei CZ-Wafern als Fangstellen für sich durch den Wafer ausbreitende Versetzungen wirken. Im Gegensatz dazu können sich Versetzungen in FZ-Wafern aufgrund des Fehlens dieser O₂-Anhäufungen deutlich leichter durch das Material ausbreiten und erhöhen so die Bruchempfindlichkeit.

Eine merkliche Reduktion der Bruchempfindlichkeit bei FZ-Material konnte durch sehr langsame Be- und Entladevorgänge der auf Raumtemperatur bzw. im heißen Ofenrohr befindlichen Scheiben erreicht werden. Eine über den Standardprozeß hinausgehende Ausnutzung von Aufheiz- und Abkühlvorgängen im Reaktionsrohr und damit Ein- und Ausfahrbewegungen des Wafers bei erniedrigter Ofenrohrtemperatur brachten eine weitere Verbesserung.

1.2 CMOS-Prozeß-Varianten für SIMMWIC

Im Mittelpunkt dieser Arbeiten stand die Übertragung des am Institut vorhandenen Alu-Gate-CMOS-Prozesses auf hochohmiges p⁻-Substrat, welches das Basismaterial heutiger Millimeterwellen-Schaltungen auf Silizium (SIMMWIC-Technologie) darstellt. Die Nebenforderung nach unmittelbar vergleichbaren Bauelementeigenschaften der MOS in Bezug auf Steilheiten und Schwellspannungen bedingt u.a. eine Beibehaltung der Gateoxid-dicke und Oberflächenkonzentration auch bei unterschiedlicher Grunddotierung des eingesetzten Substrats. Wenngleich heute eine Vielzahl deutlich fortschrittlicherer CMOS-Prozesse mit Strukturgrößen bis in den Sub-Mikrometer-Bereich in der Literatur [1.14] beschrieben werden, ist der am Institut vorhandene, recht konservative Alu-Gate-Prozeß aus Gründen der Übersichtlichkeit und des mit einer Umstellung verbundenen hohen Aufwands gewählt worden. Ein solches Vorgehen ist insbesondere deshalb vertretbar, da die Untersuchung der grundsätzlichen Integrationsfähigkeit von SIMMWIC- und CMOS-Schaltungen und nicht die Realisierung komplexer Schaltungen im Vordergrund der Arbeit steht. Diese Rahmenbedingungen führten zu den in Abs. 1.2.2 beschriebenen Ansätzen. Ein Integrationsprozeß für SIMMWIC- und CMOS-Schaltungen ist in Abs. 4.3 dargestellt.

1.2.1 Standard-CMOS

Der als interner Standard-CMOS-Prozeß am Institut für Halbleitertechnik eingesetzte Prozeß basiert auf Transistoren mit einer kleinsten nominellen Kanallänge von $L_{\text{nom}} = 10 \mu\text{m}$. Durch Unterdiffusion und Strukturvergrößerungen bei photolithographischen Prozessen und Ätzungen ergibt sich eine effektive Kanallänge von etwa $L_{\text{eff}} = 6 \mu\text{m}$. Entsprechend dieser Strukturgröße ist das thermisch gewachsene Gateoxid mit einer Dicke von $t_{\text{ox}} = 92 \text{ nm}$ und die Tiefe der hochdotierten P⁺- und N⁺-Gebiete mit etwa $x_j = 1.5 \mu\text{m}$ gewählt worden. Der PMOS befindet sich unmittelbar im n-Substrat (CZ-Material, <100> orientiert, Phosphor-dotiert mit $N_D = 1.5 \cdot 10^{15} / \text{cm}^3$). Die durch eine zweifache Bor-Implantation und anschließende Temperaturschritte auf eine Tiefe von $x_j = 5 \mu\text{m}$ eingestellte p-Wanne hat eine Oberflächenkonzentration von $N_A = 2 \cdot 10^{16} / \text{cm}^3$ und umschließt den NMOS. Der Prozeßverlauf [1.15] ist in Bild 1.4 dargestellt, und benötigt 6 aufeinander folgende photolithographische Prozesse. Die in dieser Darstellung nicht aufgeführten Reinigungsschritte umfassen:

(i) Nach jeder Strukturierung mit Photolack und BHF-Ätzung sowie nach der auf die Bor-Implantation folgenden Streuoxidätzung schließt sich eine komplette Reinigungssequenz aus Schwefelsäure- und Standardreinigung an. Diese ist in Bild 1.2 dargestellt.

(ii) Eine Reinigung in heißem Wasserstoffperoxid (H₂O₂ bei T = 80 °C) erfolgt nach der Aluminiumstrukturierung. Eine Standard-Reinigung würde das Aluminium angreifen und kann daher nicht eingesetzt werden.

Zur Überprüfung der Technologieschritte mit Hilfe von Schichtwiderstands- und Oxiddickenmessungen wurden jeweils zwei Testscheiben zusätzlich prozessiert. Die in Bild 1.4 aufgeführten Oxiddicken t_{ox} stellen die Sollwerte nach dem jeweiligen Prozeßschritt dar.

Feldoxidation O ₂ : von T ₁ =700°C auf T ₂ =1000°C rampen H ₂ O: t = 360 min bei T ₂ = 1000 °C t _{ox1} ≈ 1.1 μm N ₂ : t = 5 min bei T ₂ = 1000 °C	N ⁺ -Gebiet- Positivlack SP25 Strukturierung Oxidätzung in BHF bei T = 55 °C
p-Wannen- Positivlack SP25 Strukturierung Oxidätzung in BHF T = 55 °C	N ⁺ -Phosphor- N ₂ : t = 5 min bei T = 855 °C Vorbelegung N ₂ +O ₂ +POCl ₃ : t = 20 min bei T = 855 °C N ₂ : t = 10 min bei T = 855 °C anschließend p-etch bei Raumtemperatur
Streuoxid O ₂ : langsam einfahren bei T = 1000 °C O ₂ : t = 95 min bei T = 1000 °C t _{ox} ≈ 70 nm N ₂ : t = 5 min bei T = 1000 °C	Drive-In O ₂ : langsam einfahren bei T = 1050 °C N ⁺ H ₂ O: t = 15 min bei T ₂ = 1050 °C t _{ox1} ≈ 1.3 μm O ₂ : langsam ausfahren bei T = 1050 °C t _{ox2} ≈ 0.64 μm t _{ox3} ≈ 0.35 μm t _{ox4} ≈ 0.28 μm
p-Wannen- Bor: N _{□1} =2.0*10 ¹³ /cm ² bei E ₁ =120 keV Implantation und N _{□2} =1.5*10 ¹³ /cm ² bei E ₂ = 40keV	Gatefenster- Positivlack SP25 Strukturierung Oxidätzung in BHF bei T = 55 °C
Streuoxid- Oxidätzung in BHF bei T = 25 °C Ätzung	Gateoxidation Chlorgetterung des Oxidationsrohres O ₂ : von T ₁ =700°C auf T ₂ =1000°C rampen O ₂ : t = 130 min bei T ₂ = 1000 °C t _{ox, Gate} ≈ 92 nm N ₂ : t = 5 min bei T ₂ = 1000 °C
Drive-In N ₂ : von T ₁ =700°C auf T ₂ =1000°C rampen p-Wanne H ₂ O: t = 90 min bei T ₂ = 1000 °C N ₂ : auf T ₃ = 1200 °C rampen t _{ox1} ≈ 1.2 μm N ₂ : t = 240 min bei T ₃ = 1200 °C t _{ox2} ≈ 0.45 μm N ₂ : abkühlen lassen bis auf T ₁ = 700 °C	Kontaktloch- Positivlack SP25 Strukturierung Oxidätzung in BHF bei T = 55 °C
P ⁺ -Gebiet- Positivlack SP25 Strukturierung Oxidätzung in BHF bei T = 55 °C	Aluminiumbedampfung t _{Al} ≈ 1.0 μm
P ⁺ -Bor- N ₂ : t = 5 min bei T = 1000 °C Vorbelegung N ₂ +O ₂ +B ₂ H ₆ : t = 20 min bei T = 1000 °C N ₂ : t = 5 min bei T = 1000 °C	Aluminium- Positivlack SP25 Strukturierung Aluätzung in H ₃ PO ₄ bei T = 55 °C
Borglas- H ₂ O: t = 75 min bei T = 650 °C oxidation anschließend p-etch bei Raumtemperatur	Alutemperung N ₂ : t = 15 min bei T = 480 °C
Drive-In N ₂ : von T ₁ =700°C auf T ₂ =1000°C rampen P ⁺ H ₂ O: t = 40 min bei T ₂ = 1000 °C t _{ox1} ≈ 1.3 μm N ₂ : abkühlen lassen bis auf T ₁ = 700 °C t _{ox2} ≈ 0.55 μm t _{ox3} ≈ 0.30 μm	

Bild 1.4: Prozeßabfolge des Alugate-CMOS-Prozesses auf Standard n-Substrat (<100> orientiertes CZ-Material mit N_D = 1.5*10¹⁵ /cm³). Die Größen t_{ox1}, t_{ox2}, t_{ox3}, t_{ox4} und t_{ox, Gate} bedeuten die jeweiligen Oxiddicken des Feldoxides, des Oxides über der p-Wanne, über den P⁺- und N⁺-Gebieten sowie die Dicke des Gateoxides.

1.2.2 CMOS auf hochohmigem Silizium

Aufbauend auf dem CMOS-Prozeß auf n-Substrat, in der Folge auch als Standard-Prozeß bezeichnet, sind die vier Prozeßvarianten A, B, C und C₂ (vgl. Bild 1.5) auf hochohmigem p-Silizium (<100> orientiertes FZ-Material, Bor-dotiert mit N_A = 1.5*10¹² /cm³) entworfen worden.

Der Typ A stellt einen Einzel-Wannen-Prozeß auf p-Substrat dar und bildet das direkte Gegenstück zum Institut-sprozeß auf n-Substrat. Anstelle der p-Wanne im Prozeß auf n-Substrat kommt hierbei eine den PMOS umschließende n-Wanne zum Einsatz. Bei dem Entwurf dieser Variante stand von Anfang an nicht die Überlegung

im Vordergrund, eine optimale CMOS-Technik auf p^- -Substrat zu entwerfen, sondern das elektrische Verhalten eines direkt in das hochohmige p^- -Substrat integrierten NMOS-Transistors untersuchen zu können. Die im Vergleich zu einem konventionellen NMOS, der sich in einer ausreichend stark dotierten p-Wanne oder auf entsprechendem p-Substrat befindet, unterschiedliche elektrische Charakteristik wird in Abschnitt 2.1.1 behandelt.

Als mögliche Lösungsansätze sind die beiden Varianten B und C entworfen worden. Durch die Schaffung einer großflächigen tiefen n-Wanne, die den CMOS-Schaltungsteil komplett umschließt, entsteht die Variante B. In diese n-Wanne wiederum wird ähnlich dem Prozeßverlauf des Standard-Prozesses eine p-Wannen-Technik eingebracht. Die n-Wanne ist ohne eine zusätzliche Wanne ganzflächig eingebracht worden. Als Hauptnachteil einer solchen Lösung verbleibt die gegenseitige Abhängigkeit der Dotierniveaus der verschachtelten Wannen, da die später eingebrachte p-Wanne um mindestens eine Größenordnung höher dotiert sein muß als die n-Wanne, was für letztere ein recht geringes Dotierniveau festlegt.

Die Variante C entspricht einer Doppel-Wannen-Technik mit separaten p- und n-Wannen, so daß eine getrennte Einstellung der Oberflächenkonzentrationen beider Wannen möglich ist.

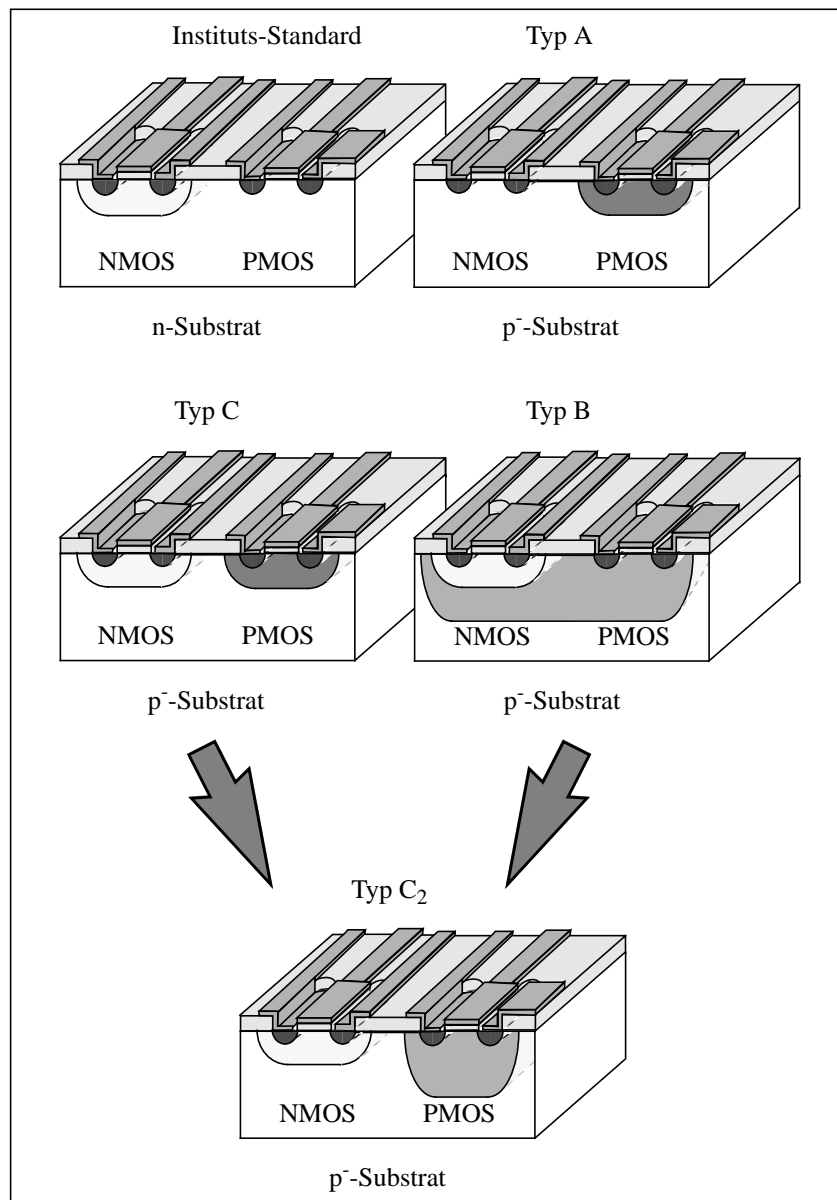


Bild 1.5: Prinzipbild der vier Varianten des CMOS-Prozesses auf hochohmigem p^- -Substrat im Vergleich zum Instituts-Prozeß auf n -Substrat. Variante A (n-Wannen-Prozeß), B (gemeinsame tiefe n-Wanne), C (Doppel-Wannen-Technik) und C₂ (Doppel-Wannen-Technik mit tiefer n-Wanne).

Aufgrund der Vorarbeiten mittels der Varianten B und C ist die Variante C₂ entworfen worden. Diese stellt eine Modifikation der Variante C dar, in der durch eine größere n-Wannentiefe, die jener in der Variante B entspricht, ein besseres Unterschwellen-Verhalten der PMOS-Fets erreicht wird. Diese Abfolge bietet die NMOS-Fets der Variante C und die PMOS-Fets der Variante B und stellt die Basis für den SIMMWIC-CMOS-Integrationsprozeß (vgl. Abs. 4.3) dar.

Durch die geeignete Wahl der Implantationsparameter wurde eine dem Standard-Prozeß weitgehend entsprechende Oberflächenkonzentration der p- und n-Wannen bei allen Prozeßvarianten erreicht. Die Tiefen der metallurgischen Übergänge und die Oberflächenkonzentration im Kanalbereich sind in Tabelle 1 aufgelistet.

TABELLE 1. Wannenparameter

	Standard-CMOS	Variante A	Variante B	Variante C	Variante C ₂
p-Wannen-Tiefe [µm]	4.4		6.1	* ₂)	* ₂)
Oberflächenkonzentration der p-Wanne [1/cm ³]	1.9*10 ¹⁶	1.5*10 ¹² * ₁)	2.3*10 ¹⁶	1.7*10 ¹⁶	1.7*10 ¹⁶
n-Wannen-Tiefe [µm]		4.7	12.3	4.7	12.3
Oberflächenkonzentration der n-Wanne [1/cm ³]	3.7*10 ¹⁵ * ₁)	2.5*10 ¹⁵	1.8*10 ¹⁵	2.5*10 ¹⁵	1.8*10 ¹⁵

*₁) Bei der Variante A und beim Standard-Prozeß sind, da es sich hierbei um Einzel-Wannen-Prozesse handelt, auch die Substratdotierungen an der Oberfläche angegeben.

*₂) Da es sich um gleichnamige Wannen in Bezug auf das Substrat handelt, existiert in diesen Fällen kein pn-Übergang.

Um die verschiedenen Varianten mit begrenztem Aufwand prozessieren zu können, wurde der in Bild 1.6 dargestellte Kombinationsprozeß entworfen, der ab der Strukturierung der hochdotierten P⁺-Gebiete dem Standard-Prozeß entspricht. Mit zwei zusätzlichen Phosphorimplantationen, die jedoch jeweils nur ein einziges Mal pro Variante erforderlich sind, und den beiden n-Masken für die Varianten A und C sowie für B und C₂ sind jeweils drei Varianten in einen Prozeß integriert worden. In späteren Durchläufen ersetzte die Variante C₂ die Variante A, die aus in Abs. 2.1.1 genannten Gründen nicht für eine SIMMWIC-CMOS-Integration in Frage kommt.

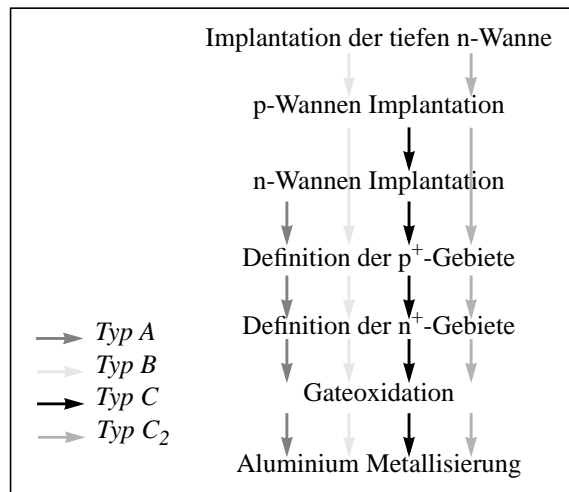


Bild 1.6: Prozeßfluß des kombinierten CMOS-Prozesses auf hochohmigem p⁻-Substrat, der die vier Varianten A (n-Wannen-Prozeß), B (gemeinsame tiefe n-Wanne), C (Doppel-Wannen-Technik) und C₂ (Doppel-Wannen-Technik mit tiefer n-Wanne) zuläßt.

Die Reinigungssequenz entspricht der im Abs. 1.1.1 erläuterten. Der Prozeßablauf der Varianten B und C₂ ist in Bild 1.7 dargestellt. Die Varianten B und C₂ unterscheiden sich nur in der Art der n-Wannen-Maske, die für die Variante B den gesamten Bereich der CMOS-Schaltung öffnet und in der Variante C₂ nur lokale Öffnungen um die PMOS-FETs schafft.

Feldoxidation $t_{ox1} \approx 0.7 \mu\text{m}$	O_2 : von $T_1=700^\circ\text{C}$ auf $T_2=1000^\circ\text{C}$ rampen H_2O : $t = 180 \text{ min}$ bei $T_2 = 1000^\circ\text{C}$ N_2 : $t = 5 \text{ min}$ bei $T_2 = 1000^\circ\text{C}$ N_2 : abkühlen lassen bis auf $T_1 = 700^\circ\text{C}$	P^+ -Bor-Vorbelegung N_2 : $t = 5 \text{ min}$ bei $T = 1000^\circ\text{C}$ $\text{N}_2+\text{O}_2+\text{B}_2\text{H}_6$: $t = 20 \text{ min}$ bei $T = 1000^\circ\text{C}$ N_2 : $t = 5 \text{ min}$ bei $T = 1000^\circ\text{C}$
n-Wannen-Strukturierung	Positivlack SP25 Oxidätzung in BHF $T = 55^\circ\text{C}$	Borglas-oxidation H_2O : $t = 75 \text{ min}$ bei $T = 650^\circ\text{C}$ anschließend p-etch bei Raumtemperatur
Streuoxid $t_{ox} \approx 60 \text{ nm}$	N_2 : von $T_1=700^\circ\text{C}$ auf $T_2=1000^\circ\text{C}$ rampen O_2 : $t = 65 \text{ min}$ bei $T_2 = 1000^\circ\text{C}$ N_2 : abkühlen lassen bis auf $T_1 = 700^\circ\text{C}$	Drive-In P^+ $t_{ox0} \approx 1.0 \mu\text{m}$ $t_{ox1} \approx 1.3 \mu\text{m}$ $t_{ox2} \approx 0.55 \mu\text{m}$ $t_{ox3} \approx 0.30 \mu\text{m}$
n-Wannen-Implantation	Phosphor: $N_D = 6.7 \cdot 10^{11} / \text{cm}^2$ bei $E = 100 \text{ keV}$	N^+ -Gebiet-Strukturierung Positivlack SP25 Oxidätzung in BHF bei $T = 55^\circ\text{C}$
Streuoxid-Ätzung	Oxidätzung in BHF bei $T = 25^\circ\text{C}$	N^+ -Phosphor-Vorbelegung N_2 : $t = 5 \text{ min}$ bei $T = 855^\circ\text{C}$ $\text{N}_2+\text{O}_2+\text{POCl}_3$: $t = 20 \text{ min}$ bei $T = 855^\circ\text{C}$ N_2 : $t = 10 \text{ min}$ bei $T = 855^\circ\text{C}$ anschließend p-etch bei Raumtemperatur
Drive-In n-Wanne $t_{ox0} \approx 0.7 \mu\text{m}$ $t_{ox1} \approx 1.2 \mu\text{m}$	O_2 : von $T_1=700^\circ\text{C}$ auf $T_2=1000^\circ\text{C}$ rampen H_2O : $t = 180 \text{ min}$ bei $T_2 = 1000^\circ\text{C}$ N_2 : auf $T_3 = 1200^\circ\text{C}$ rampen N_2 : $t = 480 \text{ min}$ bei $T_3 = 1200^\circ\text{C}$ N_2 : abkühlen lassen bis auf $T_1 = 700^\circ\text{C}$	Drive-In N^+ $t_{ox0} \approx 1.05 \mu\text{m}$ $t_{ox1} \approx 1.3 \mu\text{m}$ $t_{ox2} \approx 0.65 \mu\text{m}$ $t_{ox3} \approx 0.35 \mu\text{m}$ $t_{ox1} \approx 0.28 \mu\text{m}$
p-Wannen-Strukturierung	Positivlack SP25 Oxidätzung in BHF $T = 55^\circ\text{C}$	Gatefenster-Strukturierung Positivlack SP25 Oxidätzung in BHF bei $T = 55^\circ\text{C}$
Streuoxid $t_{ox} \approx 60 \text{ nm}$	N_2 : von $T_1=700^\circ\text{C}$ auf $T_2=1000^\circ\text{C}$ rampen O_2 : $t = 65 \text{ min}$ bei $T_2 = 1000^\circ\text{C}$ N_2 : abkühlen lassen bis auf $T_1 = 700^\circ\text{C}$	Gateoxidation Chlorgetterung des Oxidationsrohres O_2 : von $T_1=700^\circ\text{C}$ auf $T_2=1000^\circ\text{C}$ rampen O_2 : $t = 130 \text{ min}$ bei $T_2 = 1000^\circ\text{C}$ $t_{ox, Gate} \approx 93 \text{ nm}$ N_2 : $t=30 \text{ min}$, dabei sofort abkühlen lassen
p-Wannen-Implantation	Bor: $N_D=1.5 \cdot 10^{13} / \text{cm}^2$ bei $E_1=120 \text{ keV}$ und $N_D=1.5 \cdot 10^{13} / \text{cm}^2$ bei $E_2=40 \text{ keV}$	Kontaktloch-Strukturierung Positivlack SP25 Oxidätzung in BHF bei $T = 55^\circ\text{C}$
Streuoxid-Ätzung	Oxidätzung in BHF bei $T = 25^\circ\text{C}$	Aluminiumbedampfung $t_{A1} \approx 1.0 \mu\text{m}$
Drive-In p-Wanne $t_{ox0} \approx 0.9 \mu\text{m}$ $t_{ox1} \approx 1.2 \mu\text{m}$ $t_{ox2} \approx 0.45 \mu\text{m}$	O_2 : von $T_1=700^\circ\text{C}$ auf $T_2=1000^\circ\text{C}$ rampen H_2O : $t = 90 \text{ min}$ bei $T_2 = 1000^\circ\text{C}$ N_2 : auf $T_3 = 1200^\circ\text{C}$ rampen N_2 : $t = 240 \text{ min}$ bei $T_3 = 1200^\circ\text{C}$ N_2 : abkühlen lassen bis auf $T_1 = 700^\circ\text{C}$	Aluminium-Strukturierung Positivlack SP25 Aluätzung in H_3PO_4 bei $T = 55^\circ\text{C}$
P^+ -Gebiet-Strukturierung	Positivlack SP25 Oxidätzung in BHF bei $T = 55^\circ\text{C}$	Alutemperung N_2 : $t = 15 \text{ min}$ bei $T = 480^\circ\text{C}$

Bild 1.7: Prozeßabfolge der Varianten B und C₂ des Alugate-CMOS-Prozesses auf hochohmigem p⁻-Substrat (<100> orientiertes FZ-Material mit $N_A = 1.5 \cdot 10^{12} / \text{cm}^3$). Die Größen t_{ox0} , t_{ox1} , t_{ox2} , t_{ox3} , t_{ox4} und $t_{ox, Gate}$ bedeuten die jeweiligen Dicken des Oxides über der n-Wanne, des Feldoxides, des Oxides über der p-Wanne, über den P⁺- und N⁺-Gebieten sowie die Dicke des Gateoxides. Allen feuchten Oxidationen (H₂O-Atmosphäre) ist eine fünfminütige trockene Oxidation (O₂) bei Prozeßtemperatur vorgeschaltet.

1.3 Prozeßmeßtechnik unter besonderer Berücksichtigung der niedrigen Dotierung

Das grundsätzliche Verhalten der Diffusionsvorgänge kann recht genau anhand der Fick'schen Diffusionsgesetze beschrieben werden. Es führt für den Grenzfall einer unerschöpflichen Dotierquelle auf einen resultierenden Profilverlauf gemäß einer komplementären Fehlerfunktion und bei einer konstanten Gesamtdotierstoffmenge zu einem gaußförmigen Profil. Problematisch ist jedoch die Berücksichtigung von Redistributionseffekten aufgrund der unterschiedlichen Löslichkeit der Dotierstoffe in Silizium und dessen thermisch gebildetem Oxid. Diese als pile-up bzw. pile-down bekannten Effekte führen gegenüber dem berechneten Verlauf zu einer Erhöhung bzw. Erniedrigung der Konzentration in der Nähe der Siliziumoberfläche. Durch Oxidation erhöhtes Diffusionsvermögen, sog. transient enhanced diffusion (TED), aufgrund der Generation von zusätzlichen Leerstellen und Zwischengitteratomen während einer Oxidation, führt ebenfalls zu Abweichungen von obigen Verteilungen, die analytisch sehr schwer zu bestimmen sind. Gängige Praxis ist daher heute der Einsatz von Prozeßsimulatoren, wie etwa SUPREM, die numerisch unter Berücksichtigung der verschiedenen Effekte Dotierverläufe mit bestechender Genauigkeit bestimmen können. Auch in diesem Fall verbleibt jedoch die Notwendigkeit eines sauberen Abgleichs zwischen den Eingabeparametern und meßtechnisch ermittelten Größen. Im folgenden Abschnitt stehen daher die im Rahmen dieser Arbeit eingesetzten Verfahren zur Bestimmung der Dotierungsverläufe im Vordergrund.

1.3.1 Vierpunkt-Messung und Schrägschliff

Das grundlegende Verfahren zur Bestimmung des spezifischen Widerstands ρ von Halbleiterkristallen ist die sog. Vierpunkt-Messung. Eine Spannungsmessung nach dem Kelvin-Verfahren mit separat eingepprägtem Konstantstrom erlaubt die Ermittlung des Schichtwiderstands R_s . Bild 1.8 zeigt die Anordnung mit vier in Reihe befindlichen Sonden, die jeweils $s = 1 \text{ mm}$ voneinander entfernt sind. Der mittels der beiden äußeren Kontaktsonden eingepprägte Meßstrom sorgt für einen Spannungsabfall aufgrund des Widerstands des Halbleitermaterials und der Übergangswiderstände von der Metallsonde in den Halbleiter. Mit Hilfe der nahezu stromlosen Spannungsmessung über die beiden inneren Kontakte P_1 und P_2 gelingt es, den Einfluß der Kontaktwiderstände auf das Meßergebnis weitgehend zu unterdrücken. Durch die Umpolung des Meßstroms I und der arithmetischen Mittelwertbildung der bei beiden Polaritäten vorhandenen Schichtwiderstandswerte R_{s+} und R_{s-} ist auch eine Verfälschung des Ergebnisses durch Thermospannungen zu vermeiden [1.16].

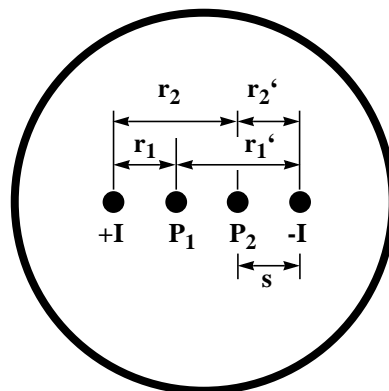


Bild 1.8: Darstellung des Meßkopfes für die Vierpunkt-Messung. Der Konstantstrom I wird an den beiden äußeren Kontakten eingepprägert und sorgt für einen Spannungsabfall aufgrund der endlichen Leitfähigkeit des Halbleiters und der Kontaktwiderstände zwischen Metallsonde und Halbleiter.

Der Wert des Schichtwiderstands ergibt sich aus den ermittelten Potentialunterschieden zwischen den Kontakten P_1 und P_2 gemäß folgender Überlegung [1.17]. Ein im Abstand r von einer punktförmigen Stromquelle entfernter Punkt P , der sich auf einer dünnen, homogenen und lateral unendlich ausgedehnten Schicht mit einem Schicht-

widerstand R_s befindet, weist das Potential

$$\phi_p = \frac{I \cdot R_s}{2\pi} \cdot \ln(r) \quad (1.2)$$

auf. Aus Bild 1.8 ergeben sich die für die Punkte P_1 und P_2 relevanten Abstände r zu

$$\begin{aligned} r_1 &= \frac{r_1'}{2} = s \\ \text{und} \\ r_2 &= 2r_2' = 2s. \end{aligned} \quad (1.3 \text{ a,b})$$

Mit Hilfe der linearen Überlagerung der Potentiale, bedingt durch die beiden Stromquellen $+I$ und $-I$, ergeben sich die Potentiale ϕ_1 und ϕ_2 für die beiden Punkte P_1 und P_2 . Die Spannung zwischen diesen Punkten beträgt

$$U = \phi_1 - \phi_2 = \frac{I \cdot R_s}{\pi} \cdot \ln(2). \quad (1.4)$$

Damit folgt für den gesuchten Schichtwiderstand R_s

$$R_s = \frac{\pi}{\ln(2)} \cdot \frac{U}{I} = 4,53 \cdot \frac{U}{I}. \quad (1.5)$$

Für endliche Verhältnisse zwischen Sondenabstand s und Waferdurchmesser d , d.h. $d/s \ll \infty$, ergeben sich von dem Faktor $\pi/\ln 2$ abweichende kleinere Multiplikatoren, die tabelliert vorliegen und der Literatur zu entnehmen sind [1.13]. Während aufgrund des geringen Sondenabstands diese Abweichungen bei ausreichender Entfernung des Meßkopfes vom Rand des Wafers zu vernachlässigen sind, zeigt sich bei der Bestimmung des Schichtwiderstands von hochohmigem p^- -Substrat ($\rho \approx 5 \text{ k}\Omega\text{cm}$) das Problem der Leitfähigkeitsmodulation des Wafers aufgrund des eingepprägten Meßstroms. Die Routinemessung mit einem am Institut vorhandenen Vierspitzen-Meßplatz mußte daher durch Messungen mit einem Präzisionshalbleitertester ersetzt werden. Diese Anordnung erlaubt mit Meßströmen deutlich unterhalb von $I = 1 \text{ }\mu\text{A}$ den Bereich einer linearen Strom-Spannungs-Kennlinie auch bei hochohmigem Silizium zu erreichen.

Der ermittelte Schichtwiderstandswert kann gemäß $R_s = \rho/w$ zur Bestimmung des spezifischen Widerstands ρ benutzt werden, hierbei ist w die Dicke der betrachteten leitfähigen Schicht und ist im allgemeinen gleich der Sperrschichttiefe x_j . Bei Annahme eines geeigneten Profilverlaufs, also gemäß einer Gauß-Funktion oder einer komplementären Fehlerfunktion, kann eine erste Abschätzung der Oberflächenkonzentration aus dem spezifischen Widerstand unter Vernachlässigung der obengenannten Redistribuitionseffekte mit Hilfe publizierter Kurven [1.18] erfolgen.

Der im Rahmen dieser Arbeit beschrittene Weg erlaubt aufgrund numerischer Integration der simulierten Dotierverläufe eine Bestimmung des Schichtwiderstands auch bei beliebigen Profilen gemäß der Vorschrift

$$R_s = \frac{1}{e \cdot \int_0^{x_j} [\mu_n(N(x)) \cdot N_d(x) + \mu_p(N(x)) \cdot N_a(x)] dx}, \quad (1.6)$$

hierbei sei $N_d(x) = \max(N_D(x) - N_A(x), 0)$ und $N_a(x) = \max(N_A(x) - N_D(x), 0)$. Die von der Gesamtkonzentration $N(x) = N_A(x) + N_D(x)$ abhängige Elektronen- μ_n und Löcherbeweglichkeit μ_p ist durch die beiden Näherungen [1.19, 1.20]

$$\mu_n(N(x)) = \frac{1360 - 92}{1 + \left(\frac{N(x)}{1,3 \cdot 10^{17}}\right)^{0,91}} + 92 \quad \text{in} \quad \frac{cm^2}{Vs} \quad (1.7)$$

und

$$\mu_p(N(x)) = \frac{468 - 49,7}{1 + \left(\frac{N(x)}{1,6 \cdot 10^{17}}\right)^{0,7}} + 49,7 \quad \text{in} \quad \frac{cm^2}{Vs} \quad (1.8)$$

berücksichtigt worden.

Neben dem Vergleich der aus Simulationsdaten berechneten Schichtwiderstände und korrespondierender Meßwerte (vgl. auch Abs. 2.3.1), die wesentlich von den oberflächennahen Beiträgen der hochdotierten Bereiche bestimmt werden, erfordert die Validitätsprüfung der durch Simulation gefundenen Profile auch eine Bestimmung der Tiefe des pn-Übergangs ($w=x_j$).

Zur Vergrößerung der Tiefenauflösung erfolgt ein Schrägschliff eines prozessierten Waferstücks durch den Einsatz eines Schleifklotzes mit einem kleinen Ansträgungswinkel α . Durch den Winkel des Schleifklotzes $\alpha = 1^\circ 9' = 1.15^\circ$ kommt es gemäß $\sin \alpha = 1/50$ zu einer 50-fachen Vergrößerung der vertikalen Ortsauflösung. Die anschließende Einfärbung des n-Gebiets mit einer sog. 'staining'-Lösung, hier einer Lösung aus 20 g $CuSO_4 \cdot 5H_2O$ mit 1 ml HF (48 %ig) in 100 ml H_2O [1.21], geschieht bei starker Beleuchtung und erlaubt das Sichtbarmachen dieses Übergangs. Allerdings reicht die Färbung nur bis zum Rand der Sperrschicht, so daß deren Ausdehnung berücksichtigt werden muß. Mittels einer numerischen Sperrschichtweitenbestimmung (vgl. Abs. 2.3.1) ist diesbezüglich eine Korrektur durchgeführt worden.

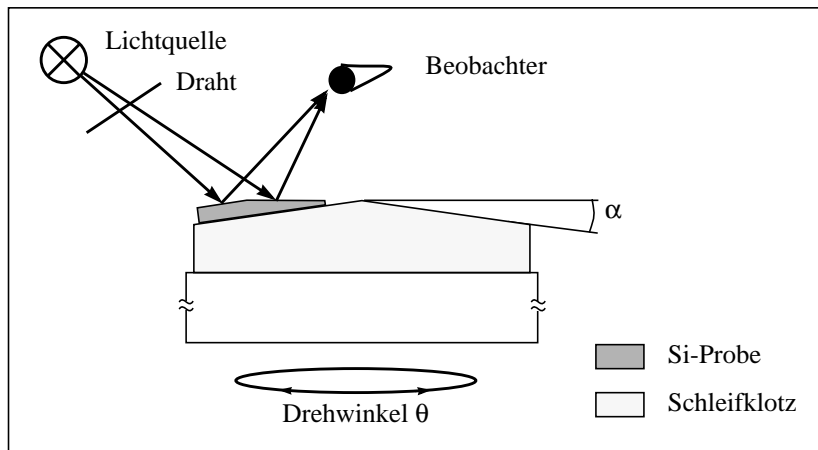


Bild 1.9: Darstellung des Strahlengangs bei der SAM-Messung.

Da sich diese Einfärbung an die im folgenden zu beschreibende Spreading-Resistance-Messung (SR) anschloß, konnte die meßtechnische Bestimmung des real erreichten Schliffwinkels α_{real} mit Hilfe des SAM-Verfahrens erfolgen, das in der SR-Apparatur zur Verfügung gestellt wird. Dieses Small-Angle-Measurement (SAM) genannte Meßprinzip basiert auf unterschiedlich langen Reflexionswegen des Schattens eines unterhalb der Mikroskoplinse befindlichen dünnen geraden Drahtes. Der Wegunterschied von der geschliffenen und der ungeschliffenen Oberfläche des Siliziumstücks wird dabei für die weitere Auswertung genutzt, indem das Siliziumstück auf seiner Vorrichtung unter Beibehaltung des konstanten vertikalen Abstands zum Draht in seiner Ebene rotiert wird (siehe Bild 1.9). Bei den beiden Winkeln θ_1 und θ_2 erfolgt eine parallele Berührung der von beiden

Oberflächen reflektierten Schatten. Mit Hilfe dieser Winkel kann gemäß [1.22]

$$\sin(\alpha) = \frac{s}{2d} \cdot \frac{1}{\sin\left(\frac{\theta_1 - \theta_2}{2}\right)} \quad (1.9)$$

der Schliffwinkel α bestimmt werden. Hierbei beschreibt s den Drahtdurchmesser und d den vertikalen Abstand zwischen der Schliffebene und dem Draht. Für die beiden möglichen Drahtdurchmesser mit den Stärken $s_1 = 5$ mil und $s_2 = 10$ mil gelten in der Fokusebene die Werte $s_1/2d_1 = 0.0048$ und $s_2/2d_2 = 0.0024$. Die üblicherweise erreichbare Genauigkeit des Verfahrens wird mit $\pm 2\%$ angegeben und daher erlaubt die Methode eine Tiefenbestimmung in vergleichbarer Präzision wie die Interferenzmethode [1.23] bei deutlich geringerem Aufwand. Weniger die eigentliche Sperrschichttiefenbestimmung durch Einfärbung als die vorangehende Spreading-Resistance-(SR)-Messung erfordert eine nahezu riefenfreie Oberfläche des Schliffbereichs [1.24]. Nach umfangreichen Untersuchungen konnten die in folgender Tabelle aufgelisteten Schleifparameter ermittelt werden, die zu nahezu rauschfreien SR-Messungen führen (vgl. Abs. 1.3.2).

TABELLE 2. Schleifparameter für Si

Schleifmittel	0.25 μm -Diamantpaste in Schleiföl gelöst
Material der Schleifunterlage	mattiertes Glas
Größe des Si-Probenstücks	ca. $3 \times 6 \text{ mm}^2$
Probenherstellung	Ritzen und Brechen des Si-Wafers von der Rückseite her
Fläche des Schliffs	ca. $3 \times 2 \text{ mm}^2$
Drehzahl und Drehrichtung der Schleifplatte	5-10 U/min, der Schliffkante entgegenlaufend
Abstand der Probe vom Scheibenmittelpunkt	ca. 10 cm

1.3.2 Spreading-Resistance Verfahren

Obgleich die Untersuchung des Ausbreitungswiderstands (engl. Spreading Resistance) zu den schon seit geraumer Zeit etablierten Verfahren [1.25] der Prozeßtechnik gehört, ist die Methode aufgrund zahlreicher Weiterentwicklungen insbesondere in Bezug auf die Tiefenauflösung [1.26] und durch Einsatz besserer Probenpräparations-Methoden und Auswertetechniken auch heute sehr aktuell.

Zu den Vorteilen der Spreading-Resistance-Technik gehören

- (i) ein sehr hoher Empfindlichkeitsbereich, d.h. Messungen bei Dotierkonzentrationen $10^{12} \text{ cm}^{-3} \leq N(x) \leq 10^{21} \text{ cm}^{-3}$.
- (ii) die Möglichkeit von Messungen bei Profilen mit mehrfachen pn-Übergängen
- (iii) und beliebig großer Tiefe.
- (iv) Bei Berücksichtigung der oben genannten Verbesserungen [1.26] ist auch die Bestimmung sehr flacher Profile ($t < 100 \text{ nm}$) bei sehr guter räumlicher Auflösung möglich.
- (v) Aufgrund der Bestimmung der Gesamtträgerkonzentration ist bei korrekter Analyse ein Maß der elektrisch aktiven Dotierstoffkonzentration zu erhalten und nicht die chemische Konzentration, die z.B. durch SIMS (vgl. Abs. 1.3.3) ermittelt wird.

Diesen Vorteilen stehen allerdings die Nachteile

- (i) einer steigenden Anforderung an die Proben- und Spitzenpräparationsqualität bei zunehmender geforderter Tiefenauflösung und
- (ii) eine aufwendige Auswertungsprozedur, die zumindest eine grobe Kenntnis der Schichtabfolge erfordert, gegenüber.

Die Messung des Ausbreitungswiderstands $R_m = V_m/I_m$ besteht in der kontrollierten Aufbringung der beiden Meßspitzen auf die geschliffene Halbleiteroberfläche und der Bestimmung des bei niedrigen Spannungen, typ. $V_m \approx 5 \text{ mV}$, fließenden Stroms I_m . Durch die geringe Spannung V_m wird im Bereich um den Spannungsnullpunkt eine lineare Strom-Spannungskennlinie anstelle eines Schottky-Verhaltens erreicht.

Der Kontaktwiderstand R_k zwischen den Meßspitzen und dem zu vermessenen Waferstück mit spezifischem Widerstand ρ liegt meßtechnisch in Reihe zu dem die Dotierkonzentration bestimmenden Parameter ρ . Insbesondere bei stark dotierten Halbleitern ist der Kontaktwiderstand dem spezifischen Widerstand gegenüber dominierend und bewirkt bei unzureichend präparierter Siliziumoberfläche bzw. schlecht konditionierten Spitzen starke Schwankungen der gemessenen Werte. Wenngleich Modelle zur Abschätzung dieses Kontaktwiderstands vorliegen, ist die Übereinstimmung mit praktisch ermittelten Größen oft unzureichend. Die vielfältigen Einflußgrößen auf diesen Widerstand R_k umfassen

- (i) den spezifischen Widerstand des Halbleiters,
- (ii) das Halbleitermaterial (Si, Ge, GaAs, etc.),
- (iii) den Ladungsträgertyp (n oder p),
- (iv) die kristallographische Ausrichtung der Oberfläche,
- (v) die Qualität des Schliffs,
- (vi) die Zeit zwischen Schliff der Probe und Messung sowie
- (vii) den Durchmesser und den Zustand der Spitzen.

Da der Kontakt der Spitzen mit dem Halbleiter nicht homogen - etwa in Form einer Teilkugeloberfläche - ist, sondern aus zahlreichen lokalen Mikrokontakten auf dieser Oberfläche besteht, ist die Modellierung des Kontaktwiderstands R_k mit einem geometrisch definierten Spitzenradius a wenig aussagekräftig. Weder das Modell des halbkreisförmigen Kontakts mit $R_{k,r} = \rho/(2\pi a)$ noch das der flachen Spitze mit $R_{k,f} = \rho/(4a)$ gelten exakt, so daß eine Formulierung mit einem effektiven Radius a_{eff} benutzt wird. Bedingung für eine sichere Messung ist daher neben der erforderlichen Schliffqualität der Probe (vgl. Abs. 1.3.1) auch die Konditionierung der Meßspitzen, um die erforderliche Konzentration an Mikrokontakten zu erreichen. Eine Nebenanforderung an die Spitzenkonditionierung ist eine gewisse Rauigkeit der Oberfläche, um das natürlich gewachsene Oxid (engl. native oxide) an der Oberfläche aufzubrechen und die Kontaktierung des Halbleiters zu ermöglichen. Eine zu große Rauigkeit wiederum bedingt ein lokal zu tiefes Eindringen in die Probe und damit Einschränkungen der Tiefenauflösung. Die vorgeschriebene Prozedur zur korrekten Aufrauung der Spitzen besteht in der Anwendung des sog. Gorey-Schneider-Probe-Grinder (GSPG), einer mit unterschiedlichen Belägen ($0.25 \mu\text{m}$ und $1.0 \mu\text{m}$ Diamantpaste) ausgestatteten Reibfläche. Nachdem die Spitzen zuerst mit dem $0.25 \mu\text{m}$ -Korn geglättet werden, erfolgt anschließend eine teilweise Aufrauung durch den gröberen Belag. Die Güte der Spitzenkonditionierung zeigt sich unmittelbar an der weitgehenden Rauschfreiheit von Messungen an homogen dotierten Silizium-Proben.

Trotz der Spitzenkonditionierung verbleibt eine Unbestimmtheit des gemessenen Schichtwiderstands R_m , der außer vom Zustand der Spitzen auch von den weiteren oben aufgelisteten Punkten abhängt. Um nun die geforderte Aussagefähigkeit des Verfahrens sicherzustellen, bedient man sich einer Vergleichsmethode mit sog. Eichproben, die genormte spezifische Widerstandswerte aufweisen. Bei einer den eigentlichen Meßproben entsprechenden Oberflächenpräparation dieser Eichproben und einem geringen zeitlichen Abstand zwischen den Eichmessungen und der Messung der unbekanntenen Proben, kann der durch die Eichproben gegebene Zusammenhang zwischen spezifischem Widerstand ρ und dem gemessenen Ausbreitungswiderstand (Spreading Resistance) für die Auswertung der unbekanntenen Proben herangezogen werden. Die für die im folgenden vorgestellten Messungen durchgeführte Charakterisierung der Eichproben ist in Bild 1.10 dargestellt.

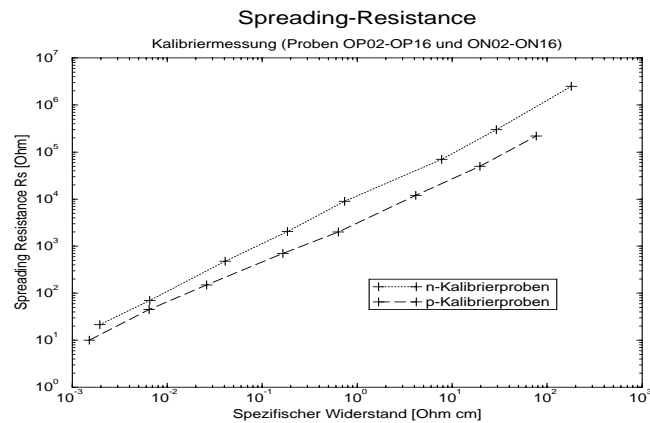


Bild 1.10: Ergebnisse der Spreading-Resistance-Messungen an den SSM-Eichproben. Die Werte des spezifischen Widerstands ρ dieser Probestücke sind sehr genau bekannt und können mit den meßtechnisch ermittelten SR-Werten in Bezug gesetzt werden, um die für die Interpretation erforderliche Eichkurve zu erhalten. Widerstandswerte außerhalb der gegebenen Bereiche können allenfalls extrapoliert werden, für die Werte zwischen den einzelnen Eichproben (mit Kreuzen markiert) ist eine logarithmische Interpolation benutzt worden.

Die bei allen Messungen benutzten geometrischen Einstellungen sind in folgender Tabelle aufgelistet:

TABELLE 3. Spreading-Resistance Parameter

seitlicher Abstand der Spitzen	$s = 100 \mu\text{m}$
Auflagegewicht der Spitzen	entsprechend je 20 g
horizontaler Vorschub pro Meßschritt	$\Delta x = 2,5 \mu\text{m}$
nomineller Schliffwinkel der Probe	$\sin \alpha = x/h = 1/50$

Anmerkung: Der Schliffwinkel wurde jeweils nach dem SAM-Verfahren (Abs. 1.3.1) gemessen.

Für die Extraktion des Dotierprofils bedarf es einer Beziehung zwischen den ortsabhängigen Werten des gemessenen Ausbreitungswiderstands $R_m(x)$ und den Dotierkonzentrationen $N_A(x)$ und $N_D(x)$. Bei der Annahme, daß sich die Konzentrationen der beweglichen Träger $n(x)$ und $p(x)$ direkt aus der Nettokonzentration $[N_A(x) - N_D(x)]$ ergeben, läßt sich aus der Poisson-Gleichung mittels des sog. Schumann-Gardner-Integrals [1.27] der Ausbreitungswiderstand R_m einer Halbleiterschicht

$$R_m = \frac{2\rho_o}{\pi a} \cdot \int_0^\infty A_n(\lambda) \left\{ \frac{J_1(a\lambda)}{a\lambda} - \frac{1}{2} \cdot J_0\left(\frac{2s}{a}\lambda\right) \right\} \sin(a\lambda) \frac{d\lambda}{\lambda} \quad (1.10)$$

berechnen. In diesem Integral bedeuten $J_0(\lambda a)$ und $J_1(\lambda a)$ die Besselfunktionen 0. und 1. Ordnung, ρ_o ist der spezifische Widerstand der obersten Schicht, s der Abstand der Meßspitzen und $a = a_{\text{eff}}$ der effektive Spitzenradius. Die von Choo [1.28] eingeführte Funktion A_n berücksichtigt den ortsabhängigen Verlauf des spezifischen Widerstands $\rho(x)$ der Probe und ist rekursiv zu lösen. Bei diesem rekursiven Verfahren wird die Probe in dünne homogene Schichten aufgeteilt, so daß eine stufenförmige Approximation des prinzipiell stetigen Widerstandsverlaufs $\rho(x)$ erfolgen kann. Die Auflösung dieser Vorschrift nach dem gesuchten Profil bei Vorgabe des Widerstands stellt die eigentliche Schwierigkeit dar.

Ferner bestehen Schwächen dieses Ansatzes in der Vernachlässigung der freien Träger in der Raumladungszone und der kompletten Nichtberücksichtigung des sog. 'carrier spillings' [1.29] auf den Oberflächen des Schrägschliffs, das zu einer Abweichung zwischen dem Konzentrationsprofil der freien Träger und dem Verlauf der Dotierung führen kann. Das Auftreten einer solchen Abweichung zeigt sich jedoch z.B. im Bereich des n-Wannen-p-Substrat-Übergangs der Variante B (vgl. Bild 1.14).

Die Lösung der Poissongleichung

$$\nabla^2 \phi = -\frac{e}{\epsilon_0 \epsilon_{Si}} \cdot [N_D - N_A + p - n] \tag{1.11}$$

unter Berücksichtigung der Trägerdichten in Abhängigkeit des Potentials ϕ

$$n(x) = n_i \cdot \exp\left(\frac{e\phi(x)}{kT}\right)$$

und

$$p(x) = n_i \cdot \exp\left(-\frac{e\phi(x)}{kT}\right), \tag{1.12 a,b}$$

beginnend vom gegebenen Ausbreitungswiderstandsverlauf $R_m(x)$, erfordert einen noch höheren mathematischen Aufwand und liefert im Ergebnis allenfalls die Nettodichte $|N_A(x) - N_D(x)|$ der Dotierung. Der umgekehrte Weg startet bei der Vorgabe eines möglichen Dotierprofils bestehend aus den einzelnen Komponenten $N_{A,1}(x), N_{A,2}(x), \dots, N_{A,i}(x), N_{D,1}(x), N_{D,2}(x), \dots, N_{D,j}(x)$ und endet in der Bestimmung des zu diesem Profil gehörenden Widerstandsprofils $R_{sr}(x)$, das mit den gemessenen Werten $R_m(x)$ durch Variation des Eingabeprofils abzugleichen ist. Bild 1.11 zeigt das Ablaufdiagramm dieser Berechnungsfolge, wie sie mit der SRP2-Software von SSM [1.30] durchgeführt wird.

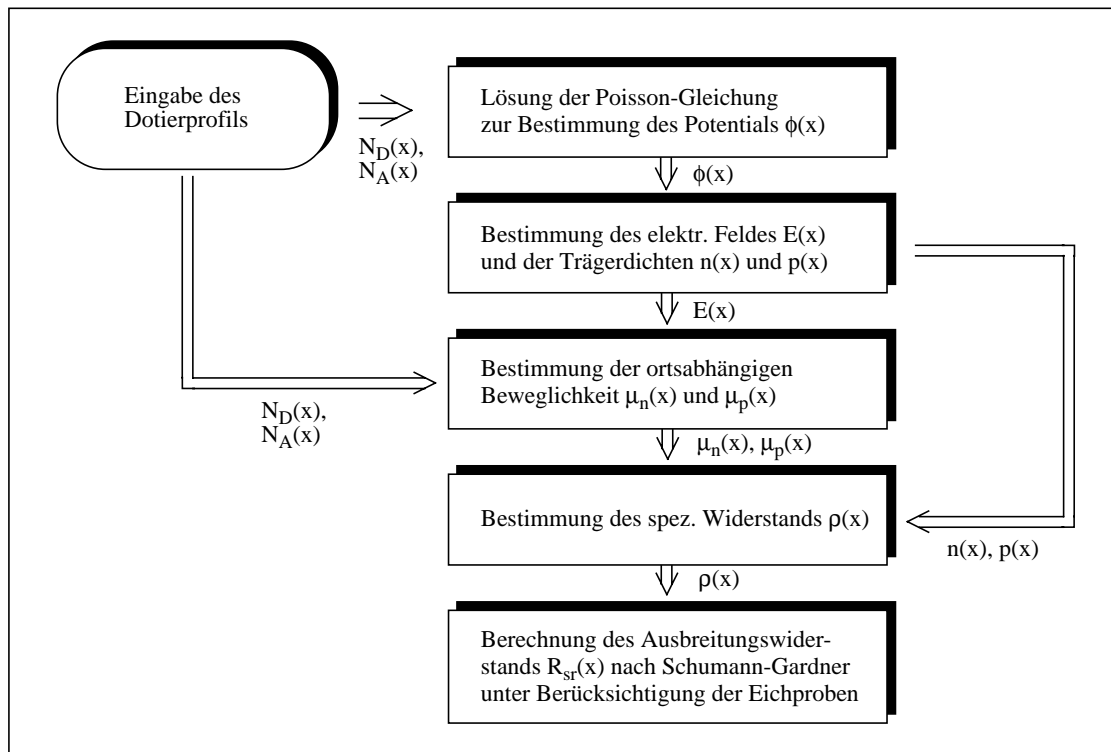


Bild 1.11: Schematische Darstellung der Berechnung des Ausbreitungswiderstands-Profiles (engl. Spreading-Resistance) mit der SRP2-Software von SSM.

Das angenommene Dotierprofil für den NMOS der CMOS-Variante C (vgl. Abs. 1.2.2) ist in Bild 1.12 zusammen mit der daraus berechneten Trägerverteilung dargestellt. Bild 1.13 zeigt den Vergleich des aus diesem Profil bestimmten Ausbreitungswiderstandsverlaufs $R_{sr}(x)$ und den meßtechnisch ermittelten Werten $R_m(x)$. Die gute Übereinstimmung der gemessenen und der berechneten Kurve ist ein Beleg für die richtige Annahme des vorgegeben Dotierprofilverlaufs, der sich wiederum mit den Ergebnissen der Prozeßsimulation (vgl. Abs. 2.3.1) weitgehend deckt.

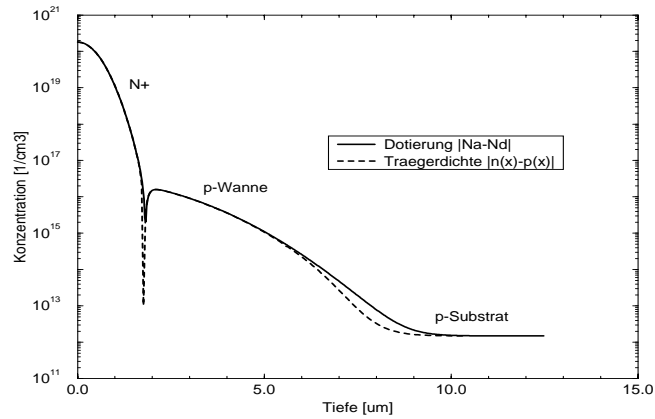


Bild 1.12: Vergleich der für die Berechnung des Ausbreitungswiderstands (SR) benutzten Nettodotierung (durchgezogene Kurve) mit der von der SRP2-Software berechneten Trägerdichte (gestrichelte Kurve) auf der Schlißfläche für den Schnitt durch die Source des NMOS der Variante C. Für die Berechnung des Dotierprofils ist sowohl für die Source- als auch für die Wannendotierung von Gaußverteilungen ausgegangen worden.

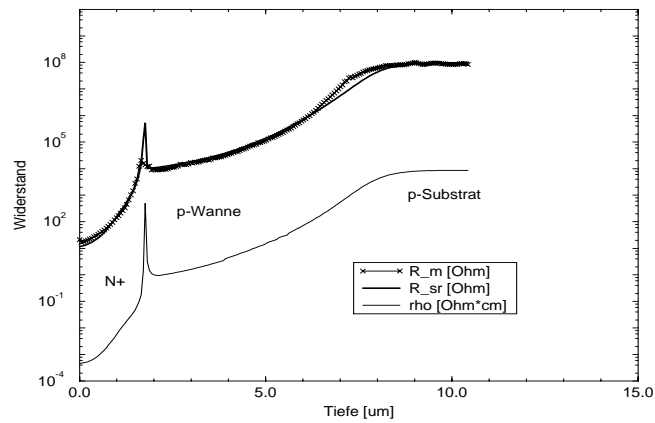


Bild 1.13: Die Darstellung zeigt die gute Übereinstimmung zwischen meßtechnisch ermittelten Werten des Ausbreitungswiderstands des NMOS der Variante C (mit R_m bezeichnete Kurve) und den aus dem Eingabe-Dotierprofil (vgl. Bild 1.12) berechneten Daten (mit R_sr bezeichnete Kurve). Die zur Berechnung des Widerstands R_sr benutzten spezifischen Widerstände ρ (rho) sind ebenfalls dargestellt.

Für den PMOS der Variante B ist das vorgegebene Dotierprofil und die daraus abgeleitete Trägerverteilung in Bild 1.14 dargestellt. Die gute Übereinstimmung der Ausbreitungswiderstände $R_{sr}(x)$ und $R_m(x)$ ist aus Bild 1.15 ersichtlich.

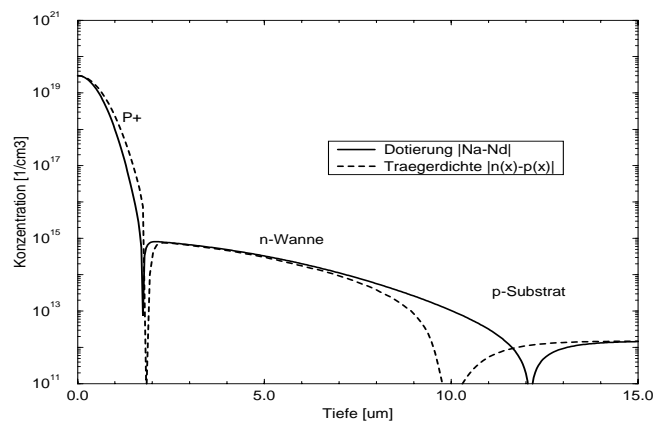


Bild 1.14: Vergleich der für die Berechnung des Ausbreitungswiderstands (SR) benutzten Nettodotierung (durchgezogene Kurve) mit der von der SRP2-Software berechneten Trägerdichte (gestrichelte Kurve) auf der Schlißfläche für den Schnitt durch die Source des PMOS der Variante B. Für die Berechnung des Dotierprofils ist sowohl für die Source- als auch für die Wannendotierung von Gaußverteilungen ausgegangen worden.

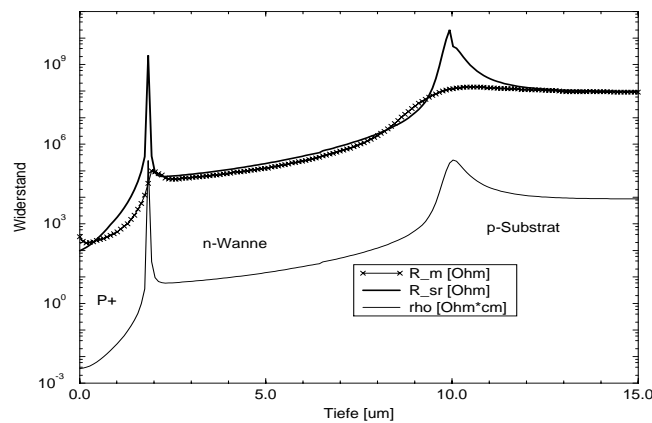


Bild 1.15: Die Darstellung zeigt die gute Übereinstimmung zwischen meßtechnisch ermittelten Werten des Ausbreitungswiderstands des PMOS der Variante B (mit R_m bezeichnete Kurve) und den aus dem Eingabe-Dotierprofil (vgl. Bild 1.14) berechneten Daten (mit R_{sr} bezeichnete Kurve). Die zur Berechnung des Widerstands R_{sr} benutzten spezifischen Widerstände ρ (ρ) sind ebenfalls dargestellt.

1.3.3 SIMS

Ein alternatives Verfahren zur Bestimmung der Dotierung von Halbleitern ist die Sekundärionen-Spektroskopie (SIMS von engl. Secondary-Ion Mass Spectrometry). Hierbei wird ein Primärionen-Strahl auf die zu untersuchende Probe gerichtet, der üblicherweise aus einfach positiv geladenen Sauerstoff- (O_2^+) oder Cäsiumionen (Cs^+) mit kinetischen Energien zwischen 0.5 keV bis 20 keV besteht. Die Energie der Ionen reicht aus, um den Halbleiter mittels Zerstäubung (engl. sputtering) schichtweise von der Oberfläche her abzutragen.

Diese schichtweise Abtragung im Zeitverlauf erlaubt die direkte Konvertierung der Meßdauer in die Tiefenkoordinate des Halbleiters. Da beim Zerstäuben der Oberfläche neben einer überwiegenden Zahl von ungeladenen Teilchen auch ionisierte herausgeschlagen werden, kann man diese geladenen Teilchen in einem Massenspektrometer analysieren und erhält dadurch ein Maß für den Gehalt dieser Teilchenart in der Probe. Um nun eine quantitative Aussage der Teilchenkonzentrationen zu erhalten, werden mit Hilfe von bekannten Eichproben, z.B. Implantationsprofilen mit genau bestimmter Gesamtdosis, Messungen durchgeführt, um dann durch den Vergleich zwischen der Größe der registrierten Ionenzahl der gesuchte Spezies und der bekannten Dosis eine Aussage zu erhalten, welcher Prozentsatz der jeweiligen Teilchenart in einem bestimmten Substrat ionisiert und damit registrierbar ist. Üblicherweise liegt dieser Anteil bei rund einem Prozent. Da dieses Verhältnis sehr genau eingehalten wird, kann man in einem etwa 5-6 Dekaden umfassenden Bereich oberhalb der Detektionsgrenze die Konzentration mit einer maximalen Abweichung von 10 % bestimmen. Durch die Erfassung eines proportionalen Anteils der Gesamtkonzentration ermittelt man durch SIMS nicht die elektrisch aktive Konzentration wie in elektrischen Meßverfahren, beispielsweise die Spreading-Resistance Methode, sondern die chemische Konzentration. Das Verfahren erlaubt ferner die Bestimmung einer Stoffkonzentration in isolierenden Materialien, wie etwa den Dielektrika von Silizium, d.h. Siliziumoxid (SiO_2) oder -nitrid (Si_3N_4). Allerdings muß der statischen Aufladung der Isolatoren durch den Ionenstrahl entgegengewirkt werden, um eine Verfälschung der Ergebnisse zu verhindern. Daß die Präsenz dünner Oxide die Ausbeute positiv-geladener Ionen, wie etwa Bor ($^{11}B^+$) erhöht, ist bei der Auswertung zu beachten [1.13]. Bei der Analyse von Si-Proben wird daher eine oxidfreie Oberfläche vorausgesetzt. Der Anstieg der Borkonzentration an der Oberfläche des Halbleiters, sowohl beim hochdotierten P^+ -Kontakt der Variante B als auch bei der p-Wanne der Variante C (vgl. Bild 1.16), ist deshalb sehr wahrscheinlich einem Restoxid bzw. der natürlichen Oxidschicht zuzuschreiben. Insbesondere, da bei den jeweiligen Oxidationsschritten eine oberflächliche Verarmung des Bors aufgrund des pile-down Effekts zu erwarten ist. Beim Phosphor zeigt sich jedoch, daß die nachfolgenden Drive-In-Schritte in inerter Atmosphäre (N_2) ausreichend lang sind, um das Profil, das in diesem Fall eine Anreicherung direkt nach der Oxidation erfährt, wieder zu verflachen.

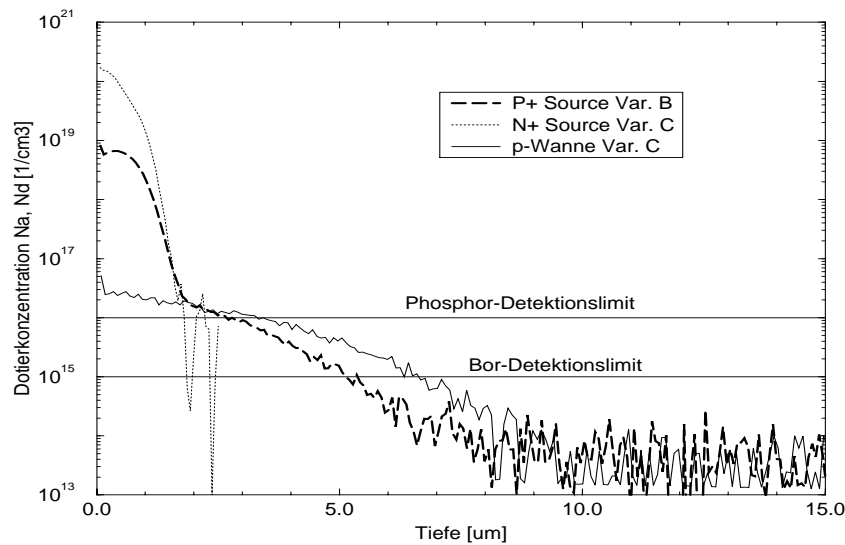


Bild 1.16: SIMS-Profile der P⁺-Source der Variante B, sowie der N⁺-Source und der p-Wanne der Variante C.

Neben einer guten Übereinstimmung der gemessenen Oberflächenkonzentrationen mit den durch Prozeßsimulation ermittelten Werten (vgl. Abs. 2.1.2 und 2.3.1), sind in Bild 1.16 am einsetzenden Meßrauschen auch die aus der Literatur bekannten Detektionsgrenzen für Bor ($N_{\min,B} = 1 \cdot 10^{15} / \text{cm}^3$) und Phosphor ($N_{\min,P} = 1 \cdot 10^{16} / \text{cm}^3$) erkennbar. Da die n-Wanne der Variante B unterhalb dieser Grenze liegt, ist auf die Darstellung der wenig aussagekräftigen Meßergebnisse verzichtet worden.

Die Identifikation der Teilchenart der Sekundärionen, die gute Tiefenauflösung und das für die meisten Anwendungen ausreichend niedrige Detektionslimit erlauben neben einer Messung der Dotierprofile auch eine Analyse unerwünschter Verunreinigungen [1.31].

1.3.4 CV-Profilung

Neben den bisher aufgeführten Verfahren, die allesamt eine zerstörende Messung der Probe beschreiben, hat sich insbesondere zur schnellen Profilmessung unter Produktionsbedingungen die Aufnahme von Kapazitäts-Spannungs-Kurven (CV-Kurven) an speziellen Teststrukturen etabliert. Als Strukturen sind pn- und Schottky-Dioden sowie MIS-Varaktoren, die aus der Schichtfolge Metallelektrode (M), Isolator (I) und Halbleiter (S) bestehen, geeignet. In allen drei Fällen wird die spannungsabhängige Weite $w(V_r)$ einer Raumladungszone für die Bestimmung der Dotierung an der Raumladungsgrenze benutzt.

Am Beispiel der einseitig abrupten p⁺n-Diode läßt sich die Dotierkonzentration $N_D(x)$ auf der niedriger dotierten Seite aus der Diodenkapazität $C(V_r)$ bei Sperrspannung V_r berechnen als

$$N_D(x) = \frac{C(V_r)^3}{e \epsilon_0 \epsilon_{Si} A^2} \cdot \frac{1}{\left| \frac{dC(V_r)}{dV_r} \right|}, \quad (1.13)$$

hierbei ergibt sich die Tiefenkoordinate aus $x = w = \epsilon_0 \epsilon_{Si} A / C(V_r)$. Die Querschnittsfläche der Diode bzw. die Fläche der hochdotierten Elektrode beträgt A. Aus dem totalen Differential der obigen Gleichung gemäß

$$\frac{d(N_D(x))}{N_D(x)} = 2 \cdot \frac{dA}{A} + 3 \cdot \frac{dC(V_r)}{C(V_r)} + \frac{d\left(\frac{dC(V_r)}{dV_r}\right)}{\frac{dC(V_r)}{dV_r}} \quad (1.14)$$

ergibt sich die starke Empfindlichkeit der Größe $N_D(x)$ auf Ungenauigkeiten dA bei der Flächen- und $dC(V_r)$ bei der Kapazitätsbestimmung. Fehler in diesen Größen beeinträchtigen die Genauigkeit des Dotierprofils mit zwei-

bzw. dreifacher Gewichtung. Genaue Flächenbestimmung und rauschfreie Kapazitätsmessungen sind daher für aussagekräftige Resultate unbedingt notwendig.

Zu den Nachteilen des Verfahrens zählen die Begrenzung der möglichen Meßtiefe aufgrund von einsetzenden Durchbrucheffekten, insbesondere bei höheren Dotierkonzentrationen, und die Limitierung der maximalen Tiefe bis zum nächstgelegenen pn-Übergang.

Die Auflösung zur Oberfläche hin ist durch die sich schon ohne äußere Spannung endlich ausdehnende Raumladungszone begrenzt.

Die Schwierigkeiten bei der Anwendung des CV-Verfahrens auf hochohmigem p⁻-Substrat sind in Anhang 3.3 dargelegt.

Bauelemente

In diesem Kapitel stehen die Bauelemente, die mittels der in Kapitel 1 beschriebenen CMOS-Prozesse hergestellt wurden, im Vordergrund der Betrachtung. Grundbaustein der CMOS-Technik ist hierbei der MOS-Feldeffekttransistor, dessen Kennlinien mit entsprechenden Bauelementen des als Referenz gewählten Instituts-CMOS-Prozesses verglichen werden. Durch diesen Vergleich mit dem bekannten Alu-Gate-Prozeß auf Standard-n-Substrat konnte gezeigt werden, daß sich Transistoren mit gleichen Schwellspannungen und Steilheiten bei moderaten Prozeßmodifikationen auch auf hochohmigem p⁻-Substrat fertigen lassen.

Um neben den Strom-Spannungs-Kennlinienfeldern auch eine Analyse des empfindlichen Gateoxides und der Grenzfläche zwischen Oxid und Halbleiter vornehmen zu können, sind Messungen der dielektrischen Integrität und Kapazitäts-Spannungs-(CV)-Meßtechnik anhand spezieller Teststrukturen, den MOS-Varaktoren, eingesetzt worden. Die auch für hochohmiges Substrat erweiterten CV-Messungen sind außerdem für die Vorarbeiten zur monolithischen Integration in Kapitel 4 zum Einsatz gekommen. Ferner wird in diesem Abschnitt auf die Problematik permanenter Oberflächeninversionskanäle aufgrund im Oxid befindlicher Ladungen eingegangen.

In einem nächsten Abschnitt sind zur Erklärung des unterschiedlichen Unterschwellenverhaltens von MOS auf Standard-n-Substrat und auf hochohmigem-p⁻-Substrat die pn-Sperrschichtübergänge der Source- und Drainkontakte analysiert worden. Bei höheren Drainspannungen hat sich ein Anstieg der Unterschwellenströme von MOS gezeigt. Dieser ist auf Leckströme des Drainkontaktes, d.h. erhöhte Sperrströme des Drain-Substrat-Übergangs, zurückzuführen. Eine bedeutsame Differenz zwischen den MOS des Standard-n-Substrats und den hier realisierten Varianten auf hochohmigem Substrat liegt damit im unterschiedlichen Verhalten der Source- und Drain-Sperrschichten begründet. Dieser Abschnitt ist jedoch allgemein den pn-Dioden gewidmet, da für die Aussagen zum Temperaturverhalten und zur Geometrieabhängigkeit profilgleiche Teststrukturen, die als großflächige quasi-vertikale pn-Dioden ausgeführt worden sind, zum Einsatz kamen.

Abschließend wird auf die für den integrierten Hochfrequenzschalter benötigten pin-Dioden eingegangen. Diese Bauelemente sind direkt in die CMOS-Prozesse auf hochohmigem Substrat integriert worden. Da diese Dioden nicht auf Standard-n-Substrat hergestellt werden können, ist in diesem Fall kein Vergleich der Bauelementeigenschaften möglich.

2.1 MOSFETs

Der Feldeffekttransistor mit isolierendem Gate aus thermisch gewachsenem Siliziumoxid ist in der Halbleitertechnik neben dem Bipolartransistor das bevorzugte aktive Dreipolelement mit einer Fülle von Anwendungen als Verstärker und Schaltelement. Im Vergleich zum Bipolartransistor bietet der MOS Vorteile aufgrund seiner lateralen Auslegung, insbesondere hinsichtlich der Integrationsfähigkeit, was neben der Möglichkeit der stromlosen Ansteuerung zum großen Erfolg der CMOS-Technik geführt hat. Nachteile ergeben sich in Bezug auf maximale Leistungsdichten und Schaltgeschwindigkeiten, so daß bei diesen Anwendungen heute oft eine Kombination von Bipolar- und MOS-Transistoren zum Einsatz kommt, die sog. BiCMOS-Technik.

2.1.1 Der NMOS-FET der Variante A auf hochohmigem p⁻-Substrat

Durch den Verzicht auf eine lokale Erhöhung der Substratdotierung in Form einer Wanne, die den Transistor umschließt, und statt dessen durch die direkte Definition der Source- und Draingebiete im p⁻-Substrat ergibt sich der NMOS-FET der Variante A. Der PMOS der Variante A erhält eine n-Wanne wie in der Doppelwannen-Variante C. Variante A verkörpert damit eine n-Wanntentechnik, die auf hochohmiges p-Substrat angewandt ist. Da der PMOS durch die n-Wanne vom hochohmigen Substrat isoliert ist, entspricht sein elektrisches Verhalten jenem der Variante C. Der NMOS hingegen unterscheidet sich sehr stark von vergleichbaren Transistoren auf konventionellen höher dotierten p-Substraten wie an den Kennlinien in Bild 2.1 deutlich wird.

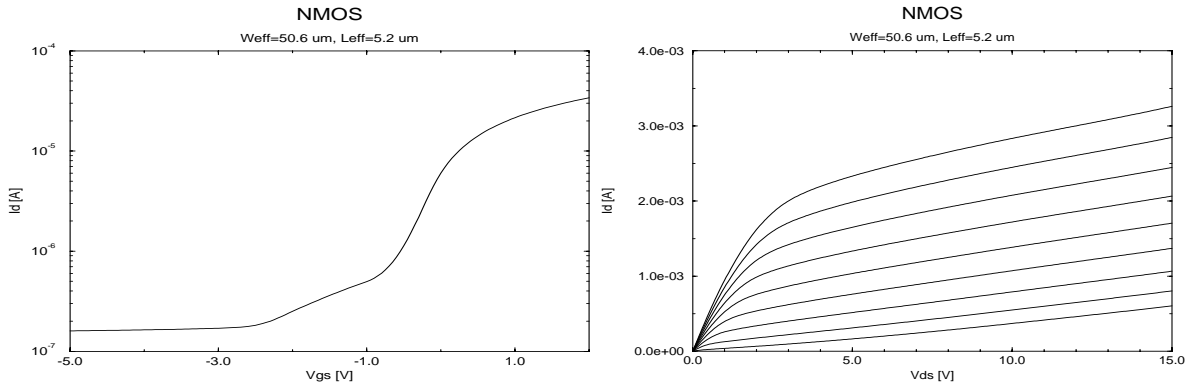


Bild 2.1: Kennlinien eines NMOS ($W_{\text{eff}} = 50.6 \mu\text{m}$, $L_{\text{eff}} = 5.2 \mu\text{m}$) auf hochohmigem p⁻-Substrat. Links ist die Transfer-Kennlinie $I_d - V_g$ im linearen Bereich bei $V_{ds} = 50 \text{ mV}$ und rechts die Ausgangskennlinie $I_d - V_{ds}$ bei verschiedenen Gatespannungen $V_g = 0, 0.5, 1.0, \dots, 4.0 \text{ V}$ dargestellt.

Die wesentlichen Unterschiede zu geometrisch identischen NMOS auf höher dotierten Substraten bzw. in p-Wannen umfassen folgende drei Punkte:

(i) Aufgrund der sehr niedrigen Substratdotierung kommt es sehr früh zu einer Inversion der Halbleiteroberfläche und damit zu einer Verschiebung der Schwellspannung V_{th} in Richtung kleinerer Spannungen. Die obige Transfer-Kennlinie zeigt einen selbstleitenden NMOS mit einer Schwellspannung $V_{th} < 0 \text{ V}$.

Ein Vergleich zwischen hochohmigem FZ-Material mit einer Dotierung $N_A = 1.5 \cdot 10^{12} / \text{cm}^3$ und CZ-Material mit $N_A = 1.0 \cdot 10^{16} / \text{cm}^3$ verdeutlicht die Verschiebung der Schwellspannung gemäß

$$\Delta V_{th} = V_{th}(FZ) - V_{th}(CZ) = -2\Delta\psi_B - \frac{\sqrt{4\epsilon_{Si}eN_A\psi_B}}{C_{ox}} \approx -1,78 \text{ V}, \quad (2.1)$$

hierbei sind $\psi_B = 0.12 \text{ V}$ für $N_A = 1.5 \cdot 10^{12} / \text{cm}^3$, $\psi_B = 0.35 \text{ V}$ für $N_A = 1.0 \cdot 10^{16} / \text{cm}^3$ und damit $\Delta\psi_B = 0.23 \text{ V}$. Die Dotierung von $N_A = 1.0 \cdot 10^{16} / \text{cm}^3$ führt bei einer Gateoxidicke von $t_{ox} = 95 \text{ nm}$ zu einer Schwellspannung $V_{th} \approx 1.3 \text{ V}$, so daß durch die berechnete Verschiebung ein selbstleitender NMOS-FET mit einer Schwellspannung

$V_{th} < 0 \text{ V}$ zu erwarten ist.

(ii) Im Unterschwellenbereich der Transfer-Kennlinie, d.h. für $V_g < V_{th}$, zeigt sich eine deutliche Erhöhung des Drainstroms I_d .

(iii) Die Ausgangskennlinie zeigt einen deutlichen Drainstromanstieg mit steigender Kanalspannung V_{ds} und damit keine klaren Sättigungseffekte. Ein solcher Anstieg wird in der Literatur auch bei einem auf hochohmigem n⁻-Substrat prozessierten PMOS-Transistor beobachtet [2.1]. Der üblicherweise als Sättigungsbereich bezeichnete Kennlinienabschnitt wird folglich als „quasi-linearer“ Bereich beschrieben, bei dem ferner eine deutliche Abweichung des quadratischen Zusammenhangs zwischen Drainstrom I_d und Gatespannung V_g beobachtet wird.

Die in der Variante A vorliegenden NMOS zeigen ebenfalls einen nur leicht überproportionalen Anstieg des Stroms und keine quadratische Funktionsabhängigkeit von der Gatespannung (siehe Bild 2.1).

Die Effekte (ii) und (iii) sind typische Merkmale von Kurzkanalverhalten in MOS-Transistoren. Bemerkenswert ist aber, daß dieses Verhalten hier bei einem NMOS mit einer großen Kanallänge von $L_{\text{eff}} = 5.2 \mu\text{m}$ also an Langkanaltransistoren beobachtet werden kann.

Wenngleich diese Transistoren als Bauelemente nicht oder nur sehr eingeschränkt, d.h. nur mit starker negativer Substratvorspannung, eingesetzt werden können, so bieten sie doch ein hervorragendes Modellsystem zur Untersuchung und Darstellung von Kurzkanaleffekten. Gerade die großen Kanalgebiete erlauben eine exakte Bestimmung ihrer geometrischen Dimensionen und vereinfachen so im Vergleich zu Kurzkanaltransistoren auf höher dotierten Substraten die Analyse beträchtlich.

Grenze zwischen Kurz- und Langkanalverhalten bei MOS

Da das Kurzkanalverhalten eine Abweichung vom nahezu idealen Langkanalverhalten und damit eine Verschlechterung der Transistorfunktion darstellt, sind minimale Kanallängen erforderlich, damit das erwünschte Langkanalverhalten sichergestellt ist. Gemäß Brews et al. [2.2] gilt in Abhängigkeit von der Oxiddicke t_{ox} und der Tiefe der Source- und Drain-Übergänge x_j für diese minimale Länge

$$L_{\text{min}} = 0,4 [x_j t_{\text{ox}} (W_s + W_d)^2]^{1/3} \quad (2.2)$$

Hierbei sind die Oxiddicke in \AA und alle übrigen geometrischen Maße in μm gegeben. Die Größen W_s und W_d stellen die Raumladungsweiten von Source und Drain unter Annahme einseitig abrupter Übergänge dar. Für den Drainübergang gilt:

$$W_d = \sqrt{\frac{2\epsilon_{\text{Si}}}{eN_A} (V_{\text{ds}} + U_d + V_{\text{bs}})}, \quad (2.3)$$

hierbei ist U_d die Diffusionsspannung des Übergangs, V_{ds} die Drain-Source- und V_{bs} die Substrat-Source-Spannung.

Für die der Variante A zugrundeliegenden Parameter $N_A = 1.5 \cdot 10^{12} \text{ cm}^{-3}$, $x_j = 1.5 \mu\text{m}$ und $t_{\text{ox}} = 925 \text{ \AA}$ ergibt sich eine minimale Kanallänge $L_{\text{min}} = 42 \mu\text{m}$ bei einer Drain-Source-Spannung $V_{\text{ds}} = 0 \text{ V}$. Daraus folgt für alle praktikablen Kanallängen ein Kurzkanalverhalten der entsprechenden Transistoren.

TABELLE 4. CMOS-Prozeßparameter

	Alugate-Prozeß	3 μm Prozeß	1.5 μm Prozeß	0.7 μm Prozeß
Minimale Kanallänge L_{min} [μm]	5	2	1.2	0.6
Dicke des Gateoxides t_{ox} [nm]	93	70	25	20
Sperrschichttiefe der Source- und Drainübergänge x_j [μm]	1.5	0.6	0.3	0.2
Versorgungsspannung $V_{\text{ds,max}}$ [V]	+/- 6	+ 5	+ 5	+ 5

In Tabelle 4 sind Daten von kommerziellen CMOS-Prozessen [2.3] dem im Rahmen dieser Arbeit angewandten institutseigenen Alugate-Prozeß gegenübergestellt. Diese Daten ergeben mittels obiger Beziehung und vorgegebener Substratdotierung die minimal mögliche Kanallänge, bei der noch kein Kurzkanalverhalten zu erwarten ist. Die Ergebnisse dieser Rechnung sind in Bild 2.2 dargestellt.

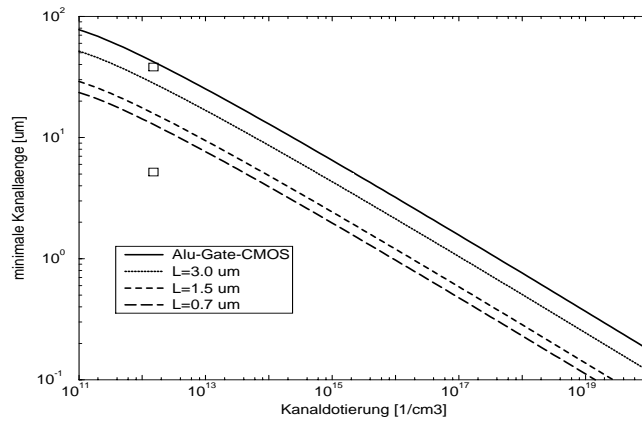


Bild 2.2: Minimale Kanallänge als Funktion der Kanaldotierung für Lang-Kanal-Verhalten von MOS. Prozeßdaten gemäß obiger Tabelle. Die Quadrate stellen die realisierten Kanallängen des NMOS der Variante A dar.

Die Rechnung zeigt, daß im Alu-Gate-Prozeß bei Transistoren auf hochohmigem Silizium ($N_A = 1.5 \cdot 10^{12} / \text{cm}^3$) Kurzkanalverhalten bis hin zu einer Kanallänge von etwa $L_{\min} = 42 \mu\text{m}$ zu erwarten ist. Die Gegenüberstellung mit den zum Vergleich angegebenen Kurven für kommerzielle CMOS-Prozesse zeigt, daß für das Erreichen der prozeßspezifisch vorgesehenen minimalen Kanallängen jeweils Dotierungen im Bereich von etwa $N = 10^{16} / \text{cm}^3$ erforderlich sind. Außerdem wird deutlich, daß auch für recht fortschrittliche Prozesse Transistoren, die ohne höherdotierte Wannen auf hochohmigem Silizium realisiert werden, Kanallängen von deutlich über $L_{\min} = 10 \mu\text{m}$ notwendig sind, um Kurzkanalverhalten zu vermeiden.

Erhöhte Drainströme im Unterschwellenbereich

Aufgrund der niedrigen Dotierung im Kanalbereich des MOS erfolgt eine starke Vergrößerung der pn-Raumladungsweiten zwischen Source und Substrat bzw. zwischen Drain und Substrat. Kommt es dabei zu einer Überlappung der Raumladungszonen um die Drain- und Sourceübergänge, so fließen zusätzliche Substratströme auch bei ausgeschaltetem Inversionskanal. Diese Leckströme und insbesondere auch ihre starke Abhängigkeit von der Substratspannung V_{bs} [2.4] bilden ein wesentliches Merkmal des Kurzkanalverhaltens von MOS-Transistoren. Bei diesen Strömen handelt es sich um einen im wesentlichen horizontalen Durchgriffstrom I_{pt} , (englisch: punch-through), der durch die Überlappung der Raumladungszonen zwischen Source und Substrat sowie zwischen Drain und Substrat entsteht.

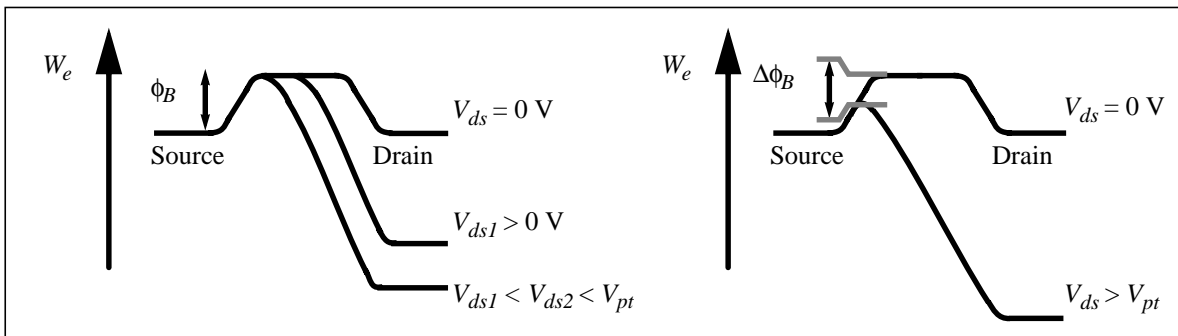


Bild 2.3: Diagramm der Elektronenenergie in einem NMOS-Transistor. Hierbei ist ein Schnitt zwischen Source und Drain unterhalb des eigentlichen Kanals gewählt worden. Bild links: Darstellung des Energieverlaufs bei verschiedenen Drainspannungen V_{ds} unterhalb der Durchgriffspannung V_{pt} . Bild rechts: Bei Überschreitung der Durchgriffspannung erfolgt eine von der Drainspannung abhängige Erniedrigung $\Delta\phi_B$ der Energiebarriere zwischen Source und Substrat.

Die bei niedrig dotierten Substraten mit zunehmender Drainspannung sehr schnell wachsende Raumladungsweite des Drainanschlusses erreicht bei der Durchgriffspannung $V_{ds} = V_{pt}$ die Raumladungszone des Sourcekontakts. Bei weiterer Erhöhung der Drainspannung erfolgt eine Absenkung der Source-Substrat-Barriere ϕ_B

und damit eine verstärkte Injektion von Elektronen aus der Source ins Substrat, welche aufgrund des elektrischen Feldes nach Überschreiten der Barriere zum Drainkontakt gezogen werden und somit zu einer Erhöhung des Stroms führen (vgl. Bild 2.3).

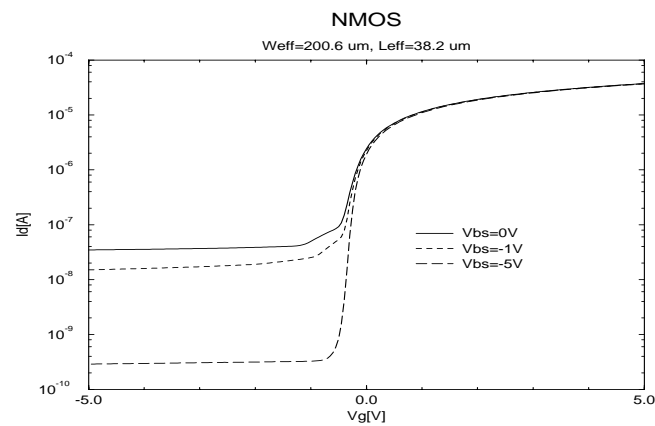


Bild 2.4: Transfer-Kennlinie $I_d - V_g$ eines NMOS ($L_{\text{eff}} = 38.2 \mu\text{m}$) auf hochohmigem p^- -Substrat. Die Aufnahme der Kennlinie erfolgte im linearen Bereich des MOS bei $V_{\text{ds}} = 50 \text{ mV}$ und verschiedenen Substratvorspannungen $V_{\text{bs}} = 0 \text{ V}$, -1 V und -5 V .

Bild 2.4 zeigt die Transfer-Kennlinie $I_d - V_g$ eines $38.2 \mu\text{m}$ -langen NMOS auf hochohmigem Substrat mit deutlichen Drainströmen I_d im ausgeschalteten Zustand. Im Unterschwellenbereich, dem sog. Off-Bereich, d.h. für $V_g < V_{\text{th}}$, befindet sich die Siliziumoberfläche im Kanalbereich in der Verarmung oder für stärker negative Gate-Spannungen in der Anreicherung, so daß kein Inversionskanal für den Stromfluß vorhanden ist. Es verbleibt jedoch ein von der Gatespannung unabhängiger Strom, der sehr stark durch die Substratvorspannung V_{bs} beeinflusst werden kann. Durch eine Senkung der Substratspannung auf $V_{\text{bs}} = -5 \text{ V}$ kann dieser Strom um zwei Größenordnungen verringert werden.

Ferner ist in diesem Bereich eine von der Höhe der Substratvorspannung V_{bs} abhängige Steigung der Stromkurve erkennbar.

Der normale Betriebsbereich des Transistors, d.h. der sog. On-Bereich für $V_g > V_{\text{th}}$, zeigt weitgehend die erwartete Linearität der Kennlinie. Hier ergibt sich die erwartete starke Abhängigkeit des Drainstroms I_d von der Gate-Spannung V_g aufgrund der Bildung und Verstärkung des Inversionskanals. Eine Wirkung der Substratvorspannung auf den gesamten Drainstrom ist kaum zu erkennen, da der vom Gatefeld beeinflusste Strom durch den Inversionskanal um einige Größenordnungen höher als der Leckstrom ist. Dieser vom Gatefeld weitgehend unabhängige Strom beträgt bei der Drainspannung von $V_{\text{ds}} = 50 \text{ mV}$ $I_d = 3 \cdot 10^{-8} \text{ A}$.

Während bei diesem recht langkanaligen Transistor durch eine moderate Substratvorspannung eine deutliche Reduktion des Leckstroms im Unterschwellenbereich erreicht werden kann, erfordert diese Reduzierung mit abnehmender Kanallänge zunehmend höhere Vorspannungen.

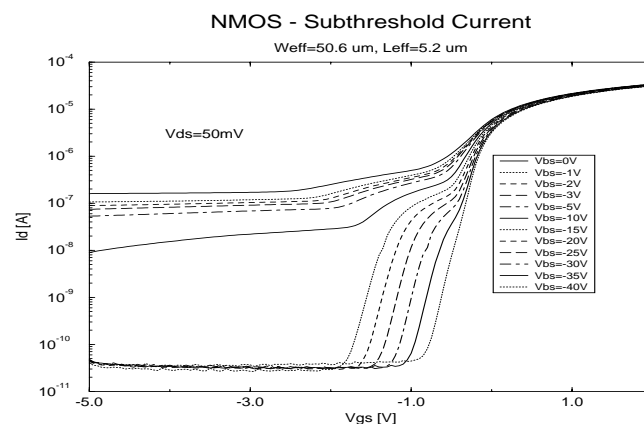


Bild 2.5: Transfer-Kennlinie $I_d - V_g$ eines NMOS ($L_{\text{eff}} = 5.2 \mu\text{m}$) auf hochohmigem p^- -Substrat. Die Aufnahme der Kennlinie erfolgte im linearen Bereich des MOS bei $V_{\text{ds}} = 50 \text{ mV}$ und verschiedenen Substratvorspannungen V_{bs} .

Bild 2.5 verdeutlicht den Einfluß einer weiteren Verkürzung der Kanallänge eines NMOS auf hochohmigem p⁻-Substrat. Der $L_{\text{eff}} = 5.2 \mu\text{m}$ kurze NMOS zeigt trotz geringerer Kanalweite $W_{\text{eff}} = 50.6 \mu\text{m}$ einen deutlich höheren Leckstrom $I_{\text{PT}} = 170 \text{ nA}$ als der langkanalige NMOS mit $W_{\text{eff}} = 200.6 \mu\text{m}$ und $L_{\text{eff}} = 38.2 \mu\text{m}$, sowie $I_{\text{PT}} = 27 \text{ nA}$ (vgl. Bild 2.4). Obwohl auch der längere NMOS schon bei der sehr geringen Drainspannung $V_{\text{ds}} = 50 \text{ mV}$ Durchgriffeffekte aufweist, verbleibt seine Source-Kanal-Barriere ϕ_{B} auf höherem Niveau als bei dem kürzeren Transistor und führt daher zu entsprechend kleineren Strömen im Off-Bereich.

Die kurze Kanallänge erfordert deutlich größere Substratvorspannungen, um den Durchgriffstrom zu unterbinden. Während eine Spannung $V_{\text{bs}} = -5 \text{ V}$ bei $L_{\text{eff}} = 38.2 \mu\text{m}$ zu einer Stromreduktion um mehr als zwei Größenordnungen führt, erfolgt bei $L_{\text{eff}} = 5.2 \mu\text{m}$ nur eine Verringerung um den Faktor 2. Eine weitgehende Unterdrückung der Punch-Through-Ströme wird bei dem kürzeren MOS bei Spannungen von $V_{\text{bs}} \leq -15 \text{ V}$ erreicht. In Bild 2.5 wird deutlich, daß der MOS erst für $V_{\text{bs}} = -40 \text{ V}$ ein typisches Verhalten mit einem exponentiellen Anstieg über mehrere Stromdekaden im Unterschwellen-Bereich zeigt.

Bei weniger negativen Spannungen ($-35 \text{ V} \leq V_{\text{bs}} \leq -15 \text{ V}$) verbleibt eine Auswölbung der Strom-Spannungs-Charakteristik im Bereich mittlerer Gatespannungen ($-3 \text{ V} \leq V_{\text{g}} \leq -1 \text{ V}$), die auf sehr oberflächennahe Durchgriffeffekte hindeutet. Erst wenn die Oberfläche des p⁻-Substrats im Kanalbereich bei diesen Substratspannungen durch eine deutlich negative Gatespannung ($V_{\text{g}} < -3 \text{ V}$) stark in die Akkumulation gezogen wird, erfolgt eine Unterbrechung dieses Strompfades und der Reststrom sinkt auf seinen Sättigungswert $I_{\text{d}} = 4 \cdot 10^{-11} \text{ A}$.

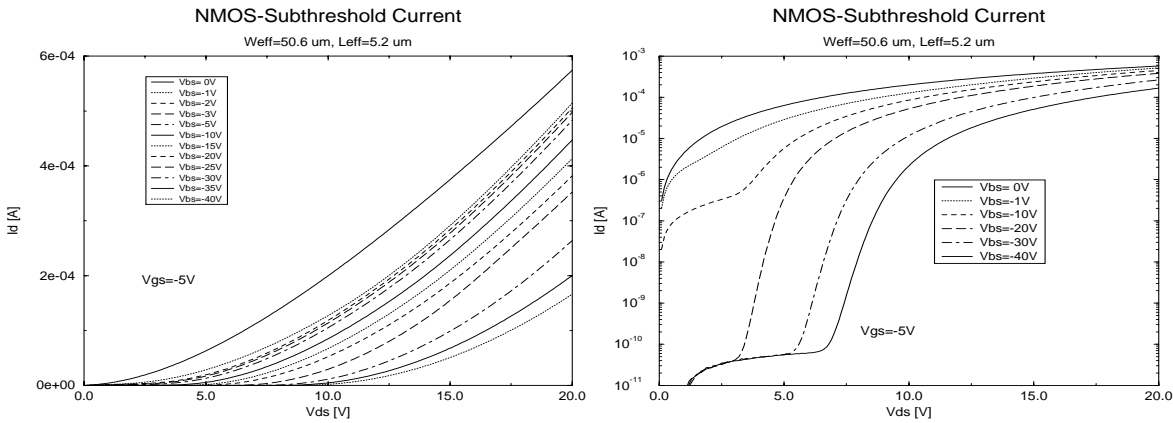


Bild 2.6: Darstellung des Durchgriffstroms $I_{\text{d}} = I_{\text{PT}}$ als Funktion der am Kanal anliegenden Spannung V_{ds} bei Betrieb des NMOS im Off-Bereich ($V_{\text{g}} = -5 \text{ V}$) mit unterschiedlichen Substratvorspannungen V_{bs} . Links lineare Darstellung des Stroms, rechts logarithmische Auftragung.

In Bild 2.6 ist der Punch-Through-Strom bei Unterdrückung des durch den Inversionskanal geführten Stroms dargestellt. Aufgrund der schon bei verschwindender Drain-Source-Spannung vorhandenen Überlappung der beiden Raumladungszonen um die Kanalschlüsse, zeigt der Transistor eine durchgehend überproportionale Stromzunahme mit der Spannung. Eitan et al. [2.5] haben die Annahme einer linearen Abnahme der Barrierenhöhe $\Delta\phi_{\text{B}}$ mit steigender Drain-Source-Spannung gemäß

$$|\Delta\phi_{\text{B}}| = c_{\text{pt}} |V_{\text{ds}}| \quad (2.4)$$

in die Literatur eingeführt. In dieser Gleichung stellt c_{pt} eine transistorspezifische Konstante dar. Ein daraus resultierender exponentieller Stromanstieg kann jedoch nur bei hohen Substratvorspannungen (vgl. Bild 2.6 rechts) beobachtet werden, da in allen anderen Fällen die schon bei sehr geringen Kanalspannungen vorhandene Überlappung zwischen Drain und Source erhalten bleibt. Die Folge dieser Überlappung ist der Übergang zu

einem raumladungsdominierten Stromverlauf [2.6] gemäß

$$I_{SCL} = \frac{9\epsilon_{Si}\mu_n A_{eff}}{8L_{eff}^2} V^2. \quad (2.5)$$

Die gleichen Meßdaten sind in Bild 2.7 als Funktion der Substratvorspannung V_{bs} mit der Kanalspannung V_{ds} als Parameter dargestellt.

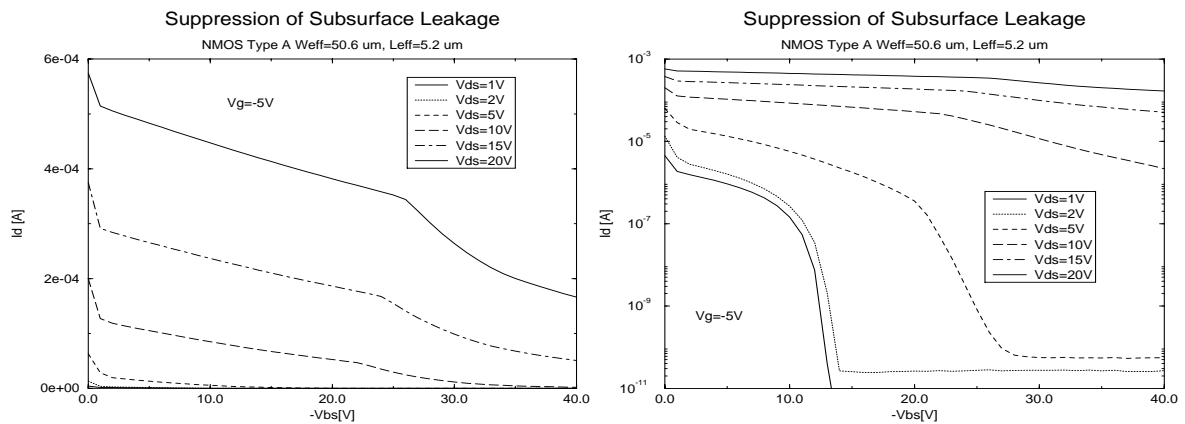


Bild 2.7: Darstellung des Durchgriffstroms $I_d = I_{PT}$ als Funktion der Substratvorspannung V_{bs} bei Betrieb des NMOS im Off-Bereich ($V_g = -5$ V) mit unterschiedlichen Kanalspannungen V_{ds} . Links lineare Darstellung des Stroms, rechts logarithmische Auftragung.

Einem Bereich (-20 V $\leq V_{bs} \leq -2$ V) mit nahezu linearer Reduzierung des Punch-Through-Stroms folgt bei weiter negativ werdenden Vorspannungen ein Bereich mit exponentieller Absenkung, in welchem die Barrierenhöhe proportional zur Substratvorspannung gemäß

$$|\Delta\phi_B| = d_{pt}|V_{bs}| \quad (2.6)$$

angehoben wird [2.1]. Durch negative Substratvorspannungen kann somit die weitgehende Unterdrückung der Ströme im Unterschwellenbereich erreicht werden.

Durchbruchverhalten des NMOS bei hohen Kanalspannungen

Abschließend ist das Durchbruchverhalten des NMOS untersucht worden, der aufgrund der niedrigen Substratdotierung eine sehr hohe Kanalspannung erlauben sollte. Die Ausgangskennlinie des NMOS ist in Bild 2.8 dargestellt, hierbei sind die Kanalspannungen V_{ds} , die den Einsatz der Lawinenmultiplikation markieren, durch eine stärkere Linie verdeutlicht.

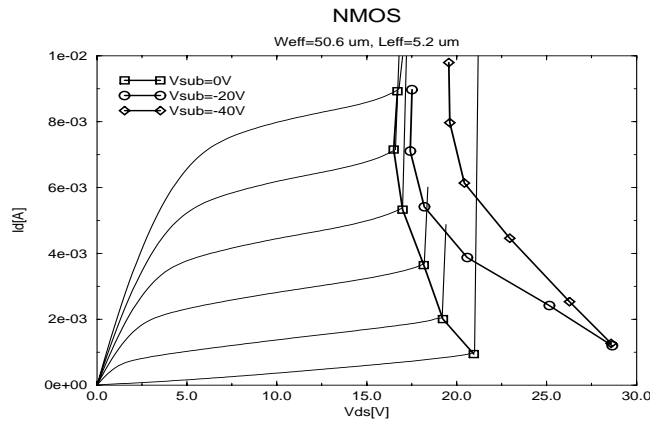


Bild 2.8: Ausgangskennlinie $I_d - V_{ds}$ des NMOS ($W_{eff} = 50.6 \mu\text{m}$, $L_{eff} = 5.2 \mu\text{m}$) mit Markierung der eingangsspannungsabhängigen Durchbruchspannungen $V_{ds,db}$ bei unterschiedlichen Substratvorspannungen V_{bs} . $V_g = 0 \text{ V}$, 2 V , 4 V , ..., 10 V .

Der Einfluß der Substratvorspannung V_{bs} auf die jeweiligen Durchbruchspannungen ist für die Spannungen $V_{bs} = -20 \text{ V}$ und $V_{bs} = -40 \text{ V}$ ebenfalls gekennzeichnet. Die maximalen Kanalspannungen, bis zu welchen der MOS betrieben werden kann, steigen von etwa $V_{ds,max} = 16 \text{ V}$ bei $V_{bs} = 0 \text{ V}$ auf $V_{ds,max} = 17 \text{ V}$ bei $V_{bs} = -20 \text{ V}$ und $V_{ds,max} = 19 \text{ V}$ bei $V_{bs} = -40 \text{ V}$.

Die Kennlinie zeigt damit nur für hohe Substratvorspannungen und geringe Gatespannungen die erwartete hohe Durchbruchfestigkeit. Diese Erhöhung der Durchbruchspannung erfolgt durch die Reduktion des Kanalstroms I_d , die durch die im folgenden genannten beiden Einflüsse bedingt ist:

- (i) Die negative Substratvorspannung bewirkt eine Verschiebung der Schwellspannung V_{th} hin zu positiveren Werten und verkleinert damit den Inversionskanal.
- (ii) Die ebenfalls durch die Substratvorspannung erfolgende Unterdrückung der Punch-Through-Ströme wird insbesondere bei Gatespannungen, die nur zu einem schwachen Inversionskanal führen, wirksam.

Eine Bestätigung des zweiten Arguments zeigt sich an der experimentellen Kennlinie durch den überproportionalen Anstieg der Durchbruchspannung bei geringen Gatespannungen; so steigt bei $V_g = 0 \text{ V}$ die maximale Kanalspannung von $V_{ds,db} = 21 \text{ V}$ bei $V_{bs} = 0 \text{ V}$ auf $V_{ds,db} = 28.5 \text{ V}$ bei $V_{bs} \leq -20 \text{ V}$.

Der hier unterstellte Einfluß der über das Substrat fließenden Ströme auf das Durchbruchverhalten steht im Einklang mit den in der Literatur vorgestellten Modellen [2.7], [2.8], insbesondere mit dem Modell der endlichen Multiplikation von Wang [2.9]. Aufgrund dieses Modells wird die insgesamt nur moderate Durchbruchfestigkeit erklärbar, da die für eine unendliche Lawinenmultiplikation [2.10] erforderliche Feldstärke bei solch niedrig dotierten Kanalgebieten erst bei deutlich höheren Spannungen erreicht werden könnte.

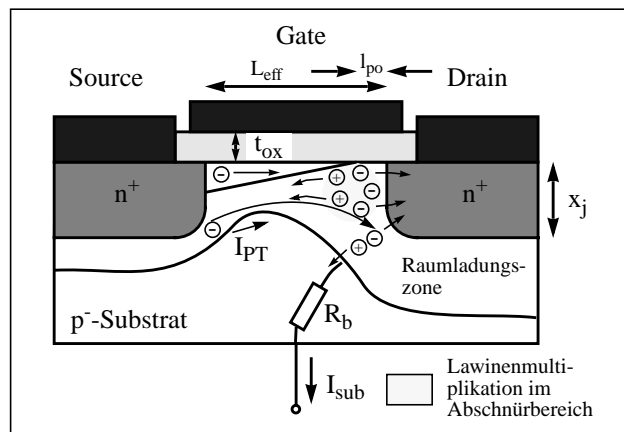


Bild 2.9: Darstellung der zum Lawinendurchbruch beitragenden Stromflüsse im NMOS. Die hier gezeichnete Darstellung entspricht weitgehend den in der Literatur üblichen (z. B. [2.9]), enthält jedoch zusätzlich den durch den Durchgriffstrom I_{PT} bedingten Anteil. Neben der Gateoxidicke t_{ox} , der Tiefe der pn-Übergänge x_j und der effektiven Kanallänge L_{eff} ist die Breite des Abschnürbereichs l_{po} gekennzeichnet.

Neben dem auf unendlicher Trägermultiplikation basierenden Prozeß der Lawinengeneration, bei dem der von

den Majoritätsträgern getragene Drainstrom auf [2.9]

$$I_d = M \cdot I_{d,sat} \quad (2.7)$$

und der durch ionisierende Stoßprozesse erzeugte Minoritätsträgerstrom auf

$$I_{sub} = (M - 1)I_{d,sat} \quad (2.8)$$

ansteigt, existiert laut Wang [2.9] die Möglichkeit des Stromanstiegs aufgrund von endlicher Trägermultiplikation und einer Rückkopplung zwischen dem Spannungsabfall im Substrat V_{sub} und dem Prozeß der Trägermultiplikation. In den obigen Beziehungen ist $I_{d,sat}$ der im Sättigungsbereich vor Erreichen des Durchbruchs fließende Drainstrom. Der Multiplikationsfaktor M ist allgemein von der Ionisationsrate α abhängig, welche ihrerseits einer exponentiellen Abhängigkeit vom elektrischen Feld im Abschnürbereich (engl. pinch-off region) folgt [2.11]:

$$M = \frac{1}{1 - \int_0^{p_0} \alpha \cdot dy} \quad \text{mit} \quad \alpha = A \cdot e^{\frac{B}{E}} \quad (2.9)$$

Hierbei gelten die Ionisationskonstanten $A = 2.45 \cdot 10^5 / \text{cm}$ bzw. $B = 1.92 \cdot 10^6 \text{ V/cm}$ für nah an der Halbleiteroberfläche befindliche Elektronen, die den Prozeß der Lawinenmultiplikation auslösen [2.12]. Bei der unendlichen Lawinenmultiplikation erfolgt der Durchbruch sobald M den Wert „unendlich“ (∞) aufgrund zunehmender Ionisierung erreicht. Diese Art des Durchbruchs ist bei MOS auf vergleichsweise niederohmigem Substratmaterial dominierend, da der Spannungsabfall im Substrat gemäß

$$V_{sub} = I_{sub} R_b = (M - 1)I_{d,sat} R_b \quad (2.10)$$

nur gering ausfällt.

Wird aufgrund höherer Substratwiderstände R_b dieser Spannungsabfall deutlich vergrößert, so steigt das Potential im Kanalbereich an und sorgt analog einer positiven Vorspannung des Substrats V_{bs} zu einer Verringerung der Schwellspannung V_{th} und führt daher in der Folge zu einem vergrößerten Drainstrom I_d . Da dieser Strom zu erhöhter Ionisierung und damit zu vergrößertem Löcherstrom zum Substratkontakt führt, besteht eine positive Rückkopplung dieses Vorgangs, der zu einem Durchbruch bei geringerer als unendlicher Lawinenmultiplikation führt ($M \ll \infty$).

Durchbruch bei endlicher Lawinenmultiplikation führt zu Durchbruchspannungen $V_{ds,db}$, die mit steigender Gatespannung abnehmen, wie in Bild 2.8 dargestellt. Demgegenüber zeichnet sich eine Lawinenmultiplikation aufgrund unendlicher Multiplikation durch eine mit der Gatespannung zunehmende Durchbruchspannung aus. Grundsätzlich wäre auch im Fall der endlichen Multiplikation mit einer Zunahme der Durchbruchspannung zu rechnen, da mit zunehmender Gatespannung V_g die Feldstärke E im Pinch-Off-Bereich sinkt. Die Zunahme der Gatespannung bedingt eine erhöhte Sättigungsspannung $V_{ds,sat} = V_g - V_{th}$ und damit einen verringerten Spannungsabfall $\Delta V = V_{ds} - V_{ds,sat}$ im abgeschnürten Kanalbereich; die Folge ist eine verringerte Feldstärke E . Der durch den Substratspannungsabfall V_{sub} bedingte Drainstromanstieg fällt jedoch größer aus, so daß es in der Summe zu einem Absinken der Durchbruchspannungen kommt. Aufgrund der deutlich niedrigeren Ionisierungsraten von Löchern [2.13] ist dieser Durchbrucheffect bei PMOS-Transistoren auf n⁻-Substrat nicht zu beobachten [2.9].

2.1.2 Der NMOS-FET in den Prozeßvarianten B und C

Sowohl in den beiden Varianten B und C auf hochohmigem p-Silizium ($<100>$ orientiertes FZ-Material mit $N_A = 1.5 \cdot 10^{12} / \text{cm}^3$) als auch bei dem zum Vergleich herangezogenen Standard-CMOS-Prozeß auf n-Substrat ($<100>$ orientiertes CZ-Material mit $N_D = 1.5 \cdot 10^{15} / \text{cm}^3$) befindet sich der NMOS-Transistor in einer durch Implantation hergestellten p-Wanne. Der Vergleich der jeweiligen Dotierprofile in Bild 2.10 zeigt die in allen drei Fällen etwa gleich hohe Dotierkonzentration im Kanalbereich an der Oberfläche, die jeweils bei etwa $N_A = 2 \cdot 10^{16} / \text{cm}^3$ liegt.

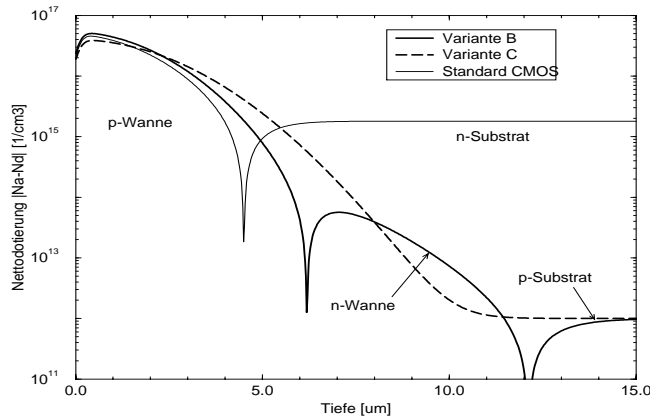


Bild 2.10: SUPREM IV-Prozeß-Simulationsergebnisse [2.14] der NMOS-Kanalgebiete. Es ist jeweils ein Vertikalschnitt von der Oberfläche ausgehend in der Mitte des Kanalgebiets zwischen Source und Drain dargestellt. Die Variante B mit der großflächigen n-Wanne und die Doppel-Wannen-Variante C sind dem Standard-CMOS-Prozeß auf n-Substrat gegenübergestellt.

Aufgrund der leicht unterschiedlichen Oberflächenkonzentrationen stellen sich geringfügig voneinander abweichende Schwellspannungen V_{th} ein (vgl. Tabelle 5).

In der Doppel-Wannen-Variante C ergibt sich ein kontinuierlicher Übergang von der p-Wanne ins p-Substrat. Da das Dotierniveau der p-Wanne in der Variante C das des Standard-Prozesses - mit Ausnahme der sehr oberflächennahen Bereiche - übersteigt, werden hier - wie im Standard-Prozeß - keinerlei Durchgriffeffekte durch die Wanne beobachtet. Die p-Wanne der Variante B zeigt ebenfalls keinen Durchgriff, jedoch erfolgt bei Anlegen der Wannepotentiale $V_{p-Wanne} = -V_{CC} = -6 \text{ V}$ und $V_{n-Wanne} = +V_{CC} = +6 \text{ V}$ ein Durchgriff des pn-Übergangs unterhalb der p-Wanne durch die n-Wanne. Die Folge ist ein Leckstrom zwischen den Wannen aufgrund der Betriebsspannungen.

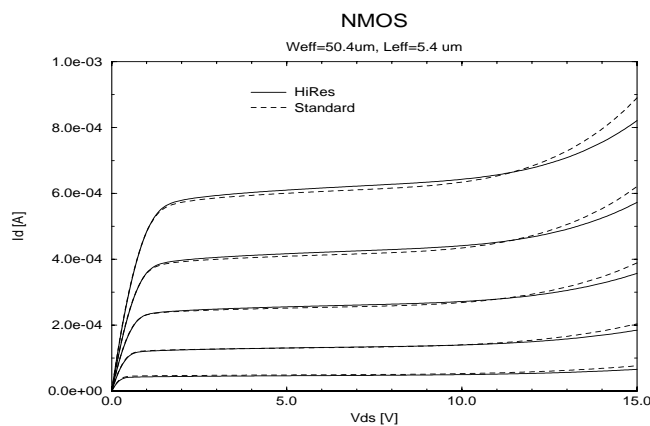


Bild 2.11: Ausgangskennlinienfeld I_d - V_{ds} von 5.4 µm-langen NMOS in der p-Wanne der Variante C und dem Standard-Prozeß. Vergleich eines Transistors auf hochohmigem FZ-Substrat (durchgezogene Linie) und konventionellem CZ-Substrat (gestrichelt). $V_g = 2.0, 2.5, \dots, 4.0 \text{ V}$.

Da der Transistor der Variante C keine parasitären Leckstromeffekte zeigt, ist er für den Vergleich mit dem NMOS des Alu-Gate-Standard-Prozesses gewählt worden. Für die folgende Untersuchung sind jeweils identisch gezeichnete NMOS der Nominalmaße $W_{nom} = 50 \text{ µm}$ und $L_{nom} = 10 \text{ µm}$ gewählt worden. Durch elektrische

Messungen der Kanalströme sind für beide Transistoren effektive Kanaldimensionen von $W_{\text{eff}} = 50.4 \mu\text{m}$ und $L_{\text{eff}} = 5.4 \mu\text{m}$ ermittelt worden. Hierbei wurden die in Anhang A1.1 ermittelten Kanalgeometrieänderungen ΔW und ΔL benutzt.

Im Hinblick auf eines der späteren Hauptanwendungsgebiete ist die maximale Drain-Source-Spannung der MOS interessant. Durch die recht konservativ gewählten Prozeßparameter, d.h. recht große Kanallängen und moderate Kanaldotierungen, ist eine Kanalspannung $V_{\text{ds}} = 12 \text{ V}$ möglich. Wie das Bild 2.11 zeigt, erfolgt ein wesentlicher Anstieg des Drainstroms aufgrund von Durchbruchmechanismen erst oberhalb von 12 V , so daß ein direkter Betrieb an einer im Automobil eingebauten Batterie denkbar ist.

In Bild 2.12 ist die Kennlinie der Durchbruchspannung $V_{\text{ds,max}}$ als Funktion der Gatespannung V_{g} dargestellt. Die Durchbruchspannung sinkt für Werte zwischen $0 \text{ V} \leq V_{\text{g}} \leq 7 \text{ V}$ auf ein Minimum bei $V_{\text{ds,max}} = 17 \text{ V}$ und steigt für höhere Gatespannungen, wie im Fall eines Lang-Kanal-MOS zu erwarten.

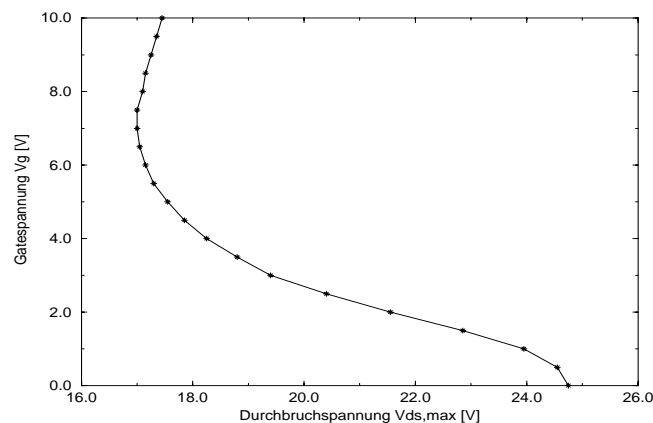


Bild 2.12: Durchbruchkennlinie $V_{\text{ds,max}}$ als Funktion der Gatespannung V_{g} eines $5.4 \mu\text{m}$ -langen NMOS in der p-Wanne der Variante C.

Das anfängliche Absinken der Durchbruchspannung mit zunehmender Gatespannung deutet auf das Einsetzen einer positiven Rückkopplung hin. Der durch Ionisationseffekte im abgeschnürten Kanalbereich bedingte Substratstrom I_{sub} und die damit verbundene Schwellspannungsabsenkung ΔV_{th} bewirken eine weitere Erhöhung des Kanalstroms I_{d} und folglich Lawinenmultiplikation bei einem nicht unendlich großen Multiplikationsfaktor M . Ein solcher Effekt wurde schon bei den NMOS der Variante A auf hochohmigem Silizium beobachtet. Obwohl hier die p-Wanne einen im Vergleich zum p^- -Substrat geringen Serienwiderstand R_{b} aufweist und der Substratstrom I_{sub} demzufolge nur einen kleinen Spannungsabfall V_{sub} in der Wanne bedingt, kommt es aufgrund des großen Backgateeffekts - der Substratfaktor γ liegt in der Größenordnung von zwei - zu einer signifikanten Reduktion der Schwellspannung und damit zu dem beobachteten Verlauf des Durchbruchverhaltens. Erst für größere Gatespannungen wird der durch das Ansteigen der Sättigungsspannung $V_{\text{ds,sat}}$ bedingte geringere Spannungsabfall über der abgeschnürten Zone $\Delta V = V_{\text{ds}} - V_{\text{ds,sat}}$ wirksam und bewirkt eine leichte Erhöhung der Durchbruchspannung.

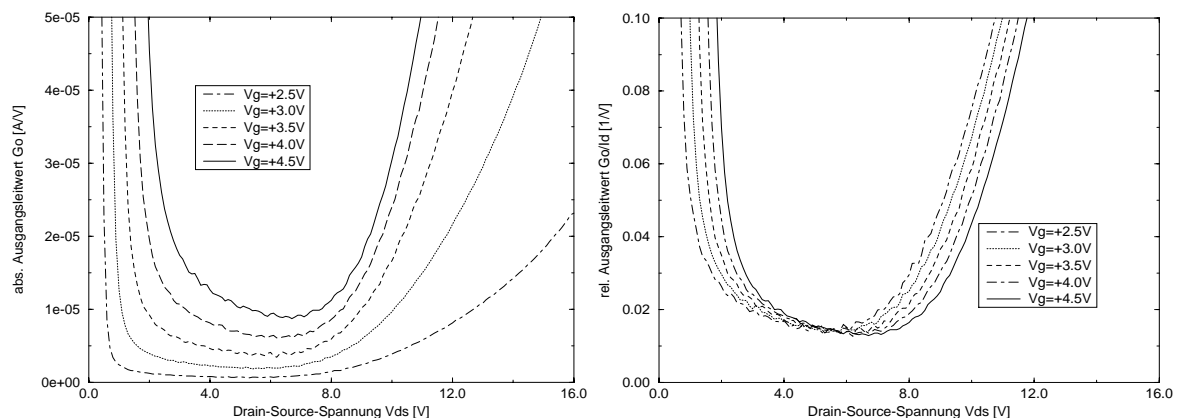


Bild 2.13: Links: Darstellung des Ausgangsleitwertes g_{o} des NMOS $W_{\text{nom}} = 50 \mu\text{m}$ und $L_{\text{nom}} = 10 \mu\text{m}$ der Variante C. Rechts: Darstellung des auf den jeweiligen Drainstrom I_{d} bezogenen Ausgangsleitwertes $g_{\text{o}}/I_{\text{d}}$.

Eine wichtige Größe für das analoge Schaltverhalten der MOS und damit auch die Leerlaufverstärkung A_v (vgl. Anhang A5.1) stellt der Ausgangsleitwert g_o , insbesondere im Sättigungsbereich $V_{ds} \geq V_g - V_{th}$, dar. Ein möglichst geringer Ausgangsleitwert führt zu einer hohen Leerlaufverstärkung. Bild 2.13 zeigt sowohl den absoluten Ausgangsleitwert g_o eines typischen NMOS der Variante C (links im Bild) als auch den auf den Ausgangsstrom bezogenen Leitwert g_o/I_d (rechts im Bild). Die Darstellung zeigt das Ansteigen des Leitwerts im Bereich der kleinen Drain-Source-Spannungen V_{ds} , wie im linearen Bereich des MOS erwartet. Bei den größeren Kanalspannungen zeigt sich der Einfluß des Stromanstiegs aufgrund der Kanallängenmodulation und bei noch höheren Spannungen treten Durchbrucheffekte auf.

Während eine exakte Modellierung des Ausgangsleitwerts sehr aufwendig und insbesondere für den Schaltungsentwurf recht unergiebig ist [2.15], kann eine für den Erstentwurf ausreichende Genauigkeit durch folgende - für den hier interessierenden Sättigungsbereich gültige - Formel erreicht werden:

$$g_o = \lambda' \cdot \frac{I_d}{L_{eff}} \quad \text{mit} \quad \lambda' \sim \sqrt{\frac{1}{N_A}} \quad \text{und} \quad I_d = \mu_o C_{ox} \frac{W_{eff}}{L_{eff}} (V_g - V_{th})^2 . \quad (2.11)$$

Wie durch obige Beziehung ausgedrückt und im Bild 2.13 rechts gezeigt, führt der Quotient g_o/I_d zu einem gate-spannungsunabhängigen Minimum - etwa in der Mitte zwischen den beiden oben genannten Bereichen sinkenden Ausgangswiderstands. Sowohl der Referenztransistor des Standard-Prozesses mit den Dimensionen $W_{nom} = 50 \mu\text{m}$ und $L_{nom} = 10 \mu\text{m}$ als auch jener der Variante C weisen das Minimum bei etwa $g_o/I_d = 1.5 \cdot 10^{-2} \text{V}^{-1}$ auf, die etwas höhere Kanaldotierung der Variante B führt hingegen zu einem leicht verringerten bezogenen Ausgangsleitwert von $g_o/I_d = 1.3 \cdot 10^{-2} \text{V}^{-1}$. Eine Umrechnung dieser Werte auf Transistoren anderer Geometrien kann durch Berücksichtigung der oben ermittelten Kanallängenverkürzung um $2 \Delta L$ erfolgen.

Die in Bild 2.14 dargestellte Übertragungskennlinie zeigt den exponentiell steigenden Drainstrom I_d im Unterschwellenbereich $V_g < V_{th}$. Der diesen Anstieg charakterisierende Parameter S ist dabei als Gatespannungsänderung ΔV_g zur Erhöhung des Drainstroms I_d um eine Dekade definiert [2.16]

$$S = \frac{dV_g}{d(\log I_d)} \quad \text{in} \quad \frac{\text{mV}}{\text{Dekade}} . \quad (2.12)$$

In Ermangelung eines deutschen Begriffs wird diese Größe als „gate swing“ oder „subthreshold swing“ bezeichnet. Der Transistor der Doppel-Wannen-Variante C zeigt eine Steigung $S = 138 \text{mV/dec}$ und liegt damit geringfügig unter dem für den NMOS des Referenzprozesses auf n-Substrat, für den $S = 154 \text{mV/dec}$ gilt. Die leichte Abweichung zwischen beiden Werten liegt in der kritischen Abhängigkeit des Parameters a von der Kanaldotierung N_A und der Gateoxididicke t_{ox} begründet.

Der „subthreshold swing“ S steigt monoton mit dem Parameter a , der wie folgt definiert ist:

$$a = \sqrt{2} \frac{\epsilon_{Si} t_{ox}}{\epsilon_{ox} \lambda_D} \sim t_{ox} \cdot \sqrt{N_A} , \quad (2.13)$$

hierbei ist λ_D die Debye-Länge der Majoritätsträger, die durch

$$\lambda_D = \sqrt{\frac{U_t \epsilon_{Si}}{e N_A}} \quad (2.14)$$

im Fall eines n-Kanal MOS gegeben ist.

Für beide Transistoren auf CZ und auf FZ-Substraten bleibt der erforderliche „gate swing“ S über etwa vier Stromdekaden weitestgehend konstant. Die hier ermittelten Werte in der Größenordnung von $S = 140\text{mV}$ pro Dekade sind bei Gateoxidicken $t_{\text{ox}} = 92.5\text{ nm}$ und einer Dotierung von $N_{\text{A}} = 2 \cdot 10^{16}\text{ cm}^{-3}$ zu erwarten [2.17].

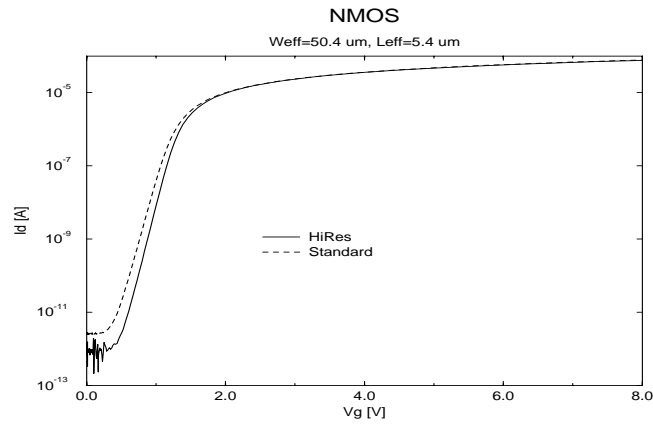


Bild 2.14: Transfercharakteristik $I_{\text{d}}-V_{\text{g}}$ eines $5.4\text{ }\mu\text{m}$ -langen NMOS im linearen Bereich. $V_{\text{ds}} = 50\text{ mV}$. Der NMOS der Variante C (durchgezogene Linie) ist dem NMOS des Standard-Prozesses (unterbrochene Linie) gegenübergestellt.

Für den NMOS der Variante B kann ein „gate swing“ S nicht exakt bestimmt werden, da aufgrund der die Messung überlagernden Leckströme, die durch den frühen Durchgriff der n-Wanne bedingt sind, nur eine Beobachtung des exponentiellen Verhaltens über etwa 2-3 Dekaden möglich ist.

TABELLE 5. NMOS-FETs

	Variante B	Variante C	Standard-CMOS
Schwellspannung	1.60	1.30	1.28
V_{th} [V] * ¹⁾			
maximale Beweglichkeit μ_{n} [cm^2/Vs] bei $V_{\text{ds}}=50\text{ mV}$ * ¹⁾	705	715	710
Maximale relative Steilheit $g_{\text{m}}/I_{\text{d}}^{1/2}$ [$\text{A}^{1/2}/\text{V}$] bei $V_{\text{ds}} = 12\text{ V}$ für $W_{\text{nom}} = 50\text{ }\mu\text{m}$ und $L_{\text{nom}} = 10\text{ }\mu\text{m}$ * ¹⁾	$1.5 \cdot 10^{-2}$	$1.5 \cdot 10^{-2}$	$1.5 \cdot 10^{-2}$
Minimaler relativer Ausgangsleitwert $g_{\text{o}}/I_{\text{d}}$ [1/V] bei $V_{\text{ds}} = 12\text{ V}$ für $W_{\text{nom}} = 50\text{ }\mu\text{m}$ und $L_{\text{nom}} = 10\text{ }\mu\text{m}$ * ¹⁾	$1.3 \cdot 10^{-2}$	$1.5 \cdot 10^{-2}$	$1.5 \cdot 10^{-2}$
Kanaldotierung an der Oberfläche N_{A} [$1/\text{cm}^3$] * ²⁾	$2.3 \cdot 10^{16}$	$1.7 \cdot 10^{16}$	$1.9 \cdot 10^{16}$
Substratkonstante γ [$\text{V}^{1/2}$] * ¹⁾	2.11	1.82	1.78
Steigung im Unterschwellenbereich S [mV/dec] * ¹⁾	* ³⁾	138	154

*¹⁾ aufgrund von elektrischen Messungen ermittelt

*²⁾ aufgrund von SUPREM-Simulationen ermittelt

*³⁾ aufgrund von Leckströmen nicht bestimmbar

Anmerkung: Die aus der Substratkonstanten γ bestimmbare Wannendotierung N_{A} weicht geringfügig von den durch die SUPREM-Simulation ermittelten Werten ab, da das meßtechnische Verfahren jeweils die unterhalb der Verarmungszone befindliche Dotierung bestimmt und nicht die Oberflächendotierung. Während für die Bestimmung der Schwellspannung V_{th} die meßtechnisch ermittelte Größe relevant ist, wird für den Vergleich auf die - durch den Pile-Down-Effekt des Bors erniedrigte - rechnerisch bestimmte (vgl. Bild 2.10) Oberflächenkonzentration zurückgegriffen.

Die anhand der Ausgangs- und Übertragungskennlinie für alle Typen ermittelten Größen sind in Tabelle 5 zusammengefaßt. Ferner ist die durch Variation der Substratvorspannung ΔV_{sub} ermittelte Schwellspannungsänderung ΔV_{th} zur Bestimmung des Back-Gate-Faktors γ benutzt worden.

Der nicht zusätzlich aufgeführte NMOS der Variante C₂, die als Basis des Integrationsprozesses benutzt wurde, entspricht sowohl prozeßtechnisch als auch in den elektrischen Eigenschaften dem NMOS der hier angeführten Variante C.

2.1.3 Der PMOS-FET

Bedingt durch die unterschiedliche zeitliche Abfolge der n-Wannen-Einbringung in den Varianten A, B und C ergeben sich deutlich unterschiedliche Dotierungsprofile der n-Wannen. Die Implantation der n-Wanne nach dem Drive-In der p-Wanne in der Variante C führt zu einer nur $x_{j,n\text{-Wanne}} = 4.7 \mu\text{m}$ tiefen Wanne.

Verbunden mit der für die gewünschte Schwellspannung V_{th} bedingten niedrigen Oberflächenkonzentration von $N_{\text{D}} = 2.5 \cdot 10^{15} / \text{cm}^3$ ergibt sich ein sehr früher Durchgriff durch die n-Wanne und damit verbunden erhebliche Leckströme des PMOS-FET.

Dies gilt für die beiden untersuchten Phosphorimplantationen mit $N_{\square} = 3.3 \cdot 10^{11} / \text{cm}^2$ und $N_{\square} = 6.7 \cdot 10^{11} / \text{cm}^2$, die jeweils mit einer Energie von $E = 100 \text{ keV}$ eingebracht wurden.

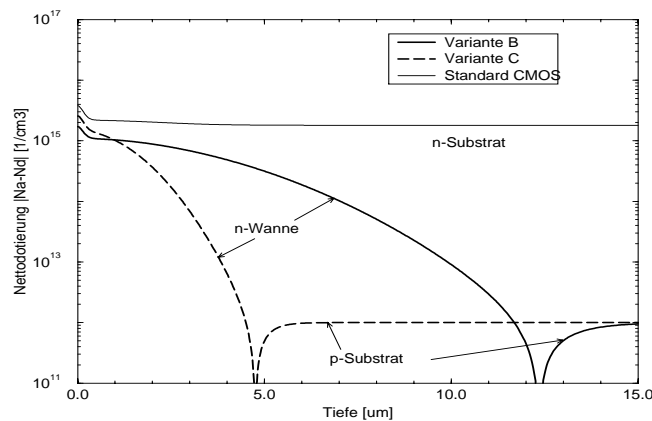


Bild 2.15: SUPREM IV-Prozeß-Simulationsergebnisse [2.14] der PMOS-Kanalgebiete. Es ist jeweils ein Vertikalschnitt von der Oberfläche ausgehend in der Mitte des Kanalgebiets zwischen Source und Drain dargestellt. Die Variante B mit der großflächigen n-Wanne und die Doppel-Wannen-Variante C sind dem Standard-CMOS-Prozeß auf n-Substrat gegenübergestellt. Sowohl in der Variante B als auch in der Variante C ist mit je zwei unterschiedlichen n-Wannendotierungen gearbeitet worden. Dargestellt als Variante B ist das durch Phosphor-Implantation von $N_{\square} = 6.7 \cdot 10^{11} / \text{cm}^2$ bei einer Energie $E = 100 \text{ keV}$ erhaltene Profil und als Variante C das durch $N_{\square} = 3.3 \cdot 10^{11} / \text{cm}^2$ und $E = 100 \text{ keV}$ erhaltene Profil.

Aufgrund einer ab der Definition der n-Wanne zur Variante C identisch erfolgenden Prozessierung und damit gleichem n-Wannen Dotierprofil liefert die Variante A vergleichbare PMOS-Transistoren, die ebenfalls schon bei niedrigen Drain-Source-Spannungen V_{ds} Kanalströme I_{d} von einigen μA im ausgeschalteten Zustand aufweisen. Erst bei Kanalspannungen $|V_{\text{ds}}| > 7 \dots 8 \text{ V}$ zeigt sich ein vergleichbarer Off-Strom bei den PMOS der Variante B mit einer Implantationsdosis $N_{\square} = 6.7 \cdot 10^{11} / \text{cm}^2$ Phosphor. Die verringerte Implantationsdosis von $N_{\square} = 3.3 \cdot 10^{11} / \text{cm}^2$, die bei der Variante B ebenfalls untersucht wurde, verschiebt den Einsatz des Stromanstiegs zu ca. 3 V geringeren Spannungsbeträgen. Bei dieser Spannung erfolgt dann ebenfalls ein Durchgriff durch die n-Wanne.

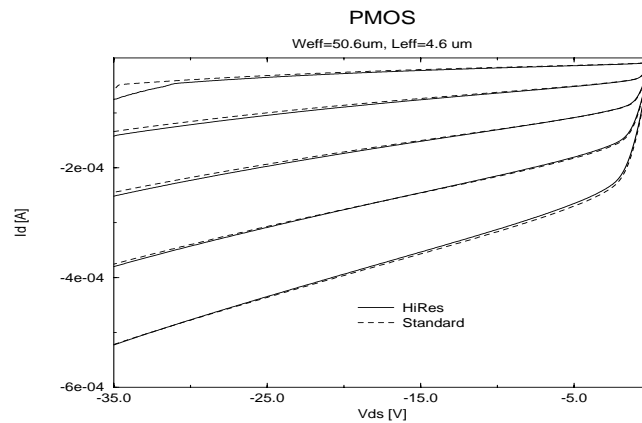


Bild 2.16: Ausgangskennlinienfeld I_d - V_{ds} eines 4.6 μm -langen PMOS in der n-Wanne der Variante B und auf dem Substrat des Standard-Prozesses. Vergleich eines Transistors auf hochohmigem FZ-Substrat (durchgezogene Linie) und konventionellem CZ-Substrat (gestrichelt). $V_g = -2.0, -2.5, \dots, -4.0$ V.

Der Vergleich der Ausgangskennlinie dieses PMOS mit einem des Vergleichsprozesses auf n-Substrat zeigt (vgl. Bild 2.16) - anders als etwa beim NMOS der Variante A - keinerlei Auswirkungen dieser recht geringen Leckströme auf das Kennlinienfeld. Die sehr gute Übereinstimmung beider Transistoren zeigt sich auch im durch Kanalmodulation bedingten Anstieg der Ausgangskennlinienäste im Sättigungsbereich. Eine geringe Oberflächenkonzentration von $N_D = 1.8 \cdot 10^{15} / \text{cm}^3$ in der Variante B bedingt diesen starken Effekt.

Die daraus resultierende Stromzunahme bei Erhöhung der Kanalspannung wird ebenfalls durch einen recht hohen Ausgangsleitwert g_o ausgedrückt (vgl. Bild 2.17). Aussagekräftiger als der Absolutwert des Ausgangsleitwerts, der bei vergleichbar hohen Beträgen der Gatespannung V_g ähnlich ausfällt wie im Fall des NMOS (vgl. Bild 2.13), ist der bezogene Leitwert g_o/I_d . Der PMOS erreicht einen etwa doppelt so großen Wert wie ein vergleichbarer NMOS-FET, d.h. $g_o/I_d = 3.5 \cdot 10^{-2} \text{ V}^{-1}$ für den PMOS und $g_o/I_d = 1.5 \cdot 10^{-2} \text{ V}^{-1}$ für den NMOS. Neben der geringeren Kanaldotierung und daraus resultierender höherer Kanallängenmodulation wirkt sich außerdem die kleinere effektive Kanallänge des PMOS auf die Größe des Ausgangsleitwerts aus. Die geringere effektive Kanallänge ist neben den unterschiedlichen Unterdiffusionsweiten von Phosphor und Bor und damit anderen metallurgischen Abständen zwischen Source und Drain bei NMOS und PMOS nicht zuletzt durch die größere Ausdehnung der Drain-Raumladungszone des PMOS bedingt.

In einem Spannungsbereich $-7 \text{ V} \geq V_{ds} \geq -20 \text{ V}$ zeigt sich die erwartete Konstanz des bezogenen Leitwerts g_o/I_d bei den untersuchten Gatespannungen.

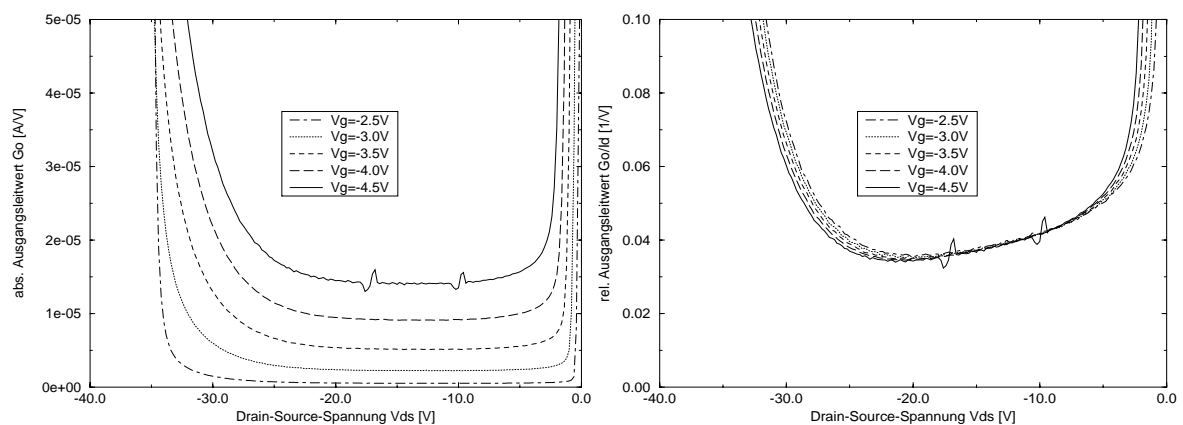


Bild 2.17: Links: Darstellung des Ausgangsleitwerts g_o des PMOS $W_{\text{nom}} = 50 \mu\text{m}$ und $L_{\text{nom}} = 10 \mu\text{m}$ der Variante B. Rechts: Darstellung des auf den jeweiligen Drainstrom I_d bezogenen Ausgangsleitwerts g_o/I_d .

Aufgrund der niedrigen Kanaldotierung ist ein spätes Einsetzen der Lawinenmultiplikation zu erwarten. Zusätzlich verschiebt das im Vergleich zu Elektronen geringere Ionisationsvermögen der Löcher diesen Effekt hin zu größeren Kanalspannungen. Die in Bild 2.18 gezeigte Kennlinie der Durchbruchspannung in Abhängigkeit der

Gatespannung verdeutlicht dieses nachdrücklich. Erst bei Drain-Source-Spannungsbeträgen über 40 V erfolgt ein deutliches Ansteigen des Kanalstroms.

Wie bei Langkanaltransistoren erwartet, ist ein stetiges Anwachsen der Durchbruchspannung als Folge steigender Gatespannung zu erkennen. Das Absinken der Feldstärke im Abschnürbereich aufgrund der linear mit der Gatespannung zunehmenden Sättigungsspannung $|V_{ds,sat}| = |V_g - V_{th}|$ zeigt sich anfangs sehr deutlich im nahezu proportionalen Anstieg der Durchbruchspannung mit der Gatespannung. Erst bei Gatespannungen $|V_g| \geq 4$ V erfolgt die weitere Zunahme unterproportional.

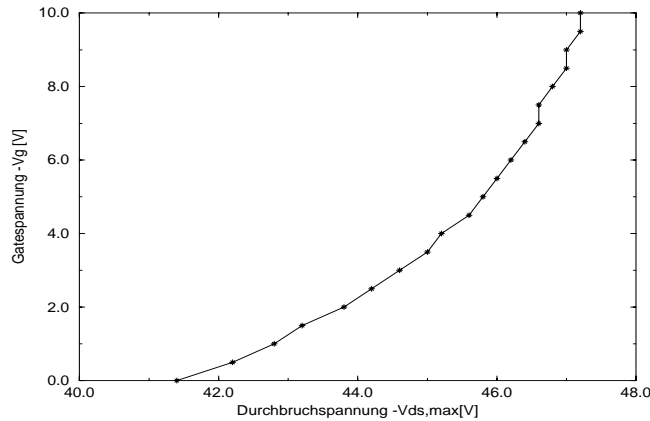


Bild 2.18: Durchbruchkennlinie $-V_{ds,max}$ als Funktion der Gatespannung $-V_g$ eines 4.6 μm -langen PMOS in der n-Wanne der Variante B.

Die Transfer- oder Übertragungs-Kennlinie in Bild 2.19 zeigt eine sehr gute Übereinstimmung der Kanalströme des MOS der Variante B mit dem Transistor des Standard-Prozesses auf n-Substrat. Über etwa fünf Stromdekaden ist der exponentielle Anstieg des Kanalstroms im Unterschwellenbereich beobachtbar. In diesem Bereich zeigen die Transistoren einen „subthreshold swing“ von etwa $S = 95 \dots 98$ mV/dec, der mit in der Literatur berichteten Werten übereinstimmt [2.17].

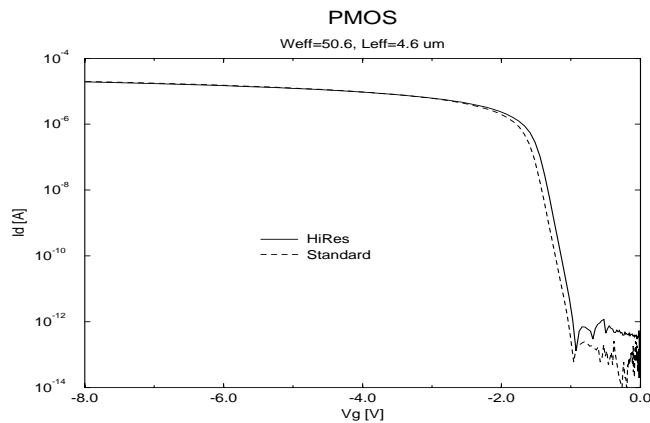


Bild 2.19: Transfercharakteristik I_d - V_g eines 4.6 μm -langen PMOS im linearen Bereich. $V_{ds} = -50$ mV.

Die beobachtete Differenz der Schwellspannungen fällt sehr gering aus, so daß der PMOS der Variante B als eine gute Annäherung an den PMOS des Standard-Prozesses in Bezug auf Schwellspannung V_{th} , Kanalstrom I_d , Steilheit g_m und Ausgangsleitwert g_o betrachtet werden kann.

Bei den Transistoren der Variante C ist auf eine Ermittlung des „gate swing“ S verzichtet worden, da auch hier der Bereich des exponentiellen Anstiegs durch die signifikanten Unterschwellenströme auf nur etwa 2-3 Größenordnungen des Stroms beschränkt wird.

TABELLE 6. PMOS-FETs

	Variante B mit $N_{\square}=3.3 \cdot 10^{11}/$ cm^2 bei $E = 100 \text{ keV}$	Variante B mit $N_{\square}=6.7 \cdot 10^{11}/$ cm^2 bei $E=100 \text{ keV}$ * ⁴)	Variante C mit $N_{\square}=3.3 \cdot 10^{11}/$ cm^2 bei $E=100 \text{ keV}$ * ⁵)	Variante C mit $N_{\square}=6.7 \cdot 10^{11}/$ cm^2 bei $E = 100 \text{ keV}$	Standard- CMOS * ⁶)
Schwellspannung V_{th} [V] * ¹)	-1.22	-1.48	-1.37	-1.67	-1.54
maximale Beweglichkeit μ_p [cm^2/Vs] bei $V_{\text{ds}} = -50 \text{ mV}$ * ¹)	230	225	225	220	225
Maximale relative Steilheit $g_m/I_d^{1/2}$ [$\text{A}^{1/2}/\text{V}$] bei $V_{\text{ds}} = -12 \text{ V}$ für $W_{\text{nom}} = 50 \mu\text{m}$ und $L_{\text{nom}} = 10 \mu\text{m}$ * ¹)	$1.0 \cdot 10^{-2}$	$1.0 \cdot 10^{-2}$	$1.0 \cdot 10^{-2}$	$1.0 \cdot 10^{-2}$	$1.0 \cdot 10^{-2}$
Minimaler relativer Aus- gangsleitwert g_o/I_d [$1/\text{V}$] bei $V_{\text{ds}} = -12 \text{ V}$ für $W_{\text{nom}} = 50 \mu\text{m}$ und $L_{\text{nom}} = 10 \mu\text{m}$ * ¹)	$7.0 \cdot 10^{-2}$	$3.5 \cdot 10^{-2}$	$7.0 \cdot 10^{-2}$	$3.5 \cdot 10^{-2}$	$3.2 \cdot 10^{-2}$
Kanaldotierung an der Oberfläche N_D [$1/\text{cm}^3$] * ²)	$1.1 \cdot 10^{15}$	$1.8 \cdot 10^{15}$	$2.5 \cdot 10^{15}$	$3.9 \cdot 10^{15}$	$3.7 \cdot 10^{15}$
Substratkonstante γ [$\text{V}^{1/2}$] * ¹)	0.29	0.58	0.36	0.74	0.64
Steigung im Unterschwel- lenbereich S [mV/dec] * ¹)	95	96	* ³)	* ³)	98

*¹) aufgrund von elektrischen Messungen ermittelt

*²) aufgrund von SUPREM-Simulationen ermittelt

*³) aufgrund von Leckströmen nicht bestimmbar

*⁴) zu den obigen Kennlinien gehörige Werte

*⁵) entspricht dem PMOS der Variante A

*⁴), *⁵) und *⁶) zu den Profilen in Bild 2.15 gehörige Werte

Anmerkung: Die aus der Substratkonstanten γ bestimmbare Wannendotierung N_D weicht geringfügig von den durch die SUPREM-Simulation ermittelten Werten ab, da das meßtechnische Verfahren jeweils die unterhalb der Verarmungszone befindliche Dotierung bestimmt und nicht die Oberflächendotierung. Während für die Bestimmung der Schwellspannung V_{th} die meßtechnisch ermittelte Größe relevant ist, wird für den Vergleich auf die - durch den Pile-Up-Effekt des Phosphors erhöhte - rechnerisch bestimmte (vgl. Bild 2.15) Oberflächenkonzentration zurückgegriffen.

Der nicht zusätzlich aufgeführte PMOS der Variante C₂, die als Basis des Integrationsprozesses benutzt wurde, entspricht sowohl prozeßtechnisch als auch in den elektrischen Eigenschaften dem PMOS der hier angeführten Variante B mit $N_{\square} = 6.7 \cdot 10^{11} / \text{cm}^2$.

2.1.4 Einfluß einer abschließenden Temperung in wasserstoffhaltiger Atmosphäre

Die nachfolgend beschriebene Temperung in wasserstoffhaltiger Atmosphäre wurde in den Instituts-CMOS-Prozeß und in die Prozesse auf hochohmigem Silizium integriert, um die Schwankungsbreite der Schwellspannung deutlich zu verringern. Durch eine bessere Konstanz dieser Größe wird die Vergleichbarkeit der Transistoren untereinander deutlich erhöht.

Ausgangspunkt war hierzu die als Teil des Metallisierungsprozesses eingesetzte Temperung in reinem Stickstoff.

Diese Temperung bei Temperaturen von etwa 400 - 480 °C dient zur Bildung ohmscher Kontakte mittels Aluminium. Das Metall wird hierbei mit den darunterliegenden hochdotierten p⁺ oder n⁺-Gebieten anlegiert [2.18].

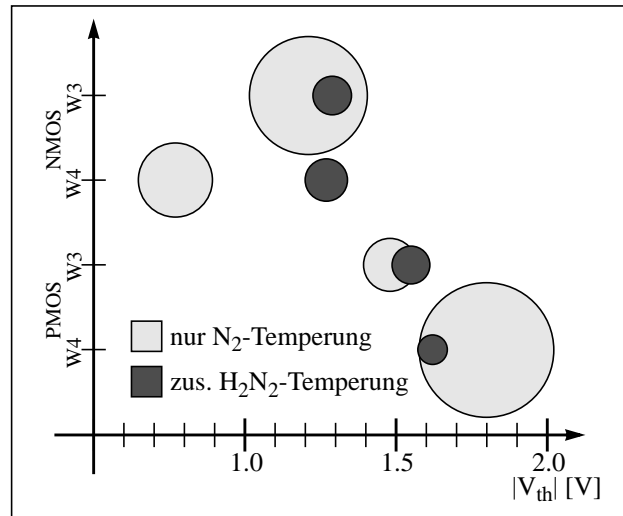


Bild 2.20: Darstellung der Verteilung der Schwellspannungen V_{th} von NMOS- und PMOS-Transistoren [2.19] vor bzw. nach einer zusätzlichen Temperung in 10 %iger H_2 -Atmosphäre ($t = 30 \text{ min}$, $T = 450 \text{ °C}$). Die Radien der Kreise sind proportional zur Standardabweichung σ bezogen auf den Mittelwert μ , die Mittelpunkte der Kreise geben in horizontaler Richtung den Mittelwert des Betrags der Spannungen an. Die Untersuchung wurde an den beiden Wafern W_3 und W_4 durchgeführt.

Neben einer Reduktion der „native oxide“-Schicht während eines solchen Temperschritts erfolgt ferner eine Verbesserung der Aluminium-Hafteigenschaften auf Silizium und Oxid. Aluminium, das als Gateelektrode das Gateoxid abdeckt, verhilft während dieses Schritts zu einer Stabilisierung der MOSFET-Eigenschaften, indem es zu einer teilweisen Absättigung der umladbaren Grenzflächenzustände D_{it} führt. Das von Deal et al. [2.20] eingeführte Modell besagt, daß sich in thermischen Oxiden Reste von Wasser befinden, selbst wenn diese mittels trockener Oxidation hergestellt werden. Wassermoleküle spalten sich während des Tempervorgangs durch chemische Reaktion mit Aluminium unter Bildung von Aluminiumoxid Al_2O_3 auf und setzen elementaren Wasserstoff frei. Ein Teil des Wasserstoffs diffundiert an die Grenzfläche zwischen Halbleiter und Oxid und kann dort durch Anlagerung an nicht-abgesättigte Siliziumatome zu einer Reduzierung der Grenzflächenzustände (engl. interface traps) führen.

Eine deutlich verbesserte Reduktion der Dichte dieser Interface Traps läßt sich durch eine abschließende Temperung in wasserstoffhaltiger Atmosphäre erreichen. Hierzu eignet sich eine N_2H_2 -Mischung mit etwa 10 % Wasserstoffanteil.

TABELLE 7. Schwellspannungen - Einfluß der N_2H_2 -Temperung

	Charge 152/ Wafer 3 NMOS	Charge 152/ Wafer 3 PMOS	Charge 152/ Wafer 4 NMOS	Charge 152/ Wafer 4 PMOS
Mittelwert μ der Schwellspannung V_{th} [V] nach N_2 -Temperung (15 min., 480 °C)	1.21	-1.48	0.77	-1.80
Standardabweichung σ der Schwellspannung V_{th} [mV] nach N_2 -Temperung (15 min., 480 °C)	188	104	75	321
relative Standardabweichung $\sigma/ \mu $ der Schwellspannung V_{th} [%] nach N_2 -Temperung (15 min., 480 °C)	156	70	98	178
Mittelwert μ der Schwellspannung V_{th} [V] nach N_2H_2 -Temperung (30 min., 450 °C)	1.29	-1.55	1.27	-1.62

TABELLE 7. Schwellspannungen - Einfluß der N₂H₂-Temperung

	Charge 152/ Wafer 3 NMOS	Charge 152/ Wafer 3 PMOS	Charge 152/ Wafer 4 NMOS	Charge 152/ Wafer 4 PMOS
Standardabweichung σ der Schwellspannung V_{th} [mV] nach N ₂ H ₂ -Temperung (30 min., 450 °C)	65	77	71	63
relative Standardabweichung $\sigma/ \mu $ der Schwellspannung V_{th} [%] nach N ₂ H ₂ -Temperung (30 min., 450 °C)	51	50	56	39

Die Transistoren wurden nach der ersten konventionell wie bisher am Institut ausgeführten Temperung (N₂-Temperung, 480 °C, 15 min.) einmal und nach der zusätzlichen N₂H₂-Temperung erneut vermessen. Der Vergleich zeigt eine deutliche Verringerung der Schwankungen der Schwellspannung V_{th} wie in Bild 2.20 und Tabelle 7 dargestellt. Neben einer geringeren Schwankungsbreite ist auch die Annäherung der Mittelwerte beider Wafer durch die Temperung deutlich zu erkennen. Die verbleibenden Schwankungen, ausgedrückt als Oxidladungsschwankungen, betragen für je eine Standardabweichung $\sigma_{V_{th}} = 63 \text{ mV} \dots 77 \text{ mV}$ gemäß $N_{ox} = Q_{ox}/e = C'_{ox} \sigma_{V_{th}}/e = 1.5 \cdot 10^{10} / \text{cm}^2 \dots 1.8 \cdot 10^{10} / \text{cm}^2$. Diese Werte sind als recht günstig zu beurteilen, führen jedoch aufgrund der großen Oxiddicke, hier $t_{ox} = 940 \text{ \AA}$, zu deutlich meßbaren Schwellspannungsschwankungen.

2.2 MOS-Varaktoren

Die Untersuchung des grundlegenden Wirkungsprinzips von MOS, d.h. der Feldeffekt durch Aufbau einer Inversionsschicht, erfolgt üblicherweise an einfachen Strukturen, den sog. MOS-Varaktoren. Diese Strukturen dienen neben der Untersuchung von Oxidqualität und -dicke auch der Charakterisierung der Si-SiO₂-Grenzfläche. Die Vorgehensweise bei der Ermittlung und Analyse von Kapazitäts-Spannungs-Kennlinien, den sog. CV-Kurven, ist im Anhang A3 erläutert. In diesem Abschnitt sind die Ergebnisse der Untersuchungen, die zur weiteren Charakterisierung der Güte der im vorherigen Abschnitt beschriebenen MOSFETs dienen, dargestellt.

2.2.1 Dielektrische Integrität

Messungen zur dielektrischen Integrität umfassen u.a. Messungen der maximalen Durchbruchfeldstärke des Oxides sowie der maximal möglichen Gateladung. Verringerungen dieser Maximalfeldstärke können z.B. bei lokalen Dünnungen des Oxides oder bei schmalen Kanälen im Oxid, sog. pinholes, auftreten.

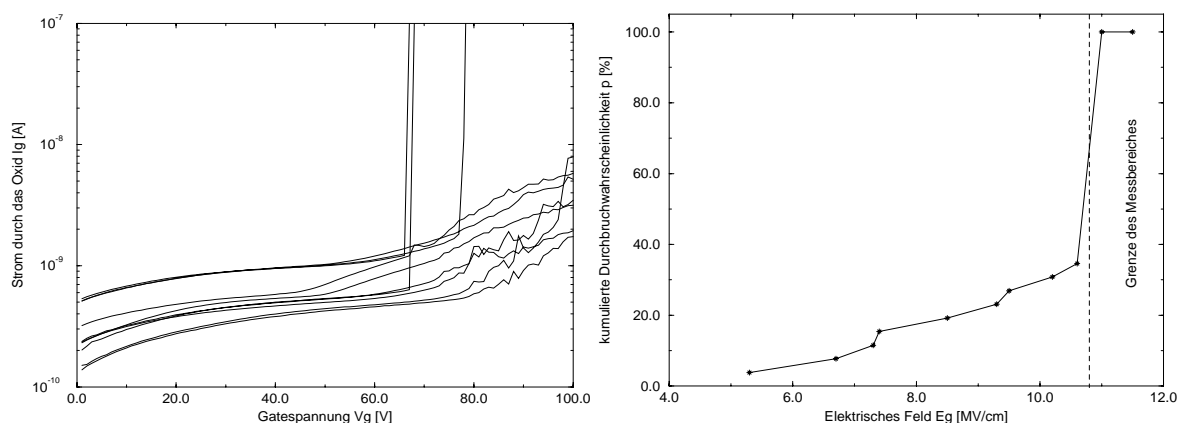


Bild 2.21: Messung der dielektrischen Integrität des Gateoxides durch Maximal-Feldstärkenbestimmung. Die Scheibe weist eine mittlere Oxiddicke $t_{ox} = 923 \text{ \AA}$ auf. Links: Verlauf von typischen Strom-Spannungskennlinien. Rechts: Kumulierte Wahrscheinlichkeit der maximal möglichen Feldstärke. Nur 34,8 % der untersuchten Varaktoren zeigen unterhalb von $V_g = 100 \text{ V}$ (entsprechend $E = V_g/t_{ox} = 10,8 \text{ MV/cm}$) Durchbrucherscheinungen.

Bei thermischer Oxidation, insbesondere in sog. trockener Atmosphäre mit O₂-Spülung und bei relativ dicken

Gateoxiden ist diesbezüglich jedoch nicht mit Problemen zu rechnen. Untersuchungen der Gateoxide des MOS-Prozesses sind durchgeführt worden und zeigen mit nur wenigen Ausnahmen die aus der Literatur [2.21] bekannten Werte, d.h. Durchbrucheffekte erst oberhalb der kritischen Feldstärke von thermisch gewachsenem Siliziumoxid $E_{\text{crit}} \approx 10 \text{ MV/cm}$. Der Anteil der weniger spannungsfesten Oxide ist durch Rauigkeit der Oxidoberfläche, bewirkt durch die Aluminiumtemperung, bedingt. Hierdurch kann es zu lokal unterschiedlichen Oxiddicken kommen. Diese Effekte lassen sich durch den Einsatz von Polysiliziumgates oder mittels Quecksilbergates vermeiden. Bei der letztgenannten Methode wird temporär eine Quecksilber-Gateelektrode gebildet, wobei die unter Druck stehende Flüssigkeit in einer isolierenden Kapillare geführt wird.

2.2.2 Inversionskanäle auf hochohmigem p-Silizium

Bei der Anwendung konventioneller CV-Meßverfahren auf MOS-Varaktoren, die auf hochohmigem p⁻-Substrat gefertigt wurden, ergaben sich zwei grundlegende Schwierigkeiten:

- (i) Aufgrund des hohen spezifischen Widerstands des Substratmaterials resultieren Serienwiderstände (vgl. Anhang A3.3), die aussagekräftige Hochfrequenzmessungen prinzipiell unmöglich machen. Eine im Rahmen dieser Arbeit entwickelte Methode zur Überwindung dieses Problems ist in Abs. 2.2.3 dargestellt.
- (ii) Eine auch durch sehr geringe Oxidladungsdichten invertierte Oberfläche des p⁻-Substrats verhindert Messungen im Nichtgleichgewicht bei Raumtemperatur. Aufgrund dieser Inversionsschicht an der Oberfläche ist unabhängig vom Betriebszustand der Varaktorstruktur jeweils ein großes Reservoir von Minoritätsträgern, d.h. hier Elektronen, in unmittelbarer Nähe des Varaktors vorhanden. Diese Minoritätsträger, die sich außerhalb der eigentlichen Varaktorstruktur befinden, gleichen nun einen Mangel an Minoritätsträger unterhalb der Gateelektrode in kürzester Zeit aus, so daß die Varaktorstruktur auch bei Einsatz gepulster CV-Meßverfahren sofort in einen Gleichgewichtszustand übergeht. In diesem Abschnitt wird ein Versuchslayout zur Isolation dieser Inversionsschicht von der Varaktorstruktur vorgestellt und eine rechnerische Bestimmung der zur Inversion der Oberfläche notwendigen Oxidladungskonzentration vorgenommen.

Versuchslayout zur Unterdrückung der Ankopplung der Inversionsschicht an den MOS-Varaktor

Eine permanente Inversionsschicht unterhalb des Oxides wirkt in erster Näherung wie eine der unterhalb der kreisförmigen Elektrode befindlichen Halbleiterkapazität $C_{\text{Si,o}} = C_{\text{RLZ}} + C_{\text{inv}} + C_{\text{it}}$ parallel geschaltete Zusatzkapazität C_{perm} . Exakt beschrieben wirken die Anteile dieser Inversionsschicht als RC-Netzwerk, wobei lateral eine resistive Komponente R in Form der Inversionsschicht die vertikal verlaufenden Kapazitäten C der Verarmungszone verbindet. Aufgrund des sehr niederohmigen Verhaltens der Inversionsschicht kann der Einfluß der Widerstandskomponenten in erster Näherung vernachlässigt werden, so daß sich der Anteil C_{perm} direkt zur Halbleiterkapazität addiert.

Eine Entkopplung dieser Inversionsschicht vom eigentlichen MOS-Varaktor läßt sich im Verarmungsfall aufgrund eines dann nicht mehr zu vernachlässigenden Kopplungswiderstands der Verarmungszone unterhalb des Varaktors annähernd erreichen. Ferner kann die laterale Isolierung der Struktur mit Hilfe einer Anreicherungsschicht außerhalb des Varaktors erreicht werden. Im Fall von p-Silizium eignet sich hierzu die in Bild 2.22 dargestellte Anordnung, in der mit Hilfe eines konzentrischen Ringes, der auf negativer Vorspannung zum Substrat gehalten wird, eine Anreicherungsschicht geschaffen wird.

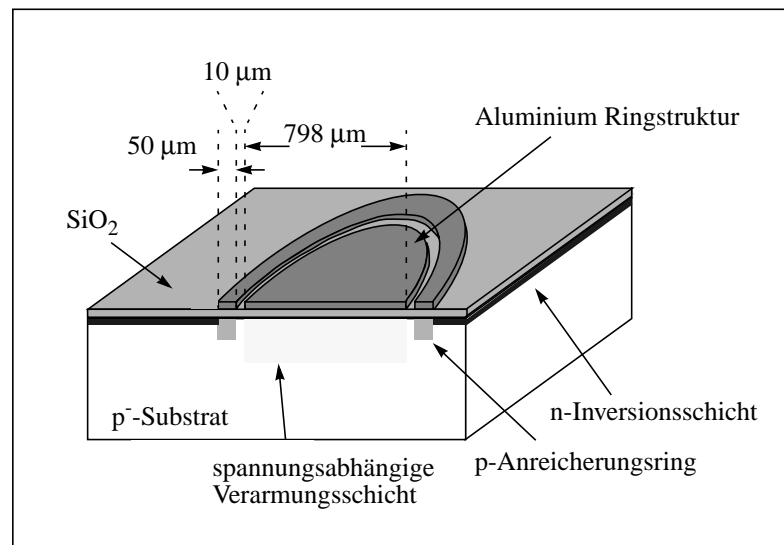


Bild 2.22: Vermeidung der lateralen Ankopplung an die permanente Inversionsschicht durch eine Akkumulationsschicht außerhalb des MOS-Varaktors.

Durch die angelegte Ringspannung V_{ring} läßt sich nun die Stärke der Anreicherung einstellen, so daß eine teilweise oder komplette Isolierung erfolgen kann. Bei Annahme einer Parallelschaltung ergibt sich damit gemäß

$$C_{Si, \text{gesamt}} = C_{Si, o} + C_{perm}(V_{\text{Ring}}) \quad (2.15)$$

die von der Ringspannung abhängige Gesamtkapazität $C_{Si, \text{gesamt}}$ des Halbleiters als Summe der Halbleiterkapazität unterhalb der kreisförmigen Elektrode $C_{Si, o}$ und des in Abhängigkeit der Ringspannung V_{Ring} angekoppelten Anteils der permanenten Inversionskapazität C_{perm} . Eine Messung unter Hochfrequenzbedingungen bei einer Frequenz von $f = 1 \text{ kHz}$ erlaubt in diesem Fall noch eine recht genaue Messung der Kapazität im Inversionsbereich auf hochohmigem p-Substrat ($\rho = 5000 \Omega \text{cm}$), wengleich hiermit jedoch - aus den in Anhang A3 genannten Gründen - keine Bestimmung der Grenzflächenzustandsdichte möglich ist.

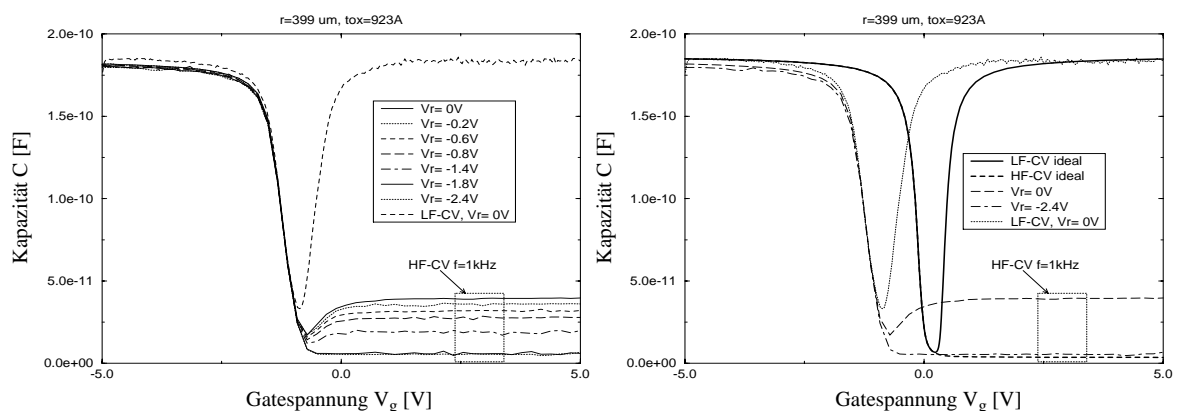


Bild 2.23: HF-CV-Messung bei $f = 1 \text{ kHz}$. Links: Die Messung des MOS-Varaktors auf hochohmigem FZ-Si (p-Typ, $\rho = 5000 \Omega \text{cm}$) zeigt Inversionskanäle unter dem Oxid. Rechts: Gute Übereinstimmung zwischen gemessenen und berechneten Kurven.

In Bild 2.23 links ist das Resultat dieser Messung bei unterschiedlichen Ringspannungen V_r dargestellt. Zum Vergleich dient eine quasistatisch aufgenommene Niederfrequenzkurve (LF-CV). Die gute Übereinstimmung der Kapazitätswerte im Anreicherungsbereich, die den theoretisch berechneten Wert der Oxidkapazität

$C_{ox} = 181 \text{ pF}$ bei einer Gateoxiddicke $t_{ox} = 923 \text{ \AA}$ erreichen, zeigt die gute Meßgenauigkeit auch bei $f = 1 \text{ kHz}$. Rechts im Bild ist der Vergleich zwischen den theoretisch berechneten und den gemessenen CV-Kurven dargestellt. Hier ist wiederum nur die erwartete Dehnung entlang der Spannungsachse, „stretch-out“ aufgrund von Grenzflächenzuständen und die aus Austrittsarbeitsdifferenz und Oxidladungsdichte ($\Delta V_{FB} = 1.1 \text{ V}$ entsprechend $Q_{ox}/e = 8 \cdot 10^{10} / \text{cm}^2$) zu erklärende Verschiebung beider Kurven in Richtung negativer Spannungen zu beobachten. Da in der idealisierten Ableitung die Existenz von Grenzflächenzuständen vernachlässigt wird, erreicht die rechnerisch bestimmte Niederfrequenzkapazität kleinere Minimalwerte der Kapazität als real gemessen werden.

Die bei den unterschiedlichen Ringspannungen beobachtete Inversionskapazität nimmt mit steigendem Betrag der negativen Ringspannung ab und erreicht schließlich den theoretisch erwarteten Minimalwert für Ringspannungen $V_{ring} < -1.8 \text{ V}$. Ein Vergleich mit der CV-Kurve in Bild 2.23 zeigt, daß sich die Ring-Struktur bei diesen Biaswerten, d.h. $V_g = V_{Ring} < -1.8 \text{ V}$, im Bereich der Verarmung bzw. für stärker negative Werte in der Anreicherung befindet.

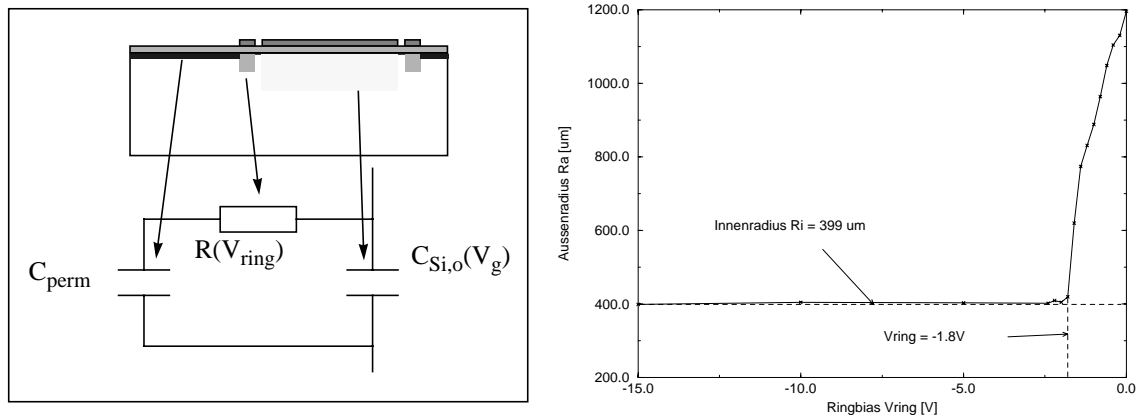


Bild 2.24: Links: Modell der spannungsabhängigen Zusatzkapazität. Rechts: Die Darstellung des effektiven Inversionsschichtradius zeigt die Entkopplung des MOS-Varaktors von der permanenten Inversionsschicht durch Schaffung einer Verarmungs- bzw. Anreicherungszone.

Die hochohmige Zone unterhalb des Ringes entkoppelt damit die Inversionsschicht außerhalb der Struktur vom eigentlichen MOS-Varaktor. Bei vollständiger Unterdrückung der Ankopplung der äußeren Inversionsschicht, d.h. $C_{perm}(V_{Ring}) = 0$, ergibt sich die Kapazität des Inversionsgebiets unterhalb des MOS-Varaktors. Mit $C_{ox} = 181 \text{ pF}$ und einer Gesamtkapazität $C = 5,44 \text{ pF}$ folgt aus Gleichung (2.16) für diese Kapazität $C_{inv} = C_{Si} = 5,61 \text{ pF}$. Der Radius der untersuchten Struktur beträgt $r_{varaktor} = 399 \mu\text{m}$, somit läßt sich eine flächenbezogene Inversionskapazität $C_{inv}/(\pi \cdot (r_{varaktor})^2)$ berechnen. Die bei Inversion unterhalb der MOS-Varaktorstruktur vorhandene Kapazität C_{Si} ist von einem etwaigen Ringbias unabhängig, so daß ein Ansteigen der Gesamtkapazität nur in Folge der teilweisen Ankopplung der lateralen Inversionsschicht erfolgen kann. Es läßt sich damit ein effektiver Radius r_{eff} definieren, der bei als konstant angenommener flächenbezogener Inversionskapazität ein Maß für die Beeinflussung darstellt. Mit der Gesamtkapazität C der verkoppelten CV-Struktur gemäß

$$\frac{1}{C} = \frac{1}{C_{ox}} + \frac{1}{C_{Si} + C_{perm}(V_{Ring})} \quad (2.16)$$

ergibt sich unter Berücksichtigung der Flächendefinition mit

$$C_{perm} = \frac{C_{inv}}{\pi r_{varaktor}^2} \pi (r_{eff}^2 - r_{varaktor}^2) \quad (2.17)$$

eine Beziehung für den effektiven Radius r_{eff} . Dieser ist rechts in Bild 2.24 dargestellt. Bei einer Ringspannung von $V_{Ring} = 0 \text{ V}$ ergibt sich ein Radius $r_{eff} = 1,2 \text{ mm}$, der die großflächige elektrische Beeinflussung durch die permanente Inversionsschicht verdeutlicht.

Berechnung der zur Oberflächeninversion führenden Oxidladungsdichte

Zur Bildung einer permanenten Inversionsschicht kommt es, wenn durch im Oxid befindliche Ladungen eine Bandverbiegung an der Oberfläche erreicht wird, die dem Beginn der schwachen (englisch: weak inversion, w.i.) bzw. starken Inversion (englisch: strong inversion, s.i.), d.h. einer Bandverbiegung von $\psi_{S,w.i.} = |E_{fi}|/e$ bzw. $\psi_{S,s.i.} = 2 |E_{fi}|/e$, entspricht.

Für ein p-Substrat gilt folgende Potentialverteilung ohne äußere angelegte Spannung:

$$\psi_S + \frac{E_{ms}}{e} = \frac{Q_f + Q_B}{C'_{ox}}, \quad (2.18)$$

d.h. die Summe aus Bandverbiegung ψ_S und Differenz der Austrittsarbeiten E_{ms} ist gleich dem Potentialabfall über dem Oxid und der Verarmungszone. Bei Beginn der schwachen Inversion ist die Oberfläche intrinsisch und es gilt $\psi_{S,w.i.} = |E_{fi}|/e$. Mit $E_{fi} = -kT \ln(N_A/n_i)$ und $E_{ms} = e\phi_{Alu} - (\chi_{Si} + E_{gap}/2 + |E_{fi}|)$ folgt

$$Q_f = \frac{C'_{ox}}{e} (e\phi_{Alu} - (\chi_{Si} + E_{gap}/2)) - Q_B \quad (2.19)$$

für die Flächenladungsdichte Q_f , die zur intrinsischen Oberfläche führt. Hierbei ist

$$Q_B = -\sqrt{2\epsilon_o\epsilon_{Si}eN_A \left| \frac{E_{fi}}{e} \right|} \quad (2.20)$$

die Ladungsdichte in der Verarmungszone.

Der Fall der intrinsischen Oberfläche und der Beginn der starken Inversion ist in Bild 2.25 dargestellt. Berechnet wurden diese Kurven für die beim Instituts-CMOS-Prozeß üblichen Oxiddicken mit darüberliegendem Aluminium. Im Fall eines 5000 Ωcm ($N_A = 4 \cdot 10^{12} / \text{cm}^3$) p-Substrats wird starke Inversion selbst bei extrem geringen Oxidladungsdichten $Q_f/e = 1 \cdot 10^9 / \text{cm}^2$ und dicken Oxiden erreicht. Solch niedrige Ladungsdichten sind jedoch kaum zu erreichen; insbesondere nicht bei dicken Feldoxiden, die durch feuchte Oxidation erzeugt werden. Folglich kann sicher davon ausgegangen werden, daß unterhalb von spannungsfreien Aluminiumleiterbahnen auf hochohmigem Silizium permanente Inversionskanäle existieren.

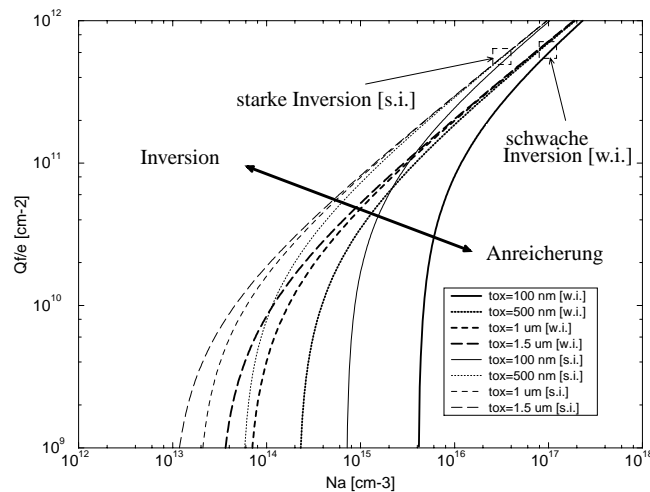


Bild 2.25: Darstellung der minimalen positiven Oxidladung Q_f , die zur Inversion einer p-Substratoberfläche bei verschiedenen Oxiddicken im Fall einer Aluminiummetallisierung führt. Berechnet für $e\phi_{Alu} = 3.95 \text{ eV}$ und $\chi_{Si} = 4.0 \text{ eV}$. [w.i.] steht für schwache (engl. weak) Inversion und [s.i.] für starke (engl. strong) Inversion der Oberfläche.

Die Wirkung der so gebildeten Oberflächenkanäle besteht neben einer Erhöhung der Dämpfung von Hochfrequenzsignalen auch in der Schaffung einer leitfähigen Verbindung zwischen benachbarten n-Wannen, wenn deren räumliche Trennung durch eine metallische Fläche, wie etwa eine Leiterbahn, überlappt wird (siehe Bild 2.26). Bei unterschiedlichen Wannenpotentialen kommt es so zu unerwünschten Leckströmen. Abhilfe bietet eine Feldimplantation, die diese Bereiche stärker p-dotiert.

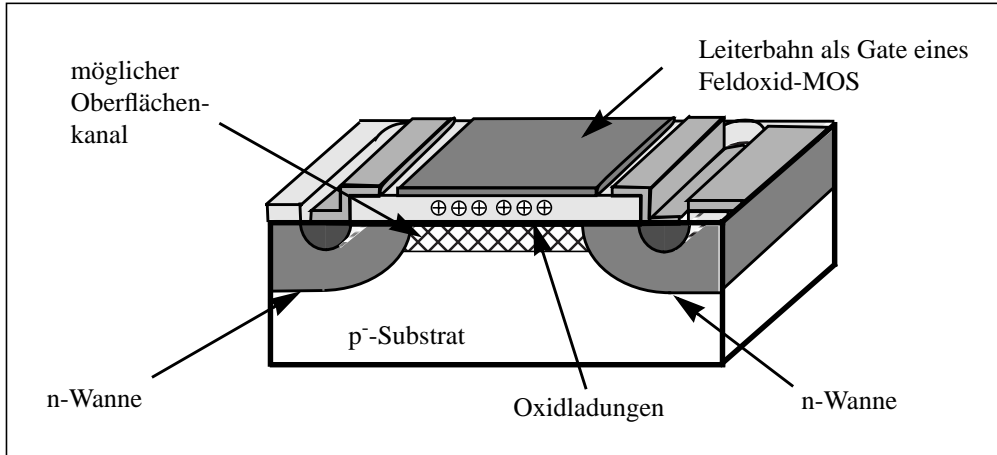


Bild 2.26: Unterhalb der Aluminium-Leiterbahn erfolgt eine Inversion der p-Bereiche, die eine elektrische Verbindung zwischen den n-Wannen darstellt

Bedeutend ist ferner der Fall eines Oxides auf der Siliziumoberfläche ohne ein darüberliegendes Metall. Dieser Fall tritt bei SIMMWIC-Schaltungen außerhalb der Wellenleiterstrukturen auf und sorgt für zusätzliche Dämpfung bei Existenz von Inversionsschichten. Das Ladungsgleichgewicht kann hier nur aus den Oxidladungen und der Ladung in der Verarmungszone bestehen [2.22], so daß

$$Q_f = -Q_B = \sqrt{2\epsilon_o\epsilon_{Si}eN_A\left|\frac{E_{fi}}{e}\right|} \quad (2.21)$$

gelten muß. Die gemäß einer Wurzelfunktion verlaufende minimale Oxidladungsdichte stellt sich im doppelt logarithmischen Maßstab als Gerade der Steigung 1/2 dar (vgl. Bild 2.27).

Eine Abhängigkeit von der Dicke der Oxidschicht existiert bei der gewählten Annahme einer Flächenladung an der Si-SiO₂-Grenzfläche nicht. Abweichungen im Inversionsverhalten bzgl. einer mit Aluminium abgedeckten Oxidschicht ergeben sich für niedrigere Substratdotierungen. Etwas höhere Oxidladungsdichten sind nun erforderlich, um Inversion zu erreichen. Bei N_A = 4*10¹²/cm³ p-Substrat führt eine Flächenladungsdichte Q_f/e = 2.7*10⁹/cm² zur schwachen und Q_f/e = 4.0*10⁹/cm² zur starken Inversion der Oberfläche. Für dicke Oxide und zunehmende Substratdotierungen verlaufen die Kurven der mit Aluminium bedeckten Oxide (vgl. Bild 2.25) asymptotisch gegen jene ohne Metallisierung. Die aus der Differenz der Austrittsarbeiten von Aluminium und p-Silizium resultierende Verstärkung der Inversion wirkt sich bei hohen Dotierungen erwartungsgemäß nur noch sehr schwach aus.

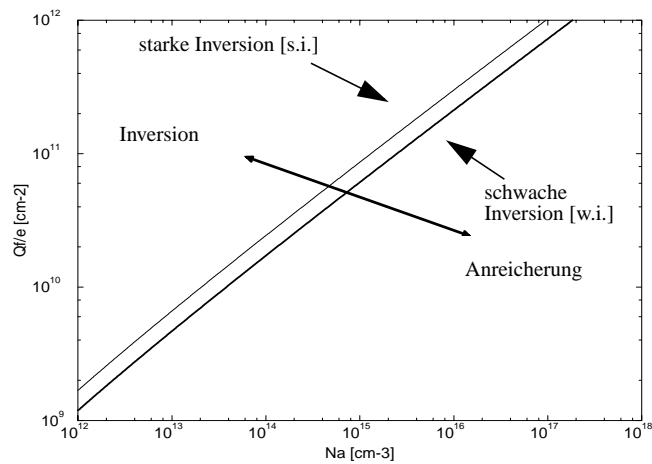


Bild 2.27: Darstellung der minimalen positiven Oxidladung Q_f , die zur Inversion einer p^- -Substratoberfläche unterhalb einer beliebig dicken Oxidschicht bei Annahme einer an der Grenzfläche liegenden positiven Oxidladung führt. Die Kurven sind für $e\phi_{\text{Au}} = 3.95 \text{ eV}$ und $\chi_{\text{Si}} = 4.0 \text{ eV}$ berechnet worden.

Inversion bei Goldmetallisierung

Aufgrund der größeren Austrittsarbeit von Gold, $e\phi_{\text{Au}} = 5.06 \text{ eV}$, erfolgt die Inversion der p^- -Substratoberfläche erst bei deutlich höheren Konzentrationen an positiven Ladungen im Oxid. Erst dies erlaubt die Realisierung von verlustarmen HF-Wellenleiterstrukturen.

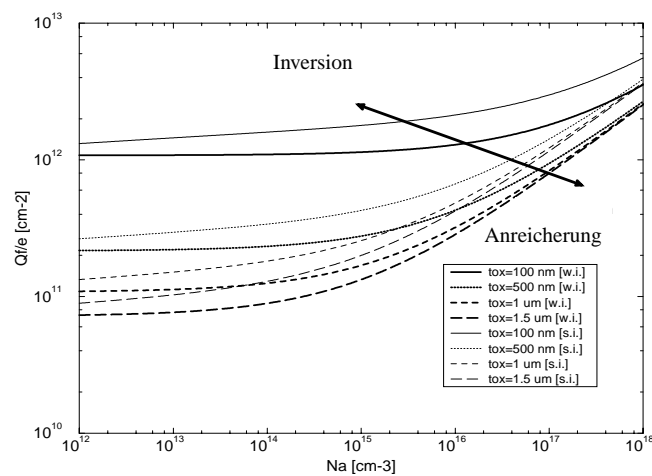


Bild 2.28: Darstellung der minimalen positiven Oxidladung Q_f , die zur Inversion einer p^- -Substratoberfläche bei verschiedenen Oxiddicken im Fall einer Goldmetallisierung führt. Berechnet für $e\phi_{\text{Au}} = 5.06 \text{ eV}$ und $\chi_{\text{Si}} = 4.0 \text{ eV}$. [w.i.] steht für schwache (engl. weak) Inversion und [s.i.] für starke (engl. strong) Inversion der Oberfläche.

In Bild 2.28 sind die notwendigen Ladungskonzentrationen Q_f dargestellt, die zur schwachen oder starken Inversion führen. Es zeigt sich eine nur geringe Abhängigkeit dieser Größe von der Dotierkonzentration unterhalb von $N_A = 10^{14} / \text{cm}^3$, so daß auch für sehr schwach dotierte Substrate bei Oxidladungskonzentration kleiner als $Q_f/e = 7 \cdot 10^{10} / \text{cm}^2$ eine Inversion der Oberfläche vermieden werden kann. Diese Konzentration kann durch eine sorgfältige Prozeßführung in einer üblichen CMOS-Linie erreicht werden und liegt in der Größenordnung der bei den am Institut prozessierten Wafern gefundenen Ladungsdichten.

2.2.3 CV-Messungen bei Tieftemperatur

Obwohl es möglich ist, bei mittleren Frequenzen von $f = 1$ kHz und Unterdrückung der Ankopplung der permanenten Inversionsschicht an den MOS-Varaktor die Hochfrequenz-CV-Kurve recht genau zu messen, eignet sich diese Methode aus zwei Gründen nicht für die Bestimmung der Grenzflächenzustandsdichte des hochohmigen Siliziums.

(i) Die Messung ist trotz langer Integrationszeiten und damit weitgehender Reduktion des Meßrauschens nicht genügend präzise. Wie in Anhang A3.2 dargestellt, basiert die Datenextraktion aus der numerischen Differenzbildung der Meßkurven unter Hoch- und Niederfrequenzbedingungen. Hierbei sind Differenzen in der Größenordnung von wenigen Picofarad relevant, so daß im Bereich der Verarmung, d. h. insbesondere bei kleinen Bandverbiegungen, extreme Anforderungen an die Meßgenauigkeit gestellt werden müssen.

(ii) Bei Messungen mit $f = 1$ kHz kann nicht davon ausgegangen werden, daß die Grenzflächenzustände in ihrem Ladezustand nicht dem Wechselsignal folgen können. Die Folge ist eine Annäherung der Hochfrequenzkurve an die Niederfrequenzkurve und damit eine Unterschätzung der Grenzflächenzustandsdichte.

Als mögliche Lösung soll hier die quasistatische Messung beider Kurven vorgestellt werden. Die Messung der Niederfrequenzkurve bereitet keine grundsätzlichen Schwierigkeiten und erfolgt bei Raumtemperatur. Eine Hochfrequenzkurve zeichnet sich dadurch aus, daß weder Minoritätsträger noch mögliche Umladungen der Grenzflächenzustände einen kapazitiven Beitrag liefern. Diese Bedingung wird bei Raumtemperatur durch Meßfrequenzen um etwa $f = 1$ MHz recht gut eingehalten. Bezüglich der Gategleichspannung sind jedoch sowohl der Ladezustand der Interface Traps als auch die Konzentration der Minoritätsträger im Gleichgewicht.

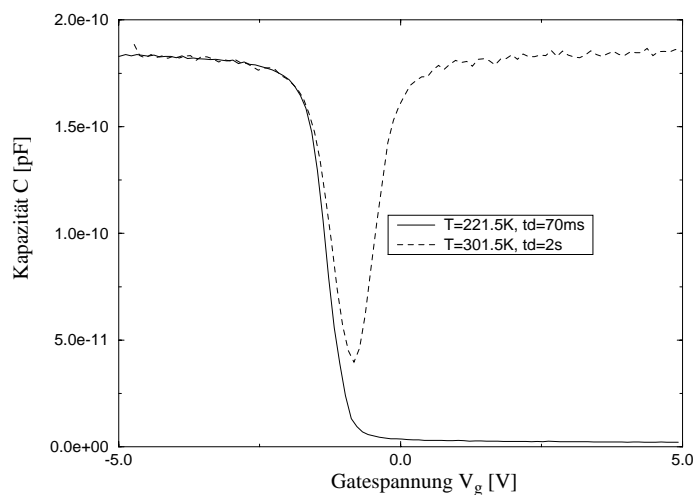


Bild 2.29: Bei Raumtemperatur $T = 28.5$ °C = 301.5 K und bei $T = -51.5$ °C = 221.5 K aufgenommene quasistatische CV-Kurven zur Bestimmung der Grenzflächenzustandsdichte in hochohmigem Silizium ($\langle 100 \rangle$ -orientiertes p^- -Substrat mit $\rho = 5000$ Ω cm). Der Varaktor hat einen Radius $r = 399$ μ m und eine Oxiddicke $t_{ox} = 923$ Å. Die Verzögerungszeiten t_d sind gerätespezifisch, jedoch auch ein Maß für die Geschwindigkeit der Gatebiasänderung.

Bei hochohmigen Substraten ist eine Erhöhung der Meßfrequenz aufgrund des großen Serienwiderstands problematisch und führt zu deutlichen Meßfehlern. Eine Unterdrückung der zyklischen Generation und Rekombination der Minoritätsträger sowie der Reaktion der Grenzflächenzustände ist jedoch auch durch eine Vergrößerung der jeweils relevanten Zeitkonstanten möglich. Diese Zeitkonstanten hängen ihrerseits von der Konzentration der Ladungsträger ab, wobei für letztere $n^*p = n_i^2$ im Gleichgewicht gilt. Die intrinsische Trägerdichte n_i ist eine stark monoton steigende Funktion der Temperatur, so daß bei Abkühlung des Substrats eine deutliche Reduktion der Trägerkonzentration erreichbar ist. Durch diesen Effekt läßt sich auch mit quasistatischen Meßmethoden eine Hochfrequenzmeßkurve ermitteln.

Bild 2.29 zeigt das Resultat einer solchen Messung. Die Niederfrequenzkurve wurde bei $T = 28.5$ °C aufgenommen, die Hochfrequenzkurve bei $T = -51.5$ °C. Als Kühlmittel wurde flüssiger Stickstoff eingesetzt. Die intrinsische Trägerdichte bei der tiefen Temperatur liegt um den Faktor $n_i(-51.5$ °C)/ $n_i(28.5$ °C) = $2.7 \cdot 10^6 / 1.7 \cdot 10^{10} = 1/6300$ niedriger als bei $T = 28.5$ °C.

Messungen bei noch tieferen Temperaturen bis $T = -80$ °C haben keine merkliche Veränderung der Charakteristik der CV-Kurve gezeigt, so daß schon bei $T = -51.5$ °C eine gute Annäherung an die Hochfrequenzbedingung

erreicht wird. Aufgrund eines weitgehenden „Einfrierens“ der Grenzflächenzustände, erfolgt die Umladung derselben bei Änderung der Gategleichspannung ebenfalls nur stark reduziert, d.h. die gemessene Kurve kann insofern Fehler aufweisen, da auch der Dehnungseffekt („stretch-out“) reduziert wird. Dieser Einfluß führt jedoch allenfalls zu einer leichten Überschätzung der Dichte der Grenzflächenzustände, so daß durch die beschriebene Meßmethode eine konservative Obergrenze ermittelt wird.

In Bild 2.30 ist das Ergebnis einer Messung auf hochohmigem Silizium (rechts im Bild), solchen auf Standard p-Substrat gegenübergestellt. Die feinere Gatespannungsabstufung im linken Fall wird mit erhöhtem Rauschen der Ergebniswerte erkauft.

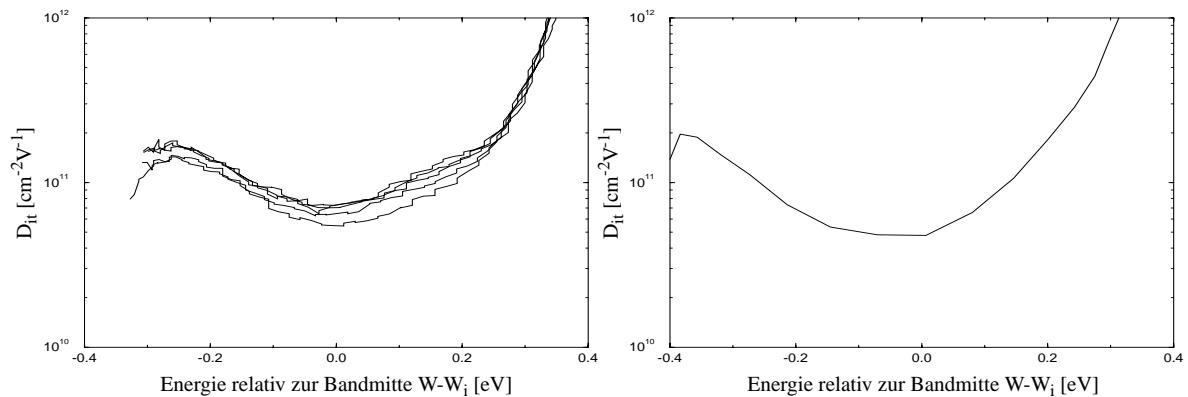


Bild 2.30: Darstellung der Grenzflächenzustandsdichte D_{it} in der Bandlücke. Links Czochralski-Silizium ((100), Bor, 2 Ωcm), rechts Floatzone-Silizium ((100), Bor, 5000 Ωcm).

Vergleichend zeigt sich ein ähnlicher Verlauf der Dichte der Grenzflächenzustände in der Bandlücke. Das Minimum liegt in der Bandmitte und zum Valenzband hin ist der durch die Annäherung der beiden CV-Kurven in der Anreicherung bedingte Abfall der Konzentration ebenso zu erkennen wie das scheinbare Ansteigen bei Annäherung an das Leitungsband. Letzterer Effekt ist durch das Einsetzen der Inversion zu erklären, die über die Beeinflussung der Niederfrequenzkurve zu einer großen Differenz der Kapazitätswerte führt, die nur zum Teil durch die Existenz von Grenzflächenzuständen bedingt ist.

TABELLE 8. Grenzflächenzustandsdichte

	CZ (100), Bor, 2 Ωcm	FZ (100), Bor, 5000 Ωcm
Mittelwert μ der Grenzflächenzustandsdichte D_{it} [$1/\text{cm}^2\text{eV}$]	$6.2 \cdot 10^{10}$	$5.2 \cdot 10^{10}$
Standardabweichung σ der Grenzflächenzustandsdichte D_{it} [$1/\text{cm}^2\text{eV}$] bei 5 Messungen	$0.85 \cdot 10^{10}$	$0.78 \cdot 10^{10}$

Die in Tabelle 8 angegebenen Werte der Grenzflächenzustandsdichte sind als sehr niedrig zu werten und belegen die gute Qualität des bei $T = 1000 \text{ }^\circ\text{C}$ in O_2 -Atmosphäre gewachsenen Gateoxides.

2.3 Source-Wannen-Dioden

Dioden als nicht-lineare gleichrichtende Zweipolelemente erfüllen eine Vielzahl von Funktionen in der Schaltungstechnik. Im Rahmen dieser Arbeit stehen jedoch weniger die schaltungstechnischen Eigenschaften im Vordergrund. Das Hauptaugenmerk liegt auf der Analyse der Source-Wannen-Dioden als Indikator von Unterschieden zwischen den CMOS-Prozessen auf niederohmigem und hochohmigem Silizium. Ferner wurden zur Schaltung von Hochfrequenzsignalpfaden bei $f = 90$ GHz benötigte pin-Dioden in den CMOS-Prozess integriert.

Die Source- und Drain-Kontakte eines MOS dienen primär der Injektion und Sammlung der den Kanal tragenden Majoritätsladungsträger. Darüber hinaus bilden sie mit ihrer im allgemeinen hohen Dotierung einen nahezu einseitig abrupten pn-Übergang mit der sie umgebenden Wanne oder dem Substrat. Im normalen Betrieb eines MOS sind diese Dioden spannungsfrei oder in Sperrichtung gepolt, da ein Stromfluß in diese Bereiche außerhalb des Kanals unerwünscht ist. Aufgrund der hohen Zahl von MOS in modernen CMOS-Schaltungen führen jedoch schon Sperrströme von einigen μA pro Bauelement zu insgesamt signifikanten Verlustströmen. Die Folge solcher Leckströme liegt im zusätzlichen Leistungsbedarf eines Chips und damit weiterer Erwärmung. Bei der Verarbeitung von Signalen mit sehr geringen Stromamplituden, wie dies an hochohmigen Schaltungspunkten der Fall ist, führen Sperrströme leicht zu einer Begrenzung des möglichen Dynamikbereichs nach unten oder auch zu einer Signalverfälschung. Die Empfindlichkeit der Diodensperrströme auf Änderungen der Prozeßführung, insbesondere bei Beeinträchtigungen der Reinheit, macht die Untersuchung derselben zu einem wichtigen Hilfsmittel.

Im Rahmen der Anpassung eines CMOS-Prozesses auf hochohmiges p^- -Substrat, bestand eine Nebenforderung darin, auch die Höhe der Sperrströme im Vergleich zum Standard-Prozeß auf n-Substrat nicht signifikant zu vergrößern. Diese Forderung ist jedoch auf die Sperrströme der Source- und Drain-Anschlüsse einzuschränken, da Sperrströme zwischen Wannen aufgrund des um drei Größenordnungen niedriger dotierten Substrats prinzipiell höher ausfallen müssen. Eine wesentliche Komponente des Sperrstroms, d.h. der durch Diffusionsprozesse bedingte Stromanteil, steigt proportional zur Dichte der Minoritätsträger an, und eben diese liegt bei p^- -Substrat mit einer Bordotierung von $N_A = 1.5 \cdot 10^{12} / \text{cm}^3$ um etwa den Faktor 1000 höher als bei dem n-Typ-Vergleichssubstrat (Phosphordotierung $N_D = 1.5 \cdot 10^{15} / \text{cm}^3$).

Im Gegensatz zu den Source- und Drainsperrströmen führen diese Wannenleckströme jedoch nur zu einer Erhöhung der DC-Leistungsaufnahme, da im Schaltungsentwurf die Wannenanschlüsse berührende Signalwege vermieden werden und somit keine Beeinträchtigungen der Signalqualität zu erwarten sind.

Die N^+p -Dioden des NMOS

Obwohl die zu betrachtenden Dioden auch laterale Einflüsse erwarten lassen, da die N^+ -Zone und der die p-Wanne kontaktierende P^+ -Bereich an der Oberfläche in einem seitlichen Abstand von etwa $25 \mu\text{m}$ liegen, kann aufgrund der großen Fläche des Übergangs - mit einem Entwurfsmaß des N^+ -Fensters von $A = 102 \cdot 28 \mu\text{m}^2$ eines typischen MOS der Größe $W_{\text{nom}} = 100 \mu\text{m}$, $L_{\text{nom}} = 10 \mu\text{m}$ - vorwiegend die vertikale Stromkomponente zur Analyse benutzt werden.

Bild 2.31 zeigt das durch Prozeßsimulation mit SUPREM IV erhaltene vertikale Dotierprofil beginnend mit dem oberflächlichen N^+ -Bereich, der in $1.4 \mu\text{m}$ Tiefe einen metallurgischen Übergang zur jeweiligen p-Wanne bildet. Während die Ausführung des N^+ -Kontakts, der als Source bzw. Drain des MOS dient, in allen drei Prozeßvarianten aufgrund identischer Prozeßführung der N^+ -Diffusion und nachfolgender Schritte gleich ausfällt, unterscheiden sich die Prozesse in Bezug auf die p-Wanne, die hier nur unterhalb des hochdotierten Kontakts betrachtet wird.

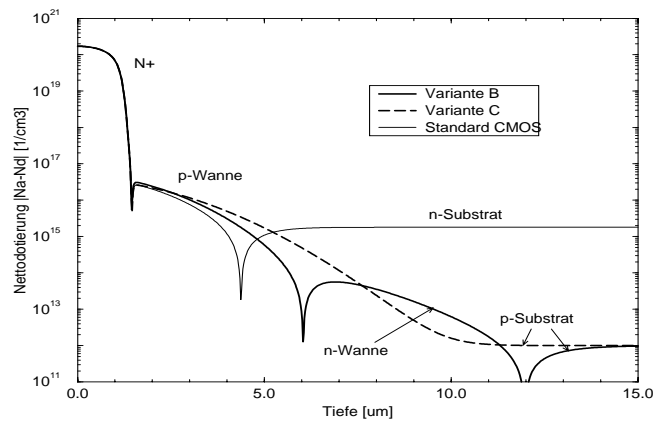


Bild 2.31: SUPREM IV Simulation des Dotierprofils vertikal durch den Source- oder Drain-Kontakt des NMOS-Transistors. Es sind die Profile für die beiden Varianten B und C auf hochohmigem p⁻-Silizium und das auf konventionellem n-Substrat (Standard-CMOS) angegeben. Abweichungen ergeben sich nicht bei der Ausführung der höherdotierten Seite (N⁺), sondern bei den unterschiedlichen p-Wannen-Profilen.

Der Schnitt durch die Source des zum Vergleich herangezogenen Standard-Prozess-NMOS zeigt eine p-Wanne, die bei einer Tiefe von 4,4 µm einen Übergang zum homogen dotierten n-Substrat bildet (Phosphor, $N_D = 1,5 \cdot 10^{15} / \text{cm}^3$).

Die Variante B besitzt im Bereich des NMOS eine in vertikaler Richtung doppelte Wannenstruktur, in der die p-Wanne, die sich von 1,4 µm bis 6,1 µm erstreckt, in die von 6,1 µm bis 12,0 µm reichende n-Wanne eingebettet ist.

Die Variante C hat eine vom N⁺-p-Wannen-Übergang bei 1,4 µm kontinuierlich in das p⁻-Substrat übergehende p-Wanne.

Auf die besonderen Eigenschaften der N⁺p-Diode in der Variante B wird im folgenden Abschnitt über das vertikale Durchgriffverhalten noch eingegangen. Für den Vergleich mit der Diode des Referenz-NMOS des Standard-Prozesses wird die Diode der Variante C benutzt, die identisch mit der in der Variante C₂ realisierten ist. Bild 2.32 zeigt diesen kontinuierlichen Übergang ins p⁻-Substrat an Hand des Bänderdiagramms der Variante C. Die beiden senkrechten Linien am Übergang zwischen N⁺-Source und p-Wanne zeigen die Ausdehnung der Raumladungszone (RLZ) im Gleichgewicht ohne äußere angelegte Spannung. Die gesamte Breite der RLZ liegt bei etwa $l_{RLZ} = 0,25 \mu\text{m}$. Mit den Dotierungsdaten des Bildes 2.35 kann numerisch der zu erwartende Schichtwiderstand der einzelnen Bereiche bestimmt werden (vgl. Tabelle 9). Mittels der Testscheiben für die Wannen und die hochdotierten Bereiche kann durch eine Vier-Spitzen-Messung auch ein experimenteller Vergleichswert bestimmt werden. Da der Wannenbereich unterhalb der N⁺-Kontakte nicht zugänglich ist, wird hier nur auf die durch numerische Integration ermittelte Größe verwiesen. Die weitgehende Übereinstimmung zwischen den Meßwerten und den auf den Simulationen basierenden Ergebnissen bestätigt die hohe Simulationsgüte in bezug auf das Dotierprofil. Dieses ist im folgenden auch Grundlage für die Feldstärkeberechnung am pn-Übergang.

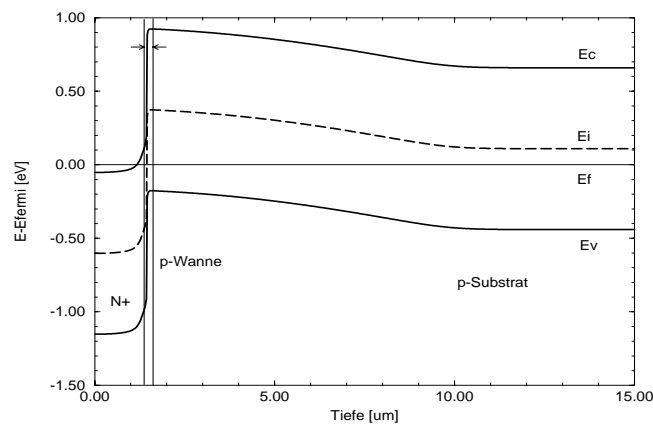


Bild 2.32: Bänderdiagramm des vertikalen Schnittes durch den Source-Kontakt des NMOS der Doppel-Wannen-Variante C. Die dünnen senkrechten Linien geben die Grenzen der Raumladungszone ohne äußere angelegte Spannung an.

TABELLE 9. Schichtwiderstände R_{sheet} NMOS

	Standard-CMOS berechnet	Standard-CMOS gemessen	Variante C berechnet	Variante C gemessen
N ⁺ -Source- und Drain-Kontakt	4.4 Ω/□	10 Ω/□	4.4 Ω/□	12 Ω/□
p-Wanne im Kanalgebiet	1850 Ω/□	2000 Ω/□	1620 Ω/□	1900 Ω/□
p-Wanne unterhalb des N ⁺ -Gebiets	5240 Ω/□		3750 Ω/□	

Mit der aus dem Dotierprofil gewonnenen Kenntnis der Breite der Raumladungszonen und der Bahngebiete läßt sich der Diodenstrom abschätzen. Im Idealfall besteht der Strom durch eine Diode aus den beiden Komponenten Diffusionsstrom [2.23]

$$I_{s1} = eAn_i^2 \left(\frac{D_p}{L_p N_D} + \frac{D_n}{L_n N_A} \right) \left(e^{\frac{V_f}{U_i}} - 1 \right) \quad (2.22)$$

und Generations-/Rekombinationsstrom

$$I_{s2} = \frac{eAn_i W_d}{2\tau} \left(e^{\frac{V_f}{2U_i}} - 1 \right) \quad (2.23)$$

Hierbei ist A die wirksame Fläche des Übergangs, D_p und D_n sowie L_p und L_n sind die Diffusionskonstanten ($D_{p,n}$) bzw. Minoritätsträgerdiffusionslängen ($L_{p,n}$) der Löcher und Elektronen, V_f ist die von außen angelegte Spannung, W_d die von dieser Spannung abhängige Ausdehnung der Raumladungszone und τ die mittlere Lebensdauer der Elektronen und Löcher in der Sperrschicht.

Die obige Formel des durch Diffusionsgradienten der Träger entstehenden Stroms I_{s1} gilt so jedoch nur im Fall einer sogenannten Langbasisdiode, d.h. wenn die Trägerdiffusionslängen klein im Vergleich zur Weite des jeweiligen Bahngebiets sind. Eine Verletzung dieser Bedingung erfolgt bei großen Diffusionslängen aufgrund hoher Trägerlebensdauern sowie bei relativ kurzen Bahngebieten.

Es ist dann in obiger Formel der Summand D_p/L_p durch $D_p/L_p \cdot \coth(d_n/L_p)$ und D_n/L_n durch $D_n/L_n \cdot \coth(d_p/L_n)$ zu ersetzen, d_n und d_p sind hierbei die neutralen Bahngebiete der n- bzw. p-dotierten Seite. Bei im Vergleich zu den Diffusionslängen sehr kurzen Bahngebieten kann in obiger Formel L_n durch d_p und L_p durch d_n ersetzt werden; dies ist der Fall der Kurzbasisdiode.

Weitere Effekte, die zu einer Abweichung vom idealen exponentiellen Verhalten beitragen, sind:

- Ein möglicher Paralleleitwert G_p , der etwa durch oberflächliche leitende Beläge oder Inversionskanäle entsteht.
- Ein Serienwiderstand R_s bedingt durch die endliche Leitfähigkeit der neutralen Bahngebiete.
- Ein Rauschanteil aufgrund von Schrotrauschen (engl. shot noise) und thermischem Rauschen [2.24] gemäß

$$\overline{I_n^2} = 2e\overline{I_d}\Delta f + 4kT\Delta f \cdot \text{Re}\{Y_d\} \quad (2.24)$$

hierbei steht $\overline{I_d}$ für den zeitlichen Mittelwert des Diodenstroms, Δf ist die betrachtete Frequenzbandbreite und Y_d der komplexe Leitwert der Diode.

Dieser Rauschanteil ist der allgemeinen Diodenkennlinie statistisch überlagert:

$$I_d = I_{s1} \left(e^{\frac{V_d}{U_t}} - 1 \right) + I_{s2} \left(e^{\frac{V_d}{2U_t}} - 1 \right) + G_p V_d. \quad (2.25)$$

Die Strom-Spannungs-Charakteristiken von N⁺p-Dioden im Durchlaßbetrieb zeigt Bild 2.33 vergleichend für die Variante C und den Standard-Prozeß. In beiden Meßkurven ist bei mittleren Spannungen von etwa $V_f = 400$ mV ein Bereich zu erkennen, in dem der Strom nahezu ausschließlich aus dem Diffusionsanteil nach Gleichung (2.22) besteht und somit den idealen exponentiellen Anstieg sehr gut zeigt. In diesem Kennlinienteil ist mittels einer Regressionsanalyse eine Exponentialfunktion gemäß

$$I_f = I_s \cdot e^{\frac{V_f}{\eta U_t}} \quad (2.26)$$

angepaßt worden, die im Bild als Gerade eingezeichnet ist.

Die ermittelten Kennwerte lauten für die Diode der Variante C $I_s = 25$ fA und $\eta = 1.04$ bei einer Temperatur $T = 34,5$ °C während der Messung (intrinsische Trägerdichte $n_i = 2.64 \cdot 10^{10}$ /cm³, $U_t = 26.5$ mV). Auf $T = 27$ °C umgerechnet entspricht dies dem Wert $I_{s, \text{kor}} = 8$ fA ($n_i = 1.50 \cdot 10^{10}$ /cm³, $U_t = 25.9$ mV). Da die Diode auf Standard n-Substrat bei $T = 34.5$ °C aufgenommen wurde, errechnet sich für $I_s = 36$ fA und $\eta = 1.04$ temperaturkorrigiert ein Sperrsättigungsstrom $I_{s, \text{kor}} = 12$ fA. Für den Idealitätsfaktor η wird dabei die Definition

$$\eta = \frac{1}{U_t} \cdot \left(\frac{d \ln I_f}{d V_f} \right)^{-1} \quad (2.27)$$

benutzt.

Um den Wert des Sättigungsstroms abzuschätzen, wird eine mittlere Lebensdauer der Elektronen von $\tau_n = 2$ µs angenommen, was recht gut den meßtechnisch ermittelten Werten entspricht (vgl. Anhang A4).

Für eine Elektronenbeweglichkeit von $\mu_n = 1000$ cm²/Vs ergibt sich damit eine Diffusionskonstante von $D_n = \mu_n \cdot U_t = 26$ cm²/s und eine Diffusionslänge von $L_n = (D_n \tau_n)^{1/2} = 72$ µm. Die Diodenfläche beträgt $A = b \cdot l = 28 \cdot 102$ µm² und die Länge des p-Bahngelands im Standard-CMOS-Prozeß

$$d_p = x_{j, p\text{-Wanne}, n\text{-Substrat}} - x_{j, n, p\text{-Wanne}} - \text{Breite}_{RLZ \text{ in der } p\text{-Wanne}} = 2,9 \mu\text{m}. \quad (2.28)$$

Damit ergibt sich mit einer intrinsischen Trägerdichte von $n_i = 1.5 \cdot 10^{10}$ /cm³ bei Raumtemperatur ($T = 27$ °C) ein Sättigungsstrom

$$I_{s1} \approx eA \cdot \frac{D_n}{L_n} \cdot \frac{n_i^2}{N_A} \coth \frac{d_p}{L_n} = 6,2 \text{ fA}, \quad (2.29)$$

der den meßtechnisch ermittelten Werten $I_s = 10 \dots 12$ fA recht genau entspricht (vgl. Bild 2.33). Hierbei wurde die Dotierung an der Grenze der Sperschicht mit $N_A = 1.5 \cdot 10^{16}$ /cm³ berücksichtigt.

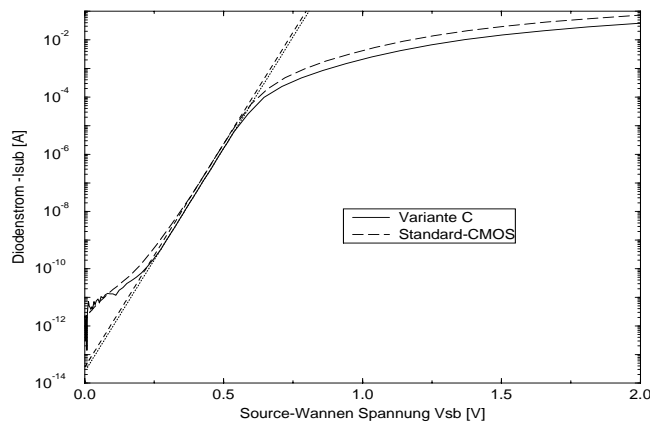


Bild 2.33: Vergleich der Vorwärtskennlinien der N^+ -Source-p-Wannen-Übergänge von NMOS der Variante C sowie auf konventionellem Silizium. Die hochdotierte N^+ -Fläche hat eine Entwurfsgröße von $102 \times 28 \mu\text{m}^2$. Die Geraden entsprechen einer idealen Diodenkennlinie mit Werten des Idealitätsfaktors η von ungefähr eins. Die Temperatur während der Messung betrug $T=34.5 \text{ }^\circ\text{C}$.

Die Abweichungen von der idealen exponentiellen Form entstammen vornehmlich Rekombinations- und Generationsströmen für kleine Spannungen und dem Einfluß des seriellen Bahnwiderstands bei hohen Spannungen. Eine Anpassung der Kennlinien bei kleinen Strömen an eine exponentielle Kurve mit dem Idealitätsfaktor $\eta = 2$, welcher - unter Vernachlässigung der Ausdehnungsänderung der Sperrschicht - bei Rekombinations- und Generationsströmen auftritt, gelingt nur recht ungenau, da das Meßrauschen, der Stromoffset und die notwendige Subtraktion der idealen Exponentialkennlinie (mit $\eta \approx 1$) von der gemessenen Kennlinie zu deutlichen Unsicherheiten führt. Für beide Kennlinien ergibt sich somit nur ungefähr ein zusätzlicher Sättigungsstromanteil $I_{s2} = 2 \times 10^{-12} \text{ A} \cdot e^{V_f/2U_t}$.

Durch Ansatz der modifizierten Kennliniengleichung

$$I_f = I_s \cdot e^{\frac{V_f - R_s I_f}{\eta U_t}} \quad (2.30)$$

läßt sich der Serienwiderstand R_s bestimmen, dessen stromabhängige Variation in Bild 2.34 dargestellt ist. Von einem Maximalwert mit etwa $R_{s,\text{max}} = 300 \Omega$ bei der Variante C auf hochohmigem Substrat und $R_{s,\text{max}} = 180 \Omega$ bei dem Vergleichs-NMOS sinkt der Widerstand zu höheren Spannungen bzw. Strömen hin ab.

Bei hohen Strömen erfolgt eine Leitfähigkeitserhöhung in den Bahngebieten durch starke Injektion von Minoritätsträgern, welche aufgrund der erforderlichen Ladungsneutralität auch ein Ansteigen der Majoritätsträgerdichte zur Folge hat. Daraus folgt eine deutliche Verringerung des Widerstands. Der resultierende Spannungsabfall am Serienwiderstand steigt daher bei hohen Strömen nicht linear - wie nach dem Ohmschen Gesetz erwartet - an, sondern mit der Quadratwurzel des Stroms [2.25]. Das Absinken der Widerstandswerte für Spannungen $V_f < 0.5 \text{ V}$ ist durch numerische und meßtechnische Ungenauigkeiten zu erklären, da die geringen Ströme zu vernachlässigbaren Spannungsabfällen am Serienwiderstand führen.

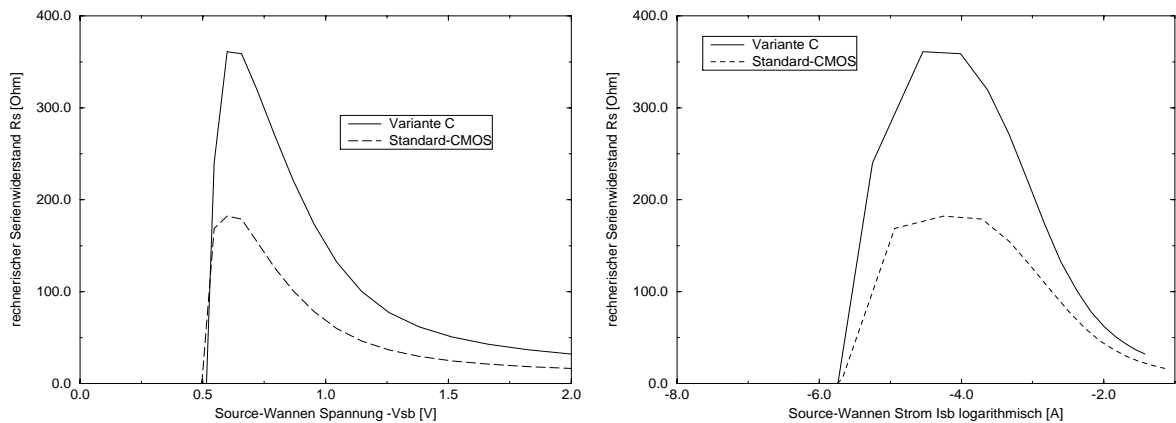


Bild 2.34: Darstellung des N^+p -Dioden-Serienwiderstands R_s bei Annahme einer optimal angepaßten idealen Diodenkennlinie. Die hochdotierte N^+ -Fläche hat eine Entwurfsgröße von $102 \times 28 \mu\text{m}^2$. Die Temperatur während der Messung betrug $T = 34.5 \text{ }^\circ\text{C}$.

Bei Raumtemperatur ist bei Anstieg der Minoritätsträgerdichte auf das Niveau der Dotierungsdichte N_A Hochinjektion gegeben, die zu einer deutlichen Erhöhung der Leitfähigkeit führt. Eine praktische Abschätzung [2.26] mit der Bedingung

$$n_p = N_A \quad \text{mit} \quad n_p = n_{po} e^{\frac{V_f}{U_T}} \quad (2.31)$$

für die Dichte n_p der Elektronen am p-seitigen Rand der Sperrschicht und $n_{po} = n_i^2/N_A$ als Minoritätsträgerdichte im thermischen Gleichgewicht führt unter Berücksichtigung von $I_f = I_s \cdot \exp(V_f/U_T)$ zum gesuchten Strom I_{hi} bei Beginn der Hochinjektion

$$I_{hi} = I_s \cdot \left(\frac{N_A}{n_i} \right)^2 \quad (2.32)$$

Im Fall der Kurzbasisdiode folgt bei eindimensionaler Betrachtung

$$I_{hi} = eA \cdot \frac{U_T \mu N_A}{d_p} \quad (2.33)$$

Bei einer Dotierung der p-Wanne von $N_A = 1.5 \times 10^{16} / \text{cm}^3$ folgt mit den Werten des Sättigungsstroms aus der Vorwärtsskennlinie $I_{hi} = 8.1 \text{ mA}$ für die Variante C und $I_{hi} = 11.6 \text{ mA}$ für die Diode des Standard-NMOS. Diese Werte beziehen sich jeweils auf $T = 34.5 \text{ }^\circ\text{C}$.

Ein Blick auf den meßtechnisch ermittelten Verlauf des Serienwiderstands zeigt jedoch eine signifikante Reduktion bei deutlich geringeren Strömen bzw. Spannungen. Bei den für die oben definierte Hochinjektionsbedingung berechneten Strömen ergibt sich in beiden Fällen ein Serienwiderstandswert von etwa 20 % des Maximalwerts $R_{s,max}$. Diese Überschätzung des Einsatzstroms I_{hi} ist bedingt durch:

- i) Die Annahme einer durchgehend exponentiell mit $\eta = 1$ ansteigenden idealen Kennlinie.
- ii) Die Vernachlässigung der in die Tiefe hin abnehmenden Wannendotierung. Die dominierenden Anteile des Serienwiderstands stammen von Gebieten, die geringer als $N_A = 1.5 \times 10^{16} / \text{cm}^3$ dotiert sind und dementsprechend früher mit Minoritätsträgern überschwemmt werden.
- iii) Die Idealisierung, daß eine Leitfähigkeitserhöhung nicht schon bei geringeren Ladungsträgerkonzentrationen als $n_p = N_A$ erfolgt.

Die Abschätzung des Serienwiderstands aus dem Transistor-Layout unter Berücksichtigung der seitlichen

Abstände zwischen den hochdotierten N⁺- und P⁺-Gebieten führt gemäß

$$R_s = \frac{L_{eff}}{W_{eff}} \cdot R_{\square} \quad \text{mit} \quad \frac{L_{eff}}{W_{eff}} = \frac{1}{6,6} \quad \text{und} \quad R_{\square} = 1620\Omega \dots 1850\Omega \quad (2.34)$$

zu Serienwiderstandswerten von $R_s = 245 \Omega$ für die Variante C und $R_s = 278 \Omega$ für den Standard-Prozeß. Beide Werte liegen nah an den meßtechnisch ermittelten Daten. Eine genauere Analyse kann mit Hilfe eines Bauelement-Simulationsprogramms auf numerischer Basis erfolgen und sollte mehr-dimensionale Effekte berücksichtigen (z.B. PISCES von der Stanford University, USA) [2.27].

Bild 2.35 zeigt den oben definierten Idealitätsfaktor η für die beiden Dioden, jeweils für $T = 34.5 \text{ }^\circ\text{C}$ ($U_t = 26.5 \text{ mV}$) berechnet. Es zeigt sich kein markanter Unterschied zwischen beiden Dioden in der Durchlaßkennlinie, die jeweils einen Bereich mit einem Idealitätsfaktor nahe eins für mittlere Ströme aufweisen. Dieser Bereich ist nach oben - zu hohen Strömen hin - durch den Serienwiderstandseinfluß und nach unten durch Rekombinations- und Generationsströme sowie Rauschstromanteile und meßbedingte Stromoffsets begrenzt.

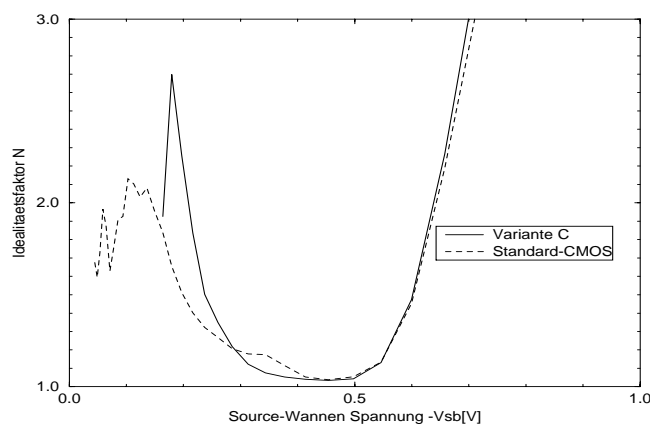


Bild 2.35: Darstellung des durch numerische Differentiation der N⁺p-Dioden-Kennlinie ermittelten Idealitätsfaktors η . Die hochdotierte N⁺-Fläche hat eine Entwurfsgröße von $102 \times 28 \mu\text{m}^2$. Die Temperatur während der Messung betrug $T = 34.5 \text{ }^\circ\text{C}$.

Da die Übergänge der Source- und Drain-Kontakte bezüglich des Substrats in Sperrichtung gepolt sind, interessieren insbesondere deren Leckströme im Sperrbetrieb, die in Bild 2.36 gezeigt sind. Beide Sperrkennlinien zeigen keine starken Auffälligkeiten, wenngleich die Diode der hochohmigen Variante C insgesamt einen etwas höheren Sperrstrom zeigt.

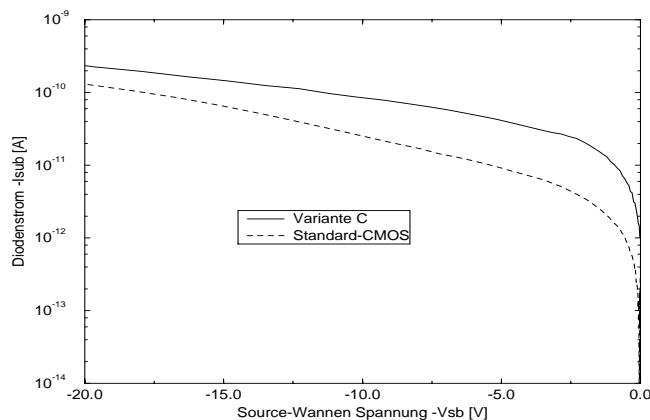


Bild 2.36: Vergleichende Darstellung der Sperrkennlinien der N⁺p-Dioden zwischen Source und Wanne in der Variante C und im Vergleichs-CMOS-Prozeß. Die hochdotierte N⁺-Fläche hat eine Entwurfsgröße von $102 \times 28 \mu\text{m}^2$. Die Temperatur während der Messung betrug $T = 34.5 \text{ }^\circ\text{C}$.

Für beide Sperrcharakteristiken zeigt sich ein - gegenüber der Theorie für einen abrupten pn-Übergang - erhöhtes Niveau des Sperrstroms. So liefert die Abschätzung für einen abrupten pn-Übergang mit einer Dotierung von $N_A = 1.5 \cdot 10^{16} / \text{cm}^3$ einen Sperrstrom aufgrund des Diffusionsanteils - wie oben abgeschätzt - von etwa 6.2 fA. Ein Wert, der um drei bis vier Größenordnungen geringer ausfällt, als der beobachtete Strom von $I_r(-5 \text{ V}) = 40 \text{ pA}$ bei Variante C und $I_r(-5 \text{ V}) = 9 \text{ pA}$ beim Standard-CMOS-Prozeß.

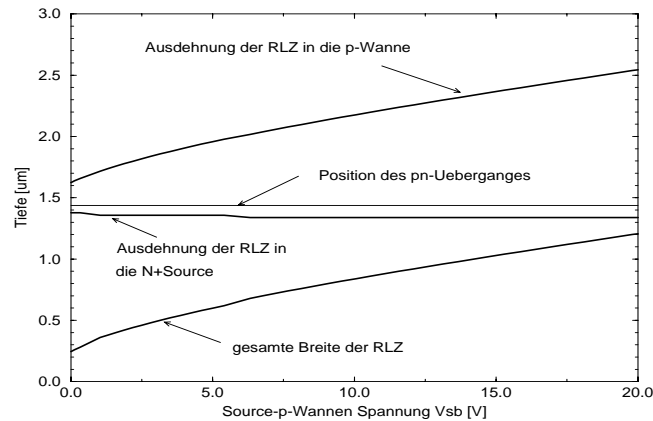


Bild 2.37: Darstellung der räumlichen Ausdehnung der Raumladungszone (RLZ) bei verschiedenen Sperrspannungen. Die Berechnung erfolgt numerisch auf Basis des simulierten Dotierprofils und berücksichtigt die Spannungsabhängigkeit der Diffusionsspannung U_d aufgrund der ortsabhängigen Dotierung.

Ein Teil dieses Unterschieds ist dadurch zu erklären, daß die Ableitung des idealen Sättigungsstroms in der gewählten Form von homogenen Dotierungen auf beiden Seiten des pn-Übergangs ausgeht, im betrachteten Fall jedoch mit der Tiefe abnehmende Wannendotierungen auftreten.

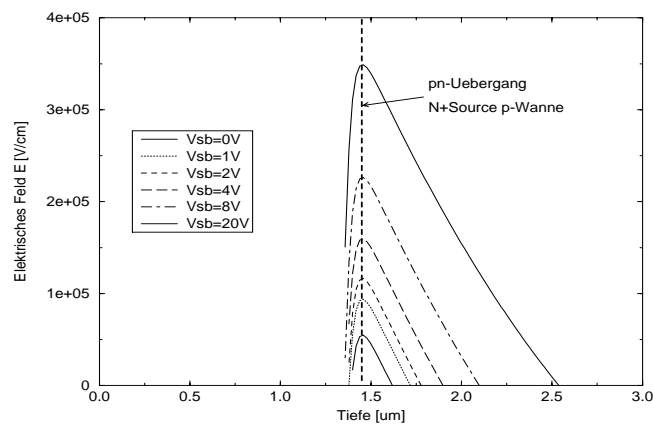


Bild 2.38: Darstellung des durch numerische Integration aus dem simulierten Dotierprofil entstandenen Feldstärkeverlaufs bei Anlegung unterschiedlicher Sperrspannungen an den N^+ -Source-p-Wannen-Übergang der Variante C.

Bild 2.37 zeigt die Ausdehnung der Sperrschicht bei verschiedenen Sperrspannungen. Die numerische Berechnung berücksichtigt auch die leicht variierende Größe der Diffusionsspannung, bedingt durch die spannungsabhängige Raumladungsweitenmodulation, die unterschiedliche Randkonzentrationswerte bewirkt. Die den Sättigungsstrom bestimmende Dotierkonzentration am p-Wannen-seitigen Rand der RLZ variiert jedoch um deutlich weniger als eine Größenordnung im Bereich von Sperrspannungen $0 \text{ V} \leq V_{sb} \leq 20 \text{ V}$ (vgl. Bild 2.31 und Bild 2.37).

Durchbrucheffekte z.B. aufgrund von Lawinmultiplikation können ebenfalls ausgeschlossen werden, insbesondere, da der Anstieg des Stroms sehr langsam erfolgt. Eine in Bild 2.38 dargestellte numerische Feldstärkeberechnung des pn-Übergangs der Variante C ergibt bei der Sperrspannung $V_{sb} = 20 \text{ V}$ eine maximale Feldstärke $|E_{\text{max}}| = 3.5 \cdot 10^5 \text{ V/cm}$, die jedoch noch unterhalb der kritischen Feldstärke für Lawinmultiplikation liegt. In der Literatur sind Durchbruchspannungen von $V_{br} = 35 \text{ V}$ für einen abrupten Übergang mit der gewählten Dotierung ($N_A = 1.5 \cdot 10^{16} / \text{cm}^3$) angegeben [2.28].

Gemäß der Gleichung (2.23) für den Generations- und Rekombinationsstrom ergibt sich rechnerisch für eine

Weite der Sperrschicht von $W_d = 0.6 \mu\text{m}$ bei $V_{sb} = 5 \text{ V}$ (vgl. Bild 2.37) eine effektive Lebensdauer der Träger in der Sperrschicht von $\tau_{\text{eff}} = 100 \text{ ns}$ bei der Variante C und $\tau_{\text{eff}} = 450 \text{ ns}$ beim Standard-Prozeß. Verunreinigungen, die aufgrund der fehlenden intrinsischen Getterung im float-zone-Silizium nicht von den Bauelementen ferngehalten werden, können - als Rekombinations- bzw. Generationszentren dienend - den Sperrstrom erhöhen.

Der in beiden Fällen insgesamt recht niedrige rechnerische Wert der Lebensdauer in der RLZ liegt z.T. auch in der Oberflächenrekombination begründet. Spezielle vertikal aufgebaute Dioden mit geringen Oberflächen-zu-Volumenverhältnissen zeigen bei zusätzlichen Getterschritten Lebensdauern von einigen $100 \mu\text{s}$ [2.29].

Eine mögliche Ursache für einen erhöhten Leckstrom ist durch die Isolation des Wafers während der Messung verhindert worden. Unerwünschte Ströme über das p^- -Substrat und den Chuck konnten so vermieden werden. Dennoch ist zu vermuten, daß die Einbettung der p-Wanne in das p^- -Substrat einen Teil des Unterschieds zwischen den beiden Stromkennlinien bewirkt. Der wesentliche Unterschied besteht in der unterschiedlichen Lage des P^+ -Wannenanschlusses, der einmal leitend mit dem p^- -Substrat verbunden ist und einmal isoliert im n-Substrat liegt. Ferner bleibt der Unterschied des Übergangs der p-Wanne in das jeweilige Substrat. Während der Übergang p-Wanne-n-Substrat sperrendes Verhalten zeigt, ist der Übergang p-Wanne- p^- -Substrat ohmscher Natur.

Die Vermutung, daß die Überlappung des P^+ -Kontakts in das Substrat zu einem erhöhten Sperrstrom im Vergleich zu einem von der p-Wanne komplett umgebenen Kontakt führt, konnte nur eingeschränkt bestätigt werden. Die in Bild 2.39 dargestellte Messung spezieller Teststrukturen zeigt, daß insbesondere im Bereich geringer Sperrspannungen keine nennenswerten Unterschiede beobachtet werden können. Aber gerade in diesem Bereich weicht die Kennlinie der Diode des hochohmigen Prozesses stärker von jener des Standard-CMOS ab (vgl. Bild 2.36). Nur für Sperrspannungen ab etwa 8 V vergrößert sich der Sperrstrom um etwa den Faktor zwei.

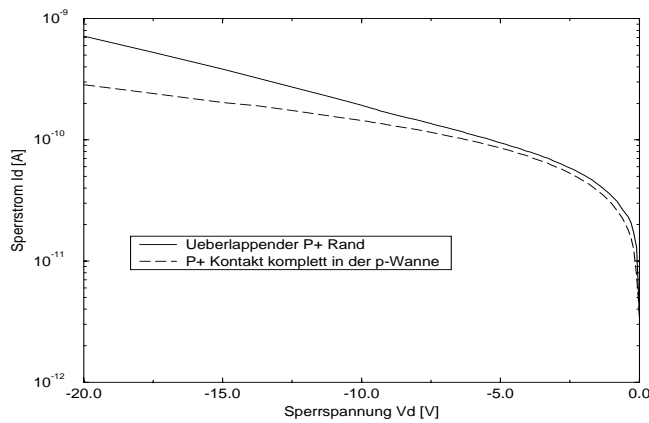


Bild 2.39: Sperrstromverhalten von großflächigen N^+/p -Wannen-Dioden in der Variante C. Die Diode mit überlappendem Rand entspricht im Querschnitt den Source- und Drainkontakten der NMOS-Transistoren. Das die p-Wanne kontaktierende P^+ -Gebiet ist insgesamt $15 \mu\text{m}$ breit und überlappt um $10 \mu\text{m}$ den Rand der p-Wanne. Diese Überlappung reduziert sich durch die deutlich größere Unterdiffusion der p-Wanne auf etwa $5 \mu\text{m}$. Bei der Vergleichsstruktur befindet sich die P^+ -Kontaktierung innerhalb der p-Wanne. Im ersten Fall beträgt das Entwurfsmaß der Wanne $200 \times 200 \mu\text{m}^2$ mit jeweils um $10 \mu\text{m}$ überlappenden P^+ -Kontakten. Im zweiten Fall wird die Wanne mit $220 \times 220 \mu\text{m}^2$ dimensioniert. Der N^+ -Kontakt hat eine Entwurfsgröße von $100 \times 100 \mu\text{m}^2$. Die Temperatur während der Messung betrug $T = 34.5 \text{ }^\circ\text{C}$.

Nimmt man jedoch als Ursache für den vergrößerten Stromanstieg der Sperrschichten auf hochohmigem Substrat einen Parallel-Leitwert G_p an, so ergibt sich für Sperrspannungen zwischen 5 V und 20 V im Fall der Variante C in Bild 2.36 $G_{p1} = 1.3 \cdot 10^{-11} \text{ S} = 1/76 \text{ G}\Omega$ und im Fall der Diode mit überlappendem Rand in Bild 2.39 $G_{p2} = 4.0 \cdot 10^{-11} \text{ S} = 1/25 \text{ G}\Omega$. Daraus läßt sich ein Flächenbelag des Leitwerts

$$\frac{G_{p1}}{A_1} = \frac{1,3 \cdot 10^{-11} \text{ S}}{102 \cdot 28 \mu\text{m}^2} = 4,5 \cdot 10^{-15} \frac{\text{S}}{\mu\text{m}^2} = \frac{1}{2,2 \text{ M}\Omega \text{ cm}^2} \quad (2.35)$$

bzw.

$$\frac{G_{p2}}{A_2} = \frac{4,0 \cdot 10^{-11} \text{ S}}{100 \cdot 100 \mu\text{m}^2} = 4,0 \cdot 10^{-15} \frac{\text{S}}{\mu\text{m}^2} = \frac{1}{2,5 \text{ M}\Omega \text{cm}^2} \quad (2.36)$$

ableiten, der in beiden Fällen in etwa die gleiche Größe aufweist. Zum Vergleich ergibt sich bei der Teststruktur, deren P⁺-Kontakt komplett in der p-Wanne (vgl. Bild 2.39) liegt, ein deutlich geringerer Leitwert von $G_{p3} = 1,3 \cdot 10^{-11} \text{ S} = 1/79 \text{ G}\Omega$ und auch der Flächenbelag

$$\frac{G_{p3}}{A_3} = \frac{1,3 \cdot 10^{-11} \text{ S}}{100 \cdot 100 \mu\text{m}^2} = 1,3 \cdot 10^{-15} \frac{\text{S}}{\mu\text{m}^2} = \frac{1}{8 \text{ M}\Omega \text{cm}^2} \quad (2.37)$$

ist entsprechend reduziert. Folglich kann bei den Source- und Drainkontakten der Variante C von einem parasitären Parallel-Leitwert ausgegangen werden, der von der Teilüberlappung des P⁺-Wannenanschlusses ins p⁻-Substrat bedingt sein könnte.

Versucht man nun die Stromdifferenz zwischen der Sperrkennlinie der Variante C und der N⁺p-Diode auf Standard-n-Substrat (vgl. Bild 2.36) als Zusatzleitwert ΔG_p zu erklären, so ergibt sich bei der Vorspannung $U_{sb} = -5 \text{ V}$ $\Delta G_{p,-5 \text{ V}} = 0,7 \cdot 10^{-11} \text{ S} = 1/140 \text{ G}\Omega$ und bei $U_{sb} = -20 \text{ V}$ $\Delta G_{p,-20 \text{ V}} = 0,6 \cdot 10^{-11} \text{ S} = 1/167 \text{ G}\Omega$. Die Differenzen beider Werte sind z.T. durch die unterschiedlichen Dotierprofile der Dioden erklärbar, so daß eine Modellierung des zusätzlichen Sperrstroms im Fall der Diode auf hochohmigem Substrat mit einem Parallel-Leitwert von etwa 150 G Ω möglich ist.

Die P⁺n-Dioden des PMOS

Die vertikal durch den Source- oder Drainkontakt gesehenen Dotierprofile der beiden Varianten B und C auf hochohmigem p⁻-Substrat und der Vergleichsversion auf n-Substrat sind in Bild 2.40 dargestellt. B und C unterscheiden sich im Bereich des PMOS in erster Linie durch die n-Wanne, die in unterschiedlicher Tiefe den Übergang ins p⁻-Substrat aufweist. Während die Oberflächenkonzentration der n-Wannen in beiden Fällen aufgrund der Forderung nach vergleichbaren Schwellspannungen V_{th} nahezu identisch gewählt wird, ist die n-Wannen-Dotierkonzentration am P⁺n-Übergang der Variante C deutlich geringer als in der Version B. Die geringere Tiefe der n-Wanne von nur 4,7 μm in der Variante C resultiert aus der Prozeßfolge, in der die n-Wanne zeitlich nach der p-Wanne eingebracht wird. Aufgrund einer demgegenüber vertauschten Reihenfolge in der Variante B resultiert ein schwächeres Gefälle der Dotierung und folglich eine größere Tiefe der Wanne von 12,3 μm . Da der NMOS der Variante B auch in der Variante C₂ realisiert wurde, wird dessen Diode mit der des Standard-CMOS-Prozesses verglichen. Die Tiefe des P⁺n-Übergangs liegt in allen drei Fällen bei 1,5 μm .

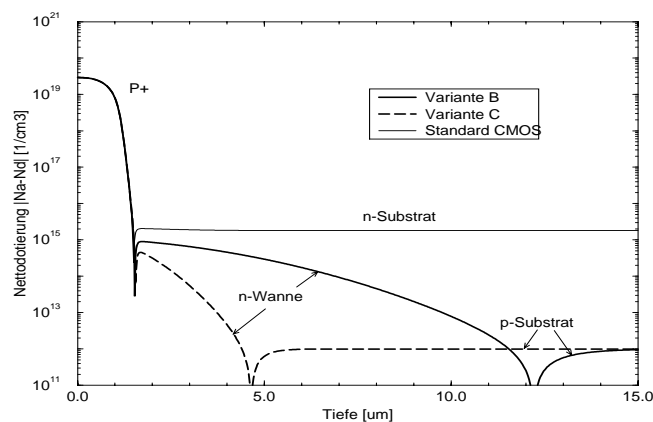


Bild 2.40: SUPREM IV Simulation des Dotierprofils vertikal durch den Source- oder Drain-Kontakt des PMOS-Transistors. Es sind die Profile für die beiden Varianten B und C auf hochohmigem p⁻-Silizium sowie auf konventionellem n-Substrat (Standard-CMOS) angegeben. Abweichungen ergeben sich nicht bei der Ausführung der höherdotierten Seite (P⁺), sondern bei den unterschiedlichen n-Wannen-Profilen.

Bild 2.41 zeigt das Bänderdiagramm der Variante B mit den zugehörigen Positionen der Raumladungszonen aufgrund der Diffusionsspannungen U_D der einzelnen Übergänge. Schon ohne äußere angelegte Spannung ist der neutrale Bereich der n-Wanne deutlich geringer als der Abstand zwischen den beiden metallurgischen Übergängen, da sich sowohl der P⁺n-Übergang (um mehr als 1 μm) als auch der n-Wannen-p⁻-Substrat-Übergang (um etwa 3.2 μm) in die n-Wanne ausdehnen.

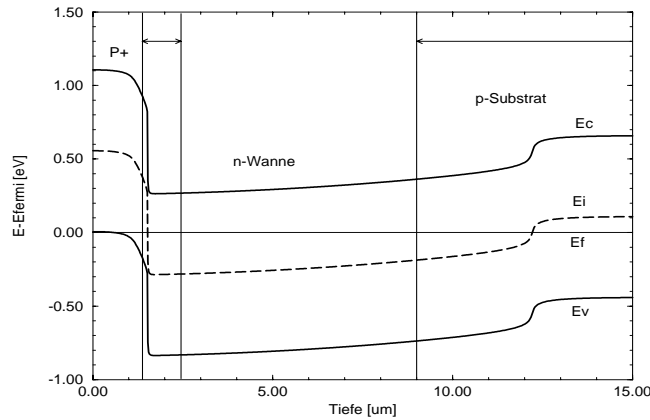


Bild 2.41: Bänderdiagramm des vertikalen Schnitts durch den Source-Kontakt des PMOS der Variante B. Die dünnen senkrechten Linien geben die Grenzen der Raumladungszone ohne äußere angelegte Spannung an.

Die in Tabelle 10 zusammengefaßten Werte der Schichtwiderstände der einzelnen Gebiete zeigen die Hochohmigkeit der n-Wanne aufgrund der geringen Dotierung und lassen einen großen Serienwiderstand R_s erwarten.

TABELLE 10. Schichtwiderstände R_{sheet} PMOS

	Standard-CMOS berechnet	Standard-CMOS gemessen	Variante B berechnet	Variante B gemessen
P ⁺ -Source- und Drain-Kontakt	42 Ω/\square	45 Ω/\square	42 Ω/\square	39 Ω/\square
n-Wanne im Kanalgebiet			9840 Ω/\square	10000 Ω/\square
n-Wanne unterhalb des N ⁺ -Gebiets			18160 Ω/\square	
n-Substrat	102 Ω/\square	105 Ω/\square		

Die Unterschiede der Vorwärtsskennlinien - wie in Bild 2.42 dargestellt - basieren zum einen auf deutlich verschiedenen Sättigungsströmen des idealen Kennlinienteils ($\eta = 1$) und zum anderen aus ungleich großen Serienwiderständen. Eine Anpassung der Kennlinien an eine ideale Exponentialfunktion im Bereich der größten Steigung führt bei der hochohmigen Variante B zu $I_s = 200 \text{ fA}$ und $\eta = 1.018$ bei einer Temperatur von $T = 36 \text{ }^\circ\text{C}$ während der Messung ($n_i(T = 309 \text{ K}) = 2.94 \cdot 10^{10} / \text{cm}^3$ und $U_t = 26.7 \text{ mV}$). Auf $T = 27 \text{ }^\circ\text{C}$ umgerechnet ergibt sich damit $I_{s,\text{kor}} = 51 \text{ fA}$.

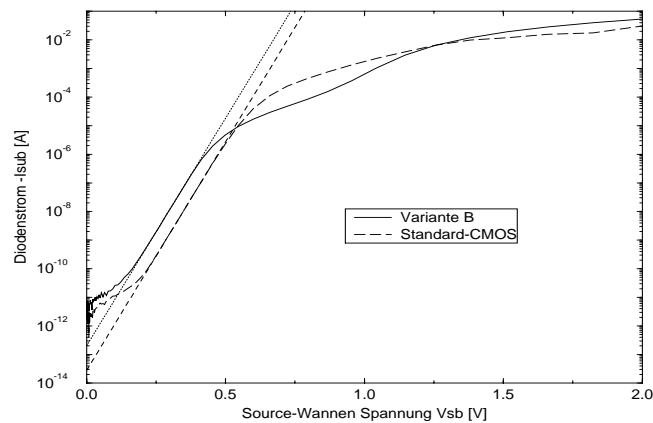


Bild 2.42: Vergleich der Vorwärtskennlinien der P⁺-Source-n-Wannen/Substrat-Übergänge eines PMOS der Variante B sowie auf konventionellem Silizium. Die Geraden entsprechen einer idealen Diodenkennlinie mit Werten des Idealitätsfaktors η von ungefähr eins. Die hochdotierte P⁺-Fläche hat eine Entwurfsgröße von $102 \cdot 28 \mu\text{m}^2$. Die Temperatur während der Messung betrug $T = 36.0 \text{ }^\circ\text{C}$.

Die für die Standard-Diode gezeichnete Exponential-Funktion hat die Werte $I_s = 28 \text{ fA}$ und $\eta = 1.012$ bei $T = 36 \text{ }^\circ\text{C}$ bzw. $I_{s,\text{kor}} = 7.3 \text{ fA}$ bei $T = 27 \text{ }^\circ\text{C}$.

Beim Vergleich mit der rechnerischen Lösung werden folgende Daten zugrunde gelegt:

TABELLE 11. Sättigungsstrom der P⁺-n-Diode

	Standard-CMOS	Variante B
Löcherlebensdauer im n-Gebiet [μs]	3	3
Löcherbeweglichkeit [cm^2/Vs]	550	550
Diffusionskonstante D_p [cm^2/s]	14.5	14.5
Diffusionslänge L_p [μm]	66	66
Dotierung des n-Gebiets am Rand der RLZ [$1/\text{cm}^3$]	$1.5 \cdot 10^{15}$	$8 \cdot 10^{14}$
Breite des n-Gebiets d_n [μm]	$>L_p$	$6.5 \text{ }^{*1)}$
Sättigungsstrom [fA]	1.1	28.7

^{*1)} vgl. Bild 2.41

Die Sättigungsstromwerte beziehen sich dabei auf eine Diodenfläche $A = 102 \cdot 28 \mu\text{m}^2$ und sind bei $T = 27 \text{ }^\circ\text{C}$ ($n_i(300 \text{ K}) = 1.5 \cdot 10^{10} / \text{cm}^3$) nach der Formel

$$I_{s1} = eA \frac{D_p}{L_p} \cdot \frac{n_i^2}{N_D} \coth \frac{d_n}{L_p} \quad (2.38)$$

bestimmt worden. Im Fall der hochohmigen Variante B zeigt sich eine recht gute Übereinstimmung, denn die Unsicherheit zwischen Meßergebnis und Rechnung liegt bei einem Faktor kleiner zwei. Der Grund der vergleichsweise großen Abweichung im Fall der Diode des Standard-Prozesses liegt auch in der Annahme, daß die Begrenzung des n-Gebiets d_n - aufgrund des homogenen n-Substrats - erst auf der Rückseite des Wafers vorliegt. Tatsächlich liegen aber die benachbarten dotierten Gebiete, wie etwa p-Wannen oder auch der P⁺-Drainkontakt, sowie die umgebende Oberfläche in deutlich geringeren Abständen. Gerade diese näher gelegenen Bereiche bestimmen aber den Konzentrationsgradienten der Löcher im n-Gebiet maßgeblich. Der bei der Messung festgestellte geringere Wert des Sättigungsstroms resultiert daher nur zum Teil aus der größeren Weite des n-Gebiets. Eine weitere Einflußgröße ist die höhere Dotierung des n-Substrats am Rand der Raumladungszone im Vergleich zur n-Wanne der Variante B.

Der in Bild 2.42 dargestellte Kennlinienverlauf der Diode der Variante B weist einen Knick in der Kennlinie bei Spannungen zwischen etwa 0.5 V und 1.0 V auf, der keinen Einzelfall darstellt, sondern typischerweise beobachtet werden kann. Als Folge der niedrigen Dotierung der n-Wanne ist - wie oben erwähnt - mit starken Serienwiderständen zu rechnen. Der dann wieder stärkere Anstieg des Stroms bei etwa 1.0 V ist daher eine Folge der deutlichen Reduktion dieser Widerstände aufgrund von starker Leitwertmodulation des n-Gebiets.

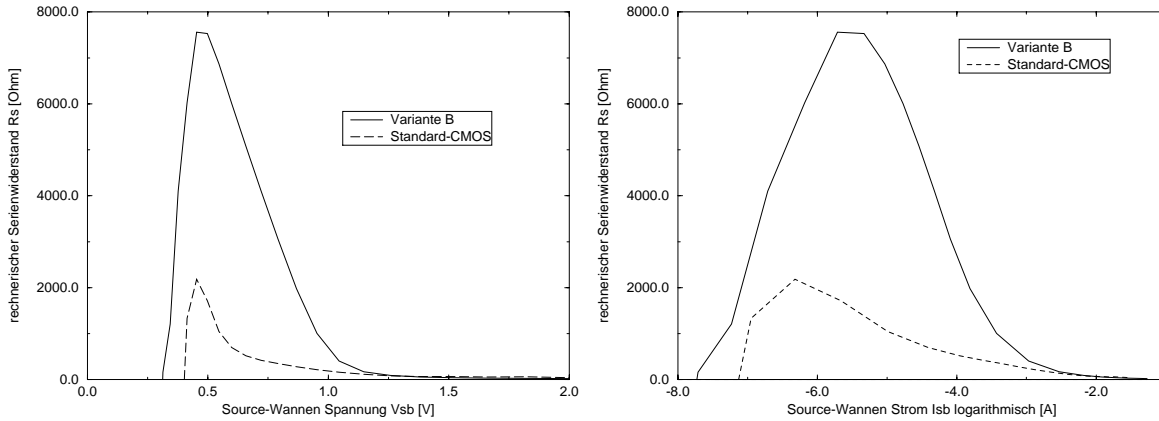


Bild 2.43: Darstellung des Dioden-Serienwiderstands R_s bei Annahme einer optimal angepaßten idealen Diodenkennlinie. Die hochdotierte P^+ -Fläche hat eine Entwurfsgröße von $102 \times 28 \mu m^2$. Die Temperatur während der Messung betrug $T = 36.0 \text{ }^\circ C$.

Diese Verringerung des Serienwiderstands wird in Bild 2.43 verdeutlicht. Ein im Vergleich zu Bild 2.34 um etwa 0.1 V bei der Standard-Diode und 0.2 V bei der Diode der Variante B früher wirksamer Serienwiderstand im Bereich einiger $k\Omega$ bedingt ein deutliches Absinken des Stromanstiegs auf niedrigerem Niveau als bei der N^+p -Diode. Der bei Variante B wieder zunehmende Anstieg des Stroms ist eine Folge der Reduktion dieser großen Widerstände auf die gleiche Größenordnung wie dies bei der Leitwertmodulation der N^+p -Diode beobachtet werden konnte. Bei etwa 1.5 V liegen sowohl die P^+n - als auch die N^+p -Dioden-Serienwiderstände bei einigen zehn bis hundert Ohm. Um für diesen Fall eine genauere Abschätzung des Stroms zu erhalten, bei welchem die Bedingung der Hochinjektion erreicht wird, soll auf einen Ansatz von Möschwitzter und Lunze [2.24], der Bahnspannungsabfälle durch große Ströme berücksichtigt, zurückgegriffen werden. Danach gilt bei hoher Injektion für die Löcherkonzentration am Rand der Raumladungszone zum n-Gebiet

$$p(x_o) = -\frac{N_D}{2} + \sqrt{\left(\frac{N_D}{2} + \frac{n_i^2}{N_D}\right)^2 + \frac{N_D I_d d_n}{e A U_t (\mu_n + \mu_p)}} \approx -\frac{N_D}{2} + \sqrt{\left(\frac{N_D}{2}\right)^2 + \frac{N_D I_d d_n}{e A U_t (\mu_n + \mu_p)}} \quad (2.39)$$

Aufgrund der Aufteilung der äußeren Spannung auf die Sperrschicht und die Bahngebiete, steigt die Trägerkonzentration nicht mehr linear mit dem Strom wie bei kleinen Strömen gemäß

$$p(x_o) = \frac{n_i^2}{N_D} \cdot \frac{I_d + I_s}{I_s} \quad (2.40)$$

sondern nur mehr mit der Quadratwurzel.

Mit den in Tabelle 11 gegebenen Daten läßt sich der Strom bei Beginn der Hochinjektion gemäß Gleichung (2.39) zu $I_{hi} = 460 \mu A$ für die Variante B bestimmen. Dieser Wert liegt etwas oberhalb des in Bild 2.42 beobachteten Beginns der Leitfähigkeitsmodulation. Die Ursache der Abweichung liegt in der Annahme eines homogenen Profils für die Gleichung (2.39), wohingegen das reale Profil aufgrund der in die Tiefe hin abnehmenden Dotierung früher in den Bereich der Leitfähigkeitsmodulation kommt.

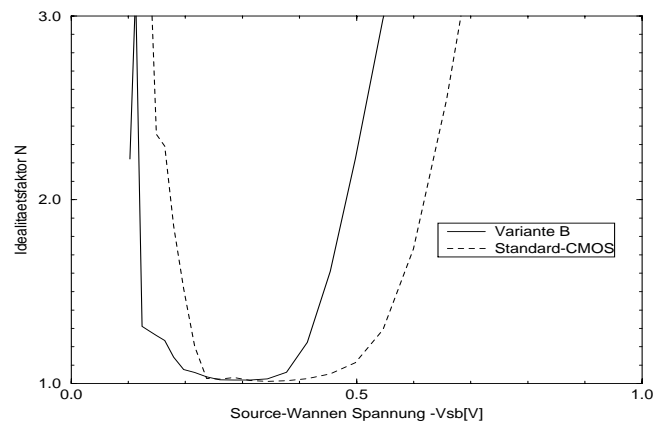


Bild 2.44: Darstellung des durch numerische Differentiation der P^+n -Dioden-Kennlinie ermittelten Idealitätsfaktors η . Die hochdotierte P^+ -Fläche hat eine Entwurfsgröße von $102 \times 28 \mu\text{m}^2$. Die Temperatur während der Messung betrug $T = 36.0 \text{ }^\circ\text{C}$.

Bild 2.44 zeigt den Idealitätsfaktor η der Dioden. Die Gültigkeitsgrenzen, d.h. $\eta = 2$, über der sich keine direkte physikalische Bedeutung mehr mit der Definition verbinden läßt, sind bei der Diode auf Standard-Substrat um etwa 0.1 V nach oben verschoben, zeigen aber einen Idealitätsfaktor nahe eins für etwa die gleichen Spannungs- (ca. 0.25 V) und Stromabstände (ca. 3 Dekaden). Der geringere Serienwiderstand der Standard-Diode verlagert dessen obere Grenze hin zu höheren Spannungen. Aufgrund des geringeren Sättigungsstroms gewinnt der Rekombinationsstromanteil bei geringen Strömen an Gewicht und dominiert demzufolge das Verhalten bei geringen Spannungen und führt zu erhöhten η -Werten.

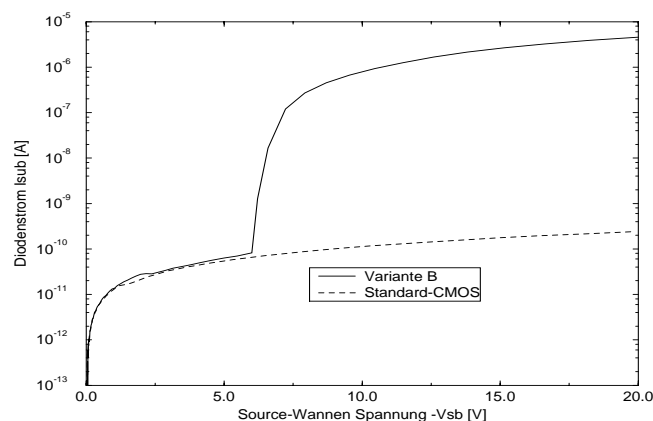


Bild 2.45: Vergleichende Darstellung der Sperrkennlinien der P^+n -Dioden zwischen Source und Wanne in der Variante B und im Vergleichs-CMOS-Prozeß. Die hochdotierte P^+ -Fläche hat eine Entwurfsgröße von $102 \times 28 \mu\text{m}^2$. Die Temperatur während der Messung betrug $T = 36.0 \text{ }^\circ\text{C}$.

Die Sperrkennlinien der P^+n -Dioden sind in Bild 2.45 gezeigt. Bis zu einer Sperrspannung von etwa 6 Volt zeigen beide Dioden ein nahezu identisches Verhalten. Während die Diode auf n-Substrat einen erwarteten Verlauf für weiter zunehmende Spannungen zeigt und bis auf $I_{\text{sub}} = 2 \times 10^{-10} \text{ A}$ bei $V_{\text{sb}} = -20 \text{ V}$ ansteigt, erfährt die Diode der Variante B eine sehr starke Stromzunahme bei etwa 6 V und erreicht Ströme im μA -Bereich bei $V_{\text{sb}} = -20 \text{ V}$. In einer linearen Darstellung dieser Kennlinie wird deutlich, daß der Stromanstieg oberhalb von etwa 7.5 V linear erfolgt und einem differentiellen Parallelleitwert $G_p = 1/2.7 \text{ M}\Omega$ entspricht. Der Verlauf beider Kennlinien unterhalb 6 V - für die Diode des Standard-PMOS auch darüber hinaus - entspricht recht genau in Größe und Anstieg dem Sperrstrom, der bei der N^+p -Diode der Variante C beim NMOS beobachtet wurde. Der Strom scheint auch hier wieder generations-/rekombinationsdominiert zu sein, da der Sättigungsstrom I_{S1} - nach Bild 2.42 - um einige Größenordnungen kleiner als der Sperrstrom ausfällt.

Für den starken Anstieg des Sperrstroms kommen in erster Linie die beiden Möglichkeiten Durchbruch oder Durchgriff in Frage. Obwohl die Modellierung des Effekts oberhalb von 6 V mit einem Parallelleitwert erfolgen kann, spricht gegen einen physikalischen Parallelleitwert die Beobachtung des sprunghaften Einsetzens, und daß unterhalb von 6 V kein erhöhter Strom erkennbar ist.

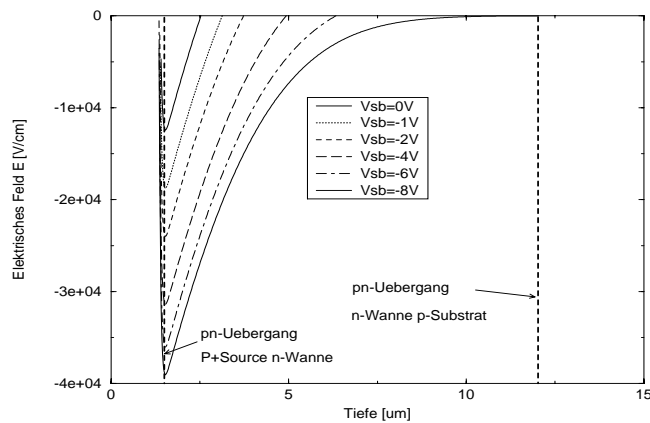


Bild 2.46: Darstellung des durch numerische Integration aus dem simulierten Dotierprofil entstandenen Feldstärkeverlaufs bei Anlegung unterschiedlicher Sperrspannungen an den P⁺-Source-n-Wannen-Übergang der Variante B.

Die numerische Feldberechnung des P⁺n-Übergangs der Variante B (vgl. Bild 2.46), basierend auf dem durch Prozeßsimulation gewonnenen Dotierprofil, zeigt maximale Feldstärken, die um etwa eine Größenordnung unterhalb der kritischen Feldstärke für Lawineneffekte liegen. Aufgrund der schwachen Dotierung der n-Wanne ist mit Durchbrucheffekten erst bei sehr hohen Spannungen zu rechnen, $V_{db} = 65 \text{ V}$ bei $N_D = 8 \cdot 10^{14} / \text{cm}^3$ [2.28].

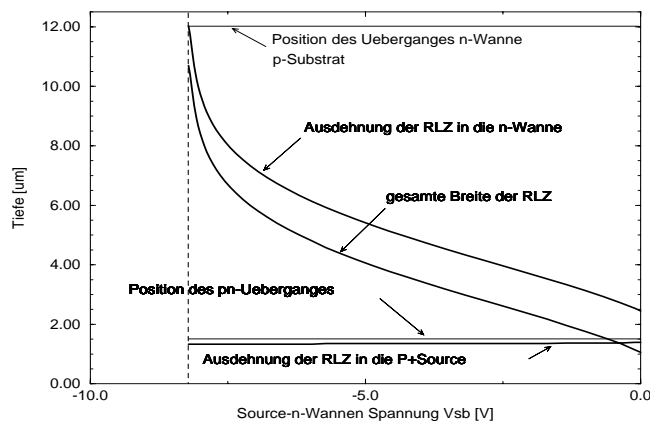


Bild 2.47: Darstellung der räumlichen Ausdehnung der Raumladungszone (RLZ) bei verschiedenen Sperrspannungen. Die Berechnung erfolgt numerisch auf Basis des simulierten Dotierprofils.

Dahingegen ist ein Durchgriff der Raumladungszone durch die n-Wanne sehr wahrscheinlich. Die Ausdehnung der Sperrschicht in Abhängigkeit der äußeren Spannung ist in Bild 2.47 dargestellt. In der numerischen Berechnung erfolgt der komplette Durchgriff bei etwa $V_{sb} = 8 \text{ V}$, also der Spannung, bei welcher der lineare Anstieg der Kennlinie beginnt. Aufgrund der guten Übereinstimmung und der Beobachtung, daß n-Wannen mit verringerter n-Dotierung (d.h. einer nur halb so großen Implantationsdosis) schon bei ca. 5 V durchbrechen und dies ebenfalls durch eine Prozeßsimulation und numerische Berechnung der RLZ-Grenzen bestätigt werden kann, ist ein Durchgriff durch die n-Wanne für den starken Sperrstromanstieg verantwortlich zu machen.

Durchgriffeffekte

Neben den schon im Zusammenhang mit dem NMOS der Variante A (vgl. Abschnitt 2.1.1) besprochenen punch-through-Effekten, kann ein solcher Durchgriff auch in niedrigdotierten Wannen auftreten. Eine Abschätzung von Kasper [2.26] verdeutlicht den Zusammenhang zwischen maximaler Dotierkonzentration, Tiefe des dotierten Gebiets und der Durchgriffspannung V_{pt} . Als Beispiel soll eine P⁺(Drain)-n(Wanne)-p⁻(Substrat) Struktur dienen, diese ist z.B. im Fall des PMOS in der Variante B gegeben.

Wenngleich eine Gaußfunktion proportional zu $\exp(-z^2/a^2)$ die realen Gegebenheiten besser widerspiegelt, soll zur rechnerischen Vereinfachung ein exponentielles Abklingen der n-Wannen-Dotierung mit der Tiefenkoordinate z vorausgesetzt werden:

$$N_D = C_o \exp\left(-\frac{z}{\lambda}\right). \quad (2.41)$$

Hierbei befindet sich der Source-Wannen-Übergang bei $z = 0$ und der Übergang zwischen Wanne und Substrat bei $z = w$. Bei steigender Kanalspannung zwischen Source und Drain vergrößert sich die Raumladungszonenbreite am Drainkontakt und damit erstreckt sich die Verarmungszone tiefer in die sie umgebende Wanne. Bei einer Sperrschichtbreite l bestimmt sich das Potential $\phi(z)$ in der Sperrschicht zu

$$\Phi(z) = \frac{eC_o\lambda^2}{\epsilon_{Si}} \left[\exp\left(-\frac{l}{\lambda}\right) \left(1 + \frac{l}{\lambda} - \frac{z}{\lambda}\right) - \exp\left(-\frac{z}{\lambda}\right) \right] \quad (2.42)$$

mit $\phi(l) = 0$. Ein Durchgriff durch die Wanne erfolgt, wenn die Breite l der Sperrschicht die Tiefe w der Wanne erreicht. Die zugehörige Drain-Source-Spannung V_{ds} bei Erreichen dieser Bedingung wird als Punch-Through-Spannung V_{pt} bezeichnet. Bei Berücksichtigung der Diffusionsspannung des Übergangs U_d folgt damit

$$\frac{\epsilon_{Si}}{eC_o\lambda^2} (U_d - V_{pt}) = 1 - \left[\exp\left(-\frac{w}{\lambda}\right) \right] \left[1 + \frac{w}{\lambda} \right]. \quad (2.43)$$

Aus der Definition des metallurgischen Übergangs bei $z=w$, d.h. $C(z) = N_D(z) = C_{sub}$ mit C_{sub} der Akzeptorenkonzentration des p-Substrats, folgt aus Gleichung (2.41) $C_{sub} = C_o \exp(-w/\lambda)$ und somit

$$\frac{\epsilon_{Si}}{eC_o w^2} \ln\left(\frac{2C_o}{C_{sub}}\right) (U_d - V_{pt}) = 1 - \frac{C_{sub}}{C_o} \left[1 + \ln\left(\frac{C_o}{C_{sub}}\right) \right]. \quad (2.44)$$

Aufgrund der Konstanz der rechten Seite obiger Gleichung folgt für die Durchgriffspannung V_{pt}

$$|V_{pt}| \cong (U_d - V_{pt}) \sim C_o w^2, \quad (2.45)$$

d.h. die Durchgriffspannung kann durch eine höhere Dotierung C_o und deutlich stärker durch eine größere Wanntiefe w erhöht werden.

Basierend auf den numerischen Daten eines beliebigen Dotierkonzentrationsprofils $N_D(z)$ lassen sich die Durchbruchspannungen V_{pt} sehr genau bestimmen. Ein Algorithmus für großflächige Übergänge, deren Verhalten von der Sperrschichtausdehnung in z -Richtung dominiert wird, wurde im Rahmen dieser Arbeit entwickelt und ist in Bild 2.48 dargestellt.

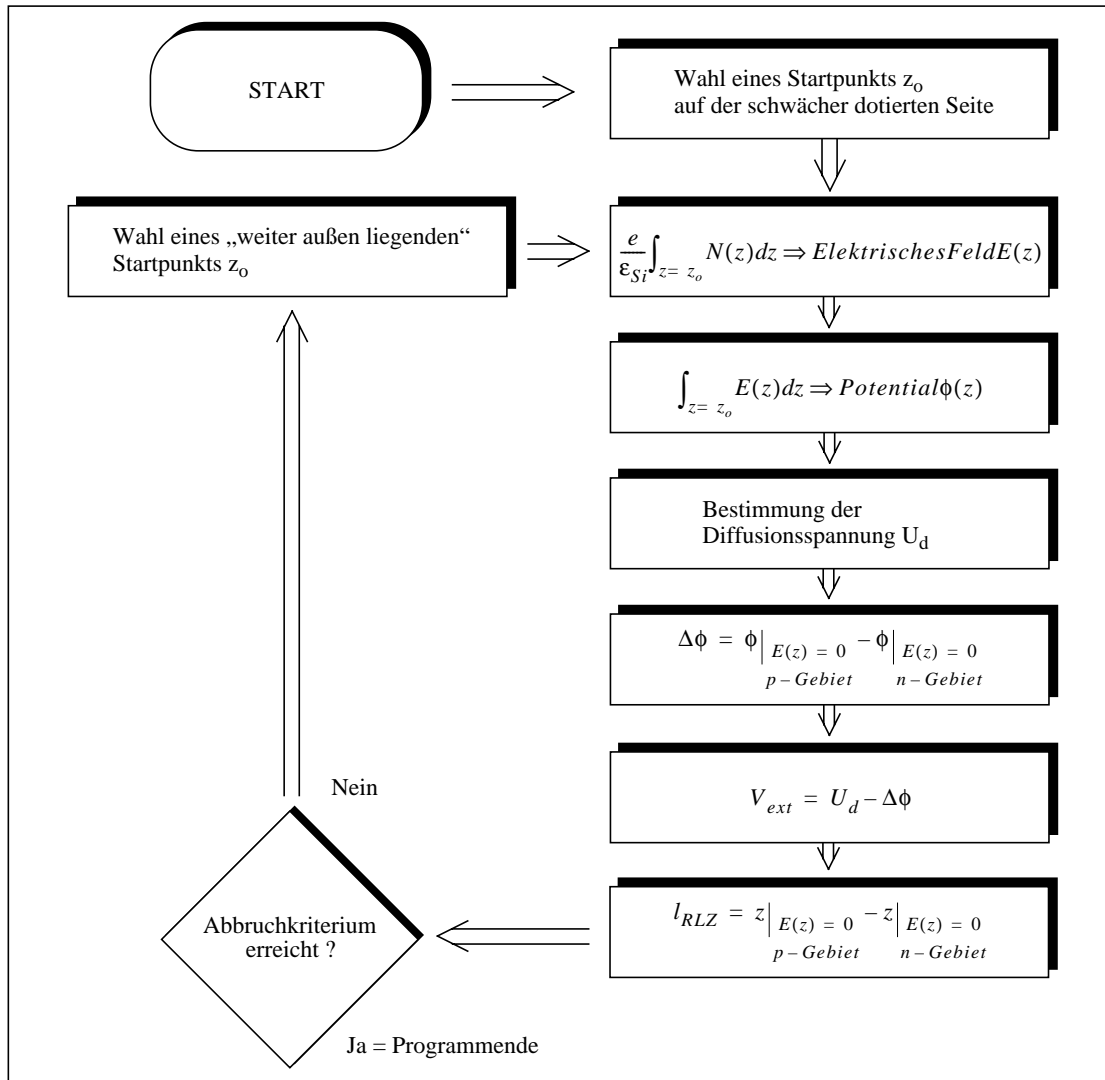


Bild 2.48: Auf numerischer Basis arbeitender Algorithmus zur Bestimmung der Ausdehnung von Sperrschichten in Abhängigkeit der äußeren Spannung bei beliebigen Dotierverläufen. Anhand dieser Ausdehnung kann auf die Durchgriffspannung V_{pt} geschlossen werden.

Beginnend an einem nahe am pn-Übergang auf der schwächer dotierten Seite liegenden ersten Startpunkt durchläuft das Programm mit sukzessive nach außen fortschreitenden Startpunkten den numerischen Teil, der zur Potentialbestimmung dient. In einem solchem Zyklus wird das elektrische Feld $E(z = z_{st})$ und das Potential $\phi(z = z_{st})$ am Startpunkt $z = z_{st}$ mit dem Wert Null angenommen. In einem ersten Durchlauf wird gemäß der Poisson-Beziehung unter Berücksichtigung der Verarmungsnäherung und eindimensionaler Verhältnisse das elektrische Feld $E(z)$ bestimmt. Abbruch-Kriterium für jeden Durchlauf ist der Vorzeichenwechsel des elektrischen Feldes auf der höherdotierten Seite, der die Begrenzung der Raumladungszone darstellt. In einem zweiten Durchlauf erfolgt auf Grundlage des elektrischen Feldes die Berechnung des Potentials $\phi(z)$ und der am Übergang anliegenden Spannung $\Delta\phi$.

Die durch die Prozeßsimulation gegebene nichtlineare Teilung der Tiefenkoordinate z wird bei diesen Integrationen ebenso wie die von der Position der Raumladungsgrenzen abhängige Diffusionsspannung $U_d(z)$ berücksichtigt. Aus der jeweiligen Diffusionsspannung $U_d(z)$ und dem ermittelten Potentialunterschied $\Delta\phi(z)$ wird schließlich die außen anliegende Spannung $V_{ext}(z)$ bestimmt. Damit wird eine Darstellung der Sperrschichtweite in Abhängigkeit der äußeren Spannung für beliebige Dotierprofile $l_{RLZ} = l_{RLZ}(V_{ext})$ möglich. Bei in die Tiefe hin abnehmenden Dotierkonzentrationen $N(z)$ erfolgt in der Nähe der Punch-Through-Spannung V_{pt} ein starker Anstieg der Sperrschichtweite l_{RLZ} (vgl. Bild 2.47); allgemein stellt das Erreichen von $V_{ext} = V_{pt}$ ein automatisches Abbruchkriterium dar.

Eine Gegenüberstellung (vgl. Tabelle 12) der durch numerische Berechnung und Prozeßsimulation ermittelten

Durchgriffspannungen V_{pt} und der nach der Formel von Kasper [2.26] ermittelten Spannungsdifferenzen $U_d - V_{pt}$ verdeutlicht die gute Genauigkeit der mit letzterem Ansatz gewonnenen Werte.

TABELLE 12. Punch-Through-Spannungen

pn-Übergang	$U_d - V_{pt}$ [V] [2.26]	V_{pt} [V] Numerisch Integriert
PMOS (Variante B) P ⁺ -n-Wanne	8	8.2
NMOS (Variante B) N ⁺ -p-Wanne	32	34
NMOS (Variante B) p-Wanne-n-Wanne	0.16	Durchgriff schon bei $V_{ext} = 0$ V
PMOS (Variante C) P ⁺ -n-Wanne	0.2	Durchgriff schon bei $V_{ext} = 0$ V
NMOS (Variante C) N ⁺ -p-Wanne-p ⁻ -Substrat	48	Begrenzung durch Lawinendurchbruch ($V_{pt} \approx 80$ V)

Die Betrachtung des PMOS-Dotierprofils der Variante B (vgl. Abs. 2.1.3 Bild 2.15 und Abs. 2.3.1 Bild 2.40) zeigt, daß der Durchgriff vertikal unterhalb des P⁺-Kontakts erfolgt. Aufgrund der höheren Dotierung an der Oberfläche ist dort ein Durchgriff unwahrscheinlich. Rechnerisch ergibt sich bei der Durchgriffspannung $V_{pt} = 8.2$ V eine Raumladungsweite von $l_{RLZ} = 3$ µm um den P⁺-Drain-Kontakt an der Oberfläche.

Durch das Erreichen des p⁻-Substrats unterhalb des Drains wird eine Vertikalstrecke erzeugt, an der zwar die für den Durchgriff erforderliche Spannung V_{pt} abfällt, die jedoch eine leitfähige Verbindung zwischen P⁺ und p⁻ darstellt. Deshalb wirkt oberhalb der Durchgriffspannung der p⁻-Substrat-n-Wannen-pn-Übergang, der aufgrund der deutlich größeren Querschnittsfläche und des niedrigeren Dotierniveaus zu weit erhöhten Sperrströmen führt.

Wie in Gleichung (2.45) dargelegt, kann eine Erhöhung der Durchgriffspannung V_{pt} durch eine verstärkte Wannendotierung N_D oder eine größere Wannentiefe x_j erfolgen, wobei letzteres aufgrund der in erster Näherung quadratischen Abhängigkeit die wirksamere Maßnahme darstellt. Mittels Simulation wurden bei einer leichten Erhöhung der Implantationsdosis für die n-Wanne und einer um 24 Stunden verlängerte Drive-In-Zeit eine 25 µm tiefe Wanne mit einer Oberflächenkonzentration $C_o = 1.9 \cdot 10^{15} / \text{cm}^3$ und Durchgriffspannungen um $V_{pt} = 35$ V ermittelt. Aufgrund der nahezu identischen Oberflächenkonzentration stellt sich eine sehr ähnliche Schwellspannung V_{th} ein, als Nachteil verbleibt jedoch der überlange Drive-In-Schritt dieser Lösung.

Alternativ kommt die Erhöhung der Oberflächenkonzentration C_o bei nahezu unverändertem Wannen-Drive-In in Frage. Aus

$$V_{th} = V_{FB} - \frac{Q_B + Q_{ss}}{C_{ox}} - \frac{2E_{Fi}}{e} \quad \text{mit} \quad Q_B = \sqrt{4\epsilon_{Si} N_D E_{Fi}} \quad \text{und} \quad E_{Fi} = eU_t \ln \frac{N_D}{n_i} \quad (2.46)$$

folgt jedoch ein deutlicher Anstieg des Betrags der Schwellspannung V_{th} , dem prinzipiell durch die beiden Maßnahmen (i) einer zusätzlichen oberflächlichen, gering dosierten (einige $10^{11} / \text{cm}^2$) Implantation mit einem Akzeptorion (sog. threshold-voltage-shift-implant [2.30]) oder (ii) einer Reduktion der Gateoxiddicke t_{ox} entgegengewirkt werden kann.

Um trotz einer Erhöhung der Dotierkonzentration N_D allein durch eine Reduktion der Gateoxiddicke t_{ox} eine identische Schwellspannung V_{th} des PMOS zu erreichen, folgt unter Berücksichtigung des totalen Differentials der obigen Gleichung

$$dV_{th} = -\frac{Q_B + Q_{ss}}{C_{ox}} \cdot \frac{dt_{ox}}{t_{ox}} - \left[\frac{Q_B}{2C_{ox}} \left(1 + \frac{1}{\ln \frac{N_D}{n_i}} \right) - 2U_t \right] \frac{dN_D}{N_D} \quad (2.47)$$

mit der Bedingung $dV_{th} = 0$ für die erforderliche Änderung der Oxiddicke dt_{ox}

$$\frac{dt_{ox}}{t_{ox}} = -\left(\frac{Q_B + Q_{ox}}{C_{ox}}\right)^{-1} \cdot \left[\frac{Q_B}{2C_{ox}} \left(1 + \frac{1}{\ln \frac{N_D}{n_i}} \right) - 2U_t \right] \frac{dN_D}{N_D} \quad (2.48)$$

bei kleinen Änderungen der Dotierkonzentration dN_D .

Eine mittlere Oxidladungsdichte, z.B. $Q_{ox}/e = 1.0 \cdot 10^{11} / \text{cm}^2$ vorausgesetzt, folgt für die bei einer Durchgriffspannung von $V_{pt} = 20 \text{ V}$ erforderliche Dotierungserhöhung um den Faktor 2.5, d.h. $C_o = 4.5 \cdot 10^{15} / \text{cm}^3$ gegenüber $C_o = 1.8 \cdot 10^{15} / \text{cm}^3$ in Variante B, eine neue Gateoxiddicke von $t_{ox} = 640 \text{ \AA}$, wenn von einer ursprünglichen Dicke von $t_{ox} = 950 \text{ \AA}$ ausgegangen wird. Obige Beziehung gilt jedoch nur für kleine Änderungen, und so ergibt eine genaue Rechnung eine neue erforderliche Oxiddicke von $t_{ox} = 700 \text{ \AA}$.

Gegen die Maßnahme einer Änderung der Gateoxiddicke spricht jedoch der Einfluß auf den Steilheitsparameter β , der umgekehrt proportional zur Gateoxiddicke verläuft. Die Vergleichbarkeit sowohl der Einzeltransistoren und noch weit stärker der Schaltungen wäre stark in Frage gestellt.

Neben einem Wannendurchgriff, der durch eine entsprechende Erhöhung der Wannenkonzentration und -tiefe ausgeglichen werden kann, verbleibt als prinzipielles Problem der in Bild 2.49 gezeigte Effekt eines Durchgriffs zwischen n-Wannen im p⁻-Substrat, sobald sich die Wannen auf unterschiedlichen Potentialen befinden.

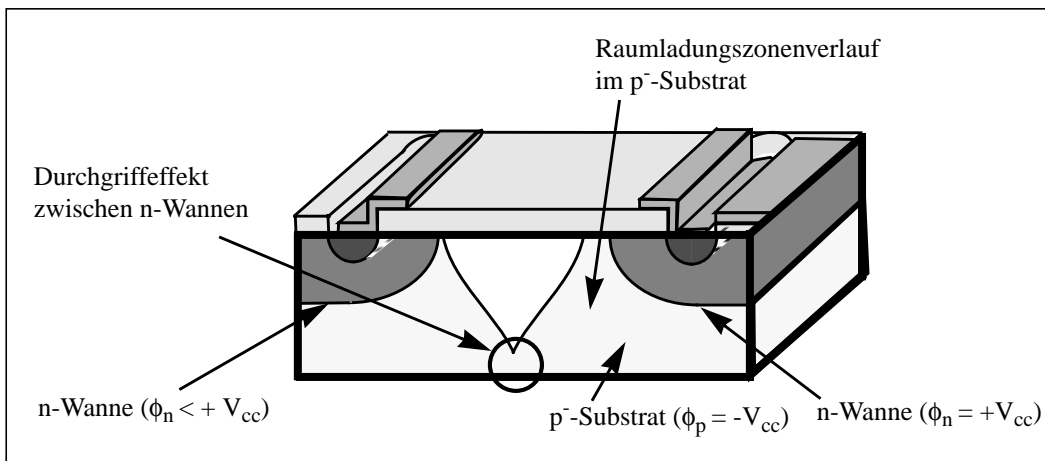


Bild 2.49: Durchgriffeffekt im Volumen bei n-Wannen auf unterschiedlichem Potential. Bei dieser Darstellung ist eine zusätzliche p-Dotierung an der Oberfläche berücksichtigt.

Grundsätzlich entstehen Leckströme zwischen potential-ungleichen p-Wannen auf p-Substrat, da zwischen diesen eine leitfähige Verbindung besteht, jedoch bietet sich bei allen Doppel-Wannen-Prozessen auf höherdotiertem Material eine Bildung von gegenüber dem Substrat umdotierten Wannen auf unterschiedlichem Wannepotential an, sofern der Schaltungsentwurf - wie etwa beim Differenzverstärker üblich - dies erfordert. Bei Transistoren mit hohem Backgatefaktor γ - wie der NMOS der Variante C (vgl. Abs. 2.1.2) - ist ein Unterschied zwischen Source- und Wannepotential mit unerwünschten Einschränkungen bei Spannungshub und -verstärkung verbunden. Auf hochohmigem p⁻-Material entsteht nun zusätzlich die Problematik, daß aufgrund des Durchgriffs zwischen Wannen auch die n-Wannen mit unterschiedlichem Potential faktisch nicht mehr isoliert sind. Eine Unterdrückung dieses Durchgriffs ist nur an der Oberfläche durch höherdotierte p-Bereiche möglich, eine Begrenzung des Volumendurchgriffs in der Tiefe ist hingegen nur mit isolierenden Schichten (z.B. SiO₂, in SOI- insbes. SIMOX-Techniken) erreichbar. Die Beschränkung auf gleichnamiges Wannepotential erfordert Transistoren mit niedrigem Backgatefaktor γ , was durch niedrige Kanaldotierungen - wie etwa im Fall des PMOS der Variante B (vgl. Abs. 2.1.3) - oder dünne Gateoxide zu erreichen ist. In dieser Arbeit ist beim Schaltungsentwurf weitestgehend auf ungleiche Wannepotentiale verzichtet worden.

Temperaturabhängigkeit des Sperrstroms

Die in Bild 2.50 gezeigten Sperrkennlinien von Dioden bei variabler Temperatur zeigen die unterschiedlichen Größen der Minoritätsträgersättigungsstromanteile, die das Sperrverhalten bei erhöhten Temperaturen dominieren. Obwohl die jeweils bei den drei Sperrspannungswerten $V_{\text{sperr}} = 2 \text{ V}$, 3 V und 4 V aufgenommenen Ströme I_{sperr} der N^+p - und der P^+n -Dioden nahezu identisch in der Nähe der Raumtemperatur erscheinen ($-3.2 \cdot 1000 \cdot 1/T$ entspricht ca. $T = 40 \text{ }^\circ\text{C}$), zeigen sich deutliche Unterschiede bei erhöhten Umgebungstemperaturen ($-2.6 \cdot 1000 \cdot 1/T$ entspricht $T = 112 \text{ }^\circ\text{C}$). Die unterschiedlich großen Dioden weisen ein Größenverhältnis zwischen P^+n - und N^+p -Diode von $A_{P^+n}/A_{N^+p} = 362 \cdot 28 \text{ } \mu\text{m}^2 / (202 \cdot 28 \text{ } \mu\text{m}^2) = 1.8:1$ auf, der den verbleibenden Unterschied der Ströme bei Raumtemperatur zu einem Teil erklärt. Aus diesen Messungen wird deutlich, daß das Sperrverhalten der Dioden auf hochohmigem Substrat bei Raumtemperatur sehr stark durch allenfalls linear mit der Temperatur ansteigende Stromanteile dominiert wird. Dies können in erster Linie Rekombinationsstromanteile oder auch Ströme über einen parallelen ohmschen Leitwert sein. Das Verhalten bei höheren Temperaturen zeigt hingegen die aus der Vorwärtskennlinie erwartete Größe des Sättigungsstroms.

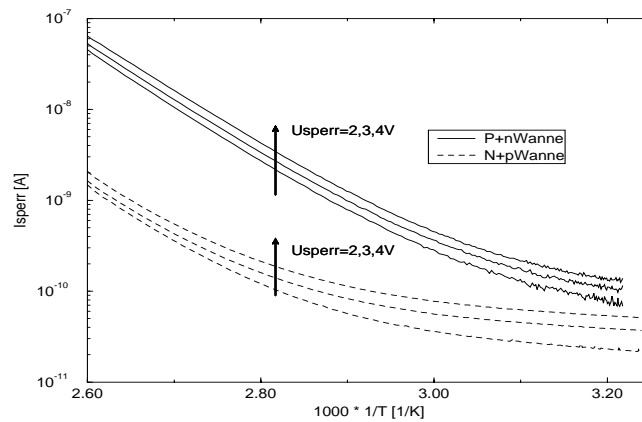


Bild 2.50: Abhängigkeit des Sperrstroms der pn-Übergänge auf hochohmigem Silizium von der Umgebungstemperatur. Die Darstellung zeigt jeweils den Sperrstrom bei $V_{\text{sperr}} = 2 \text{ V}$, 3 V und 4 V . Die Diode des PMOS der Variante B (P^+n -Wanne) hat die Entwurfsmaße $A = 362 \cdot 28 \text{ } \mu\text{m}^2$, die des NMOS der Variante C (N^+p -Wanne) $A = 202 \cdot 28 \text{ } \mu\text{m}^2$.

Betrachtet man den Sättigungsstrom bestehend aus Diffusionsstromanteil I_{s1} und Rekombinations-/Generationsanteil I_{s2} (vgl. Abs. 2.3.1 Gleichungen (2.22) und (2.23)), so ergibt sich eine Abhängigkeit von der stark temperaturabhängigen intrinsischen Dichte n_i gemäß der Form

$$I_s = I_{s1} + I_{s2} = c \cdot n_i^2 + d \cdot n_i, \quad (2.49)$$

worin die Konstanten c und d jeweils weitgehend temperaturunabhängig sind. Aus dieser Beziehung ist ersichtlich, daß der Diffusionsstromanteil I_{s1} ab einer bestimmten Temperatur die dominierende Größe darstellt. Da Silizium bei Raumtemperatur eine deutlich geringere intrinsische Dichte ($n_{i,\text{Si}} = 1.5 \cdot 10^{10} / \text{cm}^3$ bei $T = 300 \text{ K}$) als etwa Germanium aufweist ($n_{i,\text{Ge}} = 1.0 \cdot 10^{13} / \text{cm}^3$ bei $T = 300 \text{ K}$), ist dieser Übergang von rekombinations-/generationsdominiertem zu diffusionsbestimmtem Strom üblicherweise oberhalb der Raumtemperatur zu erwarten. Der Übergang wird in Bild 2.50 durch die erhöhte Kurvensteigung bei etwa $2.8 \cdot 1000 \cdot 1/T$ erreicht.

Die Temperaturabhängigkeit der quadratischen intrinsischen Trägerdichte

$$n_i^2 = N_C N_V e^{-\frac{E_{\text{gap}}}{kT}} = c' \cdot T^3 \cdot e^{-\frac{E_{\text{gap}}(T)}{kT}} \quad (2.50)$$

besteht aus den Anteilen der effektiven Zustandsdichte N_V und N_C , die jeweils proportional zu $T^{3/2}$ ansteigen,

der direkten Abhängigkeit des Exponentialterms durch kT und der leicht temperaturabhängigen Größe der Bandlücke E_{gap} . Die intrinsische Trägerdichte n_i ist in Bild 2.51 links als Funktion der absoluten Temperatur T dargestellt.

Bei alleiniger Betrachtung der dominierenden exponentiellen Abhängigkeit durch kT ergibt sich

$$\frac{d(\ln(I_s))}{d\left(\frac{1}{T}\right)} = \frac{d(\ln(n_i^2))}{d\left(\frac{1}{T}\right)} = -E_{gap} \cdot \frac{1}{k} \quad \text{mit} \quad E_{gap} \quad \text{in} \quad eV, \quad (2.51)$$

so daß die Steigung im Arrhenius-Plot (Bild 2.50) proportional zur Bandlücke E_{gap} sein sollte. Eine Berücksichtigung der weiteren Terme sorgt für geringe Abweichungen, so daß sich die Steigung wie in Bild 2.51 rechts ergibt. Für eine inverse Temperatur $1/T = 2.6 \cdot 10^{-3} \text{ 1/K}$ folgt eine Steigung $d(\ln n_i^2)/d(1/T) = -13100 \text{ K}$, die von den beiden Meßkurven in Bild 2.50 sehr genau erreicht wird. Demzufolge ist davon auszugehen, daß der Sättigungsstrom bei dieser Temperatur durch den von n_i^2 abhängigen Diffusionsstromanteil bestimmt wird.

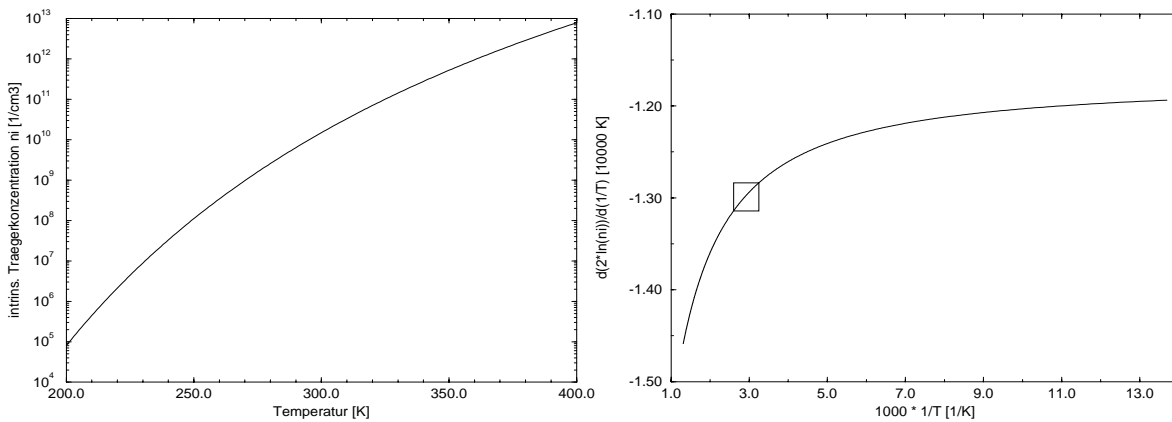


Bild 2.51: Links Darstellung der intrinsischen Trägerdichte n_i als Funktion der Temperatur. Rechts Darstellung der Ableitung des natürlichen Logarithmus von n_i^2 . Das Rechteck markiert die Steigungsvariation für den Temperaturbereich der Messung (vgl. Bild 2.50).

Eine Umrechnung der bei dieser Temperatur ermittelten Stromwerte $I_{s,P+n} = 43 \text{ nA}$ und $I_{s,N+p} = 1.35 \text{ nA}$ bei $1/T = 2.6 \cdot 10^{-3} \text{ 1/K}$ ($n_{i,1/T} = 2.6 \cdot 10^{-3} \text{ 1/K} = 3.6 \cdot 10^{12} / \text{cm}^3$) auf Raumtemperatur unter Berücksichtigung der unterschiedlichen Querschnittsflächen ergibt bei reiner Flächenabhängigkeit des Sperrstroms $I_{s,P+n} = 214 \text{ fA}$ und $I_{s,N+p} = 12 \text{ fA}$ für eine Diode der Größe $A = 102 \cdot 28 \text{ } \mu\text{m}^2$. Dieser Wert entspricht im Fall der N^+p -Diode sehr genau der aus der Vorwärtskennlinie (vgl. Bild 2.33) bestimmten Sättigungsstromdichte. Für die P^+n -Diode wird ein etwa vierfach höherer Wert ermittelt (vgl. Bild 2.42), der zum Teil aus der niedrigeren Dotierungsdichte $N_D = 5 \cdot 10^{14} / \text{cm}^3$ der n-Wanne bei $V_{sperr} = -2 \text{ V}$ gegenüber $N_D = 8 \cdot 10^{14} / \text{cm}^3$ bei $V_{sperr} = 0 \text{ V}$ (vgl. Bild 2.40 und Bild 2.47) erklärbar ist. Auf die Problematik der Vernachlässigung der Sättigungsstromabhängigkeit vom Umfang und der Annahme einer reinen Flächenproportionalität wird im folgenden Abschnitt eingegangen.

Sperrstromabhängigkeit von Umfang und Fläche des pn-Übergangs

Zur Untersuchung der Abhängigkeit des Sperrstroms I_{sperr} von der Fläche F und dem Umfang U des pn-Übergangs sind großflächige Dioden mit Hilfe eines speziellen Maskenentwurfs prozessiert worden. Beginnend mit quadratischen Strukturen mit pn-Sperrschichtflächen von $100 \cdot 100 \text{ } \mu\text{m}^2$ stehen in Abstufungen quadratische Strukturen bis $850 \cdot 850 \text{ } \mu\text{m}^2$ sowie rechteckige Strukturen bis $100 \cdot 1700 \text{ } \mu\text{m}^2$ zur Verfügung. Die Dotierungsprofile der Dioden sind vertikal identisch zu den Source- und Draingebieten der Varianten B und C ausgelegt. Die Kontaktierung der Wannen erfolgt über gleichnamige hochdotierte Gebiete.

Aufgrund der deutlich unterschiedlichen Flächen-zu-Umfangsverhältnisse kann mit Hilfe dieser Dioden die gesuchte Stromabhängigkeit von Fläche und Umfang ermittelt werden. Die Sperrstromanteile eines pn-Übergangs können in einen flächenabhängigen Anteil mit der Proportionalitätskonstante C_F und einen umfangabhängigen Anteil mit C_U aufgeteilt werden. Da sowohl die Fläche als auch der Umfang aufgrund der spannungsabhängigen

Breite der Sperrschicht beeinflusst werden und Vergrößerungen der realen Struktur gegenüber den Entwurfsmaßen durch die Prozessierung auftreten, wird ein zusätzlicher konstanter Anteil C_o hinzugefügt. Mittels dieses konstanten Anteils ist auch ein möglicher Strom über einen Paralleleitwert darstellbar. Durch den in zwei Dimensionen linearen Ansatz

$$I_{sperr} = C_F \cdot F + C_U \cdot U + C_o \quad (2.52)$$

soll eine Regressionsanalyse mit den meßtechnisch ermittelten Sperrstromwerten I_{sperr} bei einer festen Spannung $V_{sperr,o}$ erfolgen.

$$\text{Min}(f(C_F, C_U, C_o)) \quad \text{mit} \quad f(C_F, C_U, C_o) = \sum_{i=1}^n [I_{sperr} - C_F \cdot F - C_U \cdot U - C_o]^2 \quad (2.53)$$

Ziel ist hierbei die Minimierung des quadratischen Fehlers zwischen den Meßergebnissen und der gesuchten Regressionssebene.

$$\frac{\delta f(C_F, C_U, C_o)}{\delta C_F} = 0, \quad \frac{\delta f(C_F, C_U, C_o)}{\delta C_U} = 0 \quad \text{und} \quad \frac{\delta f(C_F, C_U, C_o)}{\delta C_o} = 0. \quad (2.54)$$

Die Auflösung dieser Extremwertbedingungen führt auf die Bestimmungsgleichungen für die Proportionalitätskonstanten der Fläche C_F , des Umfangs C_U und der Konstanten C_o .

Tabelle 13 zeigt die mit diesem Ansatz ermittelten Parameter für die Dioden P^+n der Variante B und N^+p der Variante C. Die Sperrströme wurden bei einer Sperrspannung $V_{sperr,o} = 2.5$ V aufgenommen.

TABELLE 13. Sperrstrom-Flächen- und Umfangsabhängigkeit

	p^+n	n^+p
C_F [fA/ μm^2]	1.62	1.86
C_U [fA/ μm]	50.7	49.2
C_o [pA]	4.57	3.82
C_U/C_F [μm]	31.3	25.4

Regressionsparameter für $I_{sperr}(V_{sperr} = 2.5$ V)

Bei den üblichen Source- bzw. Drain-Wannen-Dioden (mit den Maßen $A = 102 \times 28 \mu\text{m}^2$ bei einem MOS mit $W/L = 100 \mu\text{m} / 10 \mu\text{m}$) zeigt sich eine deutliche Dominanz des umfangsabhängigen Sperrstromanteils, der einen gegenüber der Fläche um zweieinhalb- bis dreifach höheren Anteil am Gesamtstrom aufweist. Dies wird auch in Bild 2.52 deutlich, in dem der relative Anteil der Umfangs- bzw. Flächenkomponente als Funktion des Flächen- zu Umfangsverhältnisses F/U mit dem Verhältnis der Proportionalitätskonstanten C_U/C_F als Parameter dargestellt wird.

Der obige MOS weist ein Verhältnis $F/U = 11 \mu\text{m}$ auf. Bei den hier ermittelten Werten von $C_U/C_F = 31 \mu\text{m}$ bei P^+n - und $C_U/C_F = 25 \mu\text{m}$ bei N^+p -Dioden liegt der Anteil des umfangsabhängigen Sperrstroms bei rund 70 %, wenn der konstante Anteil C_o unberücksichtigt bleibt. Die Vernachlässigung von C_o beeinflusst zwar den relativen Anteil am Gesamtstrom, nicht aber die relative Größe der beiden Anteile $C_U U$ und $C_F F$ zueinander.

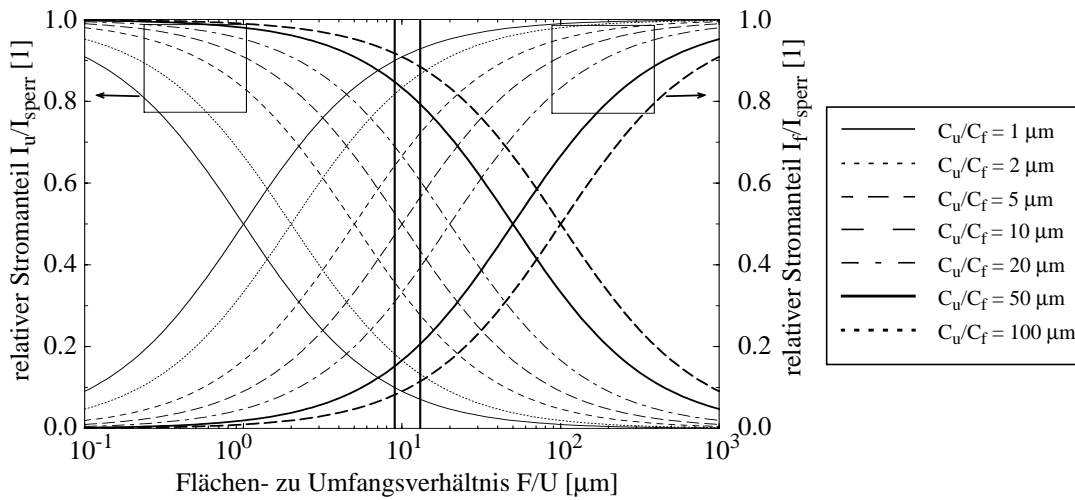


Bild 2.52: Darstellung der Flächen- bzw. Umfangsanteile des Sperrstroms von pn-Dioden als Funktion des Flächen- zu Umfangsverhältnisses. Die beiden senkrechten Linien bei $F/U = 9 \mu\text{m}$ und $F/U = 13 \mu\text{m}$ markieren die Flächen- zu Umfangsverhältnisse der MOS-Transistoren, die im Rahmen dieser Arbeit benutzt worden sind.

Dieser große Anteil des vom Umfang abhängigen Sperrstroms bedeutet, daß bei den eindimensionalen Ansätzen zur Sperrstrombestimmung, die eine reine Flächenproportionalität unterstellen, mit Abweichungen zu den meßtechnischen Resultaten zu rechnen ist.

2.4 pin-Dioden für Schalter

Pin-Dioden lassen sich ohne Beeinflussung der CMOS-Prozeßfolge auf dem SIMMWIC-CMOS-Chip herstellen. Die für die Bildung der Source- und Drain-Gebiete notwendigen N^+ - und P^+ -Diffusionen dienen gleichfalls zur Definition der p- und n-Zone dieser Dioden. Das durch das Feldoxid geschützte hochohmige p^- -Substrat bildet hierbei die intrinsische Zone.

Während die im vorherigen besprochenen pn-Dioden im wesentlichen vertikale Stromkomponenten am Übergang aufweisen, findet bei diesen planaren pin-Dioden der Stromtransport in horizontaler Richtung statt. Die hier untersuchten Dioden haben jeweils identische N^+ - und P^+ -Kontakte mit entweder $W_{N^+,P^+} = 18 \mu\text{m}$ oder $W_{N^+,P^+} = 38 \mu\text{m}$ Breite und einer konstanten Länge von $L_{N^+,P^+} = 28 \mu\text{m}$. Dazwischen liegt die i-Zone mit einer Breite $W_i = W_{N^+,P^+} = 18 \mu\text{m}$ oder $38 \mu\text{m}$ und einer Länge L_i . Da sich die Breite der hochdotierten Anschlüsse aus den Maßen der i-Zone ergibt und deren Länge konstant ist, wird in den folgenden Schaubildern nur jeweils auf die Maße der i-Zone Bezug genommen.

Im Rahmen dieser Arbeit wird nicht die niedrige Kapazität und hohe Durchbruchspannung dieser Dioden im Sperrbetrieb ausgenutzt, sondern die Eigenschaften der Diode im Vorwärtsbetrieb im Hinblick auf Anwendungen als Schalter für Hochfrequenz.

Wie bei gewöhnlichen pn-Dioden besteht auch hier der Diodenstrom aus den beiden Anteilen aufgrund der Diffusion- ($\eta = 1$) und der Generations- bzw. Rekombinationseffekte ($\eta = 2$). Bedingt durch die hohe Dotierung und damit geringe Minoritätsträgerdichte sowohl der n- als auch der p-Gebiete ist jedoch der Diffusionsstromanteil vernachlässigbar, so daß sich bei eindimensionaler Rechnung gemäß [2.31] und [2.32] ein Diodenstrom von

$$I_f = \frac{4eAn_iD_aF_L}{L_i} \cdot e^{\frac{V_f}{2U_i}} \tag{2.55}$$

ergibt. Da sowohl Löcher aus dem p- als auch Elektronen aus dem n-Gebiet im Flußbetrieb in die i-Zone injiziert werden, wird der Strom durch die ambipolare Diffusionskonstante $D_a = 2D_pD_n/(D_p+D_n)$ bestimmt. Der Faktor F_L hat eine Größe zwischen 0.01 und 0.3 und wird durch die Länge der i-Zone L_i und die Trägerlebensdauer in der i-Zone τ bestimmt [2.33].

Sobald nun die Frequenz f_{RF} des an der Diode liegenden hochfrequenten AC-Signals deutlich höher als die durch die Lebensdauer der Träger bestimmte Frequenz $f_0 = 1/(2\pi\tau)$ ($\tau = 4 \mu\text{s}$ führt auf $f_0 = 40 \text{ kHz}$, vgl. Anhang A4) wird, können die Ladungsträger in der i-Zone nicht länger dem äußeren Signal durch Rekombination und Generation vollständig folgen. Die Folge ist ein Unterbleiben der Gleichrichtung und Signalverzerrung des hochfrequenten Signals, da sich die Diode wie ein durch den Gleichstrom I_f steuerbarer Widerstand verhält. Dieser Widerstand beträgt [2.34]

$$R_{RF} = \rho \frac{L_i}{A} = \frac{L_i^2}{I_f \tau (\mu_n + \mu_p)} . \quad (2.56)$$

Diese Eigenschaft kann in Hochfrequenzschaltern und -modulatoren ausgenutzt werden, so z.B. im pin-Schalter [2.35], bei dem ein hochfrequentes Eingangssignal in jeweils einen der beiden Ausgangszweige geschaltet werden kann. In jedem der beiden Zweige ist eine pin-Diode parallel zur Leitungsimpedanz Z_0 geschaltet, die durch den hohen Widerstand der pin-Diode im Sperrbetrieb nahezu überhaupt nicht beeinflusst wird. Sorgt nun ein Vorwärtsbias für einen Stromfluß in der Diode, so kann schon mit Strömen $I_f = 1 \dots 10 \text{ mA}$ der Hochfrequenzwiderstand R_{RF} der Diode in den Bereich einiger Ohm gebracht werden. Die Folge davon ist ein Absinken der Gesamtimpedanz $Z_{\text{ges}}^{-1} = Z_0^{-1} + R_{RF}^{-1}$ des jeweiligen Zweigs und damit eine Fehlanpassung, die zu einer verringerten Transmission und erhöhten Reflexion des hochfrequenten Signals führt und damit diesen Zweig faktisch entkoppelt.

Die stetige Änderung des Hochfrequenzwiderstands kann ebenso zur Modulation von Signalen benutzt werden, wengleich durch die Ladung der Träger in der i-Zone eine schnelle Schaltzeit kaum zu erreichen ist. Die durch das Ausräumen dieser Träger erforderliche Verzögerungszeit t_d beträgt [2.23]

$$t_d = \tau \cdot \ln\left(1 + \frac{I_f}{I_r}\right) \approx \tau \cdot \frac{I_f}{I_r} \quad \text{falls} \quad I_r \gg I_f, \quad (2.57)$$

worin I_r der nach dem Umschalten auf Sperrbetrieb fließende Strom ist. Diese Zeit verlängert sich noch - wie auch bei pn-Dioden üblich - um die sogenannte Übergangszeit t_{tr} , in der die komplette Sperrwirkung erst allmählich erreicht wird. Schaltzeiten unterhalb einiger 100 ns sind damit nur mit ultrakurzen Lebensdauern τ zu erreichen, die jedoch dann zu erhöhten Strömen im Sperrbetrieb führen.

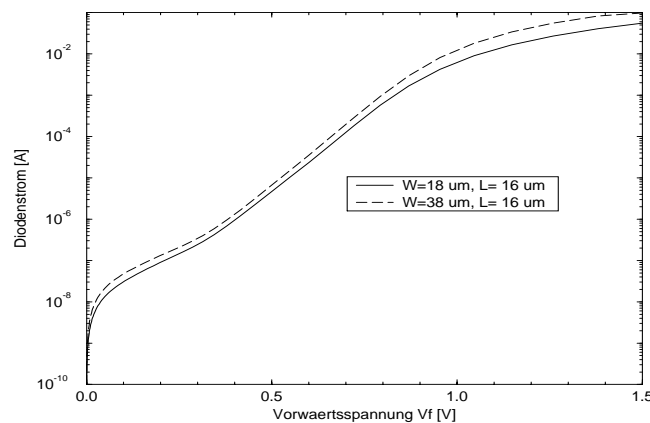


Bild 2.53: Vorwärtskennlinie von pin-Dioden auf hochohmigem p-Silizium. Die angegebenen Maße bedeuten die Breite (W) und die Länge (L) des i-Gebiets zwischen den N^+ und P^+ -Kontakten.

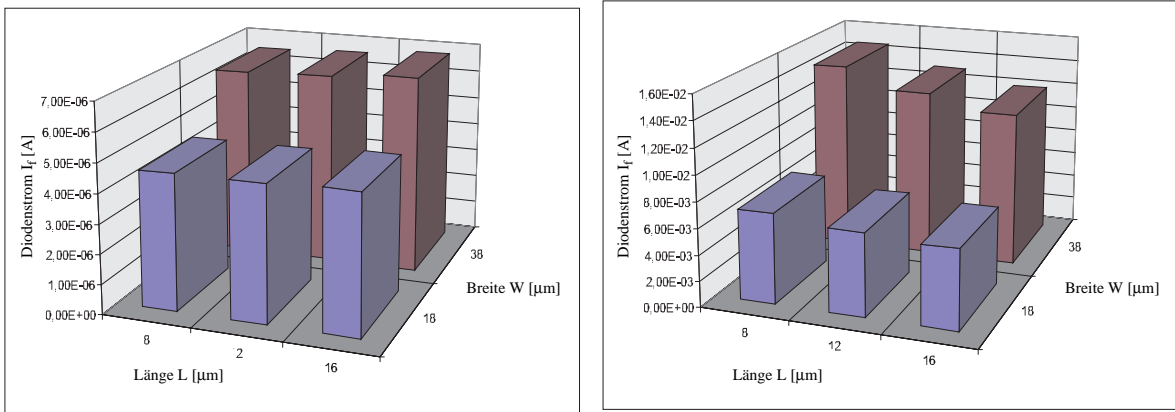


Bild 2.54: Darstellung der Abhängigkeit des Vorwärtsstroms von pin-Dioden als Funktion von Breite und Länge des i-Gebiets. Links Stromwerte bei einer Durchlaßspannung $V_f = 0.5$ V, rechts $V_f = 1.0$ V.

In Bild 2.53 ist die Vorwärtskennlinie zweier lateraler pin-Dioden mit gleicher Länge der i-Zone $L_i = 16 \mu\text{m}$ dargestellt. Die unterschiedliche Breite der n- und p-Gebiete führt zu einer unterproportionalen Erhöhung des Stroms I_f . Durch die auch bei kleinen Breiten W vorhandene Aufweitung der i-Zone ist der den Stromfluß tragende Querschnitt größer als durch die hochdotierten Gebiete vorgegeben.

Die in Bild 2.54 aufgetragenen Stromwerte bei den festen Spannungswerten $V_f = 0.5$ V (links im Bild) und $V_f = 1.0$ V (rechts im Bild) zeigen die Abhängigkeit des Stroms I_f von unterschiedlichen Breiten $W_i = 18 \mu\text{m}$ und $W_i = 38 \mu\text{m}$ sowie Längen $L_i = 8, 12$ und $16 \mu\text{m}$ der i-Zone. Bei mittleren Spannungen, $V_f = 0.5$ V, bei denen die maximale Steilheit der Kennlinie und rekombinationsdominiertes Verhalten beobachtet werden (vgl. Bild 2.53), besteht bezüglich der Länge L_i der i-Zone weitgehend eine Konstanz des Stroms I_f ; lediglich die Erhöhung der Breite W_i führt zu einem deutlich größeren Strom. Die oben beschriebene inverse Abhängigkeit des Stroms von der Länge der i-Zone trifft hier nicht zu, da in der Ableitung der Formel (2.55) ein eindimensionaler Stromfluß vorausgesetzt wird. Der Stromfluß dieser Dioden verläuft jedoch weitgehend dreidimensional. Sowohl die durch die benachbarten Ränder der n- und p-Gebiete oberflächennah in die i-Zone injizierten Träger als auch die vertikal nach unten durch die konstante Fläche des hochdotierten Gebiets geführten Stromanteile sind signifikant.

Die für $V_f = 1.0$ V im Bereich beginnender Serienwiderstandseinflüsse beobachteten Ströme zeigen im Fall der breiteren Dioden eine geringe Abhängigkeit des Stroms von der i-Zonen-Länge, nicht aber bei den schmaleren Dioden.

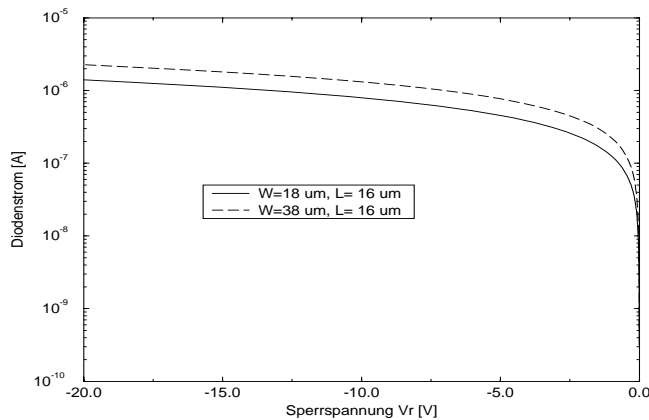


Bild 2.55: Rückwärtskennlinie von pin-Dioden auf hochohmigem p-Silizium.

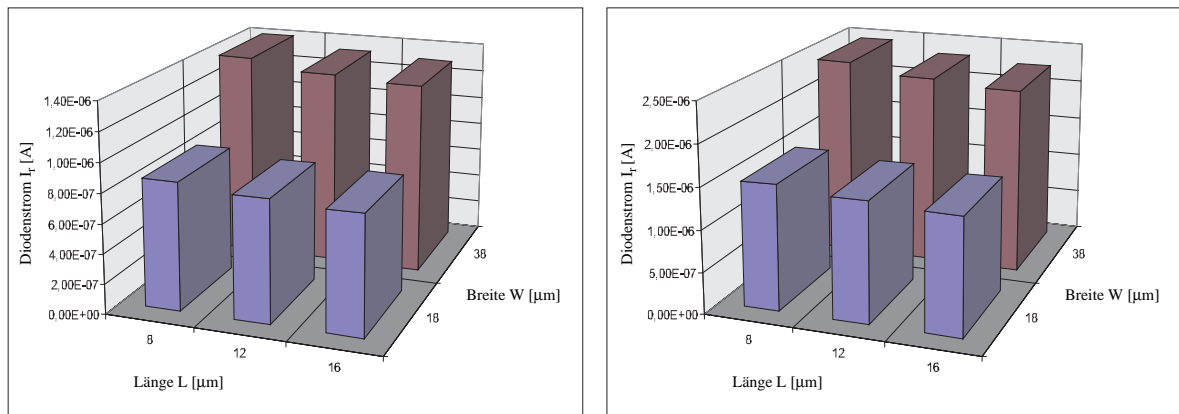


Bild 2.56: Darstellung der Abhängigkeit des Rückwärtsstroms von pin-Dioden als Funktion von Breite und Länge des i-Gebiets. Links Stromwerte bei einer Sperrspannung $V_r = 10$ V, rechts $V_r = 20$ V.

Die in Bild 2.55 gezeigte Sperrkennlinie zeigt - wie erwartet - eine signifikante Abhängigkeit des Stroms von der i-Zonen-Breite W . Der in Bild 2.56 gezeigte Vergleich der Ströme bei $V_r = 10$ V (links) und $V_r = 20$ V (rechts) zeigt bei den schmalen Dioden nahezu keinen Einfluß der i-Zonen-Länge und einen nur schwachen bei den breiteren Dioden. Aufgrund der im gesamten Bereich der i-Zone bis an die Oberfläche reichenden Sperrschicht ist eine starke Oberflächengeneration von Trägern im Sperrbetrieb zu erwarten und damit ein vergleichsweise hoher Sperrstrom der lateralen pin-Dioden.

Hybridtechnik mit Multi-Chip-Modul

Für eine Kombination von SIMMWIC- und CMOS-Schaltungen kommt neben der angestrebten monolithischen Integration grundsätzlich auch die hybride Zusammenschaltung in Frage. Im Rahmen des dieser Arbeit zugrunde liegenden BMBF-Projekts [3.1] bildete der im folgenden dargestellte hybride Aufbau eines SIMMWIC-Front-Ends mit einem CMOS-Vorverstärker einen wichtigen Meilenstein. Die hybride Kombination von Chips mit unterschiedlichen Herstellungstechnologien kann insbesondere in Fällen zum Einsatz kommen, in denen aufgrund zu geringer Stückzahlen eine monolithische Serienfertigung außer Frage steht.

3.1 Unterschiede zwischen Hybridtechnik und monolithischer Integration

Integrierte Millimeterwellenschaltungen auf Silizium-Substrat in SIMMWIC-Technologie (engl. Silicon Monolithic Millimeter-Wave Integrated Circuits) haben in den letzten Jahren als Alternative zu den bisher dominierenden III-V-MMICs, z.B. auf Gallium-Arsenid (GaAs) basierend, große Fortschritte gemacht. Die Möglichkeiten der heutigen Halbleitertechnologie erlauben die Prozessierung von hochohmigen Silizium-Wafern mit spezifischen Widerständen bis 10 k Ω cm ohne signifikante Beeinträchtigung der Leitfähigkeit. Dadurch wird die Nutzung dieses Substratmaterials für Wellenleiterstrukturen, beispielsweise in Microstrip- oder Koplanartechnologie, ermöglicht. Die Verfügbarkeit von aktiven Elementen wie Heterobipolartransistoren (HBT), IMPATT- und Schottky-Dioden in Silizium-Germanium (SiGe)-Technik ermöglicht Schaltungen mit Frequenzen bis etwa 100 GHz und umfaßt somit auch die kommerziell sehr interessanten Frequenzbänder der Nahbereichskommunikation und Automobilsensorik bei 27 GHz, 77 GHz und 93 GHz. Verschiedene Sende- und Empfangsschaltungen bis 100 GHz sowie HF-Mixer in SIMMWIC-Technologie sind von einer Daimler-Benz-Forscherguppe entwickelt und vorgestellt worden [3.2].

Für die Herstellung der passiven Wellenleiterstrukturen kommen bei MMICs überwiegend die beiden Optionen Mikrostreifen- oder Koplanartechnik zum Einsatz. Bei Mikrostreifenleitungen mit Goldmetallisierung sind auf 195 μ m dickem hochohmigem Siliziumsubstrat minimale Dämpfungswerte von 0.6 dB/cm erzielt worden [3.3]. Durch die gegenüber der Vakuumwellenlänge λ_0 reduzierte Substratwellenlänge $\lambda_{Si} = \lambda_0/\epsilon_{re}^{1/2}$ lassen sich somit - auf die Wellenlänge bezogen - extrem niedrige Dämpfungen erzielen. Bei einer Frequenz von $f = 90$ GHz und einer Microstrip-Konfiguration mit dem Geometrieverhältnis $w/h = 0.8$ folgt eine Wellenlänge $\lambda_{Si} \approx 1$ mm.

Bei der Herstellung von MMICs mit den Methoden der Halbleitertechnologie bietet die Koplanartechnik einige Vorteile aufgrund verringerter Prozeßkomplexität. Die von der Substrathöhe h weitgehend unabhängigen Wellenleiterstrukturen erlauben eine direkte Integration der aktiven Elemente und kommen ohne Durchkontaktierungen, sog. Via-Holes, und Rückseitenmetallisierung aus. Nachteilig ist die gegenüber Mikrostreifenleitungen erhöhte Leitungsdämpfung. Die bei Mikrostreifenstrukturen erforderliche Dünnung der Wafer und die Bildung von Via-Holes sind jedoch ebenso wie die Luftbrückenherstellung in Koplanartechnik als Abwandlung konventioneller Halbleiterprozeßtechnik Bestandteil der SIMMWIC-Technologie [3.4]. Nach der halbleitertechnischen Prozessierung erlauben chemisch-mechanische Polierverfahren die Dünnung der Wafer auf die bei Microstrip-technik erforderliche Substrathöhe, die üblicherweise im Bereich von $h = 100$ μ m bis $h = 200$ μ m liegt. Bei dem zum Wellenwiderstand $Z_0 = 50$ Ω gehörenden Geometrieverhältnis $w/h = 0.8$ ergibt sich damit eine Leiterbreite von $w = 80$ μ m bis $w = 160$ μ m. Diese ist ausreichend klein, um auch komplexere Hochfrequenzschaltungen integrieren zu können.

Die Hybridtechnik erlaubt nun - unabhängig von der gewählten Wellenleitertechnik - die direkte Ausnutzung dieser SIMMWIC-spezifischen Eigenschaften. Die Kombination der SIMMWIC- und CMOS-Schaltungen kann erfolgen, ohne daß eine nennenswerte Beeinflussung der Hochfrequenzeigenschaften auftritt. Zudem kann aufgrund der separaten Prozessierung der einzelnen Chips jede Schaltungstechnologie weitgehend unabhängig von der anderen optimiert werden. Durch die im Vergleich zur monolithischen Integration größere räumliche Entfernung der Chips zueinander steigt zwar der gesamte Flächenbedarf, jedoch ist eine gegenseitige elektrische Beeinträchtigung nahezu ausgeschlossen.

Die Option, eine separate Prozessierung der SIMMWIC- und der CMOS-Chips vornehmen zu können, ist solange von Vorteil wie die eingesparte Prozeßkomplexität den höheren Aufwand des Die- und Wire-Bonding überwiegt. So ist beim Einsatz von Mikrostreifenleitungen im SIMMWIC-Teil eine mechanische Dünnung des Substrats i.a. unumgänglich. Für den CMOS-Bereich ist eine solche Dünnung zwar unkritisch, aber auch unnötig und kann deshalb bei einer hybriden Kombination entfallen.

Ferner verursachen die stark unterschiedlichen Temperaturbudgets der beiden Technologien Schwierigkeiten bei der angestrebten gemeinsamen Nutzung von Prozeßmodulen im Integrationsprozeß für SIMMWIC und CMOS. Einzig die Einbringung der vergrabenen Schichten (engl. buried layer) für die Schottky-Dioden ist direkt in den CMOS-Prozeß integrierbar. Das auch bei der CMOS-Technologie bestehende Bestreben hin zu geringeren Temperaturbudgets, z.B. durch Wahl einer kurzzeitigen Source- und Drainaktivierung mittels Rapid-Thermal-Annealing (RTA), verspricht eine vereinfachte gemeinsame Ausnutzung von Prozeßschritten.

Während bei der hybriden Kombination die Verwendung von beliebigen HF-Schaltungstechnologien weitgehend unproblematisch ist, ergeben sich bei der monolithischen Integration mit Mikrostreifenleitungen Schwierigkeiten, da die Herstellung von Durchkontaktierungen und Rückseitenmetallisierungen besonderen Aufwand bedeutet.

Der Einsatz von Koplanarleitungen, insbesondere bei Schaltungen mit aktiven Elementen wie Oszillatoren und HF-Mischern [3.5], vermindert allerdings durch den Verzicht auf diese Schritte den Aufwand bei Integrationsprozessen.

Problematisch ist und bleibt jedoch die Vereinbarung der unterschiedlichen Metallisierungsanforderungen von SIMMWIC-Schaltungen mit Gold und CMOS-Schaltungen mit Aluminium. Das in der CMOS-Technik eingesetzte Aluminium mit Metallstärken $t_{\text{Alu}} = 1 \mu\text{m}$ erscheint aufgrund entsprechender Versuche (vgl. Abs. 4.5) mit deutlich höheren Leitungsverlusten behaftet zu sein, zumal eine Erhöhung der Schichtdicke auf die bei der Goldmetallisierung der SIMMWIC-Schaltungen übliche Stärke $t_{\text{Au}} = 4 \mu\text{m}$ kaum mit CMOS-Anforderungen in Einklang zu bringen ist. Eine unmittelbare Verbindung der beiden Metalle führt insbesondere zu Zuverlässigkeitsproblemen, wengleich hier eine befriedigende Lösung gefunden worden ist (vgl. Abs. 4.3). Neben der Vermeidung dieser Schwierigkeiten erlaubt der hybride Aufbau auch die Kombination mit anderen Substrattypen, z.B. GaAs, wengleich auf ähnliche Ausdehnungskoeffizienten bei Betrieb mit erhöhten Temperaturen geachtet werden muß.

3.2 Hybride CMOS-SIMMWIC-Rectenna

Die realisierten Hybridschaltkreise [3.6] basieren auf einer Inverterstufe in CMOS-Technik als Vorverstärker und einer von der Daimler-Benz Forschung stammenden aktiven Antenne. In Bild 3.1 ist der prinzipielle Aufbau gezeigt.

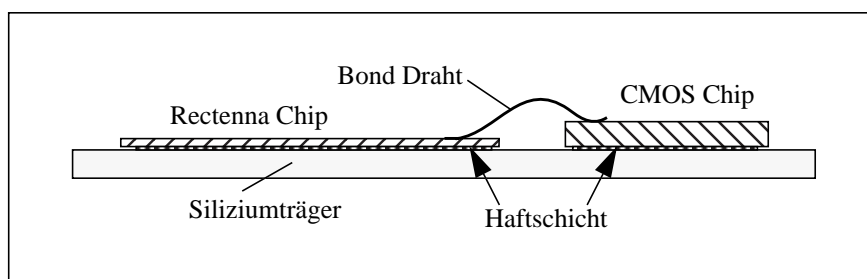


Bild 3.1: Schnittbild der hybrid aufgebauten Anordnung aus SIMMWIC-Rectenna-Chip und CMOS-Vorverstärker.

Durch thermische Oxidation wird auf dem Silizium-Grundsubstrat zu Isolationszwecken eine $t_{\text{ox}} = 1 \mu\text{m}$ dicke SiO_2 -Schicht erzeugt, auf das die beiden Einzelchips mittels eines konventionellen Chipklebstoffs montiert wer-

den. Der auf $h = 100 \mu\text{m}$ gedünnte SIMMWIC-Chip enthält eine Rectenna, kurz für receiving antenna, die aus einer in Mikrostreifentechnik aufgebauten Antenne und einer Schottky-Diode als Detektionselement besteht. Die geringe Wellenlänge im Substrat von $\lambda_0 \approx 1 \text{ mm}$ bei $f = 90 \text{ GHz}$ erlaubt die platzsparende Integration von elf $\lambda/2$ -Dipolen. Diese die Antennencharakteristik bestimmenden Strahlungselemente sind entlang einer Mikrostreifenleitung jeweils im Abstand von $\lambda/2$ zueinander angeordnet (siehe Bild 3.2).

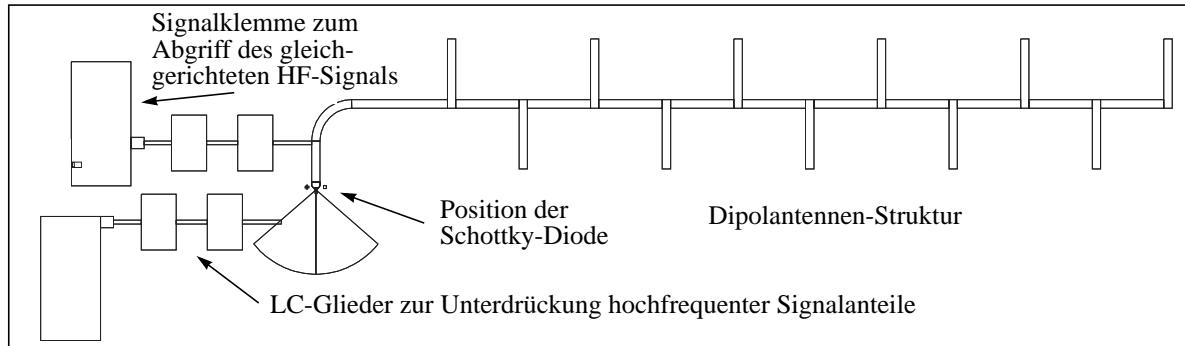


Bild 3.2: Layout der Antenne in Mikrostreifentechnik. Die zur Gleichrichtung des HF-Signals dienende Schottky-Diode befindet sich zwischen der eigentlichen Antennenstruktur und dem als Hochfrequenz-Masse dienenden viertelkreis-förmigen Pad.

Eine durch Molekularstrahlepitaxie (MBE) hergestellte Schottky-Diode dient als Detektor und wird ohne zusätzlichen Biasstrom betrieben, sog. zero bias Diode. Die zur eingestrahelten Hochfrequenzleistungsdichte S linear proportionale Detektorspannung der Diode (vgl. Bild 3.3) wird auf den hochohmigen Eingang des nachgeschalteten CMOS-Vorverstärkers geführt.

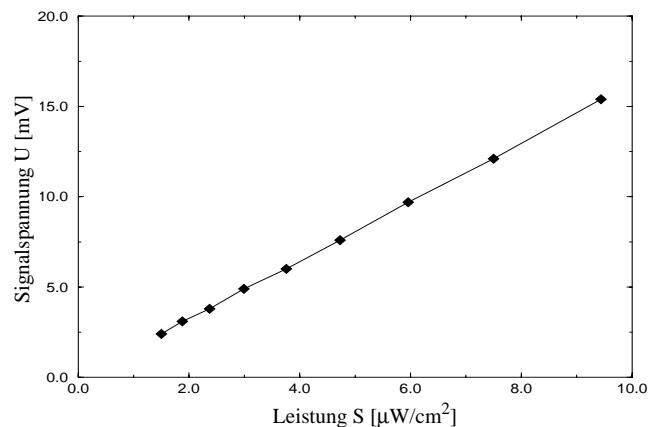


Bild 3.3: Darstellung des linearen Funktionalzusammenhangs zwischen eingestrahelter Leistung S und der Signalspannung U am Ausgang des CMOS-Vorverstärkers. Die Schottky-Diode wird ohne Biasgleichstrom betrieben (zero bias) und mit einem Hochfrequenzsignal bei $f = 94.6 \text{ GHz}$ beaufschlagt.

Die aus Bild 3.3 ersichtliche Gesamtempfindlichkeit der Anordnung beträgt $dU/dS = 1600 \text{ V}/(\text{Wcm}^{-2})$, bei Berücksichtigung des Vorverstärkungsfaktors von $A_{\text{CMOS}} = 32 \text{ dB}$ ergibt sich die Empfindlichkeit der SIMMWIC-Rectenna zu $dV_{\text{Schottky}}/dS = 38 \text{ V}/(\text{Wcm}^{-2})$. Der in Bild 3.4 dargestellte Frequenzgang des Hybridbaus zeigt das schmalbandige Verhalten der Antenne mit einer 3 dB-Bandbreite von $\Delta f_{3 \text{ dB}} = 1.6 \text{ GHz}$. Die Eignung der Rectenna zur Detektion sowohl von impulsmodulierten als auch von kontinuierlichen HF-Signalen im Bereich von etwa 95 GHz erlaubt daher einen Einsatz als frequenzselektive Feldsonde. Der Vergleich der Frequenzgänge mit und ohne Abdeckung offenbart einen Einfluß keramischer Materialien ($\epsilon_r \gg 1$, z. Bsp. Al_2O_3 mit $\epsilon_r = 8 \dots 10$) auf die Feldverteilung in unmittelbarer Nähe der Antenne. Die Beeinflussung wirkt sich als eine leichte Verschiebung des Empfindlichkeitsmaximums zu niedrigeren Frequenzen hin und als eine Empfindlichkeitserhöhung der Seitenbänder bei $f = 92.5 \text{ GHz}$ und $f = 98.0 \text{ GHz}$ aus. Durch konventionelle Kunststoffabdeckungen anstelle von Keramik kann ein ausreichender Schutz der Anordnung ohne eine merkbare Beeinflussung des Frequenzgangs erreicht werden.

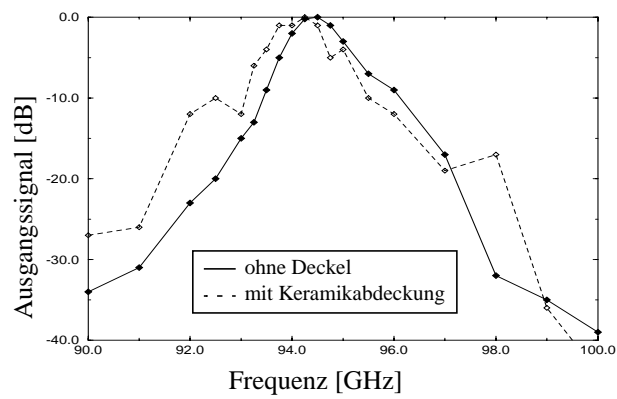


Bild 3.4: Frequenzgang der in einem 24-poligen DIP-Keramikgehäuse aufgebauten Hybridschaltung.

Die in Bild 3.5 gezeigte Richtcharakteristik offenbart eine Halbwertsbreite von $\Delta\phi_{3\text{ dB}} = 17^\circ$ und Nebenkeulen mit Pegeln von weniger als -20 dB bezogen auf den Maximalwert bei $\phi = 0^\circ$.

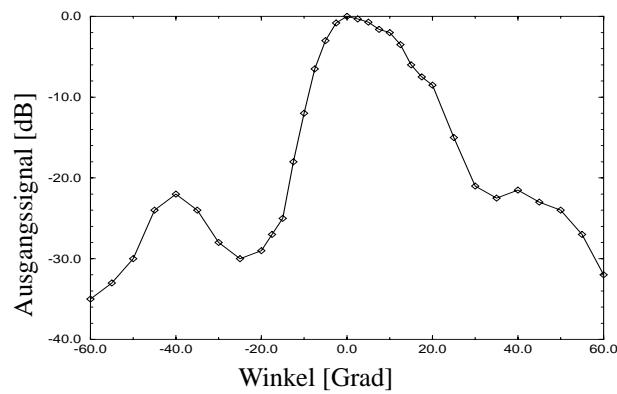


Bild 3.5: H-Ebenen-Diagramm der Anordnung bei einer festen Meßfrequenz von $f = 94.6$ GHz.

Vorarbeiten zur monolithischen Integration

In diesem Kapitel sind die im Rahmen dieses Projekts unternommenen Vorarbeiten für eine monolithische Integration von SIMMWIC- und CMOS-Schaltungen dargestellt. Die grundlegende Machbarkeit einer solchen Lösung wurde u.a. an Hand verkürzter Teilprozesse mit MOS-Varaktor-Strukturen untersucht und bestätigt. Mit Hilfe dieser Bauelemente konnte die zentrale Problematik einer Kontamination des empfindlichen Gateoxides vor allem durch mobile Ionen bewertet werden. Anschließend sind entsprechende Vergleichsuntersuchungen an einzelnen MOSFETs durchgeführt worden. So konnte gezeigt werden, daß zwar die Steilheit β der Transferkennlinie durch die Integration geringfügig beeinflusst wird, eine grundsätzliche Machbarkeit jedoch gesichert ist. Auf der Basis der so gewonnenen Erfahrungen ist eine Prozeßabfolge zur Herstellung von monolithisch integrieren CMOS-SIMMWIC-Schaltungen konzipiert worden. Hierbei ist auf eine weitgehende Übernahme etablierter Module der Einzelprozesse gesetzt worden. Die nachfolgend dargestellte Beispielschaltung ist mittels dieses kombinierten Prozesses hergestellt worden. In einem letzten Untersuchungsfeld, das sich mit der Frage der gegenseitigen Beeinflussung von SIMMWIC-HF-Schaltungsteil und der niederfrequenteren CMOS-Elektronik befaßte, konnte mit Hilfe spezieller Wellenleiter-Teststrukturen und hochfrequenter Meßtechnik ein sehr optimistisch stimmendes Ergebnis erzielt werden. Die Darstellung dieser Arbeiten beschließen das Kapitel.

4.1 Vorteile einer monolithischen Integration

Für eine monolithische Integration von SIMMWIC-Hochfrequenz-Schaltungen und CMOS-Technik sprechen insbesondere kostenrelevante Vorteile durch konsequente Anwendung der etablierten Verfahren der CMOS-Halbleitertechnologie, die sich bei einer entsprechend großen Produktionsstückzahl erreichen lassen. Gerade aber reduzierte Kosten gegenüber der heute in der Hoch- und Höchstfrequenztechnik dominierenden Halbleitertechnik mit Verbindungshalbleitern, deren wichtigster Vertreter Gallium-Arsenid ist, erlauben erst den kommerziellen Durchbruch in vielen nicht-militärischen Bereichen, wie etwa der Verkehrsleittechnik oder der Fabrikautomatisierung. Wenngleich bei Kleinserien eine Hybridtechnik, wie im vorangehenden Kapitel an Hand einer Rectenna mit nachgeschaltetem Verstärker vorgestellt, durchaus ökonomische Vorteile bietet, leidet diese Technik bei komplexen Schaltungen mit vielfachem Signalaustausch zwischen den Einzelchips an den bekannten Problemen, wie aufwendiger Kontaktierung und hohen Anforderungen an die Justagegenauigkeit. Bei zunehmenden Zwischenfrequenzen bereitet auch der Signalaustausch zwischen HF-Schaltung und Ansteuer- und Auswerteelektronik zunehmend Schwierigkeiten, da durch die Verbindungen gebildete parasitäre Elemente bedeutsam werden.

4.2 Anforderungen an einen Integrationsprozeß

Die Integration der CMOS-Technologie mit SIMMWIC-Schaltungen erfordert einen kombinierten Prozeß, der die Spezifika dieser beiden Technologien berücksichtigt. SIMMWIC-Schaltungen zeichnen sich beispielsweise durch eine hohe Temperatursensibilität der Dotierprofile aus, während die CMOS-Technik eine optimale Gateoxidqualität erfordert. Gegenüber höheren Temperaturen - zumindest im hier vorliegenden Fall einer recht konservativen Prozeßdimensionierung - ist der CMOS-Schaltungsteil deutlich weniger empfindlich als die SIMMWIC-Technik.

Die SIMMWIC-Technologie, die auf sehr kleinen Vertikalabmessungen der Bauelemente basiert, z. B. eine Bipolarbasisbreite beim HBT (Hetero-Bipolartransistor) von nur einigen 10 nm erlaubt [4.1], führt zu sehr schnellen Bauelementen. Solch abrupte Dotierübergänge sind nur zu realisieren, indem sowohl bei der Herstellung der Schichten als auch bei der späteren Weiterprozessierung der Gesamtschaltung ein geringes Temperaturbudget gewahrt wird. Heutige Technologien zur Herstellung hochqualitativer epitaktischer Schichten basieren vornehmlich auf dem Einsatz von CVD (engl. chemical vapor deposition) oder MBE (engl. molecular beam epitaxy), wobei gerade letztere sehr abrupte Dotierübergänge sowohl in Richtung auf hohe als auch niedrige Dotierkonzentrationen hin erlaubt. Im Unterschied zur CVD-Technologie ist der sogenannte Memory-Effekt, d.h. der zeitlich sich abschwächende Weitereinbau des bisherigen Dotiermaterials in die wachsende Epitaxie-Schicht, bei Abscheidung mittels MBE sehr gering, wodurch sich eine recht große Freiheit bei der Wahl des Dotierverlaufs durch diese Technik ergibt.

Typische Wachstumstemperaturen epitaktischer Schichten mittels MBE sind mit ca. 450 °C bis 750 °C einige 100 °C unterhalb der typischen CVD-Abscheidetemperaturen [4.2] angesiedelt. Diese Temperaturen liegen deutlich niedriger als die Prozeßtemperaturen in einer CMOS-Technologie. Zwar wird auch hier außer beim Drive-In der Wannsen, der bei etwa 1100 °C bis 1200 °C stattfindet, auf geringe Temperaturen geachtet, so daß möglichst flache Source- und Drain-Gebiete erreicht werden, jedoch bleiben die typischen Temperaturen in der Größenordnung von 850 °C [4.3]. Der im Rahmen dieser Arbeit zugrundeliegende CMOS-Prozeß mit Aluminium-Gate basiert auf maximalen Prozeßtemperaturen von 1200 °C und weist als letzten Hoch-Temperatur-Schritt die Gateoxidation (130 Minuten bei 1000 °C) auf.

Aus dieser Betrachtung der Prozeßtemperaturen heraus ergeben sich zwei Schlußfolgerungen: (i) Die Gateoxidation muß noch vor der Abscheidung der Epitaxieschichten erfolgen und (ii) eine Beeinflussung der Wannsen- und Source-/Drainprofile durch die anschließende MBE-Abscheidung kann sicher ausgeschlossen werden.

Damit ergibt sich die in Bild 4.1 dargestellte Prozeßfolge der Phasen I und II: Die Abscheidung der Epitaxieschichten (Phase II) erfolgt nach dem letzten Hoch-Temperaturschritt (Ende von Phase I) und würde zu einer Abdeckung des Gateoxides durch die MBE-Schichten führen.

Aus dieser Überdeckung ergeben sich zwei mögliche Probleme: (i) Eine Dotierung des Gateoxides bei unmittelbarer Epitaxie mit einer hochdotierten Schicht wäre möglich. (ii) Das empfindliche Gateoxid läge während des Transports und bei allen vor der MBE-Schichtabscheidung erforderlichen Ätzungen und Reinigungen frei. Hierzu zählt zum Beispiel die Ätzung des Feldoxides im Bereich der vergrabenen P⁺-Schichten (engl. p-buried layer) für die p-Schottky-Dioden. Um diese beiden Gefahren auszuschließen, erhalten die CMOS-Scheiben vor der anschließenden Phase II eine undotierte Poly-Silizium-Schicht-Abscheidung mittels LP-CVD (engl. low pressure CVD) bei 630 °C mit einer Sollstärke $t_{\text{Poly-Si}} = 1.0 \mu\text{m}$.

Nach der Abscheidung der Epitaxieschichten (Phase II) und vor der Aluminium-Aufdampfung und Temperung (Phase III) erfolgt die selektive Entfernung der Poly-Si-Schicht im Bereich der CMOS-Schaltung. Um das Poly-Silizium oberhalb des Gateoxides zu entfernen, bestehen zwei Möglichkeiten: (i) Der Einsatz von plasmaunterstützter Ionenätzung (RIE) und (ii) die naßchemische Entfernung.

Die abschließende Prozessierung, d.h. die Strukturierung der Epitaxieschichten und die Goldgalvanisierung der Hochfrequenz-Schaltung erfolgen im wesentlichen bei Raumtemperatur (Phase IV).

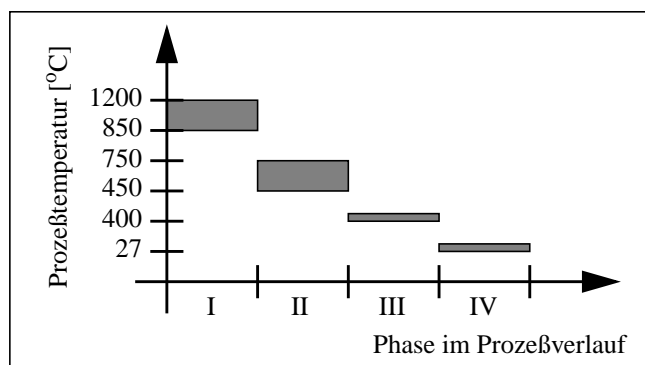


Bild 4.1: Prozeßtemperatur-Verlauf im integrierten CMOS-SIMMWIC-Prozeß. Phase I: CMOS-Prozeß bis zur Fertigstellung des Gateoxides. Phase II: Abscheidung der Epitaxieschichten mittels MBE. Phase III: Aluminium-Metallisierung für die CMOS-Schaltung. Phase IV: Strukturierung der HF-Bauelemente und Goldmetallisierung.

Beide Verfahren der Poly-Silizium-Entfernung erreichen eine Selektivität in Bezug auf thermisches Oxid. Das am Institut eingesetzte Plasmaätzverfahren mit SF₆ [4.4] erreicht beim Verhältnis der Ätzraten von undotiertem Poly-Silizium zu thermischem SiO₂ einen Wert von 10:1. Ein merklicher Abtrag des Gateoxides bei geringsten Ätzhomogenitäten aufgrund der vorgeschlagenen zehnpromtigen Verlängerung der idealen Ätzzeit ist jedoch nicht zu vermeiden. Obwohl nur eine Gleich-Vorspannung von 60 V zur Beschleunigung der Ätzionen eingesetzt wird, sind die Erzeugung von Strahlenschäden und damit zusätzliche Oxidladungen zu erwarten. Da ein Ausheilen bei erhöhten Temperaturen nicht ohne eine Schädigung des Dotierprofils der MBE-Schichten möglich ist, muß eine Poly-Silizium-Entfernung mittels RIE-Ätzung aus diesem Grund ausgeschlossen werden.

Alternativ kommt eine Ablösung durch Kali-Lauge (KOH) bei Raumtemperatur in Betracht. Die erreichbare Selektivität ist hierbei mit ca. 700:1 bei 30 %iger Lösung zwischen <100>-orientiertem Silizium und thermischem Oxid gegeben. Die Ätzrate von Poly-Silizium in KOH ist allgemein etwas höher als die von einkristallinem Silizium, so daß eine noch bessere Selektivität zu erwarten ist. Das Ende des Ätzvorgangs in KOH ist sehr gut erkennbar, da die bei der Lösung von Silizium erfolgende Blasenbildung sehr stark nachläßt. Aber auch eine deutliche Verlängerung der Ätzzeit kann ohne Schädigung des Gateoxides in Kauf genommen werden. Problematisch beim Einsatz von KOH-Lauge ist die unerwünschte Kontamination des Gateoxides mit Kaliumionen. Jedoch ist zu beachten, daß auch die bei der Photolithographie eingesetzten Lösungen alkalisch wirken und eine Verunreinigung der Oberfläche durch die Plasmaätzung ebenfalls entfernt werden müßte.

4.2.1 Untersuchungen zur Poly-Si-Abdeckung des Gateoxides an Hand von MOS-Varaktoren

Zur Untersuchung des Einflusses einer Poly-Silizium-Abdeckung auf die Gateoxidqualität ist eine Versuchsscharge prozessiert und ausgewertet worden (vgl. Bild 4.2). Diese Prozeßfolge umfaßt die Gateoxidation, eine Poly-Si-Abscheidung sowie dessen Entfernung in Kalilauge und die abschließende Gatemetallisierung.

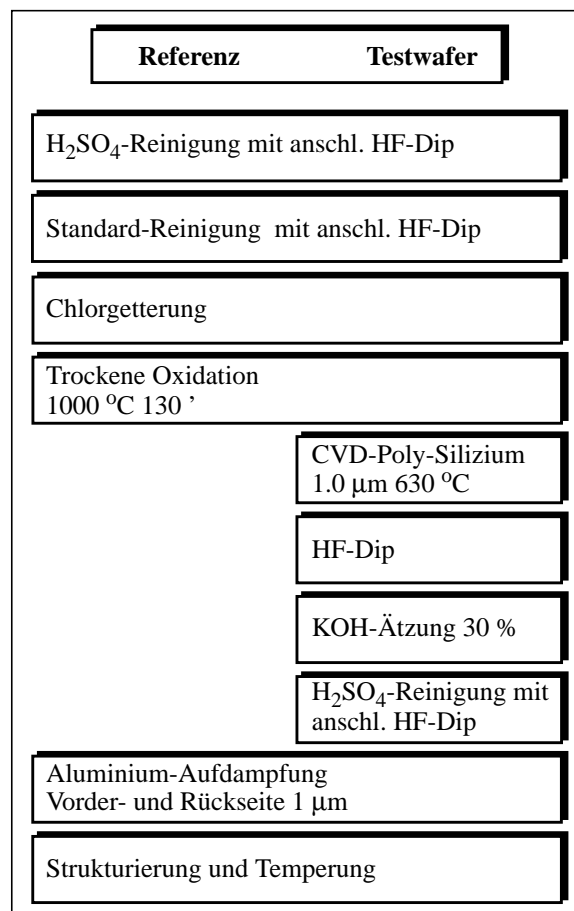


Bild 4.2: Fertigungsprozeß von MOS-Varaktoren zur Untersuchung der Auswirkung einer Überdeckung des Gateoxides mit CVD-Poly-Silizium. In der linken Hälfte ist der konventionelle Prozeßablauf und rechts im Bild der modifizierte Ablauf mit zusätzlicher Poly-Silizium-Abscheidung und Entfernung dargestellt.

Ein Teil der Scheiben, die in Bild 4.2 als Referenz bezeichnet sind, erhielten die konventionelle Prozessierung für MOS-Varaktoren und dienten als Vergleichsbasis.

Für die quantitative Bestimmung von Oxid- und Silizium-Oxid-Interface-Parametern hat sich die CV-Meßtechnik etabliert [4.5]. Die Kapazitäts-Spannungs-Kennlinie (CV) einer relativ einfach zu fertigenden MOS-Struktur erlaubt die schnelle Beurteilung der Oxid- und Grenzflächen-Qualität. Eine Beeinflussung der Oxidqualität an der Grenzfläche zum Aluminium wird zu erwarten sein, wohingegen die Beeinträchtigung des Silizium-Oxid-Interfaces ausgeschlossen werden kann. Die Prozeßtemperatur von 630 °C bei der CVD-Poly-Si-Abscheidung und die spätere Aluminium-Temperung bei 480 °C werden nicht zu einer Beeinträchtigung des Dotierverlaufs führen. Diese Überlegung erlaubt prinzipiell die Benutzung eines beliebigen Si-Substratmaterials solange die <100>-Ausrichtung des Wafers wegen des orientierungsabhängigen Oxiddickenwachstums beibehalten wird. Bedingt durch die in Anhang A 3.3 dargelegte Problematik bei CV-Messungen an hoch resistiven MOS-Varaktoren ist für diese Untersuchung daher auf Bor-dotiertes Substratmaterial mit $\rho = 2 \Omega\text{cm}$ zurückgegriffen worden. Aufgrund der Abfolge des Integrationsprozesses, in der das bereits gewachsene Gateoxid zuerst mit einem Poly-Silizium und dann anschließend mit einer durch MBE abgeschiedenen Epitaxieschicht überdeckt wird, interessiert hier vornehmlich der Einfluß des Überwachsens auf die spätere Grenzfläche zwischen Oxid und Gate-metall, in diesem Fall Aluminium. Es wird vermutet, daß es insbesondere aufgrund der KOH-Ätzung zu einer Anlagerung von Alkaliionen an dieser Stelle kommen kann und daraus resultierend zu einem Vorhandensein zusätzlicher positiver Oxidladungen. Aufgrund der guten Beweglichkeit der Alkaliionen in einem elektrischen Feld ist die im Anschluß dargestellte Messung der mobilen Ionen entscheidend für eine Aussage zur Güte der auf die Ätzung folgenden Reinigung.

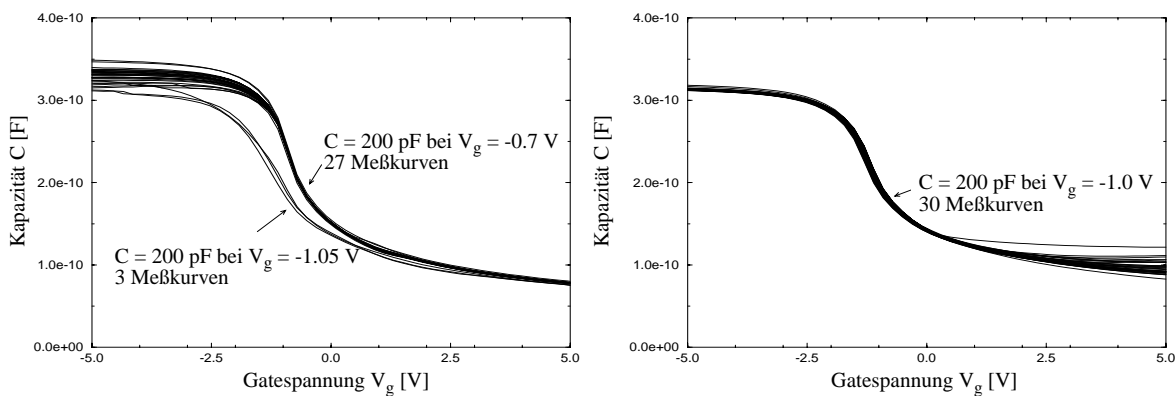


Bild 4.3: CV-Kennlinien der überwachten MOS-Kapazitäten (links) und der Referenzelemente (rechts).

Sowohl von den geätzten Testwafern als auch von den Referenzwafern wurden je 30 CV-Kennlinien unter Bedingungen der tiefen Verarmung aufgenommen. Die Variation der Spannungswerte bei einem vorgegebenen Kapazitätswert von $C = 200 \text{ pF}$ ist hierbei als Maß für die Schwankung der CV-Kennlinien entlang der Spannungsachse herangezogen worden. Eine solche Verschiebung kann einerseits aus unterschiedlichen Oxiddicken und andererseits aus dem Vorhandensein unterschiedlich großer Oxidflächenladungsdichten resultieren. Wie in Bild 4.3 links erkennbar, ist aufgrund der Entfernung der Poly-Si-Schicht, des HF-Dips und der anschließenden Reinigung eine im Akkumulationsbereich variierende Kapazität vorhanden, dies ist durch die unterschiedlich dicke Gateoxidschicht zu begründen. Als wesentlich ist jedoch herauszustellen, daß die einzelnen Kennlinien, mit Ausnahme von drei Kennlinien, die um etwa 0.35 V zu negativeren Gatespannungen hin verschoben sind, sehr gering streuen. Dies wird auch in den Histogrammen der zu einem Kapazitätswert von $C_0 = 200 \text{ pF}$ gehörenden Spannungswerte V_g deutlich (vgl. Bild 4.4).

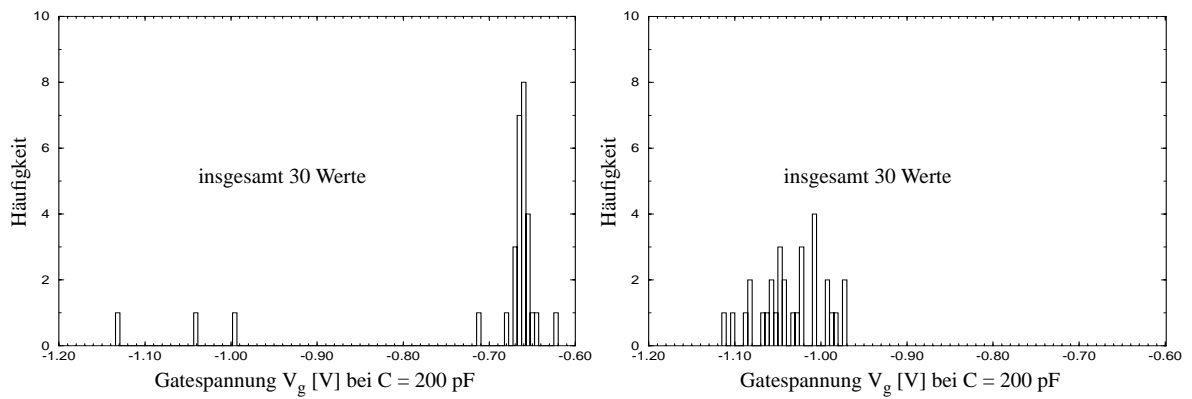


Bild 4.4: Streuung der CV-Kennlinien der überwachten MOS-Kapazitäten (links) und der Referenzelemente (rechts).

Unter Ausklammerung der drei weit entfernten Werte ergibt sich für die überwachte Scheibe eine sehr geringe Streuung, so daß von einer guten Gleichmäßigkeit des Gateoxides auch nach dem Überwachen und Entfernen des Poly-Siliziums ausgegangen werden kann.

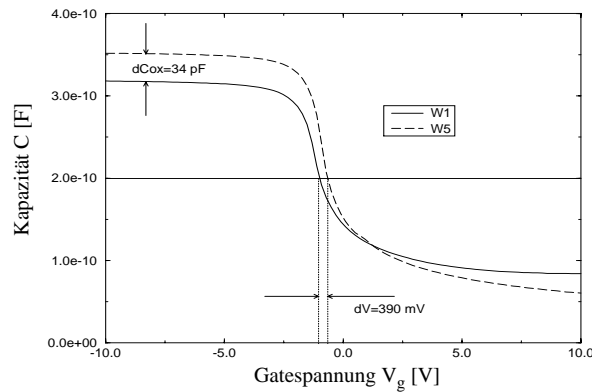


Bild 4.5: Vergleich typischer CV-Kennlinien der überwachten MOS-Kapazitäten und der Referenzelemente. Die durchgezogene Kurve stellt das Referenzelement (Wafer W1 mit $t_{ox1} = 897 \text{ \AA}$) und die unterbrochen gezeichnete Kurve den überwachten Varaktor (Wafer W5 mit $t_{ox2} = 799 \text{ \AA}$) dar.

Ein Vergleich typischer Kennlinien (vgl. Bild 4.5) zeigt eine um $\Delta C_{ox} = 34 \text{ pF}$ angestiegene Oxidkapazität des überwachten MOS-Kondensators, bedingt durch die Reduktion der mittleren Gateoxidicke von $t_{ox1} = 897 \text{ \AA}$ auf $t_{ox2} = 799 \text{ \AA}$.

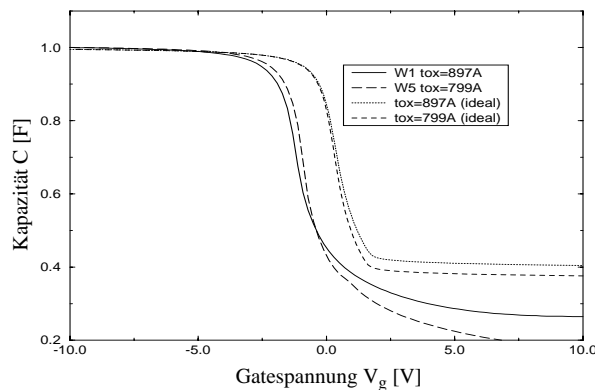


Bild 4.6: Vergleich der gemessenen normierten Kennlinien mit unter idealisierten Bedingungen berechneten (d.h. ohne Oxidladungen Q_o und unter Vernachlässigung von Differenzen in den Austrittsarbeiten E_{ms} zwischen Gatematerial und Halbleiter). Die Oxiddicke $t_{ox1} = 897 \text{ \AA}$ entspricht dem Referenzvaraktor, $t_{ox2} = 799 \text{ \AA}$ dem überwachten Varaktor.

In Bild 4.6 sind den normierten gemessenen CV-Kurven theoretisch berechnete ideale Kurven gegenübergestellt. Bei der Berechnung der idealen Kurven wird das Nicht-Vorhandensein einer Differenz der Austrittsarbeiten, also $E_{ms} = 0$, sowie ein ladungsfreies Oxid, also $Q_o = 0$, vorausgesetzt. Diese Vereinfachung bedingt eine Parallelverschiebung der CV-Kurve entlang der Spannungsachse gegenüber einer gemessenen realen Kurve und führt zu einer Flachbandspannung $V_{FB} = 0$ V. Die Verschiebung der jeweiligen Kurven zueinander läßt sich aufgrund der Spannungsdifferenz am Punkte der Flachbandkapazität C_{FB} bestimmen. Diese Kapazität berechnet sich gemäß

$$\frac{1}{C_{FB}} = \frac{1}{C_{FBS}} + \frac{1}{C_{ox}} \quad \text{mit} \quad C_{FBS} = \frac{\epsilon_{Si}}{\lambda_p}, \quad (4.1)$$

hierbei wird die extrinsische Debye-Länge in p-Silizium λ_p wie folgt definiert

$$\lambda_p = \left(\frac{\epsilon_{Si} kT}{e^2 N_A} \right)^{1/2}. \quad (4.2)$$

Bei der hier vorhandenen Dotierung von $N_A = 7 \cdot 10^{15} / \text{cm}^3$ ergibt sich für eine Gateoxiddicke von $t_{ox} = 897 \text{ \AA}$ eine bezogene Flachbandkapazität $C_{FB}/C_{ox} = 0.8459$, und für $t_{ox} = 799 \text{ \AA}$ $C_{FB}/C_{ox} = 0.8302$. Die resultierende Spannungsdifferenz zwischen gemessener und berechneter Kurve beträgt $\Delta V_{FB1} = 1.58 \text{ V}$ im Fall des Referenzwafers und $\Delta V_{FB2} = 1.26 \text{ V}$ im Fall des Testwafers.

Nach Deal et al. [4.6] ergibt sich für p-Silizium mit der gegebenen Dotierung eine Austrittsarbeitendifferenz zwischen Aluminium und Silizium von $E_{ms} = \phi_{Al} - \phi_{Si} = -0.95 \text{ eV}$. Dieser Wert ergibt sich rechnerisch aus $E_{ms} = \phi_{Al} - (\chi_{Si} + E_{gap}/2 + E_{if})$ mit $E_{if} = kT \cdot \ln(N_A/n_i)$. Die durch das Vorhandensein von Oxidladungen zu erklärende Spannungsdifferenz $Q_o/C_{ox} = \Delta V_{FB} - E_{ms}/e$ beträgt nunmehr 0.63 V entsprechend einer positiven Flächenladungsdichte $Q_o/e = 1.5 \cdot 10^{11} / \text{cm}^2$ bzw. 0.31 V entsprechend einer positiven Flächenladungsdichte $Q_o/e = 0.84 \cdot 10^{11} / \text{cm}^2$. In beiden Fällen handelt es sich um sehr niedrige Werte für die Summe aller positiven Oxidladungsbeiträge.

Aus den jeweils 30 aufgenommenen Meßkurven wurden für die Referenzscheibe $Q_o/e = 1.4 \cdot 10^{11} / \text{cm}^2$ bis $1.8 \cdot 10^{11} / \text{cm}^2$ und für die überwachsene Scheibe $Q_o/e = 0.76 \cdot 10^{11} / \text{cm}^2$ bis $0.98 \cdot 10^{11} / \text{cm}^2$ ($2.2 \cdot 10^{11} / \text{cm}^2$) ermittelt. Der in Klammern stehende Wert stellt hierbei den zu den drei stark abweichenden Kennlinien gehörigen Wert dar.

Aufgrund des für die überwachsene Scheibe ermittelten Werts, welcher noch unterhalb des direkt bedampften Wafers liegt, ist keine Erhöhung der Oxidladungen aufgrund des Überwachsens zu bemerken. In Bezug auf eine mögliche Verschiebung der Schwellspannung ist daher nur von einem Einfluß durch die unterschiedliche Oxiddicke auszugehen. Aus folgender Beziehung

$$V_{th} = \frac{E_{ms}}{e} + \frac{1}{C_{ox}} \cdot \sqrt{4\epsilon_{Si} e N_A \phi_B} + 2\phi_B - \frac{Q_o}{C_{ox}} \quad (4.3)$$

für den n-Kanal MOS ergibt sich bei den unterschiedlichen Oxiddicken t_{ox1} und t_{ox2} die Differenz der Schwellspannungen zu

$$\Delta V_{th} = \frac{t_{ox1} - t_{ox2}}{\epsilon_{ox}} \left(\sqrt{4\epsilon_{Si} e N_A \phi_B} \right) - \frac{1}{\epsilon_{ox}} (t_{ox1} Q_{o1} - t_{ox2} Q_{o2}). \quad (4.4)$$

Der erste Summand repräsentiert die Bulkladung Q_B und der zweite die Oxidladung Q_o , ϕ_B ist hierbei der Abstand zwischen intrinsischem Niveau E_i , d.h. der Bandmitte, und dem Fermi-niveau E_f im Halbleiterinneren weit entfernt von der Si-SiO₂-Grenzfläche und läßt sich aus der Dotierung gemäß $\phi_B = U_t \cdot \ln(N_A/n_i)$ bestimmen.

Im Fall des NMOS mit $N_A = 1.9 \cdot 10^{16} / \text{cm}^3$ ergibt sich bei einer Differenz der Gateoxiddicken von $t_{\text{ox}1} - t_{\text{ox}2} = 897 \text{ \AA} - 799 \text{ \AA} = 98 \text{ \AA}$ und einer Oxidladungsdichte $Q_o/e = 2 \cdot 10^{11} / \text{cm}^2$ ein $\Delta V_{\text{th}} = 0.10 \text{ V}$. Hierbei kommt es aufgrund des unterschiedlichen Vorzeichens von Bulk- und Oxidladung zu einer teilweisen Kompensation der Effekte. Im Fall einer sehr geringen Ladungsdichte von $Q_o/e = 1 \cdot 10^{10} / \text{cm}^2$ folgt $\Delta V_{\text{th}} = 0.19 \text{ V}$. Es muß also im Fall des NMOS-Fet mit einer Erniedrigung der Schwellspannung für den überwachsenen Transistor um $|\Delta V_{\text{th}}| = 100 \text{ mV} \dots 200 \text{ mV}$ gerechnet werden.

Im Fall des PMOS mit $N_D = 1.8 \cdot 10^{15} / \text{cm}^3$ erhalten beide Beiträge ein negatives Vorzeichen, so daß sich die Effekte von Bulk- und Oxidladung direkt addieren. Für $Q_o/e = 2 \cdot 10^{11} / \text{cm}^2$ folgt $\Delta V_{\text{th}} = 0.15 \text{ V}$, und für $Q_o/e = 1 \cdot 10^{10} / \text{cm}^2$ folgt $\Delta V_{\text{th}} = 0.06 \text{ V}$. Der insgesamt geringe Betrag der Bulkladung aufgrund der deutlich niedrigeren Dotierung führt zu einer Schwellwertspannungserhöhung von nur $\Delta V_{\text{th}} = 50 \text{ mV} \dots 150 \text{ mV}$. Die Beträge der Schwellspannungen von NMOS und PMOS verringern sich durch die Oxiddickenverringern, jedoch sind die zu erwartenden Schwellwertspannungsverschiebungen sowohl für den PMOS als auch für den NMOS bei mäßiger Oxidflächenladungsdichte gering.

Die weiteren Unterschiede zwischen den gemessenen und berechneten Kurven sind bedingt durch folgende zwei Punkte:

- (i) Die Messung der CV-Kurven geschieht unter Bedingungen der tiefen Verarmung, während die Rechnung unter Annahme eines thermischen Gleichgewichts erfolgt. Deshalb unterschreiten diese die berechneten Kurvenverläufe im Inversionsbereich (für positive V_g).
- (ii) Das Auftreten von Oberflächenzuständen Q_{it} sorgt für eine leichte Dehnung der Meßkurve entlang der Spannungskurve, sog. „stretch-out“-Effekt. Insgesamt ist jedoch eine sehr weitgehende Übereinstimmung von gemessenen und berechneten Kurven zu erkennen.

Mittels des in Abs. 2.2.3 beschriebenen Meßverfahrens sind sowohl für die Referenz- als auch für die Testscheibe sehr geringe Grenzflächenzustandsdichten gefunden worden. Dabei ergab sich ein mittlerer Wert von $D_{\text{it}} = 6.5 \cdot 10^{10} / \text{cm}^2 \text{ eV}$ für die Referenzscheibe und $D_{\text{it}} = 6.2 \cdot 10^{10} / \text{cm}^2 \text{ eV}$ für die Testscheibe. Demnach sind weder durch den CVD-Prozeß, noch die feuchte Ätzung oder die Aluaufbringung mittels thermischer Verdampfung zusätzliche Zustände an der Grenzfläche Silizium-Oxid erzeugt worden.

Ergebnis der mit Kalilauge (KOH) geätzten und nicht mit Flußsäure (HF) gereinigten Testscheibe

Bei Unterbleiben der Entfernung der obersten Oxidschicht in verdünnter Flußsäure zeigen die Kurven auch nach einer H_2SO_4 -Reinigung noch eine erhebliche Kontamination des Gateoxides, die sich in einer deutlich größeren relativen Verschiebung der CV-Kurven untereinander zeigt (vgl. Bild 4.7).

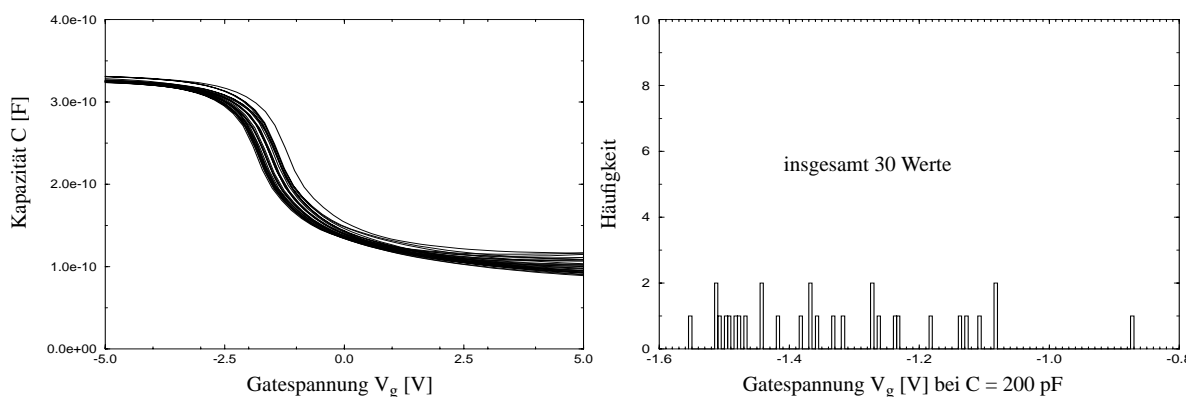


Bild 4.7: CV-Kennlinien der überwachsenen MOS-Kapazitäten und ohne abschließenden HF-Dip bedampften Testscheiben (links die CV-Kurven, rechts die Verteilung der zu $C = 200 \text{ pF}$ gehörigen Spannungswerte).

Aufgrund der starken Streuung der Ausgangskurven ist bei diesem Wafer auf die im folgenden beschriebene Messung der mobilen Ionen verzichtet worden.

Messung der mobilen Ionen

Die Kontrolle der Konzentration mobiler Ionen im Gateoxid, insbesondere von Natrium- und Kalium-Ionen, ist entscheidend, um eine zeitliche Drift der Schwellspannung V_{th} von MOS-Transistoren weitgehend zu verhindern. Die Bewegung dieser Ionen, die unter erhöhten Temperaturen und hohen Feldstärken innerhalb von Minuten zu einer Wanderung durch das komplette Gateoxid führt, kann bei üblichen Gatefeldstärken und normalen Betriebstemperaturen zu einer langsamen Verschiebung der Schwellspannung führen, und somit die Funktion der gesamten Schaltung beeinträchtigen.

Im Zusammenhang mit der Entfernung des das Gateoxid abdeckenden Poly-Siliziums interessiert die Frage, ob nach Anwendung von 30 %iger KOH-Lauge eine deutliche Erhöhung der Kaliumkonzentration auf dem Oxid verbleibt.

Die gängige Methode der Bestimmung der Konzentration mobiler Ionen basiert auf einer gewollten Beschleunigung des Ionen-Driftprozesses durch das Gateoxid aufgrund erhöhter Feldstärke und Temperatur.

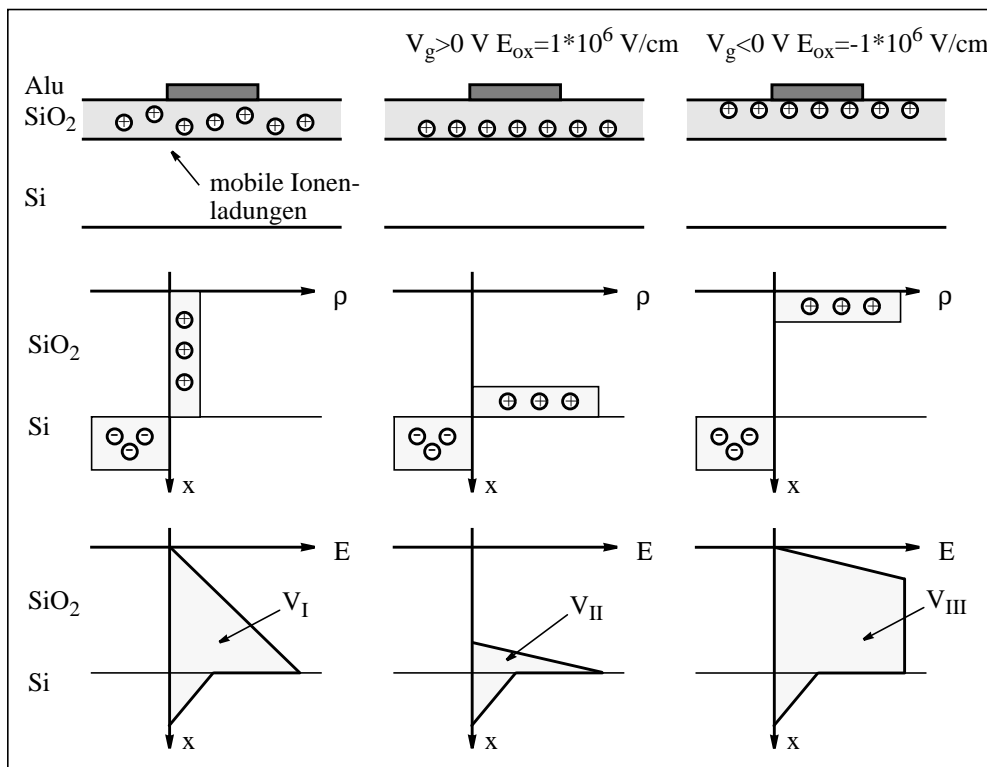


Bild 4.8: Darstellung der Verteilung der Ionen (oben), Raumladung ρ (Mitte) und Feldstärke E (unten) entlang der Tiefenkoordinate x . Links die erste Messung vor Temperatur- und Feldstärkebeanspruchung, in der Mitte nach dem positiven Streß-Zyklus, rechts nach dem negativen Streß-Zyklus.

Bei dieser Methode wird die Verschiebung der CV-Kurve entlang der Spannungsachse bei unterschiedlicher Verteilung der Ionen im Gateoxid bestimmt. In einer möglichen Ausgangslage befinden sich die Ionen gleichverteilt im Oxid, wie in Bild 4.8 links dargestellt. Eine solch nahezu homogene Verteilung ergibt sich beispielsweise durch eine Verunreinigung während der Oxidation im Ofen [4.7]. Eine Ionenkonfiguration wie in der Bildmitte zeugt von einer Verunreinigung der Siliziumoberfläche schon vor der Oxidation, während der Fall rechts eine Kontamination der Oxidoberfläche vor der Metallisierung andeutet.

Alle drei Fälle sind so gezeichnet worden, daß eine gleichbleibende Bandverbiegung, dargestellt durch eine konstante Ausdehnung der Verarmungszone, erreicht wird. Ferner wird zur Vereinfachung von einer Differenz der Austrittsarbeiten von Metall und Halbleiter sowie der Existenz von Oxidladungen abgesehen. Bei der idealen MOS-Kapazität ist für $V_g = 0 \text{ V}$ der Flachbandfall gegeben [4.5]. Die Größe der Flächen unterhalb der elektrischen Feldstärke (unten in Bild 4.8) ist hierbei ein Maß für die notwendige Gatespannung, um diesen Zustand zu erreichen. Eine festvorgegebene Weite der Verarmungszone, entsprechend einer bestimmten Kapazität, wird für Spannungen $V_{II} < V_I < V_{III}$ erreicht. Dies bedeutet, daß die zu II gehörige CV-Kurve nach links, die zu III gehö-

rige nach rechts bezüglich der Ausgangskurve I verschoben ist. Geht man von einer unendlich schmalen Ionenverteilung, d.h. einer Flächenladungsdichte, entlang des Silizium-Oxid- bzw. Aluminium-Oxid-Interfaces aus, errechnet sich die maximale Spannungsverschiebung zwischen den Kurven II und III gemäß

$$\Delta V = \frac{Q_M}{C'_{ox}} = \frac{e N_M}{C'_{ox}} \quad (4.5)$$

Hierbei steht Q_M für die Flächenladungsdichte der beweglichen Ionen mit der Flächendichte N_M . Eine Lokalisierung der Ionen an jeder der beiden Grenzflächen wird während der Messung durch ein entsprechend gerichtetes elektrisches Feld bei erhöhter Temperatur des Wafers erreicht.

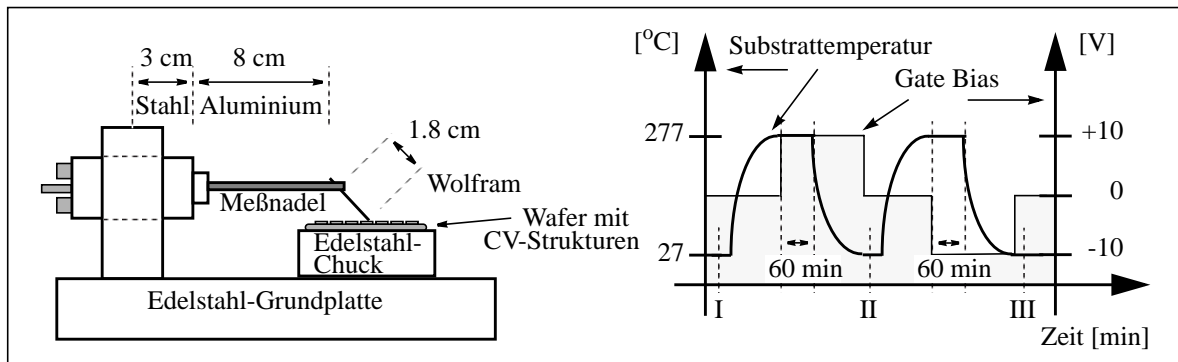


Bild 4.9: Meßaufbau zur Bestimmung der Zahl mobiler Ionen. Links im Bild der Meßaufbau und rechts das Temperatur-Bias-Zeit-Diagramm der Meßsequenz. I: Messung ohne Streß, II: Messung nach positivem Streß, III: Messung nach negativem Streß.

Für eine aussagekräftige Bestimmung der Konzentration mobiler Ionen wird verlangt, daß eine Feldstärke im Gateoxid von mindestens $E_{ox} = 5 \cdot 10^5$ V/cm bei einer Temperatur von $T = 200$ °C während einiger Minuten angelegt wird [4.5]. Eine noch höhere Feldstärke und Temperaturen bis 300 °C beschleunigen den Vorgang der Ionenbewegung. Bei Temperaturen oberhalb von 300 °C können jedoch bei negativem Gatepotential zusätzliche feste Ladungen durch Aufreißen schwacher Silizium-Silizium- sowie Silizium-Sauerstoff-Bindungen generiert werden, die zu einer Ergebnisverfälschung aufgrund einer zusätzlichen Kurvenverschiebung in negativer Richtung führen können [4.8]. Die Konzentration der mobilen Ionen würde in einem solchen Fall deutlich überschätzt. Im Versuch wurde deshalb eine maximale Temperatur von $T = 267$ °C und eine Feldstärke von $E_{ox} = 1 \cdot 10^6$ V/cm eingesetzt.

Ergebnis der Referenzscheibe

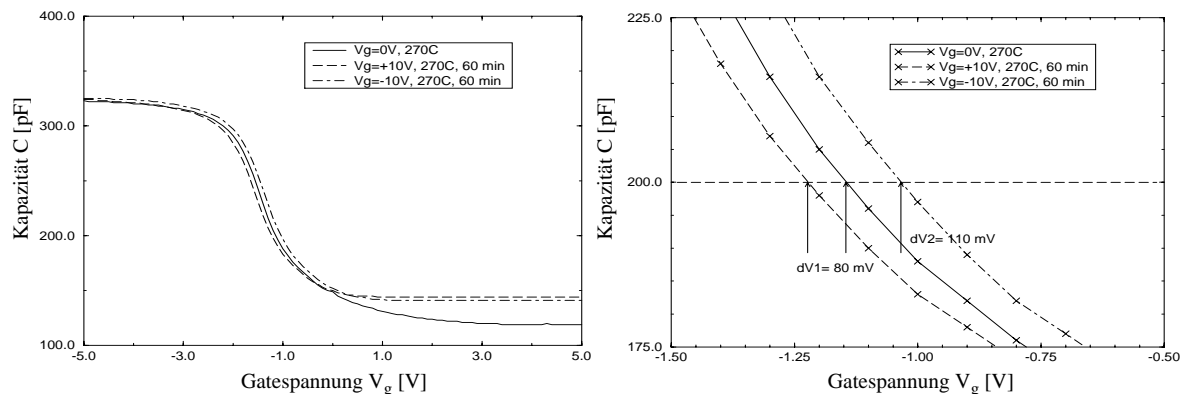


Bild 4.10: CV-Messung zur Bestimmung der Zahl mobiler Ionen im Gateoxid: Referenzscheibe. Links: Komplette CV-Charakteristik. Rechts: Vergrößerter Ausschnitt bei $C = 200$ pF.

Die CV-Kurven der direkt nach der Gateoxidation metallisierten Referenzscheibe (vgl. Bild 4.10) zeigen einen Spannungsoffset zwischen Kurve I und Kurve II von $dV_1 = 80 \text{ mV}$ und zwischen Kurve I und III von $dV_2 = 110 \text{ mV}$. Aus der Gesamtverschiebung $dV_{\text{gesamt}} = 190 \text{ mV}$ zwischen den Kurven für $V_g = -10 \text{ V}$ und $V_g = +10 \text{ V}$ errechnet sich die Flächendichte mobiler Ionen zu $N_M = 4.6 \cdot 10^{10} / \text{cm}^2$ für $t_{\text{ox}} = 897 \text{ \AA}$.

Ergebnis der KOH-geätzten Testscheibe

Die CV-Kurven der KOH-geätzten Testscheibe (vgl. Bild 4.11) zeigen einen Spannungsoffset zwischen Kurve I und Kurve II von $dV_1 = 110 \text{ mV}$ und zwischen Kurve I und III von $dV_2 = 30 \text{ mV}$. Die relative Lage der Kurven deutet hierbei stark auf eine nahezu konstante Ionenverteilung im Oxid hin. Eine zusätzliche Kontamination, wie etwa durch noch an der Aluminium-Oxid-Grenzfläche verbleibende Kalium-Ionen ist daher nicht vorhanden. Aus der Gesamtverschiebung $dV_{\text{gesamt}} = 140 \text{ mV}$ zwischen den Kurven für $V_g = -10 \text{ V}$ und $V_g = +10 \text{ V}$ errechnet sich die Flächendichte mobiler Ionen zu $N_M = 3.8 \cdot 10^{10} / \text{cm}^2$ für $t_{\text{ox}} = 799 \text{ \AA}$.

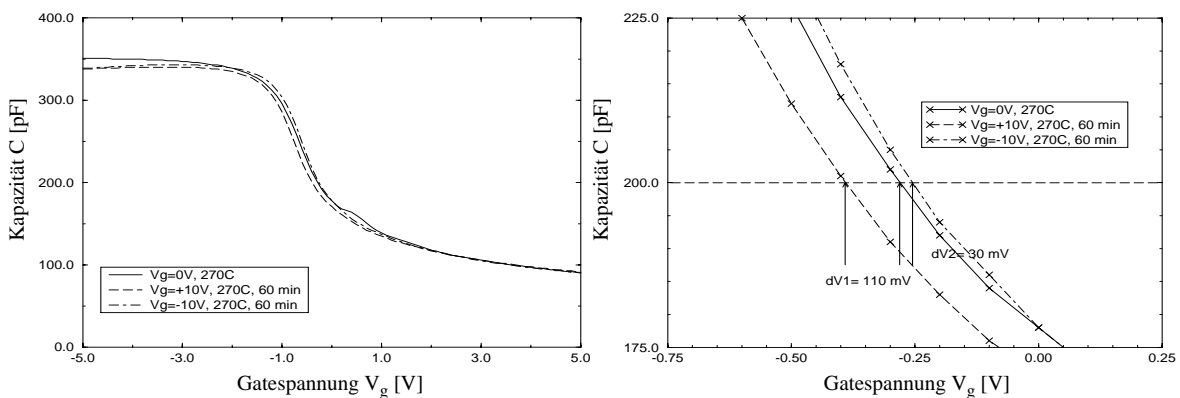


Bild 4.11: CV-Messung zur Bestimmung der Zahl mobiler Ionen im Gateoxid: KOH-geätzte Testscheibe. Links: Komplette CV-Charakteristik. Rechts: Vergrößerter Ausschnitt bei $C = 200 \text{ pF}$.

Der Unterschied der ermittelten Flächenladungsdichte zwischen Test- und Referenzwafer ist in Betracht der zu erwartenden Meßgenauigkeit nicht als signifikant zu betrachten, so daß keine Differenz der Konzentration zu konstatieren ist. Die Poly-Silizium-Ätzung mittels 30 %iger KOH-Lauge bei Raumtemperatur ist deshalb als möglicher Weg zur Integration mit Aluminiumgate-CMOS-Transistoren zu betrachten. Es verbleibt der bei diesem Verfahren unumgängliche Oxidabtrag durch den einige Sekunden dauernden Dip in verdünnter Flußsäure. Dieser ist notwendig, um die alkalische Kontamination des Gateoxides sicher zu entfernen

4.2.2 Auswirkung der Poly-Silizium-Schicht auf die Funktion der MOSFETs

Die Auswirkungen einer Abdeckung des Gateoxides von MOS-Transistoren durch Poly-Silizium ist im Rahmen einer Versuchscharge untersucht worden. Als Substratmaterial kam n-Typ-Silizium (Phosphor, 4-6 Ωcm , CZ, <100>) zum Einsatz. Nach einer für alle Scheiben gemeinsam verlaufenen Prozessierung gemäß des CMOS-Standard-Prozesses mit implantierter p-Wanne wurden vier Scheiben nach der Gateoxidation unmittelbar einer CVD-Poly-Silizium-Abscheidung ($T = 630 \text{ }^\circ\text{C}$, $d_{\text{Poly-Si}} = 1 \text{ }\mu\text{m}$) unterzogen. Diese Schicht ist anschließend bei Raumtemperatur in 30 %iger Kalilauge (KOH) abgeätzt worden. Nach der folgenden Reinigung wurden bei allen Scheiben die Kontaktlöcher geöffnet und das Aluminium aufgedampft. Die Scheiben sind nach der abschließenden N_2 -Temperung ($T = 480 \text{ }^\circ\text{C}$, $t = 15 \text{ min}$) elektrisch vermessen worden.

Einfluß auf den NMOS-Fet

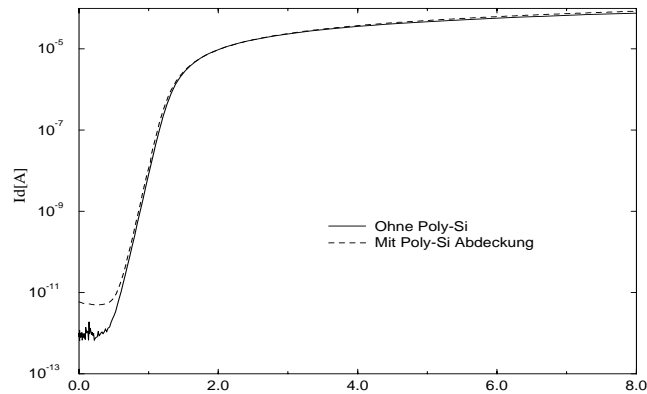


Bild 4.12: Transferkennlinie I_d - V_g von NMOS-Transistoren ($W_{\text{eff}} = 50.4 \mu\text{m}$, $L_{\text{eff}} = 5.4 \mu\text{m}$) im linearen Bereich bei $V_{\text{ds}} = 50 \text{ mV}$. Durchgezogene Kurve: MOS ohne Poly-Schicht, gestrichelte Kurve: MOS mit aufgebracht und abgeätzter Poly-Schicht.

Der Vergleich der Transferkennlinie beider NMOS-Transistoren (vgl. Bild 4.12) zeigt keine auffallenden Unterschiede, wenngleich sich ein etwas erhöhter Drainstrom I_d für den Transistor mit aufgebracht und wieder entfernter Poly-Silizium-Schicht zeigt. Durch die Entfernung der Poly-Silizium-Schicht ist eine Dünnung des Gateoxides aufgrund der Reinheitsanforderungen unumgänglich. Die untersuchte Scheibe hat eine verbleibende Gateoxiddicke von $t_{\text{ox}} = 858 \text{ \AA}$, während die nicht beschichteten Scheiben etwa $\Delta t_{\text{ox}} = 100 \text{ \AA}$ dickere Gateoxide aufweisen ($t_{\text{ox}} = 955 \text{ \AA}$).

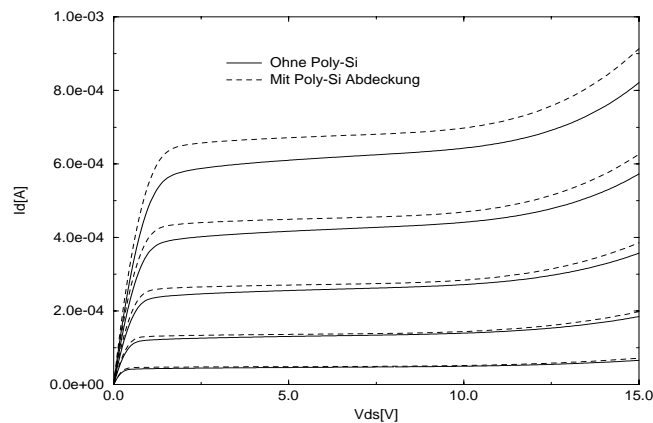


Bild 4.13: Ausgangskennlinie I_d - V_{ds} von NMOS-Transistoren ($W_{\text{eff}} = 50.4 \mu\text{m}$, $L_{\text{eff}} = 5.4 \mu\text{m}$). Durchgezogene Kurve: MOS ohne Poly-Schicht, gestrichelte Kurve: MOS mit aufgebracht und abgeätzter Poly-Schicht. $V_g = 2.0, 2.5, \dots, 4.0 \text{ V}$.

Gemäß der Definition des geometrieunabhängigen Steilheitsparameters

$$k_n = \mu_n C'_{\text{ox}} = \mu_n \frac{\epsilon_{\text{ox}}}{t_{\text{ox}}} \quad (4.6)$$

folgt bei einer als konstant angenommenen Beweglichkeit der Elektronen im Kanal μ_n eine inverse Proportionalität dieser Größe k_n mit der Gateoxiddicke t_{ox} . Als Folge davon ist ein um etwa ein Zehntel erhöhter Drainstrom zu erwarten. Mit einer Beeinträchtigung der Kanalbeweglichkeit μ_n aufgrund der Aufbringung und Entfernung der Poly-Silizium-Schicht ist nicht zu rechnen. Der erhöhte Drainstrom ist in Bild 4.13 dem Kennlinienfeld eines als Referenz dienenden Transistors gegenübergestellt.

Einfluß auf den PMOS-Fet

Für den PMOS-Fet gilt Entsprechendes wie im Fall des NMOS. Die Wirkung der Aufbringung der Poly-Si-Schicht und deren Ätzung liegt wiederum in einer leichten Erhöhung des Steilheitsparameters (vgl. Bild 4.14 und 4.17).

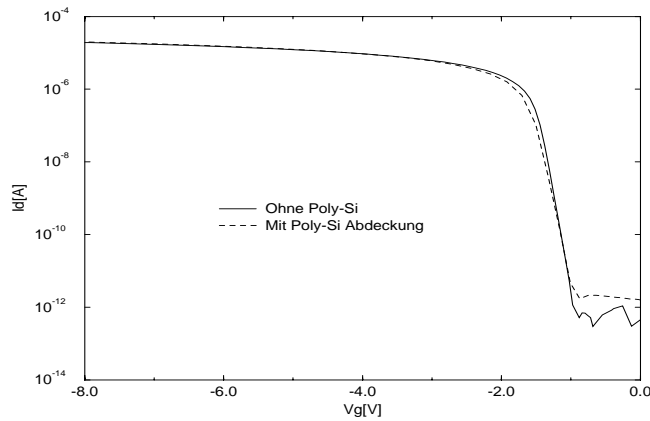


Bild 4.14: Transferkennlinie I_d - V_g von PMOS-Transistoren ($W_{eff} = 50.6 \mu m$, $L_{eff} = 4.6 \mu m$) im linearen Bereich bei $V_{ds} = -50$ mV. Durchgezogene Kurve: MOS ohne Poly-Schicht, gestrichelte Kurve: MOS mit aufgebracht und abgeätzter Poly-Schicht.

Die Erhöhung des Stroms gegenüber dem Vergleichstransistor fällt etwas größer aus als im Fall des NMOS, dies ist begründet durch die recht stark schwankende Gateoxiddicke nach der KOH- und HF-Ätzung, die Gateoxiddicke schwankt etwa $\Delta t_{ox} = 20 \text{ \AA}$ um den oben angegebenen Mittelwert. Während diese Schwankungen die Funktion des einzelnen Bauelements sowie einfacher Schaltungen nur wenig beeinflussen, sind sie für komplexe und hochpräzise Schaltungsentwürfe eventuell problematisch.

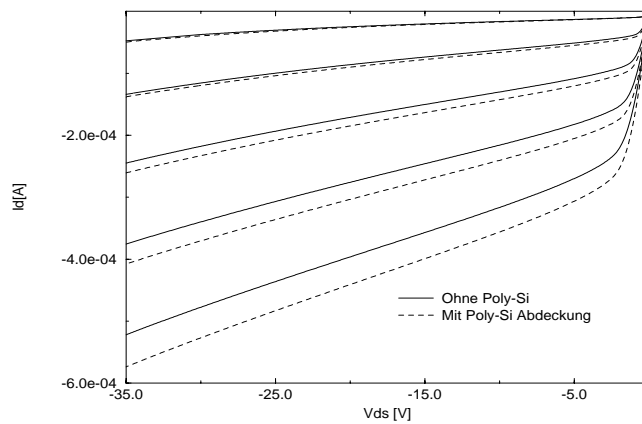


Bild 4.15: Ausgangskennlinie I_d - V_{ds} von PMOS-Transistoren ($W_{eff} = 50.6 \mu m$, $L_{eff} = 4.6 \mu m$). Durchgezogene Kurve: MOS ohne Poly-Schicht, gestrichelte Kurve: MOS mit aufgebracht und abgeätzter Poly-Schicht. $V_g = -2.0, -2.5, \dots, -4.0$ V.

4.3 Integrationsprozeß

Wie in Abschnitt 4.2 dargelegt, erfordert die monolithische Integration des Aluminiumgate-CMOS-Prozesses mit SIMMWIC-Schaltungen eine Auftrennung der Prozeßsequenzen, die im wesentlichen durch die unterschiedlichen Temperaturbudgets bestimmt wird. Der in Bild 4.16 dargestellte Prozeßzyklus zeigt diesen Wechsel zwischen den einzelnen Prozeßmodulen. Zu Beginn erfolgt die Prozessierung des CMOS-Schaltungsteils, der die höchsten Temperaturen erfordert.

Damit das Feldoxid mit maximaler Stärke als Maskierung für den SIMMWIC-Bereich erhalten bleibt, erfolgt während der CMOS-Ätzschritte jeweils eine Abdeckung dieser Gebiete durch Photolack.

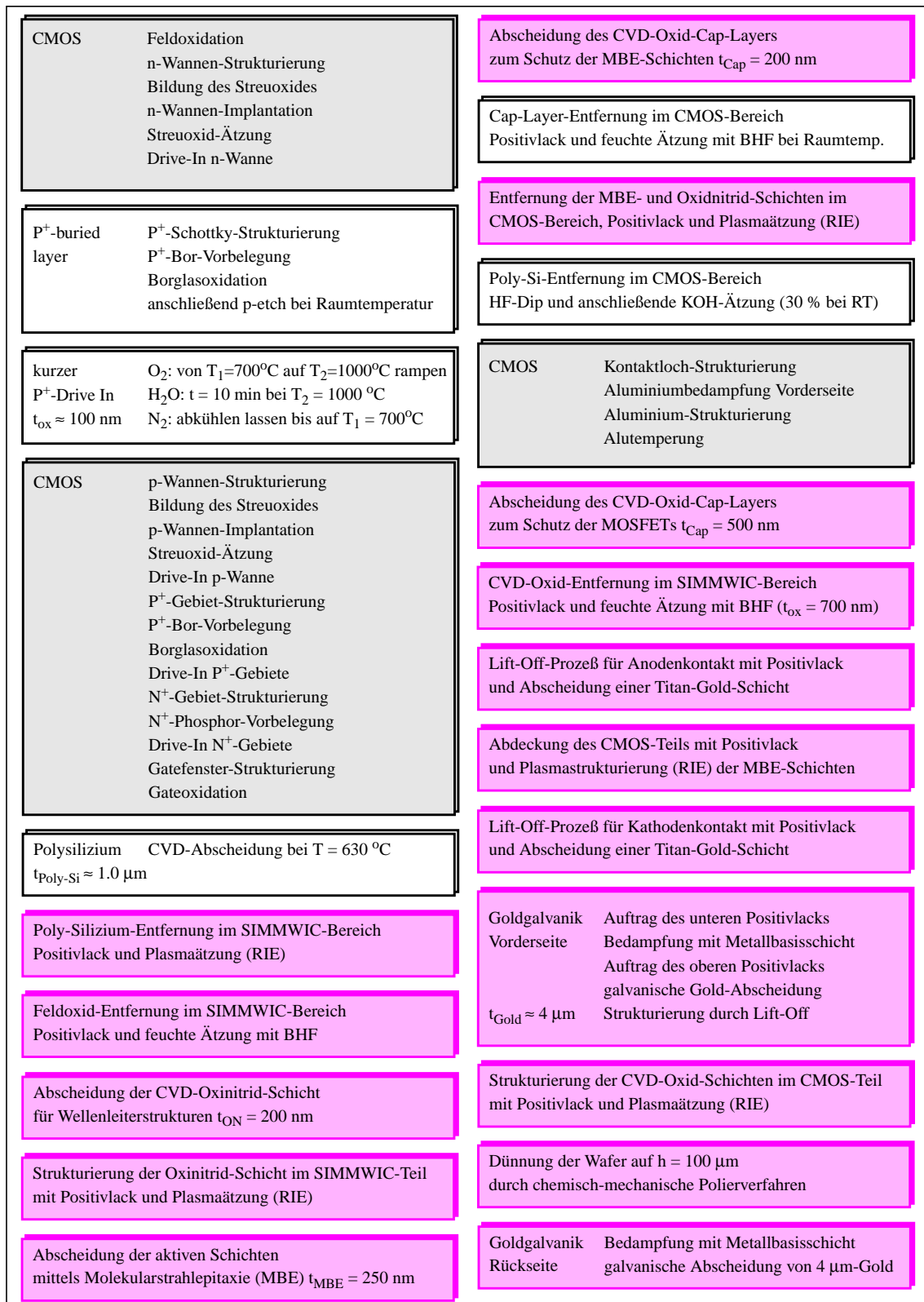


Bild 4.16: Sequenz des Integrationsprozesses. Die unverändert aus der Variante C₂ übernommenen Prozessblöcke sind leicht eingefärbt dargestellt. Die beim Projektpartner Daimler-Benz-Forschung Ulm erfolgten Prozessschritte sind in grauer Farbe und die zur Anpassung beider Technologien in Stuttgart durchgeführten Schritte sind mit weißem Hintergrund gezeichnet.

Gegenüber dem reinen CMOS-Prozeß erscheint zwischen der n- und der p-Wanne die Herstellung der vergrabe-

nen P⁺-Schicht (p-buried layer). Nach der Gateoxidation bei T = 1000 °C als letztem Hochtemperaturschritt des CMOS-Prozesses erfolgt die Abscheidung des CVD-Polysiliziums, das zum Schutz des Gateoxides dient. Diese Schicht muß vor der sich anschließenden Epitaxie-Schichtabscheidung im Bereich der SIMMWIC-Schaltung ebenso wie das darunterliegende Feldoxid entfernt werden. Durch die folgende Abscheidung eines stark sauerstoffhaltigen Siliziumnitrids, sog. CVD-Oxinitrid, wird die gewünschte Isolationsschicht unterhalb der späteren Wellenleiterstrukturen erzeugt. Die Strukturierung dieser Schicht legt nun die Gebiete der vergrabenen Schichten für die späteren Schottky-Dioden frei, so daß die Epitaxie mittels MBE erfolgen kann. Während sich im Bereich der buried layer ein monokristallines Schichtwachstum einstellt, erfolgt auf der restlichen Waferoberfläche eine amorphe Schichtabscheidung. Zum Schutz der Epitaxieschichten dient die CVD-Oxid-Abdeckung (engl. cap layer).

Die Entfernung des Schichtstapels vom Gateoxid, bestehend aus Polysilizium ($t_{\text{poly-Si}} = 1 \mu\text{m}$), Oxinitrid ($t_{\text{Oxinitrid}} = 0.2 \mu\text{m}$), Epitaxieschichten ($t_{\text{MBE}} = 0.25 \mu\text{m}$) und cap layer ($t_{\text{Cap}} = 0.2 \mu\text{m}$), erfolgt durch eine feuchte Oxidätzung in gepufferter Flußsäure (BHF), einer trockenen Ätzung (RIE) der Epitaxieschicht und des Oxinitrids und einer abschließenden Ätzung in Kalilauge (KOH). Diesem kombinierten Ätzprozeß schließt sich die gewohnte CMOS-Prozessierung mit Kontaktlochätzung und Metallisierung an. Während der nicht selektiven Ätzprozesse mit BHF und RIE wird der SIMMWIC-Schaltungsteil mittels Positivlack geschützt. Um die empfindliche Aluminiummetallisierung während der SIMMWIC-Prozessierung bei Reinigungs- und Ätzschritten abzudecken, erfolgt die Abscheidung einer CVD-Oxidschicht.

Die Fertigstellung der Schottky-Dioden und der HF-Wellenleiter erfolgt nun mittels der von Strohm et al. [4.2] entwickelten Abfolge. Eine abschließende Verringerung der Substratdicke auf $h = 100 \mu\text{m}$ durch chemisch-mechanische Verfahren und die ganzflächige Rückseitenmetallisierung erfolgen mit Schritten der SIMMWIC-Technologie.

Die ohmsche Verbindung zwischen der CMOS-Aluminiummetallisierung und der SIMMWIC-Goldmetallisierung erfolgt über die beiderseitige Kontaktierung einer die Bereiche überlappenden vergrabenen n-Schicht.

Durch die gewählte Prozeßfolge konnte eine weitgehende Übernahme etablierter Prozeßschritte der CMOS- und SIMMWIC-Technologien bei gleichzeitiger Minimierung zusätzlicher Arbeitsschritte erreicht werden. Die stark unterschiedlichen Anforderungen an die Metallisierung und die Prozeßtemperaturen begrenzen allerdings die Möglichkeiten der gemeinsamen Nutzung von Schritten für beide Technologien.

4.4 Ansteuerschaltung für einen Hochfrequenz-pin-Schalter

Ein auf dem Entwurf von Luy et al. [4.9] basierender Hochfrequenzschalter ist zur Demonstration des Integrationskonzepts gewählt worden. In diesem als SPDT-switch (single pole-double throw) bezeichneten Schalter läßt sich die am gemeinsamen Eingang (links in Bild 4.17) einlaufende Hochfrequenzwelle wahlweise in den oberen oder unteren Ausgang schalten. Die Funktion dieses in Wellenleitertechnik mit einem charakteristischen Wellenwiderstand $Z_0 = 50 \Omega$ ausgeführten Schalters wird durch die bewußte Veränderung der Impedanz eines Zweiges erreicht. Mit einem starken Vorwärtsstrom durch die pin-Dioden eines Zweiges wird dessen Impedanz soweit verringert, daß eine weitgehende Reflexion der in diesen Zweig einlaufenden Welle erreicht wird. Damit geht eine weitgehende Entkopplung des zugehörigen Ausgangs einher.

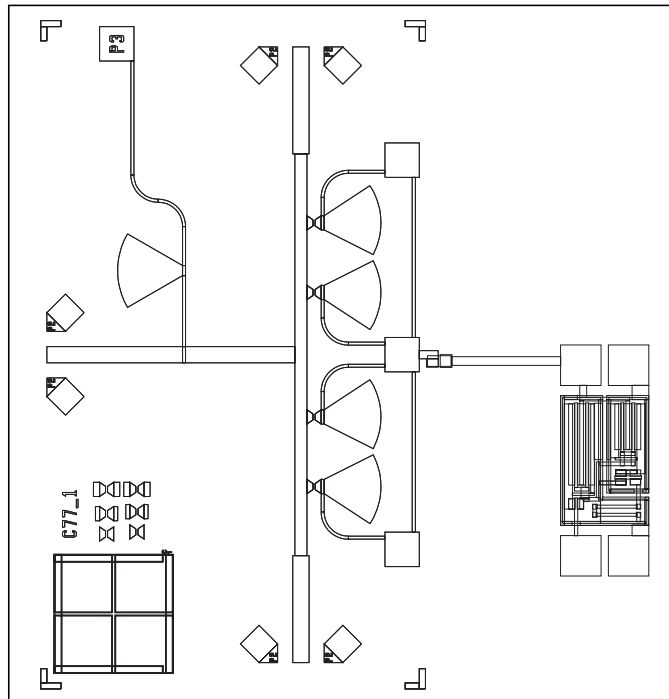


Bild 4.17: Maskenlayout des pin-Schalters mit der CMOS-Ansteuerschaltung, die rechts im Bild gezeigt ist.

Durch die Transformation der so gebildeten Kurzschlußpunkte, die sich an den Stellen der in Durchlaßrichtung geschalteten Dioden befinden, wird ein Leerlauf am Verzweigungspunkt erzeugt, so daß die Anpassung des anderen Zweiges unbeeinflusst bleibt. In diesem anderen Zweig befinden sich die pin-Dioden im Sperrbetrieb. Aufgrund ihres hohen Widerstands und der geringen i -Zonen-Kapazität beeinflussen diese die Gesamtimpedanz nicht. Die Folge davon ist ein nahezu ungedämpfter Transport der Hochfrequenzwelle zu dem dazugehörigen Ausgang.

Um diesen Entwurf in die Integrationssequenz einbinden zu können, sind folgende Veränderungen [4.10] am Entwurf vorgenommen worden:

- (i) Eine Modifikation der Leiterbahnbreite auf $w = 80 \mu\text{m}$ erlaubt den Einsatz auf Substraten mit einer Dicke von $h = 100 \mu\text{m}$.
- (ii) Die gemeinsame Zusammenführung der Dioden und die Einspeisung der Diodenvorspannung an einem Punkt erlaubt die direkte Ansteuerung der Schaltung durch einen CMOS-Schaltkreis. Ferner wurde dadurch die Überlagerung der Diodengleichspannung und des Hochfrequenzsignals am Eingang vermieden. Der gesamte Hochfrequenzpfad befindet sich aufgrund dieser Maßnahme auf Gleichspannungsmassepotential.
- (iii) Durch Änderungen der HF-Ein- und Ausgänge kann auf Durchkontaktierungen (via-holes) verzichtet werden.

Im Mittelpunkt des Interesses dieser Arbeit stand jedoch nicht der Entwurf und die Realisierung des Hochfrequenzschaltkreises, sondern die Integration der dem SIMMWIC-Schaltungsteil zugehörigen pin-Dioden mit CMOS-Schaltkreisen. Als praktisches Beispiel ist eine CMOS-Treiberschaltung gewählt worden, die bei positiven und negativen Ausgangsspannungen den geforderten Strom $|I| = 20 \text{ mA}$ zur Ansteuerung der pin-Dioden liefern kann. Diese Schaltung ist im Anhang A 5.3 beschrieben.

Mit Hilfe des in Bild 4.18 dargestellten Prozeßablaufs gelang die monolithische Integration der CMOS- und SIMMWIC-Komponenten. Nach Herstellung der p- und n-Wannen für den CMOS-Schaltkreis erfolgte die Definition der hochdotierten p- und n-Bereiche der pin-Diode gemeinsam mit den jeweiligen Source- und Draindotierungsschritten. Nach Fertigstellung des CMOS-Teils erfolgte die Metallisierung der pin-Dioden und der Wellenleiterstrukturen durch eine zweite Aluminiumbedampfung und Lift-Off-Prozeß oder alternativ durch einen Gold-Lift-Off- und Galvanisierungsprozeß. Während in Bezug auf die Kennlinie der pin-Dioden durch die unterschiedliche Metallisierung keine signifikante Veränderung beobachtet wurde, ist bei den Mikrostreifenleitungen durch eine Aluminiummetallisierung mit höheren Dämpfungsverlusten zu rechnen.

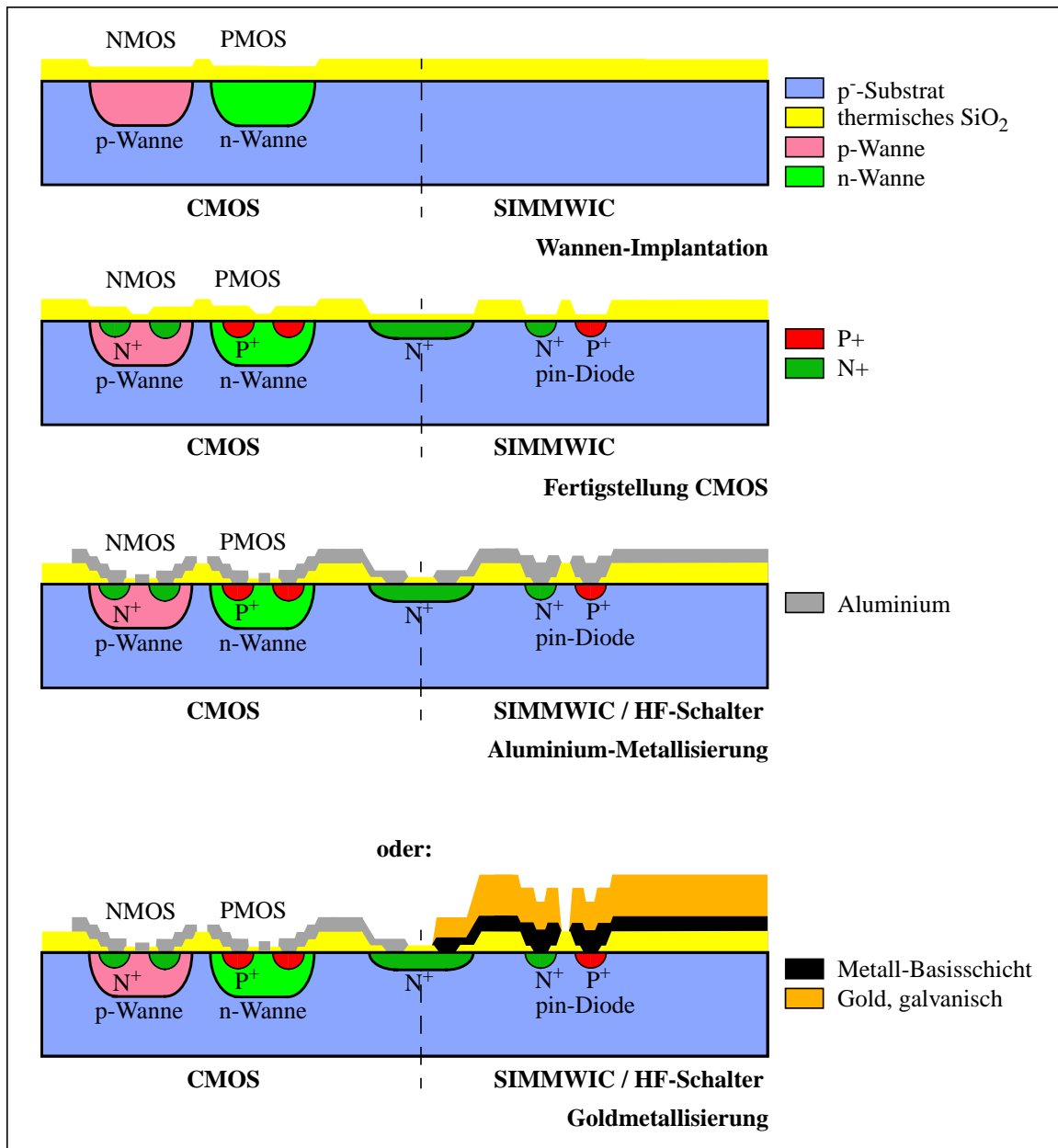


Bild 4.18: Prinzipbild der Prozeßsequenz zur Herstellung eines Hochfrequenzschalters mit pin-Dioden in Aluminium- oder Goldtechnik. Während für die CMOS-Metallisierung in beiden Fällen Aluminium erforderlich ist, kann bei den Wellenleiterstrukturen des pin-Schalters die Metallisierung wahlweise mit Aluminium oder Gold erfolgen.

Für die Herstellung der pin-Dioden kann auf zusätzliche Dotierschritte verzichtet werden, da die wesentlichen Modifikationen im Maskenentwurf vorgenommen worden sind. Dadurch lassen sich die pin-Dioden als CMOS-Bauelemente prozessieren, wenngleich sie sich direkt im Bereich der SIMMWIC-Schaltung befinden. Dies wurde durch eine mehrfache Anwendung der CMOS-Abdeckmaske erreicht, so daß beispielsweise die Öffnung der Diodenanschlüsse nach der Fertigstellung der CMOS-Metall-Strukturierung erfolgen kann. Durch die Verbindung der Treiberschaltung und des pin-Schalters entsteht bei einer Goldmetallisierung das Problem der mangelhaften Haltbarkeit einer direkten Aluminium-Goldüberlappung. Mit Hilfe einer vergrabenen N⁺-Schicht, die sowohl in den CMOS- als auch den SIMMWIC-Bereich reicht, läßt sich eine Verbindung der Schaltungsteile auch bei höheren Strömen ohne eine direkte Kombination der Metalle erreichen. Die Verbindung weist einen geringen Serienwiderstand von $R < 6\Omega$ und eine sehr gute Linearität für Ströme zwischen $-40\text{ mA} < I < 40\text{ mA}$ auf.

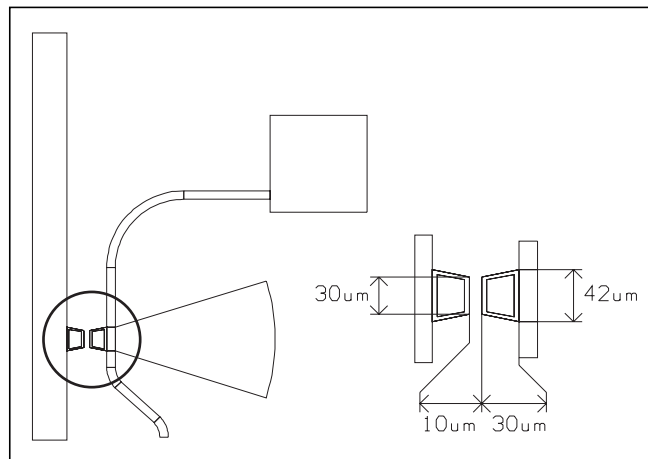


Bild 4.19: Ausschnitt des pin-Schalterlayouts mit einem pin-Dioden-Zweig. Die eigentliche pin-Diode, auf der linken Bildhälfte eingekreist dargestellt, ist rechts mit den Entwurfsmaßen dargestellt. Die Länge der i-Zone zwischen den trapezförmigen P⁺- und N⁺-Gebieten beträgt $l_i = 10 \mu\text{m}$.

In Bild 4.19 ist das Diodenlayout einer pin-Diode gezeigt. Während die Trapezform gegenüber dem Originalentwurf unverändert beibehalten wurde, ist die Länge der i-Zone gemäß der prozeßtechnischen Erfordernisse auf $l_i = 10 \mu\text{m}$ erhöht worden. Die IV-Kennlinie der vier jeweils in Paaren antiparallel geschalteten Dioden des pin-Schalters ist in Bild 4.20 gezeigt. Der in Reihe liegende ohmsche Widerstand der Verbindung zwischen den beiden Schaltungsteilen sorgt für einen kleinen zusätzlichen Spannungsabfall und die geringfügig vergrößerte Ausgangsspannung der Ansteuerschaltung. Die Symmetrie der Kennlinie belegt die gute Reproduzierbarkeit der pin-Dioden, die aufgrund der in unterschiedlichen Photolithographieschritten festgelegten p- und n-Gebiete potentiell erhöhten Längenschwankungen der i-Zone ausgesetzt sind. Die geringe Empfindlichkeit der Diodenkennlinien in Bezug auf unterschiedliche i-Zonenlängen konnte in Abschnitt 2.3.2 gezeigt werden.

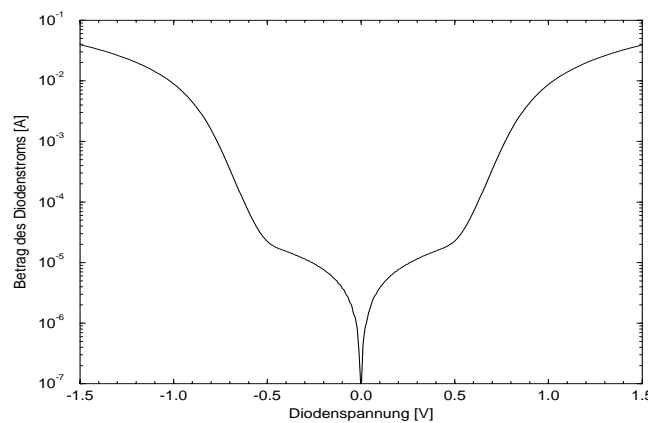


Bild 4.20: Strom-Spannungs-Kennlinie der pin-Dioden. Da die Kennlinie zwischen dem Ausgang der CMOS-Treiberschaltung und der Masse des pin-Schalters aufgenommen wurde, überlagern sich die Vorwärtsströme der beiden Dioden des einen Zweiges mit dem Rückwärtsstrom der Dioden des jeweils anderen Zweiges.

Bild 4.21 zeigt die Spannung am Ausgang der Ansteuerschaltung bei Belastung mit den vier Dioden des pin-Schalters. Aus der Stromkennlinie in Bild 4.20 läßt sich ein Diodenstrom von etwa $I_n = 40 \text{ mA}$ bei negativer und von etwa $I_p = 20 \text{ mA}$ bei positiver Ausgangsspannung ablesen.

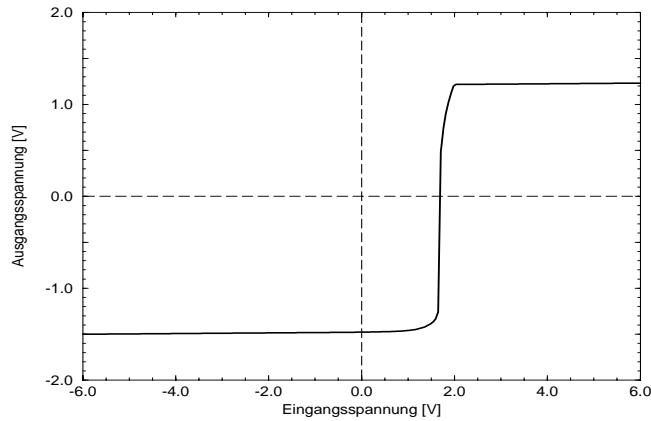


Bild 4.21: Kennlinie der Ausgangsspannung der Ansteuerschaltung im Belastungsfall.

4.5 Beeinflussung der HF-Dämpfung durch CMOS

Als mögliche Einflussfaktoren auf die Höhe der Dämpfung von Hochfrequenzsignalen in Wellenleiterstrukturen auf hochohmigem Siliziumsubstrat sind in den vorherigen Abschnitten (i) die Existenz von Inversionsschichten (vgl. Abs. 2.2.2) unterhalb der dielektrischen Oberflächenpassivierung und (ii) die unerwünschte Dotierung der Waferoberfläche (vgl. Abs. 1.1.2) erwähnt worden. Ferner ist bei einer monolithischen Integration von CMOS- und SIMMWIC-Schaltungen eine Beeinträchtigung durch die hochdotierten Gebiete der CMOS-Schaltung denkbar. Während bei den hybrid aufgebauten Empfängerchips (vgl. Abs. 3.2) keine Beeinflussung der Empfindlichkeit der Rectenna durch die Präsenz des CMOS-Verstärkers wegen der räumlichen Entfernung beider Chips (mindestens 1 mm) beobachtet werden konnte, ist bei den kleineren Abständen in der monolithischen Integration mit zusätzlichen Dämpfungsanteilen zu rechnen. Zur quantitativen Überprüfung wurden mit dem in Bild 4.22 dargestellten Maskenlayout Mikrostreifenleitungen auf hochohmigem p⁻-Silizium (<100> orientiertes FZ-Material, Bor-dotiert mit $N_A = 1.5 \cdot 10^{12} / \text{cm}^3$, $\rho_{\text{Si}} = 5000 \Omega \text{cm}$) hergestellt. Hierbei ersetzen großflächige homogen n-dotierte Bereiche die reale CMOS-Struktur.

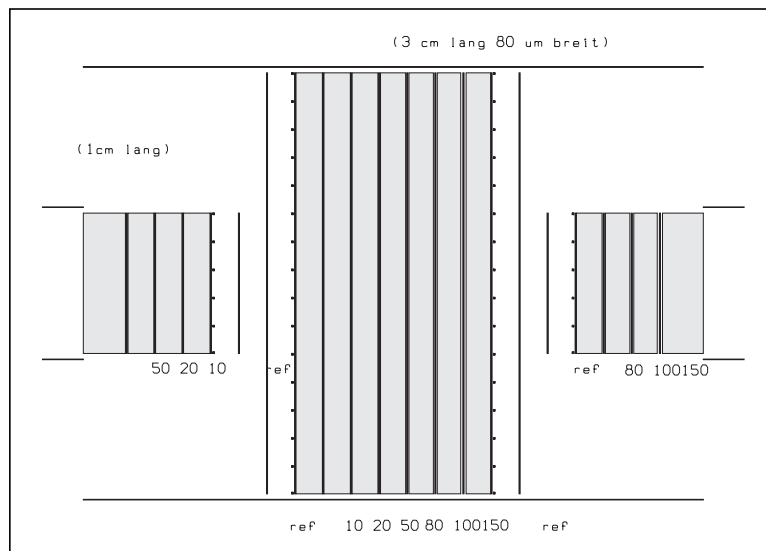


Bild 4.22: Maskenlayout zur Herstellung von Mikrostreifenleitungen auf 2-Zoll-Wafern. Die Breite der Streifenleitung beträgt $w = 80 \mu\text{m}$ und die Leitungsmitten benachbarter Microstrips liegen $a = 2 \text{ mm}$ auseinander. Zwischen den Leitungen befinden sich - im Bild grau markierte - hochdotierte N⁺-Gebiete, die variable Abstände von $s = 10 \mu\text{m}$ bis $s = 150 \mu\text{m}$ zu den benachbarten Streifenleitungen aufweisen. Als Referenzleitungen dienen die äußeren ohne benachbarte Diffusionsgebiete realisierten Strukturen.

Nach einer Reinigung, bestehend aus Schwefelsäure- und Standardreinigung, wurden die Scheiben mit einer $t_{\text{Nitrid}} = 200 \text{ nm}$ starken CVD-Nitrid-Schicht (Si_3N_4) versehen. Der Öffnung der zu dotierenden N⁺-Gebiete

durch plasmaunterstützte Ionenätzung (engl. reactive ion etching, RIE) folgte die aus dem CMOS-Prozeß bekannte Phosphor-Vorbelegung mit anschließendem Drive-In, so daß ein Schichtwiderstand $R_{sh} = 20 \Omega/\square$ und eine Sperrschichttiefe $x_j = 1.5 \mu\text{m}$ erreicht wurden. Die Metallisierung erfolgte im ersten Fall durch Aluminiumverdampfung nach dem Tiegelverfahren ($t_{\text{Alu}} = 1 \mu\text{m}$) und im zweiten Fall durch eine Titan-Gold-Haftschiicht-Aufdampfung, die durch Gold-Galvanik bis auf $t_{\text{Au}} = 4 \mu\text{m}$ verstärkt wurde. Die komplette Entfernung des Nitrids außerhalb der Mikrostreifenstrukturen, eine mechanische Dünnung der Wafer auf $h = 100 \mu\text{m}$ und die Bildung der metallischen Grundelektrode auf der gesamten Rückseite mit den erwähnten Metallisierungsprozessen be-
schlossen die Prozeßsequenz.

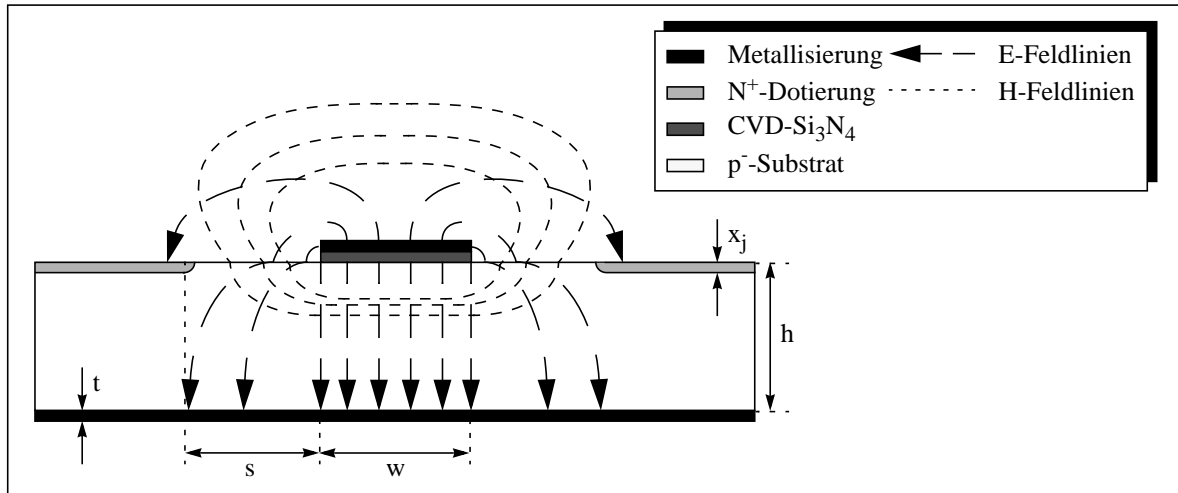


Bild 4.23: Querschnitt der Teststrukturen auf $h = 100 \mu\text{m}$ dickem p^- -Substrat. Die $w = 80 \mu\text{m}$ breite Deckelektrode befindet sich oberhalb des $t_{\text{Nitrid}} = 200 \text{ nm}$ dicken Si_3N_4 . Der Abstand s zum benachbarten hochdotierten N^+ -Gebiet mit der Tiefe $x_j = 1.5 \mu\text{m}$ ist von $s = 10 \mu\text{m}$ bis $s = 150 \mu\text{m}$ in Stufen variiert worden. Bei der als Vergleichsbasis dienenden Variante entfällt die hochdotierte Zone.

Bild 4.23 zeigt die resultierenden Strukturen im nicht maßstäblichen Querschnitt mit einigen elektrischen und magnetischen Feldlinien. Bei einer gemeinsamen Breite der Wellenleiterstrukturen von $w = 80 \mu\text{m}$ verändert sich der Wellenwiderstand Z_{om} der Mikrostreifenleitungen durch die unterschiedliche Elektrodenhöhe t nur unwesentlich [4.12]. Für die $t_{\text{Alu}} = 1 \mu\text{m}$ dicke Schicht errechnet sich $Z_{om, \text{Alu}} = 49.3 \Omega$ und für die $t_{\text{Au}} = 4 \mu\text{m}$ starke Goldschicht $Z_{om, \text{Au}} = 47.7 \Omega$.

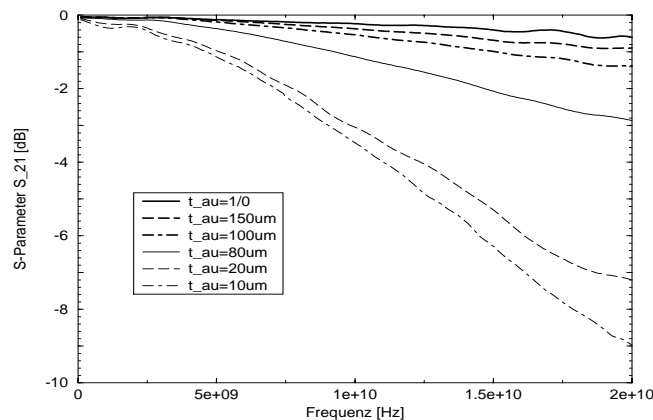


Bild 4.24: Darstellung des Betrags des Dämpfungs-Streuparameters S_{21} von $l = 1 \text{ cm}$ langen 50Ω -Gold-Mikrostreifenleitungen auf p^- -Substrat. Die Geometrie $w/h = 0.8$ ist für alle Streifen konstant gehalten worden. Nur der Abstand $s = t_{\text{au}}$ der Leitung zum hochdotierten Gebiet wird variiert. Die Leitung ohne benachbartes Diffusionsgebiet ist mit $t_{\text{au}} = 1/0$ bezeichnet.

Ein größerer Einfluß auf die Dämpfungseigenschaften als die leicht unterschiedlichen Wellenwiderstände ergibt sich aus der Präsenz der N^+ -Gebiete. Wie in Bild 4.23 gezeigt, enden einige Feldlinien in diesen Bereichen, die zu der rückseitigen Elektrode eine hochohmige Verbindung aufweisen und auf vergleichbarem Potential liegen. Wie von Plattenkondensatoren mit Luft als Dielektrika

bekannt [4.13], breiten sich signifikante Streufelder (sog. fringing fields) etwa in der Größe des Plattenabstands auch in horizontaler Richtung aus. Demzufolge ist ein Anstieg der Dämpfung für Abstände $s < h = 100 \mu\text{m}$ zu erwarten. Die Streuparameter, sog. S-Parameter, der Streifenleitungen wurden bei Frequenzen zwischen $f = 0.1 \text{ GHz}$ und $f = 20 \text{ GHz}$ ermittelt. Bild 4.24 zeigt den Betrag des S-Parameters S_{21} in Dezibel (dB) für Goldstreifenleitungen der Länge $l = 1 \text{ cm}$. Der Betrag dieser Größe ist ein Maß für das Verhältnis der am Leitungsende empfangenen zur am Leitungsanfang injizierten Leistung. Mit sich verringerndem Abstand s sinkt der Betrag dieses Parameters in monotoner Weise signifikant ab.

Da die elektrischen und magnetischen Komponenten von Wellen in Mikrostreifenleitungen nicht nur transversale sondern auch longitudinale Anteile aufweisen, mithin also keine reine TEM-Welle vorliegt, ist eine direkte Beschreibung der Strukturen mittels der aus den Feldanteilen ermittelten Spannungen und Ströme unmöglich. Die Definition der Streuparameter basiert auf den in das Tor T_1 einlaufenden Wellen a_i und den rücklaufenden Wellen b_i . Für eine Mikrostreifenleitung der Länge l mit den beiden Toren T_1 und T_2 ergibt sich daher

$$\begin{aligned} b_1 &= S_{11}a_1 + S_{12}a_2 \\ &\text{und} \\ b_2 &= S_{21}a_1 + S_{22}a_2. \end{aligned} \tag{4.7}$$

Hierbei bedeuten

$$S_{11} = \left. \frac{b_1}{a_1} \right|_{a_2=0} = r_1 \quad \text{und} \quad S_{22} = \left. \frac{b_2}{a_2} \right|_{a_1=0} = r_2 \tag{4.8}$$

die Reflexionsfaktoren r_1 und r_2 an den Toren T_1 bzw. T_2 .

Aus

$$|S_{21}| = \left. \frac{b_2}{a_1} \right|_{a_2=0} = |e^{-j\gamma(f)l}| = |e^{-j\beta(f)l} \cdot e^{-\alpha(f)l}| = 1 \cdot |e^{-\alpha(f)l}| \tag{4.9}$$

läßt sich die Dämpfungskonstante $\alpha(f)$ bestimmen, die üblicherweise in dB/cm angegeben wird, hierbei bedeutet γ die Ausbreitungskonstante und β die Phasenkonstante. Bei einer Leitung der Länge $l = 1 \text{ cm}$ - wie in Bild 4.24 - drückt der Betrag von S_{21} somit direkt die gesuchte Leitungsdämpfung α aus.

Die mit einer herkömmlichen CMOS-Metallisierung fabrizierten Aluminium-Mikrostreifenleitungen weisen im Vergleich zu den Gold-Leitungen höhere Dämpfungen auf. Dies wird am Betrag des Streuparameters S_{21} in Bild 4.25 verdeutlicht. Zusätzlich ist der Betrag der Leitungsreflexionen durch S_{11} dargestellt.

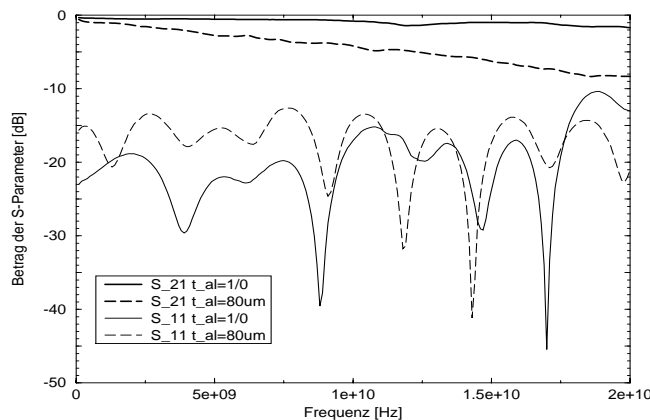


Bild 4.25: Darstellung des Dämpfungs-Streuparameters S_{21} von $l = 1 \text{ cm}$ langen 50Ω -Aluminium-Mikrostreifenleitungen auf p^- -Substrat. Zusätzlich ist der die Eingangreflexionen beschreibende Parameter S_{11} angegeben.

Eine quantitative Bestimmung der Dämpfungsverluste α_t in Mikrostreifenleitungen ist mit den folgenden Beziehungen für die additiven Komponenten der Leiterverluste α_l und der dielektrischen Verluste α_d möglich.

Die durch die endliche Leitfähigkeit der Elektroden bedingten Gleichstromwiderstände $R = \rho \cdot l/A$ betragen bei den Aluminiumleitungen $R_{Alu} = 3.58 \Omega$, da $\rho_{Alu} = 2.86 \mu\Omega\text{cm}$ mit $A = 80 \mu\text{m} \cdot 1 \mu\text{m}$ und $l = 1 \text{ cm}$, und bei den Goldleitungen $R_{Au} = 0.69 \Omega$, da $\rho_{Au} = 2.2 \mu\Omega\text{cm}$ mit $A = 80 \mu\text{m} \cdot 4 \mu\text{m}$ und $l = 1 \text{ cm}$. Zusätzliche Dämpfungen bei hohen Frequenzen, die über die durch den Gleichstromwiderstand bedingten Verluste hinaus entstehen, sind durch den sog. Skin-Effekt verursacht. Aufgrund der Verdrängung der Strompfade vom Leiterinneren an die Oberfläche verringert sich die effektive Querschnittsfläche des Leiters, so daß zusätzliche Leiterverluste auftreten. Bei Kenntnis der Stromverteilungen $J_1(x)$ und $J_2(x)$ auf dem Streifenleiter mit dem Skinwiderstand R_{s1} und der Grundlektrode mit R_{s2} lassen sich die Verluste direkt angeben [4.14]:

$$\alpha_c = \frac{R_{s1}}{2Z_{om}} \cdot \int_c \frac{|J_1(x)|^2}{|I|^2} dx + \frac{R_{s2}}{2Z_{om}} \cdot \int_{-\infty}^{\infty} \frac{|J_2(x)|^2}{|I|^2} dx. \quad (4.10)$$

Hierbei ist der erste Integrand entlang des Weges c um den gesamten Querschnitt der Streifenelektrode zu betrachten und der zweite entlang der kompletten Grundlektrode, Z_{om} ist der Wellenwiderstand der Leitung und I der Gesamtstrom. Die Größe der Skinwiderstände R_s bestimmt sich gemäß

$$R_s = \sqrt{\pi f \mu_o \rho_c}. \quad (4.11)$$

Eine hinreichend genaue Beschreibung für α_c läßt sich durch die Beziehung [4.15]

$$\alpha_c = 1,38 \cdot A \cdot \frac{R_s}{hZ_{om}} \cdot \frac{32 - (W_e/h)^2}{32 + (W_e/h)^2} \quad [dB/m] \quad (4.12)$$

erreichen. Hierin ist die auf die Substrathöhe h bezogene effektive Breite der Mikrostreifenleitung W_e laut

$$\frac{W_e}{h} = \frac{W}{h} + \frac{1,25}{\pi} \cdot \frac{t}{h} \left(1 + \ln \frac{2h}{t}\right) \quad (4.13)$$

und der Vorfaktor A durch

$$A = 1 + \frac{h}{W_e} \left(1 + \frac{1,25}{\pi} \ln \frac{2h}{t}\right) \quad [dB/m] \quad (4.14)$$

bestimmt. Die Komponente der dielektrischen Verluste α_d setzt sich aus den Anteilen der Polarisation eines Dielektrikums mit $\epsilon_r \neq 1$ und der endlichen Leitfähigkeit σ des Substratmaterials zusammen. Basierend auf Wheelers Konzept [4.16 und 4.17] eines Füllfaktors q zur Bestimmung der effektiven Dielektrizitätskonstante ϵ_{re} , d.h. der teilweisen Füllung des Feldraumes mit dielektrischem Material, ermittelten Welch und Pratt [4.18] folgende Beziehung

$$\alpha_d = 4,34 \cdot \sqrt{\frac{\mu_o}{\epsilon_o}} \cdot \sigma_{Si} \cdot \frac{\epsilon_{re} - 1}{\sqrt{\epsilon_{re}(\epsilon_r - 1)}}. \quad (4.15)$$

Der Anteil der Leitungsdämpfung aufgrund des Skinneffekts steigt proportional mit der Wurzel der Frequenz f , die dielektrischen Verluste hingegen sind nach obiger Beziehung frequenzunabhängig. Da bei der Ableitung der Substratverluste von einem homogenen Material ausgegangen wird, bietet sich der Vergleich der theoretischen Ergebnisse mit den Meßkurven des Streuparameters S_{21} bei den Referenzleitungen ohne N^+ -Dotierung an.

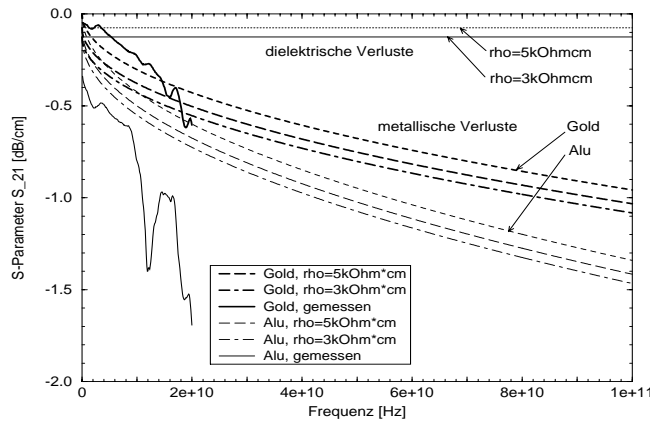


Bild 4.26: Gegenüberstellung von theoretisch erwarteten und meßtechnisch ermittelten Dämpfungsverlusten von jeweils $l = 1 \text{ cm}$ langen Mikrostreifenleitungen mit Gold- ($t_{\text{Au}} = 4 \text{ }\mu\text{m}$) und Aluminium-Metallisierung ($t_{\text{Alu}} = 1 \text{ }\mu\text{m}$). Die berechneten Komponenten sind sowohl getrennt als auch in der Summe dargestellt.

Bild 4.26 zeigt die Gegenüberstellung der Dämpfungsverluste in hochohmigem Substrat bei spezifischen Widerständen von $\rho_1 = 3000 \text{ }\Omega\text{cm}$ und $\rho_2 = 5000 \text{ }\Omega\text{cm}$ sowie den beiden Metallisierungen Gold und Aluminium. Die Meßkurven bis zu einer Frequenz $f = 20 \text{ GHz}$ zeigen bei den Goldleitungen eine im Rahmen der zu erwartenden Meßgenauigkeit gute Übereinstimmung und deuten auf keine zusätzlichen Dämpfungsanteile hin. Die Aluminiumleitungen liegen sowohl in der Rechnung als auch bei den Messungen auf insgesamt höherem Niveau und zeigen vergleichsweise größere Abweichungen zwischen rechnerischen und meßtechnischen Werten.

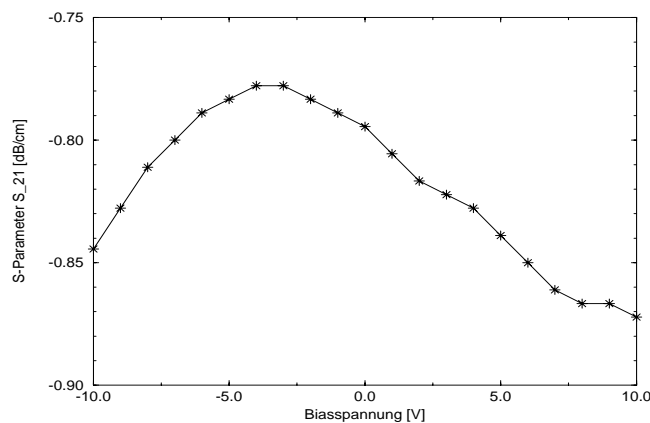


Bild 4.27: Verlauf der Dämpfung einer Aluminium-Mikrostreifenleitung bei $f = 10 \text{ GHz}$ und einer überlagerten Vorspannung an der Top-Elektrode.

Gründe für die gegenüber der Rechnung erhöhte Dämpfung liegen zum einen in einer höheren Rauigkeit des Aluminiumleiters gegenüber einer idealglatten Struktur und zum anderen - wie in Abschnitt 2.2.2 dargelegt - in der Existenz von Inversionskanälen an der Oberfläche des Wafers. Bild 4.27 zeigt den Dämpfungsverlauf einer 1 cm-langen Aluminiumleitung bei einer festen Frequenz von $f = 10 \text{ GHz}$ und einem überlagerten Gleichspannungspotential zwischen Top- und Grundelektrode. Im Bereich negativer Spannungen bei etwa $V_{\text{Bias}} = -3 \text{ V}$ zeigt die Dämpfung S_{21} ein absolutes Minimum. Hier im Bereich der Verarmung des p-Wafers unterhalb der Top-Elektrode steht die minimale Zahl von freien Ladungsträgern für die Dämpfung der Welle zur Verfügung. Aufgrund der großen Raumladungsweiten im hochdotierten Silizium ist auch die Ausdehnung der Raumladungszone in vertikaler Richtung gegenüber der Substrathöhe nicht zu vernachlässigen. Die höheren Trägerdichten in der Anreicherung bei weiter negativ werdender Vorspannung und in der Inversion bei

positiverem Bias verursachen eine Zunahme der Dämpfung. Bei den die Reflexion ausdrückenden Streuparametern S_{11} und S_{22} läßt sich eine solche Veränderung nicht beobachten. Beide liegen bei nahezu konstanten Werten von $S_{11} = -17.9 \text{ dB} \pm 0.1 \text{ dB}$ und $S_{22} = -17.1 \text{ dB} \pm 0.1 \text{ dB}$.

Die gegenüber der charakteristischen Eindringtiefe der Welle in den Aluminiumleiter δ (sog. skin depth) recht geringe Dicke der Metallschicht ($t_{\text{Alu}} = 1 \mu\text{m}$) bedingt eine zusätzliche Erhöhung der Leitungsdämpfung. Bei $f = 10 \text{ GHz}$ beträgt $\delta = 0.83 \mu\text{m}$ und bei $f = 20 \text{ GHz}$ $\delta = 0.58 \mu\text{m}$ und damit liegt die Metallschichtdicke im Bereich der von Horton et al. [4.19] beobachteten Zunahme der Leitungsdämpfung bei Reduktion der Leiterdicke t unter die doppelte Eindringtiefe δ .

Durch die Einbringung der zusätzlichen Dotierungsgebiete zeigen sich monoton zunehmende Dämpfungswerte (vgl. Bilder 4.24 und 4.25), die mit der gegenüber der reinen Mikrostreifenleitung modifizierten Feldverteilung (vgl. Bild 4.23) erklärt werden können. Die zusätzlichen Dotierungsgebiete wirken in erster Näherung wie eine nicht-ideale Koplanarleitung, deren Mittelelektrode aus der Topielektrode der Mikrostreifenleitung besteht und deren seitliche Masseflächen durch die beiden Phosphor-dotierten Bereiche nachgebildet werden. Wenngleich durch die deutlich verringerte Leitfähigkeit der N^+ -Gebiete das Verhalten dieser parasitären Leitung nur eine Annäherung an eine ideale Leitung darstellt, so kann durch die Bestimmung des Wellenwiderstands einer Koplanarleitung eine Abschätzung der Größenordnung der Anteile erfolgen. Der Wellenwiderstand Z_{ocp} läßt sich in der quasi-statischen Näherung [4.20] aus:

$$Z_{\text{ocp}} = \frac{30\pi}{\sqrt{\epsilon_{re}}} \cdot \frac{K'(k)}{K(k)} \quad (4.16)$$

mit

$$\frac{K(k)}{K'(k)} = \frac{1}{\pi} \ln \left[2 \cdot \frac{1 + \sqrt{k}}{1 - \sqrt{k}} \right] \quad 1/\sqrt{2} \leq k \leq 1 \quad (4.17)$$

und

$$\frac{K(k)}{K'(k)} = \frac{\pi}{\ln \left[2 \cdot \frac{1 + \sqrt{k}}{1 - \sqrt{k}} \right]} \quad 0 \leq k \leq 1/\sqrt{2} \quad (4.18)$$

bestimmen. Der Geometriefaktor k ist als

$$k = \frac{S}{S + 2W} \quad (4.19)$$

definiert. Die effektive Dielektrizitätskonstante ϵ_{re} einer Koplanarleitung ist gemäß [4.20] als

$$\epsilon_{re} = \frac{\epsilon_r + 1}{2} \left[\tanh \{ 1, 785 \cdot \log(h/W) + 1, 75 \} + \frac{kW}{h} \left\{ 0, 04 - \left(0, 7k + \frac{(1 - 0, l\epsilon_r)(0, 25 + k)}{100} \right) \right\} \right] \quad (4.20)$$

definiert. Hierbei ist s - wie in der Literatur üblich - die Breite des Mittelleiters mit s und die Schlitzbreite mit w bezeichnet. Bild 4.28, in dem die zu den Gold-Microstrip gehörigen Abstände der Mittelleiter zu den hochdotierten Gebieten markiert sind, zeigt die charakteristische Impedanz Z_{ocp} .

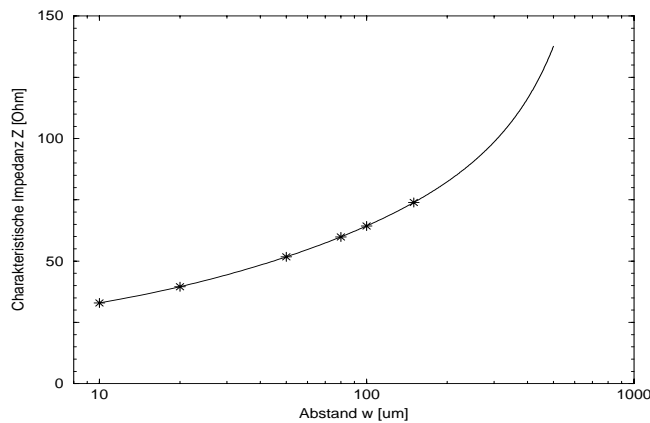


Bild 4.28: Wellenwiderstand Z_{0cp} eines koplanaren Streifenleiters. Die Sterne markieren die im Versuch verwendeten Abstände zwischen Mittelleiter und hochdotierten Gebieten.

Die Dämpfungseigenschaften der Mikrostreifenleitung werden folglich durch drei Effekte beeinflusst:

- (i) Durch die hochohmige Verbindung der hochdotierten Gebiete mit der Grundlektrode des Microstrips erfährt das Eingangssignal eine Aufteilung auf die beiden parallel geschalteten Strukturen einer Mikrostreifen- und einer Koplanarleitung. Schon bei Abständen $w = 50 \mu\text{m}$ erreicht eine ideale Koplanarleitung eine charakteristische Impedanz $Z_{0cp} = 50 \Omega$, so daß es in diesem Fall zu einer Halbierung der im Microstrip geführten Leistung käme.
- (ii) Die obige Parallelschaltung bewirkt ein deutliches Absinken der Gesamtimpedanz $Z_{ges} = (Z_{om}^{-1} + Z_{0cp}^{-1})^{-1}$, die im Fall eines einzelnen isoliert liegenden Microstrips bei $Z_{ges} = Z_{om} = 50 \Omega$ liegt und zu geringen Eingangsreflexionen der injizierten Signale führt.
- (iii) Die beobachtete Dämpfung besteht jedoch nicht allein aus diesen beiden Effekten, sondern wird durch die Kopplung der sich auf dem Mittelleiter ausbreitenden Welle in die hochdotierten Zonen weiter verstärkt. Die Kombination aus metallischem Mittelleiter und n-dotierter Zone wirkt wie ein nicht-idealer koplanarer Koppler, dessen Kopplungsvermögen sehr deutlich vom Abstand der beiden Leitungen abhängt.

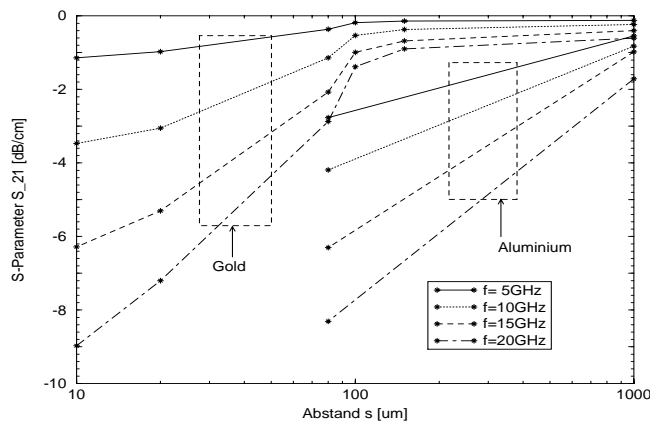


Bild 4.29: Darstellung der Dämpfungszunahme S_{21} bei Abnahme des Abstands zwischen Mikrostreifenleitung und dotiertem Gebiet.

Die in Bild 4.29 dargestellte Leitungsdämpfung zeigt einen signifikanten Anstieg bei Abständen unterhalb von $s = 100 \mu\text{m}$. Diese überproportionale Zunahme wird im ganzen Frequenzbereich zwischen $f = 0 \text{ GHz}$ und $f = 20 \text{ GHz}$ beobachtet (vgl. Bild 4.24) und stellt somit einen beim Entwurf integrierter CMOS-SIMMVIC-Schaltungen zu beachtenden Mindestabstand dar. Jedoch ist auch oberhalb dieser der Substratdicke $h = 100 \mu\text{m}$ entsprechenden Größe eine Erhöhung der Dämpfung gegenüber der Referenzstruktur zu beobachten. Der Unterschied der Dämpfung zwischen einer isoliert geführten Leitung und der sich in $s = 150 \mu\text{m}$ vom Dotiergebiet entfernten Leitung ist jedoch in den meisten Fällen zu tolerieren, so daß ein Mindestabstand s_{min} in der Größenordnung der 1.5- bis 2-fachen Substratdicke h zwischen Hochfrequenzschaltungen und der CMOS-Auswerteelektronik eingehalten werden sollte.

Wertung und Ausblick

Der Schwerpunkt der vorliegenden Arbeit lag in der Untersuchung der Integrationsmöglichkeit von Silizium Millimeter-Wellen-IC-Hochfrequenztechnologie (SIMMWIC) und CMOS-Technik. Die erforderlichen Arbeitsschritte beinhalteten die Übertragung eines Aluminium-Gate-CMOS-Prozesses auf hochohmiges float-zone Silizium (FZ-Si) sowie die Charakterisierung der Einzelbauelemente in beiden Technologien. Dabei bestand das Ziel, eine Modellierung derselben für die Schaltungssynthese zu erreichen und Unterschiede zur Vergleichsbasis auf Standard-Silizium herauszuarbeiten. Die dabei gewonnenen Erkenntnisse dienten der Konzeption eines Prozesses zur monolithischen Integration. Dieser basierte für den CMOS-Schaltungsteil auf einem modifizierten Aluminium-Gate-CMOS-Prozeß in Doppel-Wannen-Technik und für den Hochfrequenz-Schaltungsteil auf den SIMMWIC-Prozeßschritten.

Die Arbeiten zur CMOS-Prozeßentwicklung gründeten sich auf elektrische und prozeßtechnische Parameterextraktionen des Instituts-CMOS-Prozesses auf konventionellem Substratmaterial und führten zur Untersuchung von drei grundsätzlichen Prozeßvarianten A, B und C [5.1]. Das für diese Varianten eingesetzte p-Substrat wies einen spezifischen Widerstand von $\rho \approx 5000 \Omega\text{cm}$ auf. Durch die gezielte Variation der Dotierkonzentration der implantierten Wannen in den Varianten B und C sind in Bezug auf die Gleich- und Wechselspannungsparameter sehr ähnliche MOS-Transistoren auf FZ-Si hergestellt worden.

Auf Durchgriffeffekte der niedrig dotierten Wannen zurückzuführende erhöhte Leckströme zwischen den Source- und Drain-Wannendioden erforderten die Kombination der PMOS-Fets aus der Variante B mit den NMOS-Fets der Variante C. Diese zusätzliche Doppelwannenversion, als Variante C₂ bezeichnet, verfügte über die vergrößerte Eindringtiefe der n-Wanne aus der Variante B und führte zu einer CMOS-Technik mit vergleichbar guten Leckströmen bei Betriebsspannungen bis etwa 7 Volt.

Aufgrund von Prozeßsimulationen und meßtechnischen Profilbestimmungen mittels SIMS- und Spreading Resistance-Messungen konnte der verbleibende Durchgriffeffekt durch numerische Berechnung der spannungsabhängigen Raumladungszonenweite sehr exakt bestimmt werden. Änderungen des Verhaltens der Vorwärtscharakteristik der Source-Wannen-Dioden in Bezug auf den Vergleichsstandard konnten mit erhöhten Bahnwiderständen aufgrund der in die Tiefe hin abfallenden Wannenkonzentrationen und einsetzender Leitfähigkeitsmodulation erklärt werden.

Die in der Variante A direkt auf dem hochohmigen FZ-Substrat realisierten NMOS-Fets zeigten selbstleitendes Verhalten, d.h. $V_{th} < 0 \text{ V}$. Diese Transistoren könnten zusammen mit den in der n-Wanne befindlichen PMOS eine Nischenanwendung als Treiber für Oszillatordioden, z.B. IMPATT-Dioden (IMPact ionisation Avalanche Transit Time), finden. Hierbei ist die für übliche CMOS-Prozesse kritische zwischen 20 - 30 V liegende Diodenspannung als Bias des Substrats nutzbar und dient dabei der weitgehenden Unterdrückung der auf Kurzkanaleffekten beruhenden Drainströme des NMOS im Unterschwellenbereich ($V_{gs} < V_{th}$). Interessanter jedoch sind diese Transistoren als Modellsystem zur Untersuchung von Kurzkanaleffekten anhand großer Strukturen.

Die Konservierung der hochohmigen Charakteristik des Substratmaterials im Bereich der SIMMWIC-HF-Schaltung während der CMOS-Prozessierung ist durch ausreichend dimensionierte Oxid-Maskierungsschichten erreicht worden. Spreading-Resistance-Messungen zeigten sowohl unterhalb der CMOS-Strukturen als auch unterhalb des maskierenden Oxides die Beibehaltung des hohen spezifischen Widerstands.

Nur sehr nah an der Oberfläche ist unterhalb des Maskieroxides eine geringe Reduktion dieses Widerstands gefunden worden. Diese war durch eine Elektronen-Inversionsschicht bedingt, die durch die im Oxid vorhandenen positiven Ladungen induziert wurde. Rechnerisch konnte gezeigt werden, daß unterhalb einer das Oxid überdeckenden Aluminiumleiterbahn ein leitender Oberflächenkanal im hochohmigen p-Silizium auch bei potentialfreier Leiterbahn vorliegt. Erst bei einer extrem niedrigen Ladungskonzentration im Oxid könnte eine Inversion des hochohmigen p-Siliziums ($\rho \approx 5000 \Omega\text{cm}$) an der Oberfläche vermieden werden. Solch geringe Konzentrationen

sind jedoch technologisch mit den heutigen Methoden der Siliziumtechnik nicht zu erreichen. Experimentell konnte die Präsenz einer Oberflächeninversion anhand von Messungen an MOS-Varaktoren eindeutig nachgewiesen werden.

Die Bestimmung der Grenzflächenzustandsdichten (D_{it} -Bestimmung) am Silizium-Siliziumoxid-Übergang hat sich bei den vorliegenden hochohmigen Siliziumsubstraten als außerordentlich schwierig gestaltet. Erst die Anwendung eines quasistatischen Meßverfahrens und die Absenkung der Minoritätsträgerkonzentration durch eine Abkühlung der Meßproben auf ca. $-50\text{ }^{\circ}\text{C}$ erlaubte die Bestimmung der CV-Kurven unter Hochfrequenz-Bedingungen. Die damit gefundenen niedrigen Konzentrationswerte bestätigten die sehr gute Qualität des Gateoxides der MOS-Transistoren auf hochohmigem p-Silizium.

Die Entwicklung des Integrationsprozesses erfolgte unter Verwendung des Doppel-Wannen-CMOS-Prozesses auf hochohmigem p-Silizium und den SIMMWIC-Prozessschritten. Insbesondere den Verbindungspunkten zwischen den an sich nicht kompatiblen Metallisierungen, Aluminium für CMOS und Gold für SIMMWIC, ist dabei Beachtung geschenkt worden. Die Minimierung der gegenseitigen Beeinflussung der Prozeßteile, aufgrund von Temperaturbudgetrestriktionen und ätztechnischen Erfordernissen, führte zu der gewählten Folge der Prozeßabschnitte. Die CMOS-Aluminium-Gate-Technologie erforderte einen temperaturfesten Schutz des Gateoxides bei den nachfolgenden Prozeßschritten, insbesondere während der Schichtabscheidung mittels Molekularstrahl-epitaxie. Eine sputtertechnische Deposition dieser Schutzschicht, wie etwa durch gesputtertes Siliziumnitrid (Si_3N_4) sowie eine spätere trocken-chemische Entfernung mittels RIE (reactive ion etching) fielen als Lösung aufgrund der auftretenden Strahlenschäden aus. Die Abscheidung eines undotierten CVD-Poly-Siliziums bei ca. $630\text{ }^{\circ}\text{C}$ und die spätere Entfernung mittels hochselektiver Kalilauge (KOH) wurde als eine erfolgversprechende Lösung erprobt. Hierdurch konnten die Schäden minimiert werden, die mit der Aufbringung und Entfernung dieser Schicht verbunden waren. Eine Untersuchungsserie an MOS-Varaktoren und einzelnen MOS-Transistoren zeigte die Machbarkeit dieser Methode und die Auswirkungen auf das Bauelementverhalten.

Unter Ausnutzung der bei diesen Untersuchungen gewonnenen Erfahrung konnte ein Integrationsprozeß für CMOS- und SIMMWIC-Schaltungen konzipiert werden. Die Realisierung eines pin-Hochfrequenz-Schalters mit Ansteuerschaltung hat die grundsätzliche Machbarkeit einer monolithischen Integration gezeigt.

Mit Hilfe spezieller Wellenleiterstrukturen konnte eine quantitative Aussage zur Beeinflussung der Dämpfungseigenschaften von Millimeter-Wellensignalen auf Silizium-Substrat bei Präsenz von CMOS-Schaltungen auf demselben Chip getroffen werden [5.2]. Der notwendige Mindestabstand zwischen CMOS- und SIMMWIC-Schaltungsteilen mit Mikrostreifenleitertechnik bei der monolithischen Integration konnte quantifiziert werden. Die Versuche haben gezeigt, daß ein Abstand in der Größe der ein-einhalbfachen bis doppelten Substrathöhe unbedingt erforderlich ist.

Durch die Verwendung von Wellenleiterstrukturen in Gold- und Aluminiumausführung konnten signifikant erhöhte Wellendämpfungswerte bei Alu-Mikrostrips gezeigt werden. Verursacht wurden diese Verluste zu einem großen Teil durch die unterhalb der Leiterbahnen vorhandenen Inversionsschichten.

In dieser Arbeit konnte durch die geschilderten Untersuchungen die grundsätzliche Integrationsfähigkeit von CMOS- und SIMMWIC-Schaltungen gezeigt werden. So konnten anhand der Variante A die Probleme der Nutzung hochohmigen Siliziums als Substrat für MOS-Transistoren dargestellt werden. Hierbei stehen insbesondere die Verschiebung der Schwellspannung hin zu selbstleitenden Transistoren und die ausgeprägten Kurzkanaleffekte mit deutlichen Strömen im Unterschwellenbereich einer breiten Anwendung im Wege.

Jedoch konnten bei einer herkömmlichen Alugate-CMOS-Techniken eng verwandten Prozessierung in den Varianten B, C und C_2 MOS-Transistoren mit Parametern hergestellt werden, die den elektrischen Kennwerten von MOS auf konventionellem n-Substrat sehr ähnlich sind. Untersuchungen zeigten desweiteren, daß durch die geringen Unterschiede in den Parametern die Übernahme bestehender analoger und digitaler Schaltungskonzepte möglich ist, so daß zeitintensive Neuentwicklungen in der Zukunft vermieden werden können.

Jedoch zeigte sich bei den Varianten B und C bei jeweils einem Transistortyp die Problematik eines vertikalen Wannendurchgriffs, der bei höheren Drain-Source-Spannungen ebenfalls zu erhöhten Leckströmen der MOS-FETs führte. Aufgrund der doppelt-verschachtelten Wannenstruktur erscheint eine komplette Unterdrückung dieser Effekte bei dem NMOS-Transistor der Variante B als aussichtslos. Als Folge davon wurde eine der Variante C sehr ähnliche Doppelwannentechnik mit der vergrößerten n-Wannentiefe der Variante B entworfen. In dieser Variante C_2 konnten Durchgriffeffekte beim NMOS bis zur Betriebsspannung von 12 V vermieden werden. Auch der PMOS-Transistor zeigte bis zu Drainspannungen von etwa 7 V keine besonderen Auffälligkeiten. Als zukünftige Aufgabe ist daher eine Erhöhung der Durchgriffspannung des PMOS zu erreichen. Auf Basis der hier zugrundeliegenden Alugate-CMOS-Technik wäre dies durch eine erhöhte Kanaldotierung oder eine noch tiefere n-Wanne zu erreichen. Da jedoch schon jetzt eine recht lange Drive-In Zeit zu einer verhältnismäßig tiefen n-Wanne führt, sollte der Weg einer Erhöhung der Dotierung im Kanalbereich gewählt werden. Ein solcher Schritt

würde allerdings eine aufwendige Modifikation der Prozeßparameter bedeuten.

Damit in Zukunft kostengünstige CMOS-SIMMWIC-Schaltkreise kommerziell verfügbar gemacht werden können, ist deshalb nicht zuletzt aufgrund dieses Aufwands ein Übergang vom hier benutzten Alugate-Prozeß auf einen modernen Submikron-CMOS-Prozeß sinnvoll. Dies würde die mögliche Komplexität der CMOS-Schaltungen deutlich erhöhen und vertikale Durchgriffeffekte bei entsprechend höheren Wannendotierungen verhindern. Auch bei einem solchen Prozeß sind jedoch laterale Durchgriffe zwischen Wannern mit unterschiedlichem Wannepotential grundsätzlich möglich, so daß mit diesem Effekt und der Gefahr des Latch-Up aufgrund des hochohmigen Substrats unterhalb der Wannern gerechnet werden muß.

Ein moderner CMOS-Prozeß würde jedoch durch die damit einhergehenden Poly-Silizium-Gates zu einer Vereinfachung der Integrationssequenz führen, da die Entfernung der Epitaxieschichten oberhalb des CMOS-Teils einfacher und prozeßsicherer zu erreichen wäre. Um eine Schädigung der MOS-Bauelemente durch ionisierende Strahlung am sichersten zu vermeiden, wäre es empfehlenswert, wie in dieser Arbeit, auf die plasmaunterstützte Ätzung zu verzichten und vorzugsweise naßchemische Verfahren zur lokalen Entfernung dieser Schichten anzuwenden.

Allerdings verbleibt bei jeder monolithischen Lösung die Problematik der Verschmelzung der von sich aus sehr unterschiedlichen Prozeßmodule der CMOS- und der SIMMWIC-Technik. Insbesondere der zentrale Prozeßschritt der SIMMWIC-Technik, d.h. die prozeßtechnisch sehr aufwendige Epitaxieschichtabscheidung, ist mit der CMOS-Technik grundsätzlich nicht kompatibel, so daß diese auf einem großen Teil der Waferoberfläche wieder komplett entfernt werden muß.

Gerade aus dieser Überlegung heraus und nach den positiven Erfahrungen mit der hybrid aufgebauten SIMMWIC-Rectenna mit CMOS-Vorverstärker sollte auch in Zukunft eine zweigleisige Vorgehensweise gewählt werden. Neben der Suche nach einem verbesserten monolithischen Integrationsprozeß auf Basis einer modernen CMOS-Technologie sollte dem Bereich der Multi-Chip-Technik besondere Aufmerksamkeit geschenkt werden. Der hybride Integrationsansatz eröffnet in Bezug auf den zu erwartenden Prozeßaufwand einen guten Mittelweg und erlaubt gleichzeitig die individuelle Optimierung der CMOS- und der SIMMWIC-Technik.

Anhang

A 1 Entwurfsdaten aktiver und passiver Bauelemente

In diesem Anhang sind die Maskenentwurfsdaten der verwendeten aktiven und passiven Bauelemente zusammengefaßt. Ferner sind im ersten Teil über MOS-Bauelemente die mit einem kombinierten meßtechnisch-rechnerischen Verfahren ermittelten Kanallängen- und -breitenveränderungen aufgeführt. Der zweite Teil handelt von passiven Bauelementen. Eine ähnliche Abweichung der elektrisch wirksamen Geometrie von der maskentechnisch vorgegebenen führt zu den bei ohmschen Widerständen relevanten Korrekturgrößen. Im dritten Teil wird die Notwendigkeit von Isolationen und Mindestabständen beim Maskenentwurf erläutert.

A 1.1 Geometrische Kenngrößen von MOS-Transistoren

Ein grundsätzlicher Unterschied zwischen Bipolartransistoren und MOS-Transistoren liegt in der dominierenden Richtung des Stromflusses. Während dieser beim Bipolartransistor weitgehend vertikal verläuft, ist der MOS ein vorwiegend lateral arbeitendes Bauelement. Deshalb ist die laterale Dimensionierung von MOS-Transistoren die wesentliche Entwurfsgröße zur Einstellung der Stromführungs- und Verstärkungseigenschaften. Der Layout-Entwurf und die Übertragung in real vorhandene Größen, sowie deren Extraktion für die Schaltungsberechnung ist daher Thema der beiden folgenden Abschnitte.

Layout-Entwurf für MOS-Transistoren

Zur Definition der Öffnungen im Oxid für Dotierungsprozesse und Kontaktlöcher sowie für die Bildung des Gateoxides sind bei dem hier eingesetzten Alu-Gate-CMOS-Prozeß insgesamt sieben Masken erforderlich. Als Illustration ist hier die Abfolge für den Doppel-Wannen-Prozeß der Variante C angegeben. Bei den Einzel-Wannen-Prozessen der Varianten A und B entfällt jeweils eine Maske zur Wannendefinition, im ersten Fall die p-Wanne und im zweiten Fall die n-Wanne. Die folgende Tabelle 14 listet die Masken in der Prozeßabfolge auf und beschreibt ihre Funktion:

TABELLE 14. Maskensatz für MOS-Transistoren im Doppel-Wannen-Prozeß

Maskenebene	Name	Funktion
1	n-Wanne	Öffnung des Oxides für die n-Wannenimplantation; Definition von hochohmigen Widerständen
2	p-Wanne	Öffnung des Oxides für die p-Wannenimplantation; Definition von hochohmigen Widerständen

TABELLE 14. Maskensatz für MOS-Transistoren im Doppel-Wannen-Prozeß

Maskenebene	Name	Funktion
3	p ⁺ -Diffusion	Source- und Drain-Anschlüsse des PMOS; Kontaktierung der p-Wanne und der p-Typ-Widerstände; Bildung der Gegelektrode bei integrierten Oxidkapazitäten
4	n ⁺ -Diffusion	Source- und Drain-Anschlüsse des NMOS; Kontaktierung der n-Wanne und der n-Typ-Widerstände; Bildung der Gegelektrode bei integrierten Oxidkapazitäten
5	Gatefenster	Öffnung der Bereiche, in denen das Gateoxid gebildet wird; Öffnung aller Kontaktlöcher
6	Kontaktlöcher	Öffnung aller Kontaktlöcher
7	Alu-Metallisierung	Definition der Gates und Anschlußbahnen

Die bei der Prozessierung eines MOS-Transistors erforderliche Lithographie-Präzision ist in erster Linie für die Definition der Bereiche Source und Drain (Maske 3 oder 4), Gatefenster (Maske 5) und Alu-Metallisierung (Maske 7) gefordert. Dejustagen in diesen Ebenen zeigen einen großen Einfluß auf das Bauelementverhalten. Die Positionierung der Wannen ist aufgrund großzügig dimensionierter Sicherheitsabstände zu den Transistoren unkritisch. Da die Ätzung der Kontaktlöcher schon mit der Gatefenstermaske erfolgt, ist der Abstand zwischen Gateoxid und Kontaktloch fixiert. Die Öffnung vor der Aluminiumbedampfung - nach erneuter Abdeckung durch die Gateoxidation - erfolgt zwar durch die separate Maske 6, jedoch kommt es aufgrund der geringen Oxiddicke bei der Gateoxidation nur in den schon vorher bestimmten Positionen - wie gewünscht - zu einer Öffnung bis auf die Siliziumoberfläche.

Zur Positionierung der einzelnen Masken wurde am Institut für Halbleitertechnik bisher jeweils pro aufeinanderfolgendes Maskenpaar eine aus einem Quadrat und einem Justierkreuz bestehende Kombination angewandt, die in Bild A1.1 auf der linken Seite gezeigt ist. Hierbei wird das Quadrat jeweils in der unteren und das Justierkreuz in der oberen Maske dargestellt und ggf. durch Ätzung bleibend auf den Wafer übertragen. Der Nachteil dieser einfachen Justiermöglichkeit liegt in der Fehlerfortpflanzung durch eine von Maskenebene zu Maskenebene fortschreitende Fehlersummutation gemäß [A1.1]

$$\Delta_t = \left(\sum_i \Delta_i^2 \right)^{1/2}, \quad (\text{A1.1})$$

wobei sich die Einzelfehler Δ_i im besten Fall gegenseitig kompensieren, so daß auch die letzte Maskenebene mit ausreichender Genauigkeit plaziert wird. Es besteht jedoch auch die Gefahr stetig anwachsender Fehler, die nur bei Betrachtung des Gesamtentwurfs aufgedeckt werden können.

Alternativ besteht die Möglichkeit, durch Platzierung aller Quadrate auf die unterste Maskenebene eine Justage durchzuführen, die sich jeweils auf den gleichen Referenzpunkt bezieht. Eine gewisse Unübersichtlichkeit ist die Folge, da pro Maskenebene ein Quadrat schon auf dieser ersten Maske vorgesehen werden muß. Bei einer je nach Prozeßzyklus gering veränderten Maskenabfolge, wie dies zur Erstellung der drei Prozeßvarianten A, B und C erforderlich ist, sind zusätzliche Justierkreuze notwendig.

Die in der Mitte und rechts in Bild A1.1 dargestellten Justierstrukturen sind neu entworfen worden und erlauben die Justage relativ zu einem in der untersten Ebene durch Ätzung entstandenen Kreuz. Die hier dargestellte Folge wurde für die fünf Masken 3,4,5,6 und 7 eingesetzt, ist aber prinzipiell auch für mehr Ebenen nutzbar, wenngleich dies zu einem quadratisch mit der Anzahl der Maskenebenen steigenden Flächenbedarf führt.

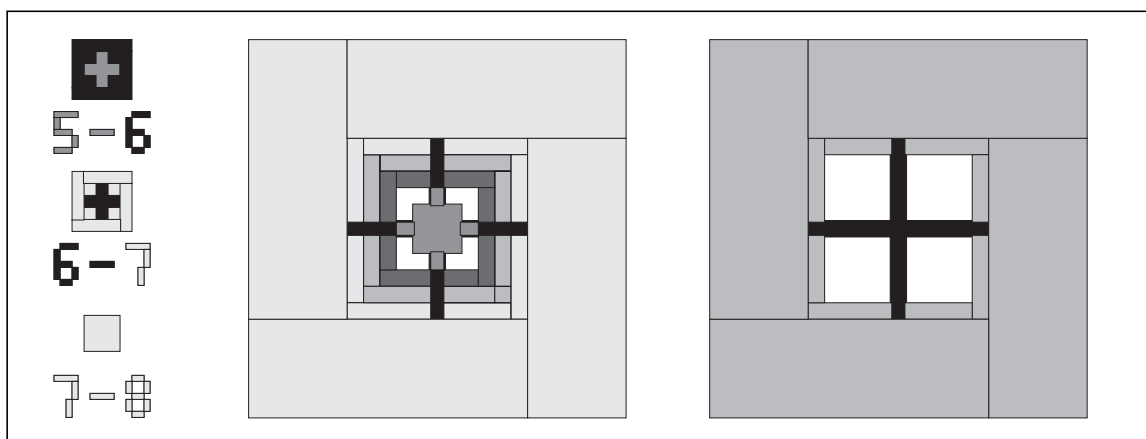


Bild A1.1: Links konventionelle Justierkreuze, bei welchen eine quadratische Öffnung des vorherigen Lithographieschritts (unten) die Bezugsbasis für das obenliegende Kreuz bildet (Mitte und oben). Mitte und rechts neuentworfene Justierkreuze, die eine Justierung aller Masken in Bezug auf einen gemeinsamen Punkt erlauben. In der Mitte die Gesamtstruktur für fünf Masken (P⁺-Diffusion, N⁺-Diffusion, Gatefenster, Kontaktlöcher und Alu-Metallisierung). Rechts sind zur Verdeutlichung nur die erste (schwarz - führt zu einer kreuzförmigen Ätzöffnung mit der Stegbreite von 10 µm im Oxid) und zweite Maske (hellgrau - beläßt die weiß gezeichnete Oxidstruktur für nachfolgende Maskenjustagen) gezeichnet.

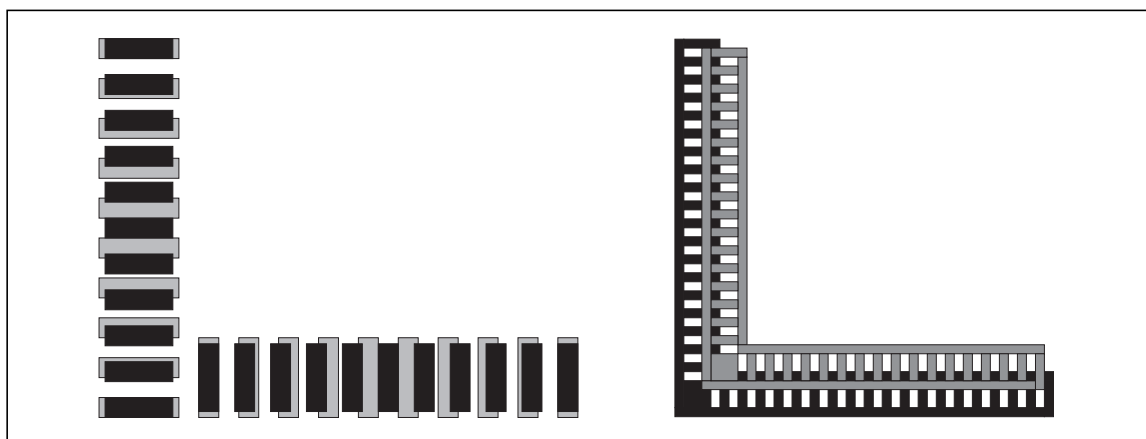


Bild A1.2: Links Nonius-Struktur, die durch partielle Überlappung der oberen Rechteckstrukturen (Größe 10 µm * 34 µm mit je 8 µm Abstand) bzgl. der untenliegenden (Größe 10 µm * 40 µm mit je 10 µm Abstand) ähnlich dem Prinzip einer Mikrometerschrauben-Teilung eine deutlich genauere Maskenjustage erlaubt. Rechts wechelseitig versetzte Balkenstruktur (Größe jedes Balkens 5 µm * 12 µm), die auf allen Maskenebenen befindlich, zur Bestimmung des Gesamtversatzes zweier beliebiger Masken dient.

Neuartige Justierstrukturen zur Feinpositionierung sind in Bild A1.2 dargestellt. Die links gezeigte Nonius-Struktur ist in dieser Form für ein Maskenpaar einsetzbar und wurde für die Masken 5 (Gatefenster) und 7 (Alu-Metallisierung) benutzt. Die rechts abgebildete Struktur, in der jeweils rechteckige 5 µm * 12 µm große Balken verbleiben, kann auf beliebig viele Maskenebenen erweitert werden, erlaubt jedoch nur eine reduzierte Genauigkeit im Vergleich zur ersten Struktur.

Die elektrischen Bauelementeigenschaften werden von unvermeidlichen Fehljustagen bei den Lithographieschritten in zweifacher Weise beeinträchtigt:

(i) Das durch die Maske 5-Gatefenster definierte Gateoxid muß durch das mittels Maske 7-Alu-Metallisierung strukturierte Gatemetall zwischen den Source- und Drain-Kontakten, definiert - je nach Typ - durch Maske 3 oder 4, komplett abgedeckt werden, da andernfalls eine Unterbrechung des Inversionskanals und damit ein Totalausfall der MOS-Transistoren auftreten könnte. Eine sichere Überlappung wird durch eine großzügige laterale Dimensionierung des Gateoxides und ein entsprechend vergrößertes Gatemetall erreicht. Dies erfolgt jedoch zu Lasten der Schaltgeschwindigkeit, da erhöhte Gate-Source- und Gate-Drain-Kapazitäten auftreten.

(ii) Fehljustagen geringeren Ausmaßes führen zu unerwünschten Parameterschwankungen. Hierbei liegt der Einfluß in einer Kanalquerschnittsänderung ΔW und damit in einer Änderung des Kanalstroms bei sonst glei-

chen Spannungen am Bauelement. Die im Maskenentwurf relevanten Justiertoleranzen sind in der Darstellung in Bild A1.3 gezeigt.

Da der Abstand zwischen Source und Drain durch eine einzige Maske - entweder 3 oder 4 - festgelegt ist, wirkt sich eine geringfügige Dejustage in Längsrichtung, die zwischen dieser Maske und den beiden den Gatebereich definierenden Masken 5 und 7 auftritt, allenfalls in einer geänderten Verteilung der Überlappungskapazitäten aus.

Bei einer Dejustage in Querrichtung ist eine Verkleinerung der Kanalbreite W möglich, wenn das Gatefenster außerhalb des Bereichs zwischen Source und Drain zu liegen kommt und somit die vergleichsweise ungenau bestimmbare Diffusionsfläche der MOS-Anschlußgebiete die Kanalbreite festlegt. Durch die Wahl einer gegenüber dem Gatefenster vergrößerten Source- und Drainbreite läßt sich dieser Einfluß auf den Kanalstrom minimieren, da in diesem Fall die höhere Genauigkeit der Gatefensterdefinition bestimmend wird. In beiden Fällen der Kanalbreitenfestlegung, d.h. entweder durch die Source-Drain-Kontakte oder durch das Gatefenster, erfolgt eine Aufweitung der jeweiligen Gebiete durch die Oxidätzung gegenüber dem ursprünglichen Maß der Maske. Im Fall der Kanalbreitenbestimmung durch die Source- und Draingebiete erfolgt zusätzlich eine Vergrößerung durch die Weite der Unterdiffusion, die von Eintreibzeiten und Vorbelagungskonzentrationen abhängig ist.

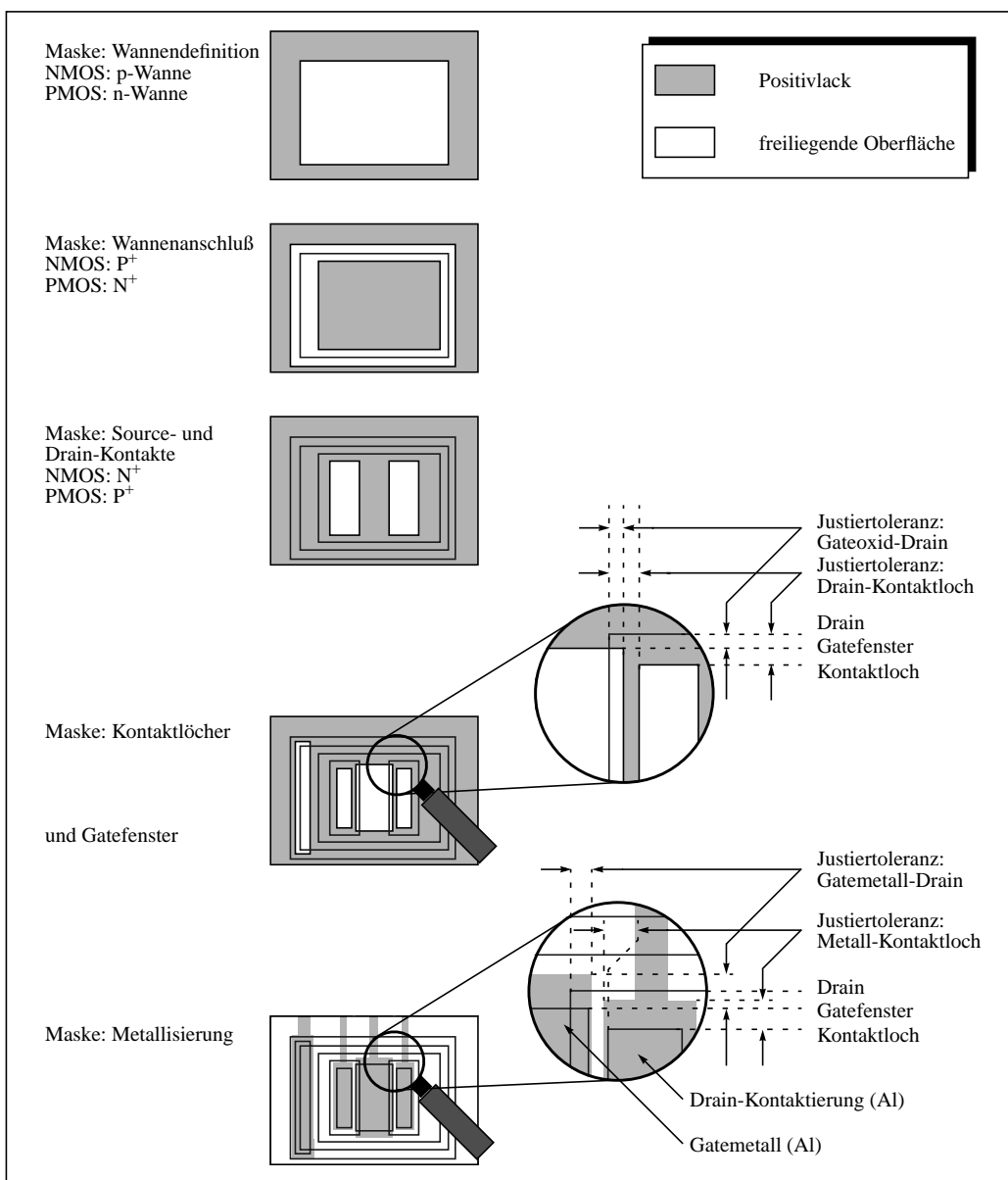


Bild A1.3: Photomasken für Positivlack zur Herstellung einer MOS-Struktur. Darstellung der Waferoberfläche nach Auftragung des Photolacks. Nach vorheriger Strukturierung verbleibende Oberflächenstufen sind in den darauffolgenden Schritten ebenfalls mit eingezeichnet.

Die für den MOS-Entwurf relevanten Maße sind in der folgenden Tabelle 15 zusammengefaßt und in Bild A1.4 dargestellt.

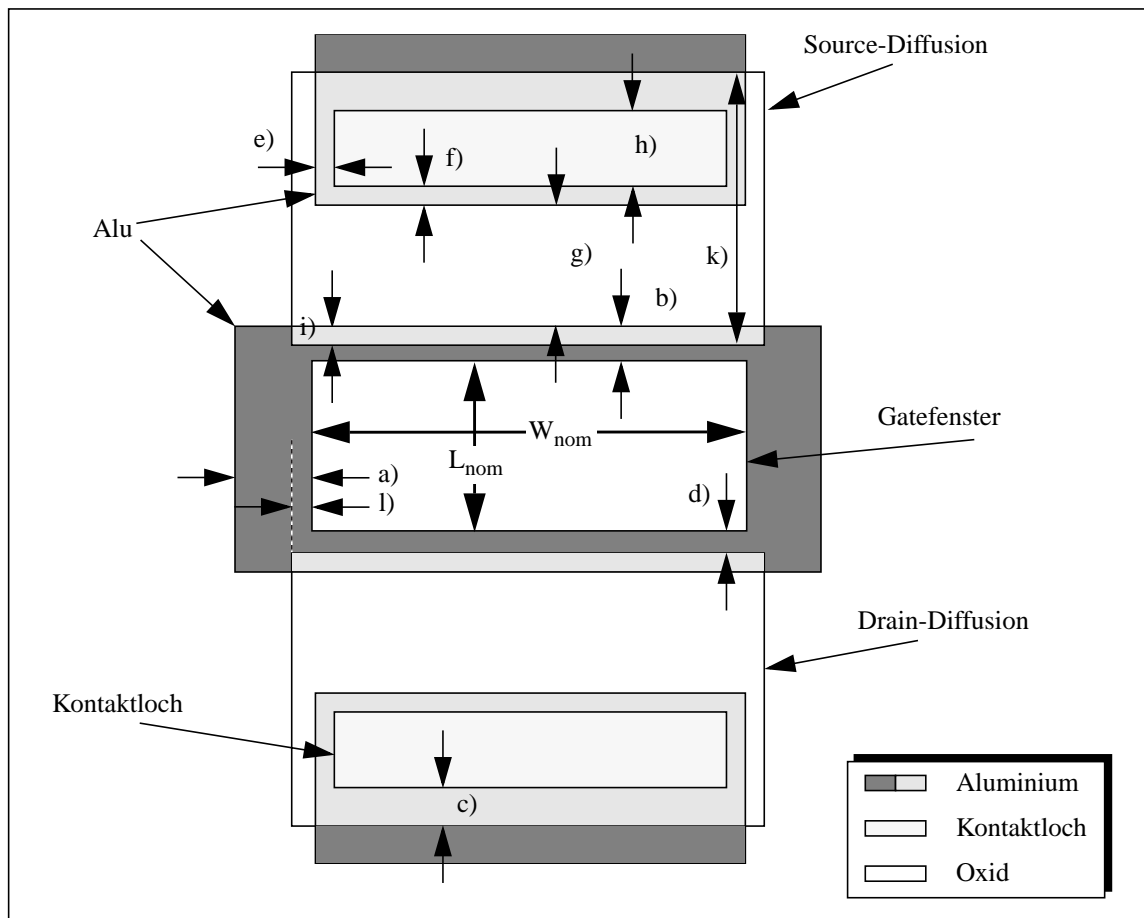


Bild A1.4: Darstellung aller wichtigen Abmessungen und Abstände der entworfenen MOSFETs.

Die Maße a) bis l) in der obigen Darstellung und in der folgenden Tabelle gelten für NMOS und PMOS-Transistoren unterschiedlicher Kanaldimensionierung. Bei einer gewünschten Kanalgeometrieänderung erfolgt eine Variation der Entwurfsparameter nominelle Kanallänge L_{nom} und -weite W_{nom} unter Beibehaltung aller übrigen Größen.

TABELLE 15. Entwurfsmaße der MOS-Lithographie-Masken

Entwurfsmaß des MOS ist gleich der Größe der Gateoxidätzung	$W_{nom} * L_{nom}$	Überlappung des Metalls über Kontaktlöcher in der Länge	$f = 4 \mu\text{m}$
Überlappung des Metalls über das Gateoxid in der Breite	$a = 5 \mu\text{m}$	Abstand zwischen Aluminiumbahnen *)	$g = 8 \mu\text{m}$
Überlappung des Metalls über das Gateoxid in der Länge	$b = 3 \mu\text{m}$	Größe des Kontaktlochs in Längsrichtung *)	$h = 10 \mu\text{m}$
Abstand des Kontaktlochs vom Diffusionsrand	$c = 4 \mu\text{m}$	Überlappung des Gatemetalls über Source- und Drain-Diffusion	$i = 2 \mu\text{m}$
Abstand Diffusionsrand-Gatefenster	$d = 1 \mu\text{m}$	Breite der Source- und Drain-Diffusion *)	$k = 28 \mu\text{m}$
Überlappung des Metalls über Kontaktlöcher in der Breite	$e = 5 \mu\text{m}$	Seitliche Breite des Diffusionsfensters über das Gatefenster hinaus	$l = 1 \mu\text{m}$

*) Die mit Stern bezeichneten Werte stellen Mindestwerte des Entwurfs dar.

Bestimmung der effektiven Kanaldimensionen W_{eff} und L_{eff}

Eine Abweichung zwischen der gewünschten Geometrie und der nach der gesamten Prozessierung erhaltenen begründet sich aus mehreren Effekten:

(i) Die durch photolithographische Methoden hergestellte Maske zeigt je nach geforderter Auflösung unterschiedlich ausgeprägte Vergrößerungen der transparenten und lichtundurchlässigen Flächen. Bei den im Rahmen dieser Arbeit verwendeten Masken kommt es generell zu einer geringfügigen Vergrößerung der transparenten Flächen im Vergleich zu identisch gezeichneten schwarzen Bereichen. Der Einfluß dieser Geometrieänderung Δx_1 ist jedoch in Bezug auf die gewählte Strukturgröße zu vernachlässigen.

(ii) Durch die Belichtung und Entwicklung des Positivlacks auf der zu strukturierenden Waferoberfläche erfolgt eine weitere Aufweitung der freigelegten Gebiete. Dieser mit Δx_2 bezeichnete Beitrag ist weitgehend durch Lichtreflexion und Beugung erklärbar.

(iii) Durch die im Prozeß benutzten feuchten SiO_2 -Ätzungen mittels gepufferter Flußsäure (BHF) erfolgt ein isotroper Ätzangriff, so daß neben der gewünschten Ätzung in vertikaler Richtung eine zusätzliche Aufzehrung des Maskieroxides in lateraler Richtung erfolgt. Dies führt zu einer Verbreiterung der geätzten Gebiete um Δx_3 .

(iv) Für die Unterdiffusion der Dotieratome in lateraler Richtung gilt $\Delta x_4 = 0.8 \cdot x_j$, wobei mit x_j die Tiefe des pn-Übergangs innerhalb der geöffneten Struktur bezeichnet ist.

Während z.B. für die Definition des Gatefensters die Einflüsse $\Delta x_1 + \Delta x_2 + \Delta x_3$ beachtet werden müssen, wirkt sich bei diffundierten Gebieten zusätzlich Δx_4 aus.

Auf das Verhalten eines MOS wirken sich diese Geometrieänderungen als Verkürzung des Kanalgebiets gemäß

$$L_{eff} = L_{nom} - 2\Delta L \quad \text{mit} \quad \Delta L = \sum_i a_i \Delta x_i \quad (\text{A1.2})$$

und als Verbreiterung des Kanalgebiets gemäß

$$W_{eff} = W_{nom} + 2\Delta W \quad \text{mit} \quad \Delta W = \sum_i b_i \Delta x_i \quad (\text{A1.3})$$

aus. Hierbei drücken die Koeffizienten a_i und b_i den jeweils unterschiedlich großen Einfluß der einzelnen Beiträge aus. Da die Kanalverkürzung ΔL durch die beiden Diffusionsgebiete Source und Drain beeinflusst wird, zeigt diese einen Einfluß durch Δx_4 . Demgegenüber wird die Kanalverbreiterung ΔW diesen Einfluß nicht zeigen, da diese durch das Gatefenster definiert ist. Der Betrag von ΔW wird demzufolge deutlich geringer ausfallen als ΔL .

Für die konkrete Bestimmung der Kanallängenverkürzung $\Delta L < 0$ und die Kanallängenverbreiterung $\Delta W > 0$ wurden MOS der Designmaße $W_1 = 50 \mu\text{m}$ und $L_1 = 10 \mu\text{m}$, $W_2 = 160 \mu\text{m}$ und $L_2 = 10 \mu\text{m}$ sowie $W_3 = 200 \mu\text{m}$ und $L_3 = 43 \mu\text{m}$ verwendet. Nach Bestimmung der transistorspezifischen Schwellspannung V_{th} wird der Drainstrom I_d im linearen Bereich bei geringen Drainspannungen V_{ds} , hier $V_{ds} = 50 \text{ mV}$, vermessen. Aus

$$I_d = \beta \left[(V_{gs} - V_{th}) V_{ds} - \frac{V_{ds}^2}{2} \right] \quad (\text{A1.4})$$

läßt sich der Steilheitsparameter β bzw. die prozeßspezifische Kenngröße k_p bestimmen

$$\beta = k_p \frac{W_{eff}}{L_{eff}} \quad \text{mit} \quad k_p = \mu_n C'_{ox} \quad . \quad (A1.5)$$

Da k_p - nur aus der Trägerbeweglichkeit μ_n und der spezifischen Oxidkapazität C'_{ox} bestehend - für Transistoren in räumlicher Nähe als konstant angenommen werden kann, besteht die Möglichkeit, durch den Vergleich der Drainströme von MOS mit unterschiedlicher Geometrie die Kanalgeometrieänderungen ΔW und ΔL , die für alle Transistoren eines Typs als gleich anzunehmen sind, zu bestimmen. Für zwei Transistoren T_1 und T_2 gilt dann für die Weitenänderung

$$\Delta W = \frac{1}{2} \cdot \frac{\beta_2 W_1 (L_2 - 2\Delta L) - \beta_1 W_2 (L_1 - 2\Delta L)}{\beta_1 (L_1 - 2\Delta L) - \beta_2 (L_2 - 2\Delta L)} > 0 \quad (A1.6)$$

und die Längenänderung

$$\Delta L = \frac{1}{2} \cdot \frac{\beta_1 L_1 (W_2 + 2\Delta W) - \beta_2 L_2 (W_1 + 2\Delta W)}{\beta_1 (W_2 + 2\Delta W) - \beta_2 (W_1 + 2\Delta W)} < 0 \quad . \quad (A1.7)$$

Bei einer direkten Lösung dieser Gleichungen für nur zwei Transistoren ergeben sich deutliche Ungenauigkeiten, die zu umgehen sind, wenn mindestens drei unterschiedliche Geometrien eingesetzt werden, wie dies hier mit den oben genannten Transistoren T_1 , T_2 und T_3 der Fall ist. Mit Hilfe eines Tabellenkalkulationsprogramms (z.B. Microsoft EXCEL) kann eine bessere Lösung gesucht werden. Für NMOS der Variante C und PMOS-Transistoren der Variante B wurden die in Tabelle 16 aufgeführten Geometrieänderungen ermittelt

TABELLE 16. Geometrieänderungen

	NMOS	PMOS
Kanallängenverkürzung ΔL [μm]	2.27	2.70
Kanallängenverbreiterung ΔW [μm]	0.15	0.31

A 1.2 Geometrische Kenngrößen passiver Bauelemente

Neben den aktiven Bauelementen wie MOS-Transistoren und Dioden erfordern analoge integrierte Schaltungen auch passive Strukturen, wie etwa Kondensatoren und Widerstände.

Widerstände

Beim Entwurf integrierter analoger Schaltungen in CMOS-Technik besteht das Bestreben in einer weitgehenden Vermeidung von ohmschen linearen Widerständen, da diese als Lastelement einer Verstärkerstufe einen sehr engen Kompromiß zwischen dem möglichen Hub der Ausgangsspannung und der maximalen Verstärkung bedeuten. Allerdings eignen sich Widerstände sehr gut zur Bereitstellung von definierten Spannungsabfällen bei der Dimensionierung von Konstantstromquellen.

In dieser Arbeit sind laterale rechteckige Widerstandsstrukturen realisiert worden, deren typische Dimensionierung in Bild A1.5 dargestellt ist.

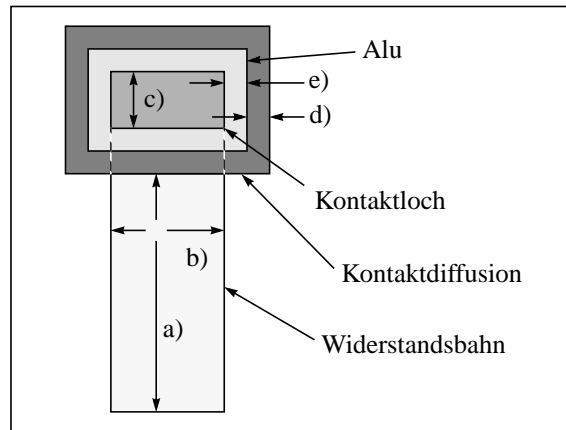


Bild A1.5: Dimensionierungsregeln für lineare ohmsche Widerstände. Die Länge L der Widerstandsbahn entspricht dem Vertikalmaß a, die Breite W dem Maß b. Die Kontaktlöcher haben eine konstante Länge c = 10 µm und eine Breite wie die Widerstandsbahn - mindestens aber 10 µm. Das Aluminium überlappt das Kontaktloch jeweils um d = 3 µm und ist wiederum e = 2 µm kleiner als die Kontakttdiffusion.

Mit Hilfe dieser Strukturen sind Widerstände von $R_{p\text{-Wanne}} = 5 \text{ k}\Omega \dots 40 \text{ k}\Omega$ durch die p-Wannen-Implantation realisiert worden. Deutlich geringere Werte sind durch den Einsatz der Bor- und Phosphordiffusionsschritte erreicht worden (einige 10 Ω bis einige 100 Ω). Vielfache und Teiler dieser Werte lassen sich durch Serien- bzw. Parallelschaltung jeweils separater designgleicher Widerstände erreichen. Auf überlange Strukturen ist aus Ausbeute- und Platzersparnissen verzichtet worden. Strukturen mit gewundenen Widerstandsbahnen lassen sich demgegenüber - wie auch in [A1.2] dargelegt - nur ungenau entwerfen. Die gemessenen Werte von drei verschiedenen Widerstandsstrukturen mit Längen- zu Breitenverhältnissen von $L/W = 3 \dots 22$ sind an die Längen- und Breitenabweichungen erfassende Widerstandsdefinition

$$R = R_{SH} \cdot \frac{L + \Delta L}{W + \Delta W} \tag{A1.8}$$

angepaßt worden. Die Daten des jeweiligen Schichtwiderstands R_{SH} , der Längen- und Breitenveränderung ΔL und ΔW dazu sind in Tabelle 17 aufgelistet.

TABELLE 17. Parameter des Widerstandsdesigns

	ΔL	ΔW	$R_{SH, \text{rechnerisch}}$	$R_{SH, 4 \text{ Punkt-Messung}}$	Prozeßschritt
Bor implantiert *)	-7.6 µm	+2.1 µm	2035 Ω	1900 Ω	p-Wanne
Bor diffundiert *)	+17.5 µm	+7.1 µm	20 Ω	39 Ω	p ⁺
Phosphor diffundiert *)	+23.4 µm	+13.7 µm	15 Ω	12 Ω	N ⁺

*) Prozeßparameter wie in Variante C

Die Längenänderung ΔL zeigt bei der implantierten Struktur die erwartete Verkürzung, da der hochdotierte P⁺-Kontakt durch Unterdiffusion die effektive Länge der implantierten Struktur verringert. Die Breitenvergrößerung $\Delta W = 2 \text{ µm}$ ist ebenfalls physikalisch sinnvoll. Größere Korrekturwerte ergeben sich für die diffundierten Strukturen. Deren positive Längenänderung ist durch den zusätzlichen Widerstand des Kontaktierungsbereichs bedingt, der sich auf dem gleichen vergleichsweise niedrigen Widerstandsniveau wie die Widerstandsbahn befindet. Es bleibt jedoch zu bemerken, daß mit diesen Werten eine möglichst hohe Entwurfsgenauigkeit und weniger eine direkte physikalische Relevanz angestrebt wird. Ein darauf basierender Entwurf von implantierten 50 kΩ-Widerständen erreichte den Zielwert bis auf 5 % genau.

Die Linearität der Widerstandswerte ist sehr hoch - die Widerstandswerte ändern sich im Bereich der Versorgungsspannungen von ± 6 V um höchstens 1 %. Die mittlere Schwankung der Widerstandswerte um den Mittelwert liegt bei implantierten Strukturen bei ± 2 % und bei den diffundierten Strukturen bei ± 5 %.

Kondensatoren

An Kondensatorstrukturen für den Einsatz in integrierten Schaltungen werden Forderungen nach (i) möglichst vernachlässigbaren Parallelleitwerten, (ii) geringer Größe und (iii) guter Reproduzierbarkeit gestellt. In dem hier benutzten CMOS-Prozeß mit einer einzelnen Metallisierungsebene als obere Kondensatorplatte bietet sich eine Struktur an, in der eine dotierte Schicht unterhalb des als Dielektrikums genutzten Gateoxides die Gegenelektrode bildet. Mögliche Dotierungsschichten müssen eine hohe Konzentration aufweisen, um der Forderung nach ideal leitenden Platten möglichst nahe zu kommen. Ferner kann so Spannungsabhängigkeit der Kapazität im betrachteten Spannungsbereich zwischen den beiden Versorgungsspannungen ± 6 V vermieden werden.

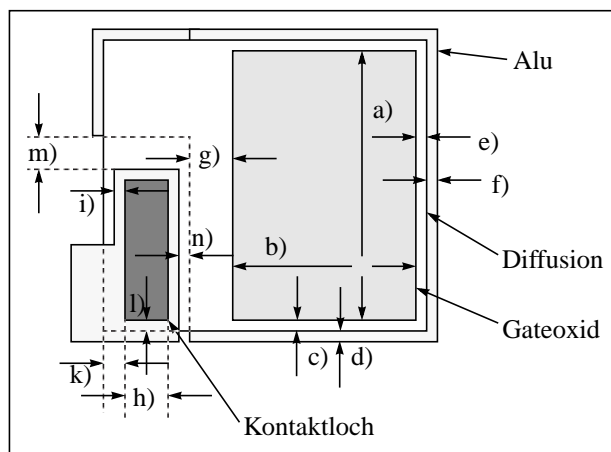


Bild A1.6: Dimensionierungsregeln für Kondensatorstrukturen mit Gateoxid als Dielektrikum. Die zur Kapazität beitragende Fläche des Gateoxides hat die Länge a und die Breite b . Für die weiteren gezeichneten Maße gelten die Werte: $c \dots f = 1 \mu\text{m}$, $g = 2 \mu\text{m}$, $h = 10 \mu\text{m}$, $i = 2 \mu\text{m}$, $k \dots l = 6 \mu\text{m}$, $m = 20 \mu\text{m}$ und $n = 10 \mu\text{m}$.

Mit dem in Bild A1.6 gezeigten Entwurf ergeben sich die Kapazitätswerte

$$C = \epsilon_0 \epsilon_{\text{SiO}_2} \cdot \frac{a \cdot b}{t_o \cdot t_{\text{ox, Gate}}}, \quad (\text{A1.9})$$

wobei für $t_{\text{ox, Gate}}$ die Gateoxidstärke über den p- oder n-Wannen einzusetzen ist. Während sich niedrig dotierte Bereiche ohne merklichen Einfluß auf die Oxidstärke bei trockener Oxidation erweisen, unterscheidet sich die während der Gateoxidation erreichte Dicke des Oxides im Bereich der hochdotierten N^+ - und P^+ -Zonen. Dies kommt durch den Korrekturfaktor t_o zum Ausdruck, der die Werte $t_{o, \text{P}^+} = 1.02$ für die Bereiche der Borddiffusion und $t_{o, \text{N}^+} = 1.2$ für die Bereiche der Phosphordiffusion annimmt. Bei niedrigdotierten Bereichen, z.B. oberhalb von Wannen, wird $t_o = 1$. Für die Oxidstärke x als Funktion der Zeit t gilt

$$x^2 + Ax = B(t + \tau), \quad (\text{A1.10})$$

worin τ eine Modellierung der frühen Phase der Oxidation erlaubt. Bei der Gateoxidation ($T = 1000 \text{ }^\circ\text{C}$ und $t = 130 \text{ min}$ in trockenem Sauerstoff) kann dieser Summand vernachlässigt werden. Je nach Oxidationsdauer t kann hieraus ein linearer Zusammenhang $x = (B/A) \cdot t$ für kleine Zeiten t und $x = (Bt)^{1/2}$ für große Zeiten t abgeleitet werden. Je nach Dauer der Oxidation ist also entweder die sogenannte lineare Wachstumsratenkonstante B/A oder die parabolische B entscheidend. Die lineare Rate drückt die Reaktionsgeschwindigkeit der Sauerstoff- und der Siliziumatome am Interface aus, während die parabolische Rate ein Maß für die Diffusionsgeschwindigkeit des Sauerstoffs oder Wasserdampfes durch die sich bildende SiO_2 -Schicht darstellt. Beide Raten wachsen nicht nur mit steigender Temperatur an und hängen vom Oxidationsmedium, d.h. trockene oder feuchte Oxidati-

on, und der Substratorientierung ab, sondern werden auch von der Dotierung des Siliziums beeinflusst. Bor, aufgrund seines kleinen Segregationsverhältnisses in Silizium zu Siliziumoxid ($S_{Si:SiO_2} \approx 0.15 \dots 0.3$), lagert sich vermehrt im Oxid an, was neben dem bekannten Pile-down Effekt, d.h. der oberflächennahen Verringerung der Borkonzentration im Silizium, zu einer Schwächung der Bindungen im Oxid und damit zu einer beschleunigten Diffusion des Sauerstoffs oder Wasserdampfes führt. Die Folge davon ist ein Anwachsen der parabolischen Rate B, während die lineare Rate B/A nahezu unverändert bleibt [A1.3]. Phosphor mit einem großen Segregationsverhältnis ($S_{Si:SiO_2} \approx 20$), lagert sich verstärkt am Interface an und wird kaum in das Oxid eingebaut. Die Folge ist ein Ansteigen der linearen Rate, da die Reaktionsgeschwindigkeit an der Grenzfläche zunimmt. Die Diffusionsgeschwindigkeit durch das Oxid verbleibt aufgrund des geringen Einbaus von Phosphor in das Oxid nahezu unbeeinflusst. Daher ändert sich die parabolische Rate allenfalls unwesentlich [A1.4].

Während der Gateoxidation in trockener O_2 -Atmosphäre überwiegt das lineare Oxidwachstum und somit erfolgt eine merkbare Wachstumsbeschleunigung nur durch die Erhöhung der linearen Wachstumsrate B/A. Folglich vergrößert sich die Oxiddicke nur bei der Phosphor- N^+ -Diffusion deutlich.

Kapazität der Sperrschichtübergänge

Obwohl nicht für den verlustarmen Einsatz als Kondensator geeignet, ist das kapazitive Verhalten der pn-Sperrschichten der Source- und Drainanschlüsse jedoch für die Modellierung der Wechselstrom-Eigenschaften der MOS-Transistoren (vgl. Anhang A 2) erforderlich. Mit Hilfe des schon für die Sperrstromuntersuchung in Abschnitt 2.3.1 benutzten Layouts mit großflächigen Rechteckstrukturen ist das spannungsabhängige Verhalten dieser Dioden untersucht worden. Eine pn-Sperrschicht mit abruptem Übergang der Dotierkonzentration weist bei Vorwärtsspannungen eine Diffusionskapazität C_D gemäß [A1.5]

$$C_D = \frac{e}{2U_i} \left(L_p \cdot \frac{n_i^2}{N_D} + L_n \cdot \frac{n_i^2}{N_A} \right) e^{\frac{V_f}{U_i}} \quad (A1.11)$$

auf. Die ebenfalls vorhandene Sperrschichtkapazität C_d [A1.5]

$$C_d = \frac{\epsilon_{Si}}{l_{RLZ}} \approx \sqrt{\frac{e \cdot \epsilon_{Si} \cdot \min\{N_A, N_D\}}{2(U_d + V_r)}} \quad (A1.12)$$

ist demgegenüber zu vernachlässigen. Jedoch stellt der Vorwärtsbetrieb der Source-/Drain-Wannen-Übergänge keinen erwünschten Betriebsfall dar, so daß auf dessen Modellierung verzichtet werden kann. Bei Sperrpolung hingegen läßt sich die Kapazität des Übergangs allein durch C_d beschreiben. Analog wie im Fall der Diodensperrströme (vgl. Abs. 2.3.1) lassen sich die Kapazitätsbeiträge der Diodenflächen F und -umfänge U separat erfassen. Aufgrund der Spannungsabhängigkeit dieser Kapazität erfolgt die Modellierung mit Hilfe zweier Strukturen C_1 und C_2 , die deutlich unterschiedliche Flächen- zu Umfangsverhältnisse F_i/U_i aufweisen:

$$\begin{aligned} F_1 C_F(V_r) + U_1 C_U(V_r) &= C_1(V_r) \\ F_2 C_F(V_r) + U_2 C_U(V_r) &= C_2(V_r) \end{aligned} \quad (A1.13 \text{ a,b})$$

Die im SPICE MOS-Modell „MOS Level 2“ definierten Parameter C_j und C_{jsw} sind die Flächen- und Umfangsparameter C_F und C_U , jeweils bei $V_r = 0$ V. Die Beschreibung der Spannungsabhängigkeit erfolgt durch [A1.6]

$$C_{BS, BD} = \frac{C_j F_{S, D}}{\left(1 - \frac{V_{BS, BD}}{U_d}\right)^{M_j}} + \frac{C_{jsw} U_{S, D}}{\left(1 - \frac{V_{BS, BD}}{U_d}\right)^{M_{jsw}}} \quad (A1.14)$$

Nach Bestimmung der Werte $C_F(V_r)$ und $C_U(V_r)$ für jeden Spannungswert können die Exponenten M_J und M_{JSW} aus

$$M_J = \frac{\log\left(\frac{C_F(V_r)}{C_J}\right)}{\log(U_d) - \log(V_r + U_d)}$$

$$M_{JSW} = \frac{\log\left(\frac{C_U(V_r)}{C_{JSW}}\right)}{\log(U_d) - \log(V_r + U_d)}$$
(A1.15 a,b)

bestimmt werden. Die meßtechnisch ermittelten Koeffizienten sind in Tabelle 18 aufgeführt.

TABELLE 18. AC-Parameter der Source-/Drain-Kapazitäten

	PMOS Variante B	NMOS Variante C
C_J [fF/ μm^2]	0.10	0.23
M_J [1]	0.51	0.31
C_{JSW} [fF/ μm]	0.90	9.8
M_{JSW} [1]	0.15	0.67

A 1.3 Isolationen und Mindestabstände

In diesem Abschnitt sind die Isolation zwischen Bauelementen und die beim Layout zu beachtenden Mindestabstände zwischen hochdotierten Gebieten erläutert. Während erstere zur Vermeidung von Inversionskanälen unter metallischen Leiterbahnen notwendig ist, bestimmen letztere die Integrationsdichte bei sonst unveränderten Bauelementgrößen. Diese Untersuchungen wurden anhand modifizierter MOS-Strukturen vorgenommen.

Durch eine metallische Signal- oder Versorgungsleitung, die zwei gleichnamige hochdotierte Zonen mit unterschiedlichen Spannungspotentialen überkreuzt, kann ein parasitärer MOS-Transistor entstehen. Entsprechendes Potential auf der das Gate bildenden Leitung vorausgesetzt, kann der MOS mit einem Feldoxid als Gate einen leitfähigen Kanal aufweisen. Die Folge sind unerwünschte Ströme, welche die Funktion der Schaltung beeinträchtigen können.

Da, wie bei einem konventionellen MOS, der Stromtransport durch einen solchen parasitären Transistor erst oberhalb der Schwellspannung V_{th} einsetzt, sind diese Einschaltspannungen an Hand modifizierter MOS bestimmt worden. NMOS der Maße $L = 43 \mu\text{m}$ und $W = 200 \mu\text{m}$ und PMOS mit $L = 41 \mu\text{m}$ und $W = 360 \mu\text{m}$ weisen bei einem durch ein Feldoxid ersetzten Gateoxid minimale Beträge der Schwellspannungen von $V_{th,NMOS} = 9.3 \text{ V}$ und $|V_{th,PMOS}| = 9.7 \text{ V}$ auf. Beide Werte liegen noch innerhalb des maximalen Spannungshubes $|+V_{cc}(-V_{cc})| = 12 \text{ V}$, so daß ein Anschalten bei einigen dieser Strukturen zu erwarten ist. Bei zusätzlicher Einbringung eines Schutzstreifens (P^+ beim NMOS und N^+ beim PMOS) der Breite $B = 12 \mu\text{m}$, der zwischen Source- und Drainanschluß über die gesamte Transistorweite verläuft, erhöhen sich die Minimalwerte der Schaltspannungen deutlich auf $V_{th,NMOS} = 29.7 \text{ V}$ und $|V_{th,PMOS}| > 40 \text{ V}$. Unerwünschte Ströme zwischen Diffusionsgebieten können so durch eine selektive Erhöhung der Wannendotierung während der Bor- und Phosphordiffusionsschritte sicher ausgeschlossen werden.

Der minimale Abstand zwischen zwei gleichdotierten Gebieten ist durch eine Transistorstruktur mit nominalen Entwurfsängen des Kanals von $L = 5 \mu\text{m}$, $L = 8 \mu\text{m}$, $L = 10 \mu\text{m}$ und $L = 43 \mu\text{m}$ bzw. $L = 41 \mu\text{m}$ untersucht worden. Die minimal beobachteten Durchbruch- oder Durchgriffspannungen $V_{ds,max} = \min(V_{db}, V_{pt})$ liegen bei Transistoren mit einer minimalen Kanallänge von $L = 10 \mu\text{m}$ und schwebendem Gate im Fall des NMOS bei $V_{db} = 24.5 \text{ V}$ und beim PMOS bei $|V_{db}| = 28.3 \text{ V}$. Für die größeren Kanallängen ist keine Steigerung der maximal möglichen Spannungen zu erkennen. Bei Kanallängen von $L = 8 \mu\text{m}$ und kürzer liegen die kritischen Spannungen, bei denen ein Anstieg des Stroms erfolgt, z.T. bei weniger als $|V_{ds}| = 10 \text{ V}$. Aufgrund der geringen Spannungswerte ist ein Durchgriff zwischen Source und Drain als Ursache anzunehmen.

Ein minimaler Entwurfsabstand zwischen dotierten Gebieten von $d = 10 \mu\text{m}$ kann aufgrund dieser Untersuchung als ausreichend angesehen werden.

A 2 Modellierung der MOS-Transistoren in SPICE

Das Mitte der siebziger Jahre an der University of California in Berkeley/USA konzipierte Simulationsprogramm SPICE (Simulation Program with Integrated Circuit Emphasis) stellt heute die etablierte Basis zur Simulation elektronischer Schaltungen dar. Besondere Bedeutung hat dieses Programm bei dem Entwurf integrierter Schaltungen erlangt, da hier der bei diskreten Schaltungen sonst übliche Laboraufbau aus bekannten Gründen entfällt. Auch im Rahmen dieser Arbeit wurde der Schaltungsentwurf (vgl. Kap. 3) durch Simulationen des Gleich- und Wechselstromverhaltens verifiziert.

Während für diskrete Schaltungen eine Fülle von fertigen Bauteil-Modellen verfügbar ist, kommt beim Entwurf einer integrierten Schaltung der Modellierung des Einzelelements, hier des MOS, besondere Bedeutung zu, da diese prozeßspezifisch erfolgen muß. Dieser Notwendigkeit wird in der Literatur, [A2.1], [A2.2] und [A2.3], jedoch nur sehr unvollkommen entsprochen, da der Extraktion der z.T. aus anderen meßtechnisch ermittelten Größen abgeleiteten Parameter nur sehr selten Aufmerksamkeit geschenkt wird. Eine gute Erläuterung des hier benutzten sog. LEVEL2-Modells erfolgt in [A2.4]. Neben diesem für Transistoren mit Kanallängen oberhalb von $L_{\text{eff}} = 1\mu\text{m}$ geeigneten Modell, existieren ebenfalls Modelle für den Submikronbereich, z.B. BSIM2 und BSIM3, die jedoch einen ungerechtfertigt hohen Extraktionsaufwand für die am Institut realisierbaren Transistoren bedingen.

Die Extraktion der meisten Gleichstromparameter, z.B. V_{T0} , K_P , GAMMA und NSUB , erfolgt durch Messung der Transferkennlinie im linearen Bereich. Andere wie etwa TOX und XJ erfordern auch nicht-elektrische Messungen, wie Ellipsometrie und Schrägschliff. Für eine Kleinsignalsimulation sind neben den oben angeführten Gleichstrom-Konstanten auch Kapazitätsparameter, wie z. B. CGSO und CJ , notwendig.

An dieser Stelle soll nur kurz auf die Extraktion der Schwellspannung V_{th} eingegangen werden, da in der Literatur unterschiedliche Methoden mit zum Teil deutlich voneinander abweichenden Resultaten genannt werden [A2.5]. Allgemeines Kriterium für die Definition der Schwellspannung ist die Forderung, daß bei dem dieser Spannung entsprechenden Oberflächenpotential $\phi_s = \phi_s(V_g)$ der weitere Zuwachs der Ladung in der Inversionschicht dQ_i dem Zuwachs an Ladung in der Verarmungszone dQ_B entspricht.

$$\left. \frac{dQ_i}{d\phi_s} \right|_{V_g = V_{\text{th}}} = \left. \frac{dQ_B}{d\phi_s} \right|_{V_g = V_{\text{th}}} \quad (\text{A2.1})$$

Zu den etablierten meßtechnischen Verfahren gehören:

- (a) Die Gatespannung, bei welcher ein festgelegter kleiner Drainstrom erreicht wird, als Schwellspannung anzusehen, ist gängige Praxis unter Produktionsbedingungen, da hierdurch eine sehr schnelle Messung und Auswertung ermöglicht wird. Aufgrund der prinzipiell gegebenen Überschätzung der Schwellspannung, ist diese Methode jedoch zur Parameterextraktion weitgehend ungeeignet.
- (b) Die sogenannte „Split-CV“-Methode [A2.6] basiert auf einer quasistatischen Messung der Gate- und Substratströme. Neben deutlich erhöhtem Meßaufwand sind die üblicherweise realisierten Strukturen für eine solche Messung zu klein, um ausreichende Meßgenauigkeiten zu erlauben.
- (c) Weitere Methoden, wie etwa die Fowler-Hartstein Methode [A2.7], basieren auf der Differenzierung des Drainstroms bzw. der Steilheit als Funktion der Gatespannung. Die Tangente des Stroms am Ort der maximalen Steilheit schneidet nach dieser Definition die Spannungsachse bei der gesuchten Schwellspannung. Der Hauptnachteil dieser Methoden liegt allerdings in einer Zunahme der Schwankungsbreite der Ergebnisse bei einem identischen MOS aufgrund des durch die erforderliche Differenzierung der Meßwerte bedingten numerischen Rauschens.

Die hier für die Schwellspannungsextraktion entwickelte Meßmethode umfaßt folgende Schritte:

- (i) Messung der Transferkennlinie bei $V_{\text{ds}} = 50 \text{ mV}$ für NMOS im linearen Bereich,
- (ii) erste ungefähre Bestimmung der Schwellspannung $V_{\text{th,app}}$,
- (iii) Extraktion des Stroms I_{do} bei $|V_g - V_{\text{th,app}}| = 3 \text{ V}$,
- (iv) Bestimmung der Geraden zwischen den Punkten $I_{\text{d1}} = 0.2 \cdot I_{\text{do}}$ und $I_{\text{d2}} = 0.9 \cdot I_{\text{do}}$ der Transferkennlinie durch lineare Regression und
- (v) Berechnung des Schnittpunkts V_{gs^*} dieser Geraden mit der Spannungsachse bei $I_{\text{d}} = 0\text{A}$. Die Schwellspannung ergibt sich daraus als $V_{\text{th}} = V_{\text{gs}^*} - V_{\text{ds}}/2$.

Die Methode berücksichtigt unterschiedlich hohe Kanalströme bei Änderung der Kanalgeometrie und bietet eine die ASTM-Norm [A2.8] deutlich übertreffende Wiederholbarkeit bei mäßigem Aufwand.

Die Tabelle 19 faßt alle zur Modellierung der MOS-Transistoren erforderlichen Parameter des CMOS-Prozesses der Variante C₂ auf hochohmigem p⁻-Substrat zusammen.

TABELLE 19. Modellierungsparameter für SPICE3 Level2-Modell

	Schwellspannung V_{th} {VT0}	Steilheitsparameter {KP}	Substratkonstante γ {GAMMA}	Oberflächeninversionspotential {PHI}
NMOS	1.30 V	$27.1 \cdot 10^{-6} S$	$1.82 V^{1/2}$	0.72 V
PMOS	-1.48 V	$7.7 \cdot 10^{-6} S$	$0.58 V^{1/2}$	0.59 V
	Sättigungsstrom der Source- und Drain-Übergänge {IS}	Diffusionsspannung der Source- und Drain-Übergänge {PB}	Schichtwiderstand der Source- und Drain-Anschlüsse {RSH}	Gateoxiddicke {TOX}
NMOS	$1.0 \cdot 10^{-14} A$	0.91 V	200 Ω	92.5 nm
PMOS	$1.0 \cdot 10^{-14} A$	0.84 V	200 Ω	92.5 nm
	Substratkonzentration {NSUB}	Gatematerial {TPG}	Tiefe der pn-Übergänge {XJ}	Laterale Unterdiffusion {LD}
NMOS	$1.7 \cdot 10^{16} cm^{-3}$	0 (Aluminium)	$1.5 \cdot 10^{-6} m$	$2.27 \cdot 10^{-6} m$
PMOS	$1.8 \cdot 10^{15} cm^{-3}$	0 (Aluminium)	$1.5 \cdot 10^{-6} m$	$2.7 \cdot 10^{-6} m$
	Kritische Gate-Kanal-Feldstärke bei Beweglichkeitsmodulation {UCRIT}	Exponent der Beweglichkeitsmodulation {UEXP}	Gate-Source- und Gate-Drain-Überlappungskapazität {CGSO} und {CGDO}	Gate-Substrat-Überlappungskapazität {CGBO}
NMOS	52000 V/cm	0.11	$0.85 \cdot 10^{-9} F/m$	$1.45 \cdot 10^{-9} F/m$
PMOS	45000 V/cm	0.20	$1.0 \cdot 10^{-9} F/m$	$1.45 \cdot 10^{-9} F/m$
	Sperrschichtkapazität Flächenparameter {CJ}	Exponent des Flächenparameters {MJ}	Sperrschichtkapazität Umfangparameter {CJSW}	Exponent des Umfangparameters {MJSW}
NMOS	$233 \cdot 10^{-6} F/m^2$	0.31	$9.8 \cdot 10^{-9} F/m$	0.67
PMOS	$96 \cdot 10^{-6} F/m^2$	0.51	$9.4 \cdot 10^{-8} F/m$	0.15

A 3 CV-Charakterisierung der Grenzfläche Si-SiO₂

Die Untersuchung des grundlegenden Wirkungsprinzips von MOS, d.h. der Feldeffekt durch Aufbau einer Inversionsschicht, erfolgt üblicherweise an einfachen Strukturen, den sog. MOS-Varaktoren. Diese Strukturen dienen neben der Untersuchung von Oxidqualität und -dicke auch der Charakterisierung der Si-SiO₂-Grenzfläche. In diesem Anhang wird zuerst die Ableitung und Anwendung der Kapazitäts-Spannungs-Kennlinien (CV-Kennlinien) eines MOS-Varaktors dargestellt. Danach wird die Bestimmung der Grenzflächenzustandsdichte erläutert. Auf die besonderen Schwierigkeiten bei der Durchführung von CV-Messungen an MOS-Varaktoren auf hochohmigem Substrat wird im letzten Teil eingegangen.

A 3.1 Kapazitäts-Spannungs-Kennlinien des MOS-Varaktors

Zur Abschätzung der Oxidladungsdichte und der Grenzflächenzustandsdichte ist der Vergleich der durch Messung erhaltenen CV-Kurven des MOS-Varaktors mit unter idealisierten Bedingungen berechneten CV-Kurven erforderlich. Der Bestimmung dieser Kurven liegen folgende Idealisierungen zu Grunde:

- (i) Die Poissongleichung wird auf ein eindimensionales Problem, senkrecht zur Grenzfläche in den Halbleiter hinein gerichtet, angewendet. Aufgrund der bei MOS-Varaktoren gewählten Geometrie d.h. einer deutlich größeren lateralen Ausdehnung der Struktur (einige 100 µm) gegenüber einem nur Bruchteile eines Mikrometers dicken Oxid, bildet diese Annahme eine unkritische Vereinfachung.
- (ii) Sowohl die Berücksichtigung einer Differenz der Austrittsarbeiten zwischen Metall und Halbleiter sowie eventuell vorhandene Oxidladungen werden vernachlässigt. Diese Vereinfachung führt jedoch nur zu einer Verschiebung der CV-Kurve entlang der Spannungsachse und zu keinerlei Verzerrungen. Kritischer hingegen ist die Vernachlässigung der umladbaren Grenzflächenzustände, da diese zu einer Verzerrung der Meßkurve gegenüber der gerechneten führen. Bei kleinen Dichten dieser Zustände sind diese Abweichungen jedoch noch recht gering und die für die nachfolgenden Untersuchungen erforderliche Größe, d.h. die Verschiebung der Flachbandspannung läßt sich auch bei dieser Vernachlässigung noch sehr genau ermitteln.
- (iii) Die Dotierung des Halbleiters wird bis zur Oberfläche hin als konstant angenommen, d.h. durch gewollte Dotierung oder thermische Redistribution bedingte Änderungen der Konzentration werden vernachlässigt. Aus dieser Annahme ergeben sich Fehler bei Verfahren, welche die gesuchten Größen durch den quantitativen Vergleich mit der idealisierten CV-Kurve bestimmen. Die hier gewählte Kombination von Hochfrequenz- und Niederfrequenzmessungen zur Bestimmung der Zustandsdichte D_{it} vermeidet diese Fehler.

Zur Ableitung der Gleichungen werden in Anlehnung an [A3.1] folgende Definitionen eingeführt:

- (i) Das Potential ϕ wird als Funktion der Ortskoordinate x in das Halbleiterinnere hinein definiert. $\phi(x) = (E_F - E_i(x))/e$ ist proportional zur Differenz des Ferminiveaus E_F und des ortsabhängigen intrinsischen Niveaus $E_i(x)$. ϕ_B ist hierbei der im ungestörten Halbleitervolumen „weit von der Grenzfläche entfernt“ vorliegende Wert des Potentials. Bei einem p-Halbleiter folgt damit $\phi_B = -U_t \cdot \ln(N_A/n_i)$.
- (ii) $p = n_i \cdot \exp(-\phi(x)/U_t)$ beschreibt die Majoritätsträgerdichte und $n = n_i \cdot \exp(+\phi(x)/U_t)$ die Minoritätsträgerdichte im p-Halbleiter als Funktion des Potentials.

Die Poissongleichung lautet im p-Halbleiter

$$\frac{d^2 \phi(x)}{dx^2} = \frac{\rho(x)}{\epsilon_{si}} = -\frac{e[p(x) - n(x) + N_D - N_A]}{\epsilon_{si}}, \quad (A3.1)$$

wobei $p(x) - n(x) = n_i(\exp(-\phi(x)/U_t) - \exp(+\phi(x)/U_t)) = -2n_i \sinh(\phi(x)/U_t)$.

Weit entfernt von der Grenzfläche gilt $p(x) - n(x) = N_A - N_D = n_i(\exp(-\phi_B/U_t) - \exp(+\phi_B/U_t)) = -2n_i \sinh(\phi_B/U_t)$. Analog läßt sich mit ϕ_S der Potentialwert an der Grenzfläche $x = 0$ definieren. Nach einmaliger Integration der Poissongleichung von der Grenzfläche bis ins Volumen an der Stelle x ergibt sich die elektrische Feldstärke zu

$$F(x) = \text{Sgn}(\phi_B - \phi_S) \sqrt{2} \frac{U_t}{\lambda_i} \sqrt{\frac{\phi_B - \phi(x)}{U_t} \sinh \frac{\phi_B}{U_t} - \cosh \frac{\phi_B}{U_t} + \cosh \frac{\phi(x)}{U_t}} = \text{Sgn}(\phi_B - \phi_S) \cdot \frac{U_t}{\lambda_i} \cdot F(\phi(x), \phi_B). \quad (A3.2)$$

Hierbei ist λ_i die intrinsische Debye-Länge des Halbleiters gemäß:

$$\lambda_i = \sqrt{\frac{\epsilon_o \epsilon_{Si} U_t}{2en_i}} . \quad (\text{A3.3})$$

Für die an der Siliziumoberfläche gespeicherte Flächenladungsdichte gilt $Q_s = \epsilon_s F_s$ gemäß des Satzes von Gauß. Daraus folgt für die flächenspezifische Kleinsignalkapazität des Halbleiters, definiert als die vom Oberflächenpotential ϕ_s abhängige Änderung der Oberflächenladungsdichte Q_s ,

$$C_s(\phi_s) = -\frac{dQ_s}{d\phi_s} = -\text{Sgn}(\phi_B - \phi_s) \cdot \frac{\epsilon_{Si}}{\lambda_i} \cdot \frac{\sinh \frac{\phi_s}{U_t} - \sinh \frac{\phi_B}{U_t}}{F(\phi_s, \phi_B)} . \quad (\text{A3.4})$$

Aufgrund der Reihenschaltung von Halbleiter- und Oxidkapazität folgt damit für die gesamte flächenspezifische Varaktorkapazität $1/C_{ox}' = 1/C_s(\phi_s) + 1/C_{ox}'$, hierbei ist $C_{ox}' = \epsilon_{ox}/t_{ox}$ die flächenbezogene Oxidkapazität. Der gesamte Spannungsabfall zwischen Gatemetall und Substratkontakt verteilt sich bei Vernachlässigung der Differenz der Austrittsarbeiten und der Existenz von Oxidladungen auf die Anteile des Spannungsabfalls über dem Oxid V_{ox} und des Anteils im Halbleiter V_{Si}

$$V_g = V_{ox} + V_{Si} = -\frac{Q_s}{C_{ox}'} + (\phi_B - \phi_s) . \quad (\text{A3.5})$$

Mit Hilfe der obigen Gleichungen läßt sich die Kapazität $C_s = f(\phi_s)$ und die Gatespannung $V_g = g(\phi_s)$ in Abhängigkeit des Oberflächenpotentials und mithin die CV-Kurve mittels $C_s = f(g^{-1}(V_g))$ bestimmen. Eine explizite Darstellung dieser Funktion ist jedoch nicht möglich, wohl aber eine graphische Darstellung wie im folgenden häufig benutzt.

Die Gleichungen erlauben direkt die Berechnung der Niederfrequenzkapazität, in der auch Minoritätsträger zur Kleinsignalkapazität beitragen. Eine Annäherung an die Hochfrequenzkurve kann mit den gegebenen Gleichungen ebenfalls erfolgen, wengleich im Bereich der Inversion auf den Beitrag der Minoritätsträger verzichtet werden muß. Für eine höhere Genauigkeit muß jedoch berücksichtigt werden, daß zwar die Dichte der Minoritätsträger in der Inversionsschicht nur vom Gleichspannungsbias am Gate abhängt und daher im Fall der Hochfrequenzkurve kein Beitrag derselben zu erwarten ist, aber dennoch eine räumliche Umverteilung innerhalb der Inversionsschicht erfolgt, die der überlagerten Wechselfspannung folgt und einen wengleich geringen Kapazitätsbeitrag liefert. Die im nachfolgenden angegebene genauere Beziehung [A3.1], welche diese periodische Umverteilung der Minoritätsträger berücksichtigt, ist Grundlage der in dieser Arbeit berechneten Hochfrequenzkurven.

$$C_s = \frac{1}{\sqrt{2}} \text{Sgn}(\phi_s - \phi_B) \frac{\frac{\epsilon_{Si}}{\lambda_p} \left[1 - \exp\left(-\frac{\phi_s - \phi_B}{U_t}\right) \right]}{\sqrt{\frac{\phi_s - \phi_B}{U_t} - 1 + \exp\left(-\frac{\phi_s - \phi_B}{U_t}\right)}} , \quad \text{falls } \phi_s < \phi_m . \quad (\text{A3.6})$$

Die Sättigung der Hochfrequenzkurve bei Erreichen der starken Inversion wird durch $C_s(\phi_s) = C_s(\phi_m)$ für $\phi_s > \phi_m$ beschrieben. Nach [A3.2] gilt für die Bandverbiegung an diesem Punkt $\psi_s = \phi_s - \phi_B = 2.10 \phi_B + 2.08 U_t$.

Klassifizierung der Oxidladungen

Die Präsenz von Ladungen im Oxid ist für die Funktion von MOS-Transistoren von großer Bedeutung, da diese direkt zu einer Verschiebung der Flachbandspannung V_{FB} und damit auch der Schwellspannung V_{th} führen. Eine konstante und möglichst geringe Konzentration dieser Oxidladungen ist insbesondere bei MOS-Fets mit dicken Gateoxiden notwendig, um einen definierten und stabilen Betrieb zu ermöglichen. Der Einfluß der

Ladungsdichte Q_o auf die charakteristischen Spannungen am Gate wächst proportional zur Oxiddicke an. Es gilt

$$\frac{dV_{th}}{dQ_o} = -\frac{1}{C'_{ox}} = -\frac{t_{ox}}{\epsilon_{ox}}. \quad (A3.7)$$

Die Konzentration der verschiedenen Oxidladungen gering zu halten, ist hierbei eine der vornehmlichen Aufgaben der Prozeßtechnik. Ein großer Teil der Bemühungen der CMOS-Technologie zielt daher auf eine weitestgehende Reduktion dieser Ladungen. Da eine komplette Unterdrückung jedoch unmöglich ist, wird versucht, zumindest eine hohe Konstanz dieser Ladungsanteile zu erreichen. Während es möglich ist, durch entsprechende Berücksichtigung in den Entwürfen die Wirkung konstanter Ladungskonzentrationen einzubeziehen, führen Schwankungen derselben direkt zu unerwünschten Parametervariationen der Bauelemente.

Nach Deal [A3.3] können vier Arten von Oxidladungsbeiträgen unterschieden werden:

- (i) Die sogenannte feste Oxidladung Q_f entsteht bei der Oxidation. Sie ist in einer sehr dünnen Schicht an der Grenzfläche lokalisiert und wirkt als positive Flächenladung.
- (ii) Getrappte Ladungen im Oxid Q_{ot} entstehen durch Strahlenschäden im weitesten Sinne, so bei der Ionenimplantation durch ionisierte Teilchen oder durch Injektion heißer Löcher und Elektronen während des Betriebs des MOS, sowie durch ionisierende Strahlung.
- (iii) Die Grenzflächenladung Q_{it} , die sog. schnellen Oberflächenzustände, führen zu elektrisch aktiven Zuständen innerhalb des verbotenen Bands des Halbleiters direkt an der Grenzfläche zum Oxid. Abhängig von der Lage des Fermi-niveaus an der Grenzfläche werden diese Zustände mit beweglichen Trägern gefüllt. Es treten sowohl Störstellen mit Akzeptor- als auch mit Donatorcharakter auf. Da ihre Ladung spannungsabhängig ist, gibt man hier nicht die absolute Flächenladungsdichte, sondern die auf die energetische Verteilung im verbotenen Band bezogene Flächendichte an (Angabe als D_{it} in $\text{cm}^{-2}\text{eV}^{-1}$).
- (iv) Aufgrund des Einbaus von Alkaliionen während des Oxidationsvorgangs, insbesondere von Natrium und Kalium, befinden sich bewegliche Ionenladungen Q_m im Oxid.

Technologische Schritte erlauben es, einige dieser Ladungsanteile zu reduzieren. So läßt sich durch die Temperung des Halbleiters nach Prozeßschritten, die Strahlenschäden verursachen, z.B. die Ionenimplantation, eine Reduktion der getrappten Ladungen Q_{ot} erreichen. Äußerste Reinheit führt zu geringer Kontamination durch alkalische Ionen und damit zu geringen Anteilen mobiler Ladungen Q_m . Die Höhe der Grenzflächenzustandsdichte D_{it} ist im wesentlichen durch nichtabgesättigte Si-Atome an der Grenzfläche zum Oxid bedingt. Diese, sog. „dangling bonds“, lassen sich durch eine abschließende Temperung in einer wasserstoffhaltigen Atmosphäre verringern. Weniger gut beeinflussbar, aber dadurch auch recht konstant bleibend, ist die Größe der festen Oxidladungen Q_f .

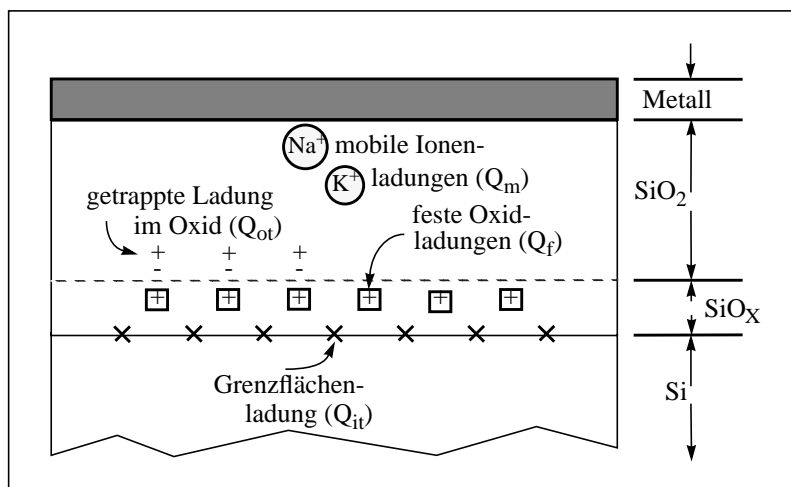


Bild A3.1: Darstellung der Oxidladungen im MOS-Varaktor. Die mit SiO_x ($1 \leq x \leq 2$) bezeichnete Schicht stellt den Übergangsbereich zwischen Silizium und Oxid mit nicht vollständig oxidierten Si-Atomen dar.

Kritische Indikatoren einer guten Prozeßführung sind daher die mobilen Ionen und die Grenzflächenzustandsdichte. Während aus einer Messung der Hochfrequenzkapazitäts-Spannungskennlinie eines MOS-Varaktors, der sog. HF-CV-Kurve, durch Vergleich mit einer theoretisch berechneten CV-Kurve mit ladungsfreiem Oxid eine Aussage über die Summe aller dieser Oxidladungsbeiträge erhalten werden kann, ist die Extraktion der mobilen Ionen und der Grenzflächenzustandsdichte mittels zusätzlicher CV-Messungen möglich.

Die Verschiebung der Flachbandspannung bezüglich einer idealisiert gerechneten CV-Kurve beträgt

$$\Delta V_{FB} = - \frac{Q_f + Q_{ot} + Q_m + Q_{it}}{C'_{ox}} + \frac{E_{ms}}{e}, \quad (\text{A3.8})$$

wobei E_{ms} die Differenz der Austrittsarbeiten von Metall und Halbleiter bedeutet.

Hier soll das Verfahren für die Bestimmung der Grenzflächenzustandsdichte D_{it} , insbesondere in seiner Modifikation für hochohmiges Substratmaterial dargestellt werden. Die Bestimmung der Anzahl mobiler Ladungen Q_m erfolgt in Abschnitt 4.2.1.

A 3.2 Grenzflächenzustandsdichte

Die in der Bandlücke des Halbleiters (für Silizium $E_{\text{gap}} = 1.12$ eV bei $T = 300$ K) befindlichen Zustände werden neben ihrer energetischen Lage relativ zu den Bandkanten noch weiter nach ihrer räumlichen Verteilung in sog. Grenzflächen- bzw. Volumenzustände unterteilt. Ein solcher energetischer Zustand kann nun je nach Lage des Ferminiveaus mit einem Elektron oder Loch besetzt oder auch elektrisch neutral sein und somit seine Ladung in Abhängigkeit vom Ferminiveau einstellen.

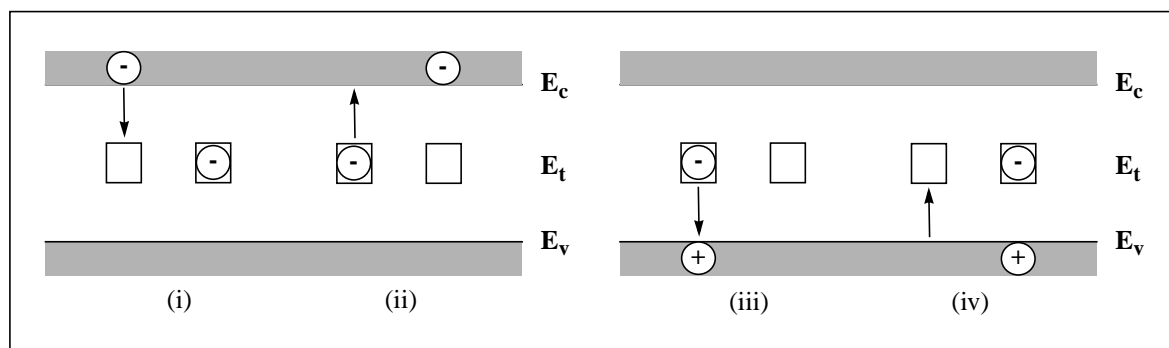


Bild A3.2: Darstellung der Teilprozesse zur Generation und Rekombination mittels Störstellen in der Bandlücke. Elektronenabsorption (i) und -emission (ii), Löcherabsorption (iii) und -emission (iv). Ein kompletter Rekombinationsprozeß besteht hierbei aus (i) und (iii). Die Vorgänge (iv) und (ii) bilden den Generationsprozeß.

In Bild A3.2 sind nun - der Darstellung von Grove [A3.4] folgend - die Emissions- und Absorptionsprozesse mittels Zuständen im Band gezeigt. Elektronen aus dem Leitungsband (E_c) können dabei von einem Energiezustand (E_t) in der Bandlücke absorbiert (Fall i) und auch wieder emittiert werden (Fall ii). In analoger Weise können Löcher aus dem Valenzband (E_v) absorbiert (Fall iii) bzw. emittiert werden (Fall iv). Erfolgt nun die Absorption eines Loches direkt nach der Absorption eines Elektrons vor dessen Re-Emission in das Leitungsband, liegt ein sogenannter Rekombinationsprozeß vor, in dessen Folge ein Elektron vom Leitungsband in das Valenzband unter Zuhilfenahme einer Störstelle gewandert ist und dort die Anzahl der Löcher um eins verringert.

In umgekehrter Weise ist nun die Generation eines Ladungsträgerpaares, d.h. die Erhöhung der Elektronenanzahl im Leitungsband bei gleichzeitiger Erhöhung der Löcheranzahl im Valenzband möglich. Diese statistisch ablaufenden Prozesse sorgen für eine Annäherung des Halbleiters an das thermische Gleichgewicht nach erfolgter Störung desselben. In diesem Gleichgewicht ist dann die Generationsrate von Ladungsträgern gleich der

Rekombinationsrate, so daß eine zeitlich stabile Trägerkonzentration gemäß $p^*n = n_i^2$ erreicht wird. Eine solche Störung des Gleichgewichts erfolgt nun beim MOS-Varaktor durch Veränderung der am Gate angelegten Spannung, in dem z.B. der Varaktor von der Anreicherung in die starke Inversion gepulst wird und dort bei konstantem Gatebias verbleibt.

Das durch Ausdehnung der Verarmungszone mittels ionisierten Dotieratomen (z.B. N_A^- im Falle des p-Si-Varaktors) erreichte elektrische Gleichgewicht stellt nun ein thermisches Ungleichgewicht dar, da es zu einer gegenüber dem Gleichgewicht reduzierten Zahl der Minoritätsträger (hier Elektronen) führt, d.h. $p^*n < n_i^2$. In diesem Fall ist nun die Anzahl der zur Rekombination zur Verfügung stehenden Minoritätsträger ebenfalls reduziert, so daß es zu einem Überwiegen der Generation kommt. Die erzeugten Minoritätsträger führen in der Folge zur Bildung einer oberflächennahen Inversionsschicht. Ladungsneutralität wird nun durch die Summe aus ionisierten Dotieratomen und diesen Minoritätsträgern gebildet, so daß sich die Verarmungszone auf ihren Gleichgewichtswert hin verkleinert.

Die Effizienz einer Störstelle als Rekombinations- bzw. Generationszentrum hängt von deren relativer Lage in der Bandlücke ab. Da die Emissions- und Absorptionsraten einer Störstelle stark von deren Abstand zur jeweiligen Bandkante abhängen und für einen Rekombinations- bzw. Generationsprozeß jeweils eine Absorption bzw. Emission von Trägern aus beiden Bändern erfolgen muß, sind in der Nähe der Bandmitte liegende Zustände effizient. Anschaulich kann dies damit erklärt werden, daß bei einer z. B. in der Nähe des Leitungsbands liegenden Störstelle zwar eine effiziente Elektronenabsorption erfolgen kann. Da aber aufgrund des großen Abstands zum Valenzband die Löcherabsorption sehr unwahrscheinlich ist, wird vor Eintreten der Rekombination eine Re-Emission des Elektrons in das Leitungsband erfolgen. Schroder [A3.5] zeigt, daß nur Störstellen, die wenige kT oberhalb oder unterhalb der Bandmitte lokalisiert sind, wesentlich zur Generation und Rekombination beitragen.

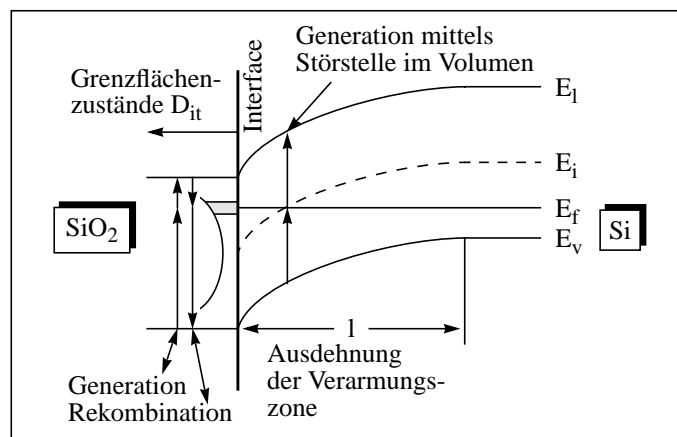


Bild A3.3: Generation und Rekombination mittels Grenzflächenzuständen und Störstellen in der Verarmungszone am Beispiel von p-Silizium-Substrat in starker Inversion.

Bild A3.3 verdeutlicht diese Ladungsträgererzeugungs- bzw. -vernichtungsprozesse mittels Störstellen an der Grenzfläche und im Volumen. Da die Besetzungswahrscheinlichkeit der Störstellen nur in unmittelbarer Nähe des Fermi-niveaus in der Folge einer Variation der Gatespannung eine signifikante Veränderung erfährt, wird deutlich, daß nur im Volumen in der energetischen Nähe von einigen kT um den Kreuzungspunkt zwischen Fermi- und intrinsischem Niveau (E_i) gelegene Zustände bei diesen Prozessen Einfluß nehmen können. Aufgrund der beiden Bedingungen: (i) Lage in der Nähe der Bandmitte und (ii) Lage in der Nähe des Fermi-niveaus sind nur Volumenstörstellen und nicht Oberflächenzustände, die sog. Grenzflächenzustände, bedeutsam am Auf- und Abbau der Inversionsschicht beteiligt. Grenzflächenzustände, insbesondere jene in der Nähe der Bandkanten, ändern ihren Ladungszustand indessen nur vergleichsweise langsam. Diese Eigenschaft wird für die Bestimmung der Grenzflächenzustandsdichte mittels hoch- und niederfrequenten CV-Messungen ausgenutzt.

Die Umladung von Grenzflächenzuständen bei Variation der Gategleichspannung führt zu einer schwächeren Variation der Bandverbiegung als im Fall eines MOS-Varaktors ohne diese Zustände. Eine Änderung der Gate-spannung δV_G bewirkt eine Variation der Gateladung δQ_G , die im Ladungsgleichgewicht durch die Summe aus

Grenzflächenladungsänderung δQ_{it} und Ladungsänderung im Halbleiter δQ_{Si} ausgeglichen werden muß:

$$\delta Q_G = -(\delta Q_{it} + \delta Q_{Si}) \quad . \quad (A3.9)$$

Bei gleicher Gatespannungsänderung erfolgt damit eine verringerte Änderung der Ladungen im Halbleiter δQ_{Si} und damit eine Dehnung der CV-Kurve gegenüber einer idealisiert gerechneten CV-Kurve entlang der Spannungsachse. Dieser als stretch-out bekannte Effekt tritt bei Umladung der Grenzflächenzustände infolge einer langsamen Änderung des Gleichspannungsanteils der Gatespannung auf und betrifft somit sowohl eine CV-Kurve unter Hochfrequenz- als auch unter Niederfrequenzbedingungen in gleicher Weise.

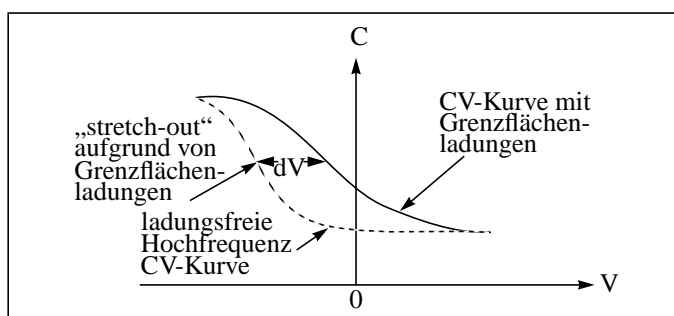


Bild A3.4: Dehnung der CV-Kurve eines Varaktors mit Grenzflächenladungen bzgl. einer idealisierten ladungsfreien Hochfrequenz-CV-Kurve.

Können nun die Grenzflächenzustände ihren Ladungszustand in Folge einer überlagerten Wechselspannung ändern, so ergibt sich neben dem Kapazitätsbeitrag aufgrund der Halbleiterkapazität C_{Si} auch ein Anteil aufgrund von Grenzflächenzuständen C_{it} , der im elektrischen Ersatzschaltbild parallel zu C_{Si} liegt. Der Anteil C_{it} ist frequenzabhängig, da die Übergangsraten von Elektronen und Löchern aus dem Leitungs- bzw. Valenzband, wie oben erläutert, stark vom Abstand der Störstelle zu diesen abhängt. Diese Raten sind proportional zur maximal möglichen Umladefrequenz, so daß $C_{it} = C_{it}(\omega)$ gilt. Die Niederfrequenz-CV-Kurve ist definiert für Bedingungen, unter denen sowohl die Minoritätsträgergeneration und -rekombination als auch die Umladung der Grenzflächenzustände der Wechselspannung am Gate folgen können. Es ergibt sich folglich für die Niederfrequenzkapazität:

$$C_{NF} = (C_{Si} + C_{it}) \frac{C_{ox}}{C_{ox} + C_{Si} + C_{it}} \quad . \quad (A3.10)$$

Hierbei tritt zusätzlich die in Reihe liegende Oxidkapazität C_{ox} auf. Da sehr geringe Meßfrequenzen erforderlich sind, werden heute vornehmlich quasistatische Verfahren oder Gleichspannungs-Rampen für die Messung benutzt [A3.6], [A3.7].

Die Definition einer Hochfrequenzkurve erfordert nun ein völliges Unterbleiben der Reaktion von Minoritätsträgern und Grenzflächenzuständen auf die Wechsignale am Gate. Meßfrequenzen um 1MHz erlauben eine sehr gute Annäherung an dieses Ideal, wengleich eine komplette Reaktionsunterdrückung erst bei etwa 100 MHz zu erreichen ist. Solch hohe Frequenzen sind jedoch für die meisten MOS-Varaktorstrukturen meßtechnisch ungeeignet.

Die unter Hochfrequenzbedingungen ermittelte Kapazität läßt sich als Serienschaltung aus Oxid- C_{ox} und Halb-

leiterkapazität C_{Si} beschreiben:

$$C_{HF} = \frac{C_{ox} C_{Si}}{C_{ox} + C_{Si}} \quad (A3.11)$$

Da nun die Halbleiterkapazität nur eine Funktion der Bandverbiegung und nicht der Meßfrequenz ist, die Oxidkapazität und der Dehnungseffekt entlang der Spannungsachse, der sog. „stretch-out“, ebenfalls nicht durch die Meßfrequenz beeinflusst werden, läßt sich aus einer Kombination von Messungen unter ausreichend niedrigen und ausreichend hohen Frequenzen die durch die Grenzflächenzustände bedingte zusätzliche Kapazität C_{it} bestimmen:

$$C_{it} = \left(\frac{1}{C_{NF}} - \frac{1}{C_{ox}} \right)^{-1} - \left(\frac{1}{C_{HF}} - \frac{1}{C_{ox}} \right)^{-1} \quad (A3.12)$$

Die auf ein Energieintervall bezogene Flächendichte der Grenzflächenladungen D_{it} , auch Interface-Trap-Dichte genannt, ergibt sich für Bandverbiegungen, die den Fermilevel einige kT von den jeweiligen Bandkanten entfernt halten, in Abhängigkeit der Gatespannung V_g zu:

$$D_{it}(V_g) = \frac{1}{e} \cdot C_{it}(V_g) \quad (A3.13)$$

Entscheidend ist die Dichte der Grenzflächenladungen als Funktion der Energie relativ zur Bandmitte. Da nur die jeweils in der Nähe des Fermineveaus E_f liegenden Interface-Traps einen Beitrag zur Kapazität C_{it} liefern können, ergibt sich $D_{it}(E)$ durch Berechnung der Lage des Fermineveaus relativ zum intrinsischen Niveau E_i an der Grenzfläche als Funktion des Gatebias V_g .

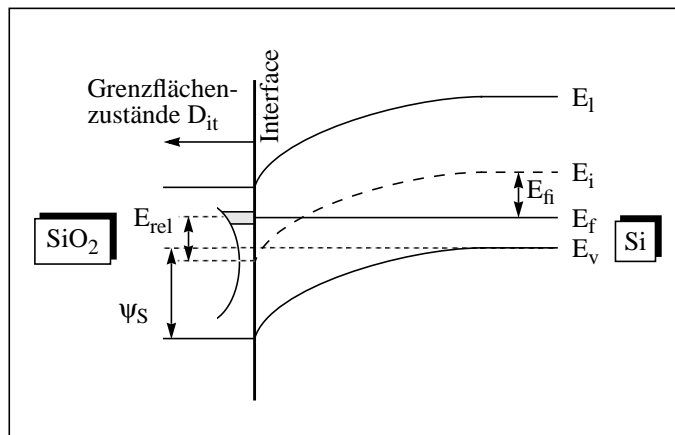


Bild A3.5: Zusammenhang zwischen Bandverbiegung ψ_S und Energie relativ zur Bandmitte E_{rel} im p-Halbleiter.

Wie in Bild A3.5 dargestellt, bestimmt sich die relative Lage der aktivierten Grenzflächenzustände bzgl. der Bandmitte gemäß

$$E_{rel} = e\psi_S + E_{fi} = e(\phi_B + \psi_S) = e\phi_s \quad (A3.14)$$

hierbei ist ψ_S die Bandverbiegung des Halbleiters und E_{fi} der Abstand zwischen Fermi-niveau und intrinsischem Niveau im Volumen. Es verbleibt nun noch die Bandverbiegung ψ_S als Funktion der Gatespannung zu bestimmen. Aufgrund der Potentialaufteilung am MOS-Varaktor gilt

$$C_{ox}(dV_g - d\psi_S) = (C_{it} + C_{Si})d\psi_S. \quad (\text{A3.15})$$

Somit folgt für die Bandverbiegung

$$\psi_S = \psi_o + \int_{V_{go}}^{V_g} \left[1 - \frac{C_{it} + C_{Si}}{C_{ox} + C_{it} + C_{Si}} \right] dV_g. \quad (\text{A3.16})$$

Hierbei ist der zweite Summand im Integranden die auf die Oxidkapazität bezogene Niederfrequenzkapazität, so daß sich in Abhängigkeit der gemessenen Niederfrequenzkapazität C_{NF} für die Bandverbiegung

$$\psi_S = \psi_o + \int_{V_{go}}^{V_g} \left[1 - \frac{C_{NF}}{C_{ox}} \right] dV_g \quad (\text{A3.17})$$

ergibt. Die Bandverbiegung ist damit experimentell aus der gemessenen Niederfrequenzkapazität C_{NF} bis auf eine additive Konstante ψ_o bestimmbar. Der Wert der Bandverbiegung ψ_o bei einer Gatespannung V_{go} kann vorteilhaft als Bandverbiegung im Flachbandfall angenommen werden.

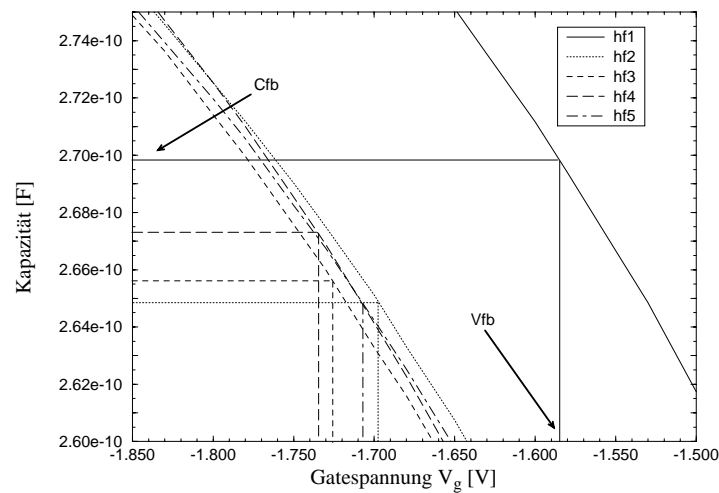
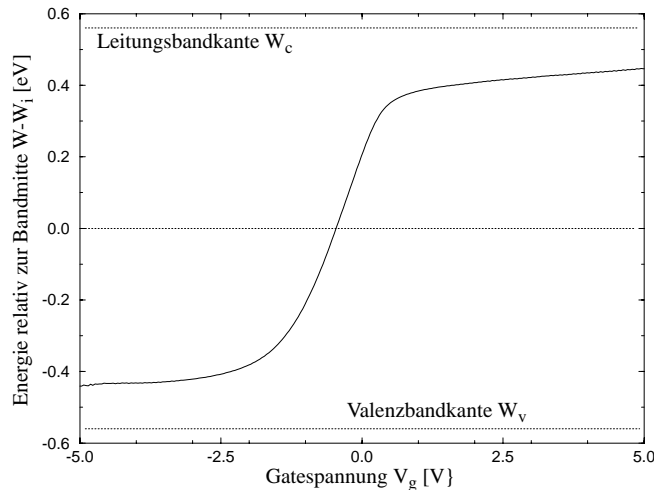


Bild A3.6: Bestimmung der Flachbandspannung bei fünf CV-Meßkurven. Die horizontalen Linien markieren die aus der gemessenen Oxidkapazität C_{ox} berechnete Flachbandkapazität C_{FB} . Die Schnittpunkte mit der jeweiligen CV-Kurve ergeben die Flachbandspannung.

Im idealen Fall eines ladungsfreien Oxides und einer verschwindenden Austrittsarbeitsdifferenz zwischen Gate-material und Silizium ergibt sich für $\psi_o = 0$ im Flachbandfall $V_{go} = 0$. Im realen Fall kann die zum Flachbandfall gehörige Gatespannung mit Hilfe einer HF-CV-Kurve als Spannungswert korrespondierend zur Flachbandkapazität $C_{FB} = \epsilon_o \epsilon_{Si} / \lambda$, mit der extrinsischen Debye-Länge λ , gefunden werden. Dies ist anhand von realen CV-Kurven in Bild A3.6 dargestellt.

Bild A3.7: Bandverbiegung relativ zur Bandmitte bei angelegtem Gatebias V_g .

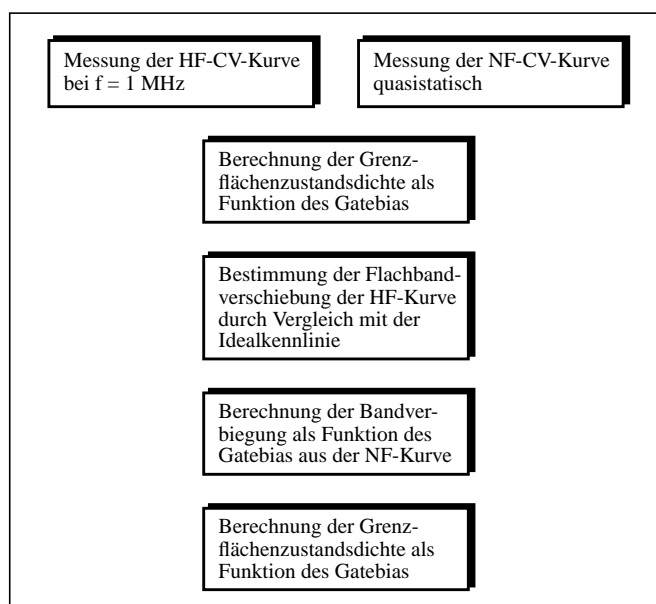
In Bild A3.7 ist die Bandverbiegung relativ zur Bandmitte, abhängig von der Gatespannung und numerisch aus einer Niederfrequenz-CV-Kurve berechnet, dargestellt. Deutlich wird dabei, daß mit praktikablen Gatespannungen nur Bandverbiegungen bis zu einem minimalen Abstand zu den Bandkanten von etwa 0.1 eV erreicht werden können. Dies hängt mit dem exponentiellen Anstieg der Trägerdichten n und p bei Annäherung an die Bandkanten zusammen:

$$n = n_i e^{\frac{E_F - E_i}{kT}} \quad \text{und} \quad p = n_i e^{-\frac{(E_F - E_i)}{kT}} \quad (\text{A3.18})$$

Die Messung der Grenzflächenzustandsdichte ist deshalb allein aus diesem Grund auf ein kleineres Intervall um die Bandmitte begrenzt. Eine weitere Begrenzung erfährt der Gültigkeitsbereich durch den Flachbandfall bei Annäherung an das Majoritätsträgerband. Aufgrund der hohen Zahl von Majoritätsträgern kommt es zu einem verstärkten Einfangen der jeweiligen Träger durch die Interface Traps und damit zu sich verkürzenden Zeitkonstanten, so daß die Voraussetzung einer Hochfrequenz-CV-Kurve, d.h. ein verschwindender kapazitiver Beitrag durch die Grenzflächenzustände, verletzt wird. Die Folge ist eine leichte Verformung der Meßkurve von einer reinen HF-Kurve hin zu einer NF-Kurve, so daß eine zu geringe Grenzflächenzustandsdichte ermittelt wird. Da die Auswertung der Grenzflächenzustandsdichte auf eine Differenzrechnung zwischen Hochfrequenz- und Niederfrequenz-CV-Kurven hinausläuft, ist der Einfluß von Meßfehlern dann am größten, wenn nahezu gleich große Werte verglichen werden. Dies ist im Bereich des Flachbands in Richtung zur Anreicherung der Fall, so daß die Auswertung in diesem Bereich allein aus Gründen der meßtechnischen Genauigkeit zu prinzipiellen Problemen führt.

Bei Einsatz der starken Inversion beträgt der durch die Niederfrequenzmessung ermittelte Anteil der Halbleiterkapazität $C_{Si} = C_{it} + C_{RLZ} + C_{inv}$, also die Summe aus Interface Trap Kapazität C_{it} , Raumladungskapazität C_{RLZ} und Inversionskapazität C_{inv} . Die Hochfrequenzmessung hingegen ermittelt weiterhin nur $C_{Si} = C_{RLZ}$, so daß die der Bestimmung der Grenzflächenzustände zu Grunde liegende Differenz um den Anteil C_{inv} überschätzt wird; ein überhöhter Wert D_{it} ist die Folge.

Aus den Ausführungen folgt, daß eine Messung nur innerhalb der Grenzen $E_{rel} = \pm |E_F - E_i|$ um die Bandmitte E_i zu genauen Resultaten führt. Als übliche Vergleichsgröße wird deshalb oft der Minimalwert der D_{it} gewählt. Dieser befindet sich nahe der Bandmitte und ist folglich innerhalb des Gültigkeitsbereichs. Das Bild A3.8 verdeutlicht zusammenfassend den Ablauf zur Bestimmung der Grenzflächenzustandsdichte D_{it} mittels MOS-Varaktoren.

Bild A3.8: Sequenz zur Bestimmung der Grenzflächenzustandsdichte D_{it} .

A 3.3 Problematik großer Serienwiderstände

Das sehr hochohmige Substratmaterial des Integrationsprozesses (<100>-orientiert, Bor, $\rho = 5000 \Omega\text{cm}$) führt zu extrem hohen Serienwiderständen R_s , welche Messungen der Kapazität unter Hochfrequenzbedingungen bei $f = 1 \text{ MHz}$ verhindern. Für die Kapazität C und den Serienwiderstand R_s aufgrund des Substratvolumens gilt [A3.8]

$$R_s = \frac{\rho}{2d}, \quad C = \frac{\epsilon A}{t_{ox}} \quad \text{und} \quad \tau = R_s \cdot C = \frac{\pi \rho \epsilon d}{8 t_{ox}}, \quad (\text{A3.19})$$

wobei ρ der spezifische Widerstand, t_{ox} die Oxiddicke sowie A und d Fläche bzw. Durchmesser der kreisförmigen Gateelektrode bedeuten. Daraus ergeben sich die in Tabelle 20 angegebenen Werte für R_s , C und τ .

TABELLE 20. Vergleich der Ersatzschaltbild-Komponenten

	<100>-Material, Bor, 2 Ωcm	<100>-Material, Bor, 5000 Ωcm
Oxiddicke t_{ox} [\AA]	897	897
Durchmesser d [μm]	1020	1020
Serienwiderstand R_s [Ω]	9.8	24500
Kapazität C [pF]	316	316
Kapazitiver Blindwiderstand $X_c = (\omega C)^{-1}$ bei $f = 1 \text{ MHz}$ [Ω]	504	504
RC-Produkt τ [s]	3.1 n	7.74 μ
Betrag des Phasenwinkels ϕ [Grad]	88.9	1.18

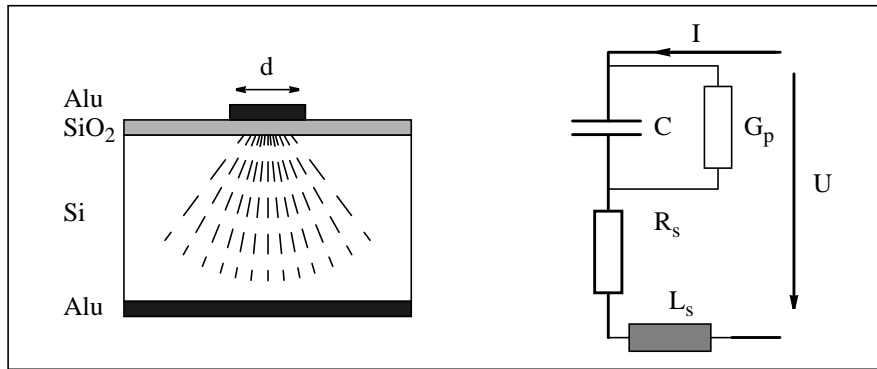


Bild A3.9: Prinzipbild und Ersatzschaltbild eines MOS-Varaktors mit Kapazität C, Serienwiderstand R_s , Parallelleitwert G_p und Zuleitungsinduktivität L_s .

Der komplexe Widerstand einer MOS-Kapazität (vgl. Bild A3.9) berechnet sich wie folgt:

$$\bar{Z} = Z \times e^{j\phi} = \frac{\bar{U}}{\bar{I}} = R_s + \frac{1}{G_p + j\omega C} + j\omega L_s = R_s + \frac{G_p}{G_p^2 + (\omega C)^2} + j \left(\omega L_s - \frac{\omega C}{G_p^2 + (\omega C)^2} \right) \quad (A3.20)$$

Eine Zunahme des Serienwiderstands R_s erhöht den Realteil des komplexen Widerstands \bar{Z} , während sich der Einfluß der Zuleitungsinduktivität L_s in einer Verringerung des Betrags des imaginären Anteils ausdrückt. Beides führt zu einem verkleinerten Phasenwinkel ϕ , ($\arctan \phi = \text{Im } \bar{Z} / \text{Re } \bar{Z}$). Eine ideale Kapazität weist den Phasenwinkel $\phi = -90^\circ$ auf. Wenn, wie im Fall des hochohmigen Substrats (vgl. Tabelle 20), ein Phasenwinkel ϕ von nur einigen Grad resultiert, ist eine sichere Extraktion des kapazitiven Anteils aus dem Meßsignal nicht mehr möglich.

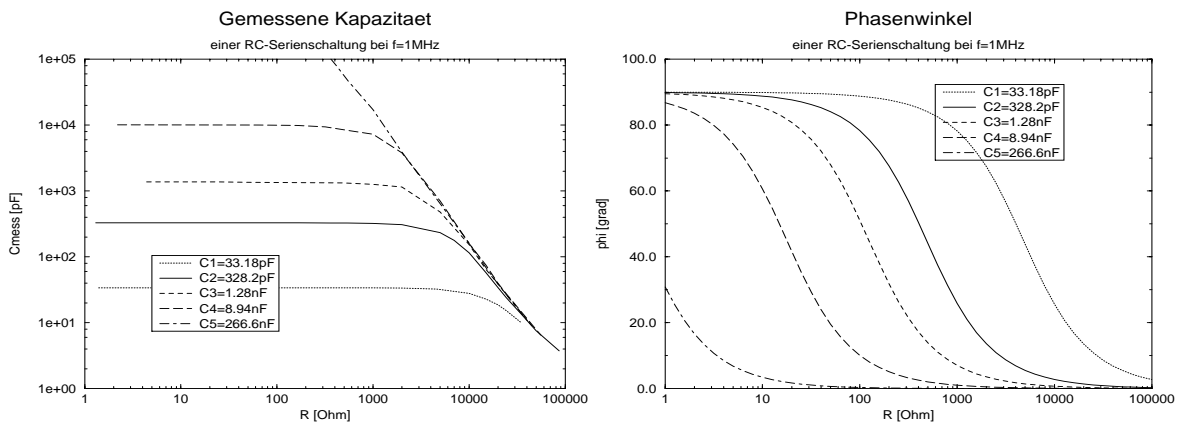


Bild A3.10: Darstellung der begrenzten Auflösungsfähigkeit bei einer C_s - R_s -Serienschaltung. Links der meßtechnisch bedingte Abfall der ermittelten Kapazitätswerte, rechts die dazugehörigen theoretisch berechneten Beträge des Phasenwinkels.

Die Darstellung in Bild A3.10 verdeutlicht, daß bei einem Abfallen des Phasenwinkelbetrags der Reihenschaltung unter etwa $\phi_{\text{krit}} \approx 80^\circ$ deutliche Meßfehler bei der Ermittlung des kapazitiven Anteils erfolgen. Dieser Fehler ist dem Meßprinzip eigen und nicht durch ein falsches Extraktionsmodell bedingt. Bei einer Reduktion des theoretisch berechneten Phasenwinkels $\phi = -\arctan(1/\omega RC)$ unterhalb dieses kritischen Werts ϕ_{krit} ist der Phasenwinkel meßtechnisch nicht mehr exakt bestimmbar, so daß eine Berechnung des kapazitiven Anteils aus dem komplexen Widerstand $\bar{Z} = Z^* e^{j\phi}$ nicht mehr erfolgen kann.

Erkennbar ist, daß bei einer MOS-Kapazität von etwa 320 pF (Kurve C_2 in Bild A3.10) eine meßtechnische Bestimmung für Serienwiderstände R_s größer als 100 Ω nicht mehr erfolgen kann. Im Falle von 5000 Ω cm p⁻Si (entsprechend etwa 24.5 k Ω Serienwiderstand) ist eine HF-CV Kurve bei $f = 1$ MHz selbst für kleine MOS-Strukturen mit $C \approx 30$ pF nicht mehr exakt durchführbar.

Aus diesem Grund können MOS-Kapazitäten auf niedrigdotiertem Substrat nur mittels statischen oder quasistatischen ($\omega \rightarrow 0$) CV-Meßverfahren untersucht werden. In Sonderfällen, d.h. wenn keine besonders hohen Anforderungen an die Meßgenauigkeit gestellt werden, kann eine Kapazitätsmessung auch bei geringeren Frequenzen geschehen (vgl. Abs. 2.2.2).

A 4 Lebensdauerbestimmung an pn- und pin-Dioden

Das thermische Gleichgewicht eines Halbleiters wird erreicht, wenn die Beziehung $n \cdot p = n_i^2$ zwischen den Trägerdichten und der intrinsischen Dichte gilt. Erfolgt durch eine äußere Beeinflussung, z.B. Lichteinfall oder elektrische Felder eine Erhöhung, d.h. $n \cdot p > n_i^2$, oder auch Erniedrigung der Trägerdichten, d.h. $n \cdot p < n_i^2$, so ist der Kristall bestrebt wieder in das thermische Gleichgewicht zurückzukehren.

Mechanismen für das Erreichen der erforderlichen Trägerdichten liegen in der Rekombination von überschüssigen Elektron-Loch-Paaren (bei $n \cdot p > n_i^2$) oder bei deren Generation, wenn $n \cdot p < n_i^2$. Rekombination erfolgt, wenn sich ein Elektron aus dem Leitungsband mit einem Loch aus dem Valenzband „neutralisiert“, die Bildung eines Elektron-Loch-Paares erfolgt bei Generation.

Diese Generations- und Rekombinationsvorgänge können

(i) durch unmittelbaren Übergang eines Elektrons aus dem Leitungsband in das Valenzband (Rekombination), bzw. eines Elektrons aus dem Valenzband in das Leitungsband unter Zurückbleiben eines Loches (Generation) erfolgen. Dieser Fall des direkten Band-zu-Band Übergangs spielt nur bei direkten Halbleitern eine wichtige Rolle (z.B. GaAs). Bei indirekten Halbleitern, wie beim hier betrachteten Silizium, ist dieser Vorgang unbedeutend und

(ii) Übergänge finden mit Hilfe sog. Fang- oder Haftzentren (engl. traps) statt. Hierbei erfolgt der Übergang über einen im verbotenen Band liegenden Energiezustand E_t .

Die Vorgänge der Rekombination und Generation laufen aufgrund thermischer Bewegungen permanent ab, bei der beobachteten Trägererzeugung und -vernichtung wird daher immer eine Nettogeneration oder -rekombination beobachtet. Die durch [A4.1] $U = R - G$ beschriebene Netto-Übergangsrate U bezogen auf die Zeit und das Volumen (Einheit in $\text{cm}^{-3}\text{s}^{-3}$) ist dementsprechend die Differenz der Rekombinationsrate R und der Generationsrate G . Gemäß der Shockley-Read-Hall Statistik gilt für diese Rate [A4.2], [A4.3]

$$U = \frac{\sigma_n \sigma_p v_{th} N_t (pn - n_i^2)}{\sigma_n \left[n + n_i \cdot \exp\left(\frac{E_t - E_i}{kT}\right) \right] + \sigma_p \left[p + n_i \cdot \exp\left(\frac{E_i - E_t}{kT}\right) \right]}, \quad (\text{A4.1})$$

worin N_t die Volumendichte der Trapzentren, σ_n und σ_p deren Einfangquerschnitte für Elektronen und Löcher und v_{th} die mittlere thermische Geschwindigkeit der Träger sind. Erkennbar ist, daß die Übergangsrate proportional zur beobachteten Abweichung vom Gleichgewicht ($pn - n_i^2$) und zur Dichte der Fangstellen N_t ansteigt. Ferner erfolgt ein Übergang über eine Fangstelle um so häufiger, je näher deren Energieniveau E_t in der Nähe der Bandmitte E_i liegt. Bei Berücksichtigung nur solcher Fangstellen in der Bandmitte folgt für U

$$U = \frac{\sigma_n \sigma_p v_{th} N_t (pn - n_i^2)}{\sigma_n (n + n_i) + \sigma_p (p + n_i)}. \quad (\text{A4.2})$$

Die Übergangsrate U läßt sich im Fall einer nur geringen Erhöhung der Minoritätsträger, d.h. bei niedriger Injektion, wenn also $n_i \ll \text{Minoritätsträger} \ll \text{Majoritätsträger}$ gilt, ebenfalls schreiben als

$$U = \frac{\Delta n}{\tau_n} \quad \text{mit} \quad \tau_n = \frac{1}{\sigma_n v_{th} N_t} \quad (\text{A4.3})$$

bei Injektion von Elektronen in einen p-Halbleiter oder

$$U = \frac{\Delta p}{\tau_p} \quad \text{mit} \quad \tau_p = \frac{1}{\sigma_p v_{th} N_t} \quad (\text{A4.4})$$

bei Injektion von Löchern in einen n-Halbleiter.

Die sogenannte Rekombinationslebensdauer τ ist umgekehrt proportional zur Häufigkeit der Netto-Trägerverrichtungen. Der umgekehrte Mechanismus liegt der Generationsträgerlebensdauer zugrunde. In analoger Weise sind die Rekombinations- und Generationsvorgänge an Oberflächen definiert, die gerade bei planaren Bauelementen einen starken Einfluß ausüben, da aufgrund der Diskontinuitäten an der Grenzfläche, z.B. zum thermischen Oxid (vgl. Anhang 3.2), aber auch durch Verunreinigungen und Kristalldefekte bedingt, eine sehr hohe Anzahl von Oberflächenzuständen N_{it} im verbotenen Band existiert.

Bei durch äußere Spannungen erzeugten Abweichungen vom thermischen Gleichgewicht bewirken diese Ausgleichsvorgänge an pn-Übergängen einen zusätzlichen Rekombinationsstromanteil bei Vorwärtsspannung und einen Generationsstromanteil bei Sperrspannung, der durch Gleichung (2.23) in Abschnitt 2.3.1 beschrieben wird.

Eine starke Verringerung der Lebensdauer führt nun zu einem Anwachsen dieses Stromanteils, beeinflusst jedoch auch über die Diffusionslängen L_n und L_p den Diffusionsstromanteil nach Gleichung (2.22). Allerdings erfolgt eine Erhöhung des Diffusionsstromanteils aufgrund einer Lebensdauerreduktion erst, wenn dieser Strom durch die Diffusionslänge $L_{n,p}$ und nicht durch die geometrische Länge der Bahngebiete $d_{n,p}$ bestimmt wird. Gilt

$$d_{p,n} < L_{n,p} = \sqrt{D_{n,p} \tau_{n,p}}, \quad (\text{A4.5})$$

so liegt eine Kurzbasisdiode vor und die Länge des Bahngebiets bestimmt den Konzentrationsgradienten und damit den Diffusionsstrom. Für eine Diffusionskonstante der Minoritätsträger von $D = 10 \text{ cm}^2/\text{s}$ und einer Länge des Bahngebiets der niedriger dotierten Seite von $d = 5 \text{ }\mu\text{m}$ folgt somit eine minimale Lebensdauer $\tau = 25 \text{ ns}$, oberhalb welcher der Strom durch die geometrischen Abmessungen bestimmt wird.

Es ist daher wünschenswert, den Vergleich der Lebensdauern nicht allein durch einen unmittelbaren Vergleich der Ströme vorzunehmen, insbesondere weil sich die Wannentiefen und damit die Längen der Bahngebiete beim Standardprozeß auf n-Substrat und bei den hochohmigen Prozeßvarianten unterscheiden.

Schroder [A4.4] beschreibt ein Meßverfahren, welches zur Bestimmung der Rekombinationslebensdauer bei pn-Dioden angewandt werden kann. Dieses Prinzip kann laut Wilson [A4.5] auch bei pin-Dioden Anwendung finden. Die Messung des zeitlichen Abklingens der Klemmenspannung V_j an einer leerlaufenden Diode (engl. open circuit voltage decay), die sich zuvor im Vorwärtsbetrieb befand, erlaubt mittels [A4.5]

$$\tau_r = -U_t \left[\frac{dV_j(t)}{dt} \right]^{-1} \quad (\text{A4.6})$$

die Bestimmung der Lebensdauer bei niedriger Injektion aus der inversen Steigung der Spannung-Zeit-Kurve $V_j(t)$.

Im Grenzfall der Hochinjektion ist das Doppelte dieses Werts anzunehmen. In den folgenden Meßkurven ist auf diesen Faktor verzichtet worden, da die Messung der Lebensdauer auch nach anderen Verfahren (z.B. Zerbst [A4.6], [A4.7] - angewandt auf Standard n-Substrat in [A4.8]) nur Resultate mit Fehlern in dieser Größenordnung liefern kann.

Die Messung erfaßt die aufgrund der in der Diffusionskapazität der Diode gespeicherten Ladungen hervorgerufene Klemmenspannung V_j . Diese Diffusionskapazität wird durch den Vorwärtsbetrieb der Diode aufgefüllt.

Die Größe der gespeicherten Ladung läßt sich gemäß Münch [A4.9] durch

$$Q_n = Aen_0L_n \quad (\text{A4.7})$$

für eine N^+p -Diode abschätzen, hierbei ist n_0 die durch die Spannung an der Sperrschicht bestimmte Konzentration der Minoritätsträger am Rand zum p-Bahngebiet. Durch Rekombinationsprozesse werden diese Überschußladungen abgebaut bis die Diode wieder ihr thermisches Gleichgewicht erreicht und die Klemmenspannung den Wert $V_j = 0$ V annimmt. Allerdings kommt es bei dieser Messung sehr leicht zu einer Verfälschung der Spannungs-Zeit-Verläufe, da bei der Spannungsmessung mit konventionellen Speicheroszilloskopen ein überwiegender Teil der Ladungen nicht rekombinieren kann, sondern über die Meßanordnung abfließt. Die praktische Realisierung erfordert daher eine hochohmige Spannungsmessung, die jedoch schnell genug sein muß, um den schnellen Signaländerungen zu folgen. Der eigens aufgebaute Meßverstärker besitzt dank eines Operationsverstärkers mit FET-Eingang einen sehr hohen Eingangswiderstand R_{in} und trotzdem eine hohe Bandbreite, die sich in der guten Kantensteilheit des Meßsignals zeigt (vgl. Bild A4.1 und Bild A4.2). Die Entladekonstante der Meßvorrichtung $\tau_{Amp} = (R_{in}C_{in})^{-1}$ muß deutlich größer sein als die des erwarteten Signals, so daß eine möglichst gute Übereinstimmung zwischen der gemessenen Zeitkonstante τ_{ges} und der realen Lebensdauer τ erreicht wird. Gemäß

$$\frac{1}{\tau_{ges}} = \frac{1}{\tau} + \frac{1}{\tau_{Amp}} \quad (\text{A4.8})$$

erfolgt eine schaltungstechnische Parallelschaltung der Zeitkonstanten, die durch die Ladungsaufteilung zwischen Diode und Meßvorrichtung bedingt ist. Die Anordnung, die auch die Stromversorgung der Diode und die Leerlaufschaltung realisiert, erreicht ohne Meßobjekt eine Entladekonstante $\tau_{Amp} = 68$ ms, die deutlich größer als die interessierenden Zeiten im μs -Bereich ist. Somit ist durch die Messung eine Beeinflussung des Spannungsverlaufs an der Diode praktisch ausgeschlossen.

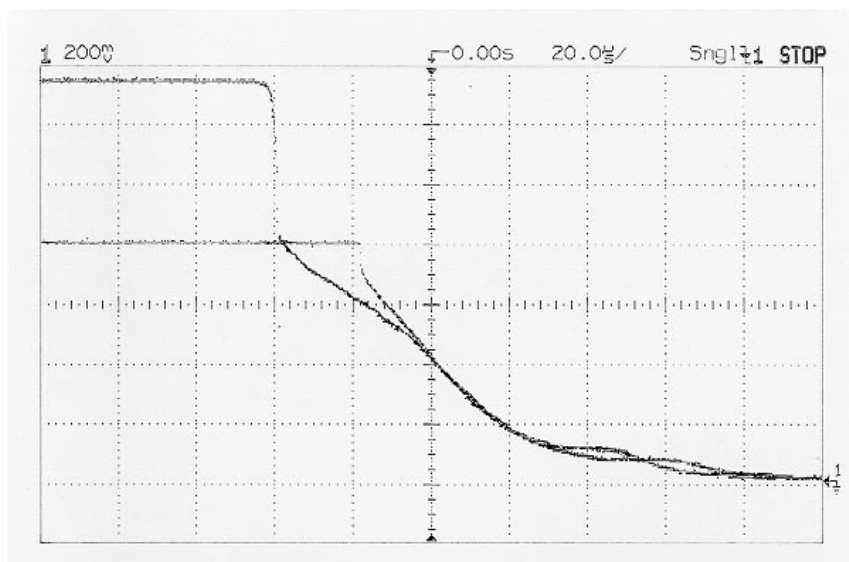


Bild A4.1: Verlauf der Klemmenspannung V_j einer N^+p -Diode der Variante C als Funktion der Zeit. Die Maße der Diode betragen $A = 102 \times 28 \mu\text{m}^2$. Die beiden Kurven stammen aus der Messung bei zwei unterschiedlichen Vorwärtsspannungen $V_j(t = t_0)$. Aufgrund der gewählten Achsenteilung beträgt die Länge der Zeitachse $\Delta t = 200 \mu\text{s}$ und die Höhe der Spannungsachse $\Delta V = 1.6$ V.

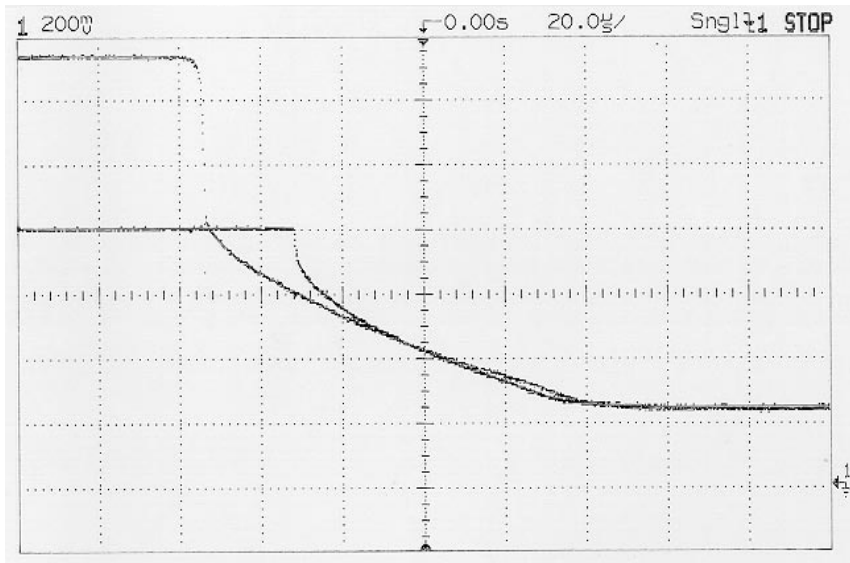


Bild A4.2: Verlauf der Klemmenspannung V_j einer P^+n -Diode der Variante B als Funktion der Zeit. Die Maße der Diode betragen $A = 102 \times 28 \mu\text{m}^2$. Die beiden Kurven stammen aus der Messung bei zwei unterschiedlichen Vorwärtsspannungen $V_j(t = t_0)$. Aufgrund der gewählten Achsenteilung beträgt die Länge der Zeitachse $\Delta t = 200 \mu\text{s}$ und die Höhe der Spannungsachse $\Delta V = 1.6 \text{ V}$.

Die Bilder A4.1 und A4.2 zeigen die $V_j(t)$ -Daten einer N^+p -Diode der Variante C und einer P^+n -Diode der Variante B auf hochohmigem p^- -Substrat. Die beiden dargestellten Verläufe in Bild A4.1 korrespondieren zu den angelegten Spannungen von $V_j(t = t_0) = 0.8 \text{ V}$ und $V_j(t = t_0) = 1.3 \text{ V}$. Die Spannung an der Diode fällt nach Auftrennen der Spannungsversorgung bei $t = t_0$ anfangs nahezu senkrecht ab. Da mit aufgetrenntem Stromkreis bei $t = t_0$ der Stromfluß durch die Diode abrupt zum Erliegen kommt, verringert sich die Klemmenspannung um den bisherigen Spannungsabfall $\Delta V = I \cdot R_s$ am Bahnwiderstand der Diode R_s . Der zu $V_j(t = t_0) = 0.8 \text{ V}$ korrespondierende Verlauf zeigt deutlich einen nahezu linearen Bereich, aus dem sich die Lebensdauer ablesen läßt. Bei der höheren Flußspannung hingegen sind zwei Steigungen zu erkennen, was aus obiger Betrachtung aufgrund der injektionsabhängigen Größe der Lebensdauer zu erwarten ist. Die Steigung, die später - bei geringen Ladungen - erreicht wird, entspricht sehr genau jener, die bei geringer Vorwärtsspannung vorliegt.

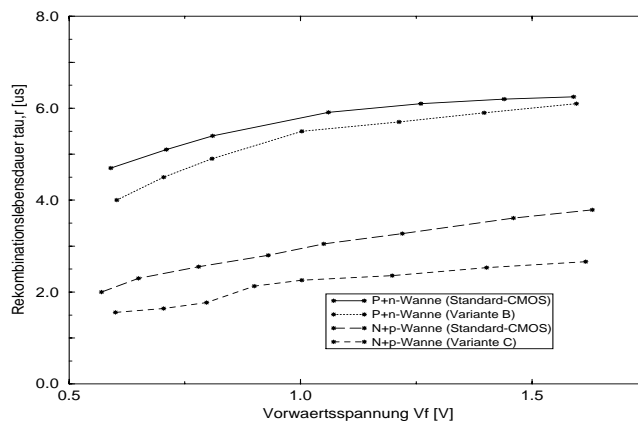


Bild A4.3: Aus V_f -t-Messungen ermittelte Rekombinationslebensdauern von pn-Dioden in Abhängigkeit der Vorwärtsspannung V_f .

Die Vorwärtsspannungen in Bild A4.2 entsprechen 0.7 V und 1.1 V . Auch hier wird der Anstieg der Lebensdauer bei stärkerer Injektion von Minoritätsträgern deutlich. Die Darstellung der Lebensdauern als Funktion der Flußspannung ist in Bild A4.3 gezeigt. Bei dieser Untersuchung weisen Löcher in etwa doppelt so hohe Lebensdauerwerte wie Elektronen auf.

Die Unterschiede zwischen Standard-Substrat und p^- -Silizium sind aufgrund der erwarteten Meßgenauigkeit als unbedeutend einzustufen. Insgesamt ergeben sich Lebensdauern im Bereich einiger μs .

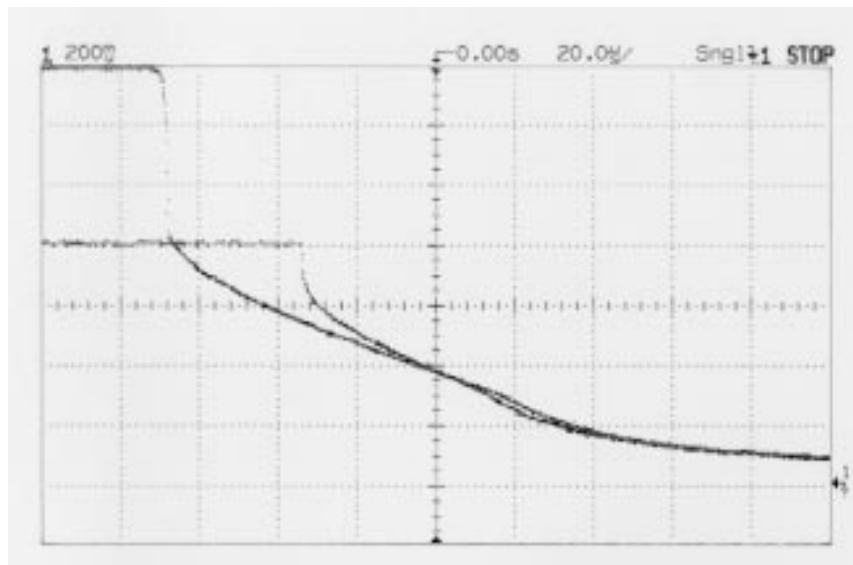


Bild A4.4: Verlauf der Klemmenspannung V_j einer pin-Diode als Funktion der Zeit. Die Maße der i -Zone betragen $W = 38 \mu\text{m}$ und $L = 12 \mu\text{m}$. Die beiden Kurven stammen aus der Messung bei zwei unterschiedlichen Vorwärtsspannungen $V_j(t = t_0)$. Aufgrund der gewählten Achsenteilung beträgt die Länge der Zeitachse $\Delta t = 200 \mu\text{s}$ und die Höhe der Spannungsachse $\Delta V = 1.6 \text{ V}$.

Zum Vergleich ist in Bild A4.4 der Spannungsverlauf an einer pin-Diode dargestellt. Ein prinzipiell ähnlicher Verlauf der Klemmenspannung und vergleichbare Lebensdauern (vgl. Bild A4.5) können beobachtet werden.

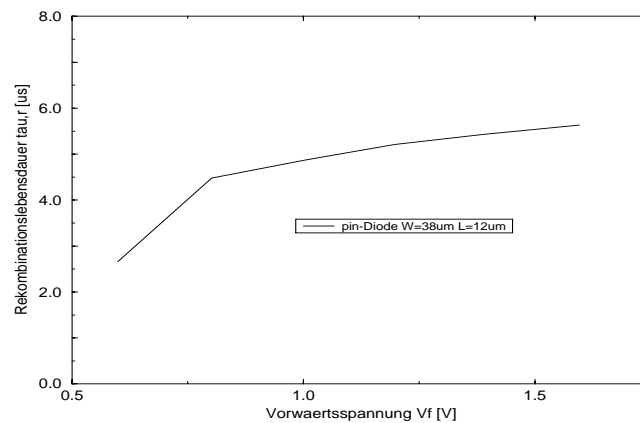


Bild A4.5: Aus V_f - t -Messungen ermittelte Rekombinationslebensdauern von pin-Dioden in Abhängigkeit der Vorwärtsspannung V_f .

A 5 CMOS-Schaltungstechnik

Der Einsatz von CMOS-Schaltungen in zukünftigen SIMMWIC-Chips ist für eine Fülle von Anwendungen wünschenswert, da eine Vielzahl von digitalen und analogen Schaltungskonzepten existiert und so eine komplette Neuentwicklung in den meisten Fällen vermieden werden kann. In dem dieser Arbeit zugrundeliegenden Konzept besteht das vornehmliche Interesse in der Realisierung von CMOS-Komponenten für künftige SIMMWIC-mm-Wellen-Empfänger.

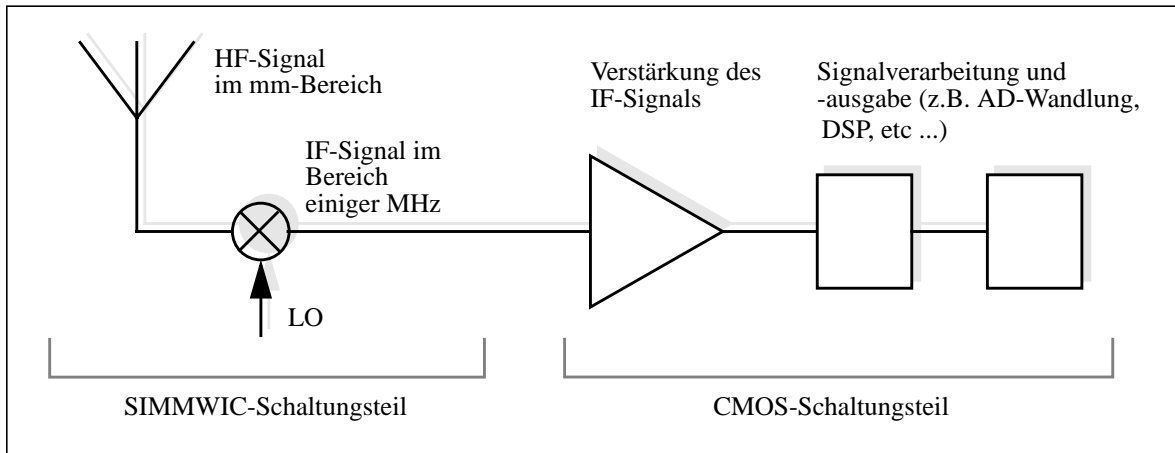


Bild A5.1: Prinzipschaltbild eines SIMMWIC-mm-Wellenreceivers. Das von der mm-Wellen-Antenne empfangene Signal wird auf eine Zwischenfrequenz (IF, englisch: intermediate frequency) unter Einsatz eines im GHz-Bereich arbeitenden Lokaloszillators (LO) heruntergemischt. Das IF-Signal im MHz-Bereich dient nach einer ersten analogen Verstärkung als Eingangssignal für die folgende Signalverarbeitung auf analogem oder nach AD-Wandlung auch digitalem Wege.

Während auf der Seite des mm-Wellen-Senders auch eine Bipolartechnik zum Einsatz kommen könnte, da diese aufgrund höherer Leistungsdichten und Schaltgeschwindigkeiten für Leistungsverstärker besser geeignet ist, kann CMOS empfängerseitig einige deutliche Vorzüge bieten. Die hohen Integrationsdichten und der geringe Leistungsverbrauch werden in Zukunft die Realisierung komplexer Schaltungen erlauben. Bei einer fortgeschrittenen Prozeßtechnologie sind Bandbreiten mittels CMOS-Technik möglich, die für die anvisierten Zwischenfrequenzen von einigen MHz als ausreichend gelten können. Die Dominanz der CMOS-Technik in der konventionellen Halbleiterfertigung hat zu einer Fülle von digitalen als auch analogen Schaltungskonzepten geführt, die bei der Signalaufbereitung, -verarbeitung und -speicherung Anwendung finden können.

Die Zielrichtung der in diesem Anhang beschriebenen Arbeiten berücksichtigt drei grundlegende Aspekte:

- (i) Die realisierten Schaltungen sind als Vorverstärker (siehe Anhang A5.2) für die Weiterverarbeitung von empfangenen und heruntergemischten Signalen ausgelegt worden und kommen z. B. in der hybriden Kombination einer SIMMWIC-Rectenna und eines CMOS-Vorverstärkers, dem sog. Hybrid-Demonstrator [A5.1, A5.2], zum Einsatz. Alternativ ist eine Ausgangstreiberschaltung (siehe Anhang A5.3) für mittlere Ströme konzipiert worden, die in einem integrierten pin-Hochfrequenzschalter zum Einsatz kommt.
- (ii) Diese Schaltungen sollen einfach genug sein, um Rückschlüsse auf die Prozeßtechnologie zu erlauben. So ist in allen Fällen ein Vergleich zwischen den Kennlinien der Schaltkreise auf dem hochohmigen p^+ -Substrat und dem n -Substrat des Vergleichsprozesses vorgenommen worden. Die zur Wechselspannungs-Charakterisierung notwendigen Meßadapter sind im Rahmen dieser Arbeit weiterentwickelt worden (Siehe Anhang A5.1).
- (iii) Aufgrund der relativ geringen Lithographieauflösung werden nur gering komplexe Schaltungen betrachtet, da andernfalls die zur Verfügung stehende Bandbreite zu stark begrenzt würde. Es sind daher nur ein- und zwei-stufige Verstärkerschaltungen entworfen worden.

A 5.1 Analoge Verstärkermeßtechnik

Frequenzgang nach Betrag und Phase

Zur Bestimmung der Übertragungscharakteristik von Verstärkerschaltungen ist ein eigens entwickelter Meßadapter eingesetzt worden [A5.3]. Diese Schaltung stellt eine Weiterentwicklung der in [A5.4] beschriebenen dar.

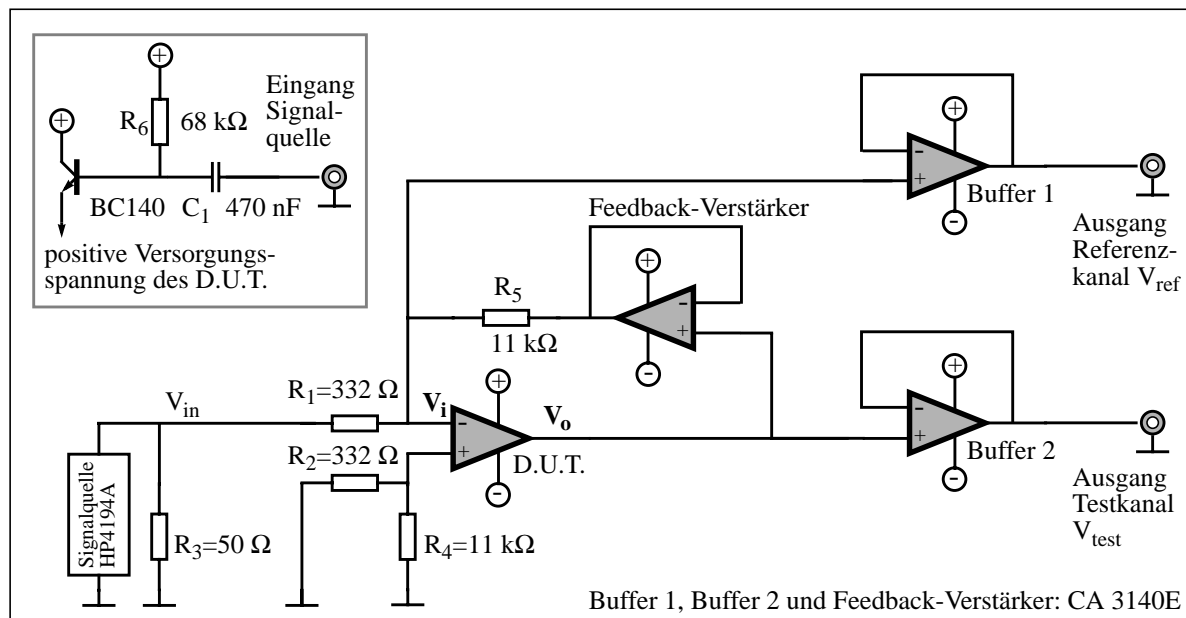


Bild A5.2: Vereinfachtes Schaltbild des Meßadapters für die Messung der Wechselspannungscharakteristik von CMOS-Verstärkern. Der zu untersuchende Verstärker ist als D.U.T. (device under test) bezeichnet. Die in dem Rahmen hinzugefügten Schaltungsteile werden bei der Bestimmung der Unterdrückung von Versorgungsspannungsschwankungen (PSRR, englisch power supply rejection ratio) eingesetzt, hier dargestellt für die positive Versorgungsspannung.

Das in Bild A5.2 dargestellte Prinzipschaltbild verdeutlicht die Konfiguration bei der frequenzabhängigen Bestimmung von Betrag A und Phase ϕ der komplexen Leerlaufverstärkung $\bar{A} = V_o/V_i = A(f) \cdot \exp(\phi(f))$. Neben den beiden Pufferverstärkern Buffer 1 und Buffer 2, die das Referenzsignal V_{ref} bzw. das Testsignal V_{test} als nicht invertierende Puffer mit einfacher Verstärkung auf ein niedrigeres Impedanzniveau transformieren, kommt auch dem Rückkopplungsverstärker (englisch: feedback) besondere Bedeutung zu. Da der Netzwerkanalysator eine komplexe Verhältnisbildung der Signale des Test- und Referenzkanals vornimmt, reduziert sich der Einfluß der als Buffer eingesetzten baugleichen Verstärker, welche die komplexen Verstärkungen $\bar{A}_1(\omega)$ und $\bar{A}_2(\omega)$ aufweisen, zu einer Multiplikation der gesuchten Verstärkung mit dem Faktor eins (vgl. Gleichung (A5.1)).

$$\bar{A} = \frac{V_o}{V_i} = \frac{V_{test}/(\bar{A}_2(\omega))}{V_{ref}/(\bar{A}_1(\omega))} = \frac{V_{test}}{V_{ref}} \cdot \frac{\bar{A}_1(\omega)}{\bar{A}_2(\omega)} \approx \frac{V_{test}}{V_{ref}} \quad (\text{A5.1})$$

Kontrollmessungen mittels einer Brücke zwischen Eingang und Ausgang des zu vermessenden Verstärkers (D.U.T., englisch: device under test) ergaben eine maximale Genauigkeitsabweichung des Betrags kleiner als 0.2 dB und der Phase kleiner als ein Grad bei Frequenzen bis $f_{max} = 3$ MHz. Kommerzielle Operationsverstärker mit Leerlaufverstärkungen von 100 dB konnten ebenfalls gemessen werden, so daß die im Rahmen dieser Arbeit entworfenen und gefertigten Verstärkerschaltungen mit einer guten Genauigkeit charakterisiert werden können. Der nichtinvertierende Operationsverstärker (OP) des Rückkopplungszweiges ist insbesondere bei hohen Frequenzen für die Messung der Leerlaufverstärkung von Bedeutung. Ersetzt man in der obigen Schaltung (siehe Bild A5.2) diesen OP durch einen Kurzschluß, so erhält man die konventionelle Schaltung zur Messung des Frequenzgangs. Der zu untersuchende Verstärker weise nun eine Spannungsverstärkung \bar{A} auf und besitze einen

Serienwiderstand \bar{Z} im Ausgangskreis. Daraus folgt für das Verhältnis der Spannungen

$$\frac{V_o}{V_i} = -\frac{\bar{A} - \frac{\bar{Z}}{R_5}}{1 + \frac{\bar{Z}}{R_5}} \approx -\bar{A}, \quad (A5.2)$$

$|\bar{A}| \gg 1, |\bar{Z}| \ll R_5$

welches nur bei großen Verstärkungsfaktoren \bar{A} und geringen Serienwiderständen \bar{Z} direkt auf den gesuchten Wert \bar{A} führt. Gerade bei höheren Frequenzen nahe der Transitfrequenz ($|\bar{A}| \rightarrow 1$) ist diese Voraussetzung nicht erfüllt.

Durch eine aktive Rückkopplung kann der Querstrom vom Ausgang des D.U.T. durch den Rückkopplungswiderstand R_5 minimiert werden. Ein an seinen Eingängen möglichst hochohmiger OP, mit einfacher Verstärkung nicht-invertierend betrieben, sorgt für eine geringe Belastung am D.U.T.-Ausgang und liefert den notwendigen Rückkopplungsstrom. Dieser Feedback-Verstärker wird durch eine Spannungsquelle mit Verstärkungsfaktor \bar{A}_{fb} und einen Serienwiderstand \bar{Z}_{afb} , sowie einen Eingangswiderstand \bar{Z}_{ifb} modelliert. Das Kleinsignalersatzschaltbild nach der Umwandlung der Spannungsquellen in gesteuerte Stromquellen ist in Bild A5.3 dargestellt.

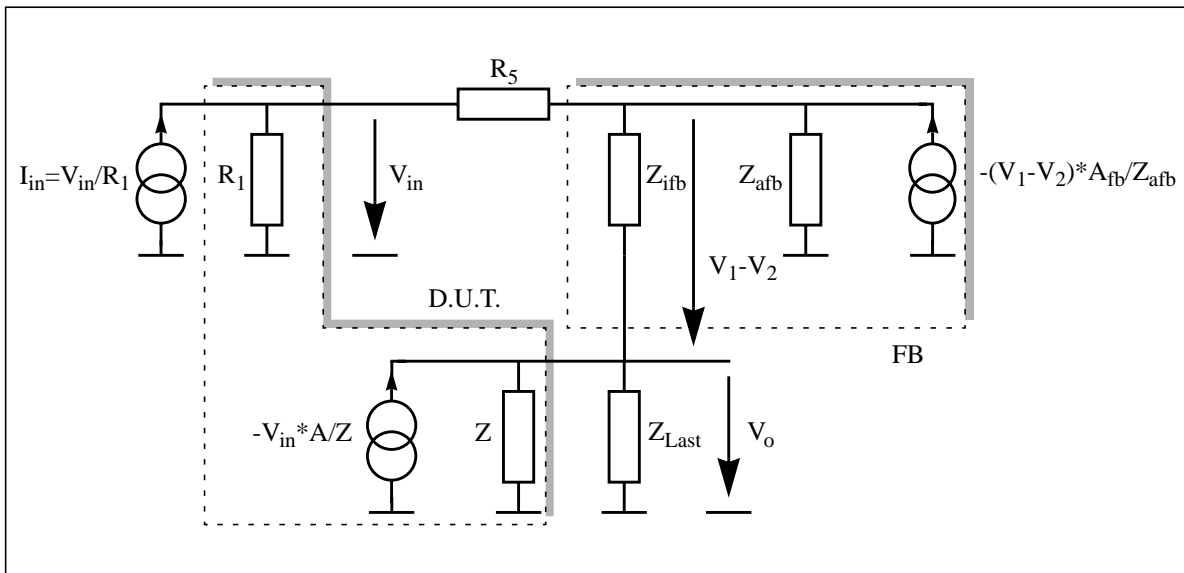


Bild A5.3: Kleinsignalersatzschaltbild der aktiven Rückkopplung aus D.U.T. und Feedback-Verstärker. $V_1 - V_2$ ist die Spannungsdifferenz an den Eingängen des Feedback-Verstärkers. Die Komponenten des D.U.T. und des Feedback-Verstärkers (FB) sind gesondert markiert, alle weiteren Größen sind wie in Bild A5.2 gewählt.

Eine Knotenanalyse [A5.5] der Schaltung führt auf

$$[\bar{I}] = [\bar{G}] \cdot [\bar{U}], \quad (A5.3)$$

wobei die Leitwertmatrix \bar{G} durch [A5.3]

$$[\bar{G}] = \begin{bmatrix} \left(\frac{1}{R_1} + \frac{1}{R_5}\right) & -\frac{1}{R_5} & 0 \\ -\frac{1}{R_5} & \left(\frac{1}{R_5} + \frac{1}{Z_{ifb}} + \frac{1}{Z_{afb}} + \frac{A_{fb}}{Z_{afb}}\right) & \left(-\frac{1}{Z_{ifb}} - \frac{1}{Z_{afb}}\right) \\ \frac{A}{Z} & -\frac{1}{Z_{ifb}} & \left(\frac{1}{Z_{ifb}} + \frac{1}{Z} + \frac{1}{Z_{last}}\right) \end{bmatrix}, \quad (A5.4)$$

der Stromvektor durch

$$[\bar{I}] = \begin{bmatrix} \left(\frac{V_{in}}{R_1}\right) \\ 0 \\ 0 \end{bmatrix}^T \quad (\text{A5.5})$$

und der Spannungsvektor durch

$$[\bar{U}] = [V_{in}, (V_1 - V_2 + V_o), V_o]^T \quad (\text{A5.6})$$

gegeben sind. Aufgrund der Pufferverstärker ist die Belastung der Ausgänge und damit der Leitwert $1/Z_{last}$ vernachlässigbar, so daß unter dieser Bedingung

$$\frac{V_o}{V_i} = \frac{Z_{ifb}(AR_5(1 + A_{fb}) + AZ_{afb}) + Z_{afb}(AR_5 - Z)}{Z_{ifb}(R_5(1 + A_{fb}) + Z_{afb}) + Z_{afb}(R_5 + Z) + ZR_5} \quad (\text{A5.7})$$

wird. Bei geeigneter Wahl des Verstärkers und des Widerstands im Feedback-Zweig ($R_5 > 100 \cdot Z_{afb}$, $Z_{ifb} > 100 \cdot Z_{afb}$ und $Z_{ifb} > 100 \cdot Z$) sowie leicht zu erfüllenden Anforderungen an die Verstärkung des D.U.T. ($A > 100 \cdot Z \cdot Z_{afb} / (R_5 \cdot Z_{ifb} + Z \cdot Z_{afb})$) [A5.4] folgt

$$\frac{V_o}{V_i} = -\bar{A} \quad (\text{A5.8})$$

mit einem Fehler kleiner als 1 %. Aufgrund des hohen Eingangswiderstands $|Z_{ifb}| = R_e = 1 \text{ T}\Omega$ sowie des niedrigen Ausgangswiderstands $|Z_{afb}| = R_a = 60 \Omega$ des Feedback-Verstärkers lassen sich so auch Verstärkungen deutlich kleiner als $A = 1$ ($= 0 \text{ dB}$) bestimmen.

Gleichtakt- und Speisespannungsunterdrückung

Eine wichtige Größe bei Verstärkern mit Differenzeingängen ist die Gleichtaktverstärkung \overline{CMG} (englisch: common mode gain) und die daraus abgeleitete Gleichtaktunterdrückung $CMRR = |\bar{A}/\overline{CMG}|$ (englisch: common mode rejection ratio) als Quotient aus Gleichtaktverstärkung \overline{CMG} und Leerlaufverstärkung \bar{A} . Die oben vorgestellte Meßvorrichtung wird zur Messung dieser Größen eingesetzt, indem nach der Bestimmung der Leerlaufverstärkung die Gleichtaktverstärkung ermittelt wird. Dies geschieht durch die Ankopplung der Signalquelle an beide D.U.T.-Eingänge. Eine sehr geringe Gleichtaktverstärkung und damit eine gute Gleichtaktunterdrückung vermindern den Einfluß der z. B. durch thermischen Drift bedingten Störsignale auf das verstärkte Nutzsignal. Analog ist die Speisespannungsunterdrückung PSRR (englisch: power supply rejection ratio) definiert. Die Speisespannungsverstärkung \overline{AP} wird hierbei auf die Leerlaufverstärkung \bar{A} bezogen: $PSRR = |\bar{A}/\overline{AP}|$. Hohe Werte von PSRR bedeuten eine gute Unterdrückung der durch die Versorgungsspannung bedingten Einflüsse. Auf den Versorgungsleitungen befindliche Wechselspannungsanteile führen bei kleinen PSRR-Werten zu einer Verfälschung des Ausgangssignals. Bei dieser Messung, die für die positive und die negative Versorgungsspannung separat ausgeführt wird, kommt es zu einer Überlagerung der Signal- und der Speisespannung durch den in Bild A5.2 eingerahmt dargestellten Schaltungssteil.

A 5.2 Analoge Verstärkerschaltungen

Die im vorherigen Abschnitt besprochenen Wechselspannungsmeßmethoden sind neben der Gleichspannungsmeßtechnik, die für Offset- und Großsignalverstärkungs-Messungen benutzt worden ist, eingesetzt worden, um die realisierten Verstärkerschaltungen auf Standard-n-Substrat und hochohmigem p⁻-Substrat vergleichend zu charakterisieren.

CMOS-Inverterstufe

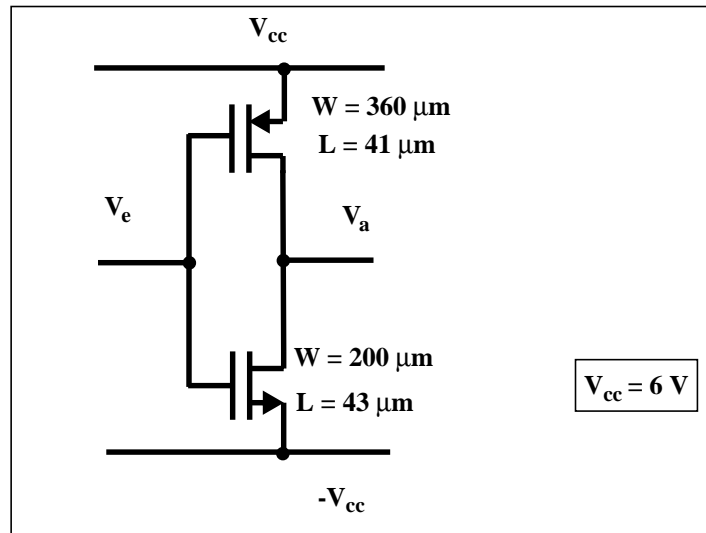


Bild A5.4: Schaltbild der Inverterstufe aus NMOS und PMOS-Fet.

Aufgrund ihrer schaltungstechnischen Einfachheit und ihrer Bedeutung als Grundbaustein der Digitaltechnik ist eine Inverterstufe realisiert worden. Die zu Vergleichszwecken mittels des Standard-CMOS-Prozesses hergestellten Inverter sind als Vorverstärker für den hybriden Aufbau einer SIMMWIC-Rectenna und eines CMOS-Schaltungsteils eingesetzt worden [A5.1], [A5.2]. Aufgrund der sehr langkanaligen MOS-Transistoren läßt sich mit dieser Schaltung bereits eine Verstärkung von mehr als 40 dB erreichen.

$$A_V = \frac{g_{m,NMOS} + g_{m,PMOS}}{g_{o,NMOS} + g_{o,PMOS}} = \frac{85 \cdot 10^{-4} \sqrt{A/V} \cdot \sqrt{I_o} + 110 \cdot 10^{-4} \sqrt{A/V} \cdot \sqrt{I_o}}{12 \cdot 10^{-4} \cdot 1/V \cdot I_o + 57 \cdot 10^{-4} \cdot 1/V \cdot I_o} = 2,82 \sqrt{A} \cdot \frac{1}{\sqrt{I_o}} = 115 \Rightarrow 41,2 \text{ dB} \quad (\text{A5.9})$$

Der Berechnung der Spannungsverstärkung liegen der Strom $I_o = 600 \mu\text{A}$ und die Transistordaten aus Kapitel 2 zu Grunde. Die Ergebnisse der meßtechnischen Charakterisierung sind in Bild A5.5 dargestellt. Der Ausgangsspannungshub überschreitet $\pm 4 \text{ V}$ und der Verlauf der Versorgungsstromkennlinie verdeutlicht die Charakteristik der CMOS-Inverter nur während des eigentlichen Schaltvorgangs eine erkennbare Leistungsaufnahme aufzuweisen. In der Digitaltechnik nutzt man diese Eigenschaft bewußt, da bei zeitlich konstanten Logik-Pegeln die Leistungsaufnahme minimiert wird. Für den Einsatz als Analogverstärker, der im Ruhebetrriebspunkt bei der Eingangsspannung $V_e = 0 \text{ V}$ gehalten wird, bedeutet dies Betrieb bei maximaler Verlustleistung der Schaltung. Die Stromaufnahme liegt jedoch in der gleichen Größenordnung wie bei der im folgenden zu diskutierenden Differenzeingangsstufe.

Die geringe Anzahl von Bauelementen im Signalweg führt trotz großer Kanallängen zu recht hohen Grenzfrequenzen und einer guten Phasenreserve von $\Delta\phi = 79,5^\circ$ bei $f_{\text{0dB}} = 1,94 \text{ MHz}$.

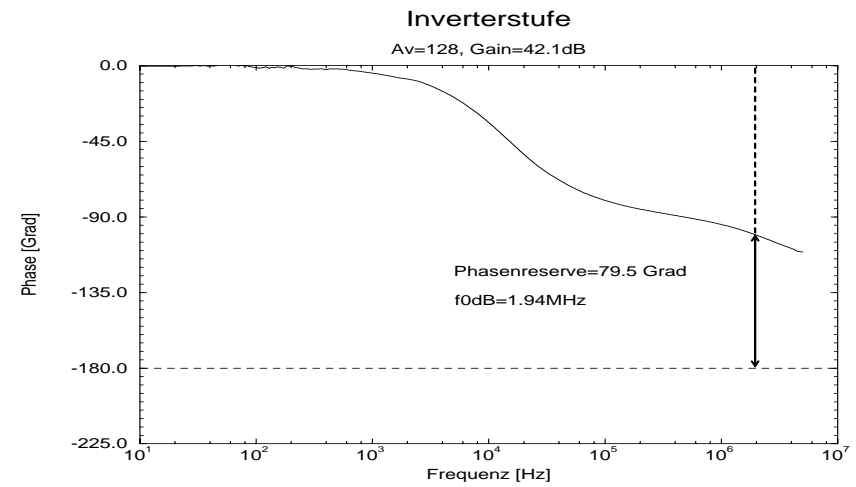
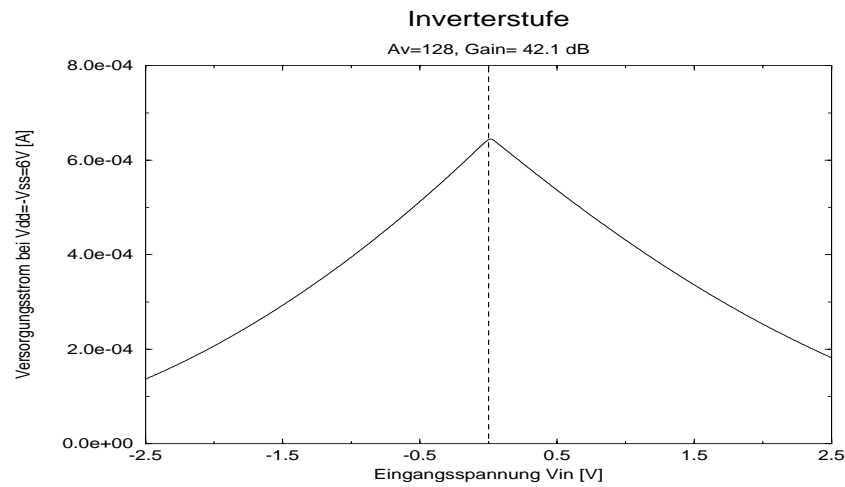
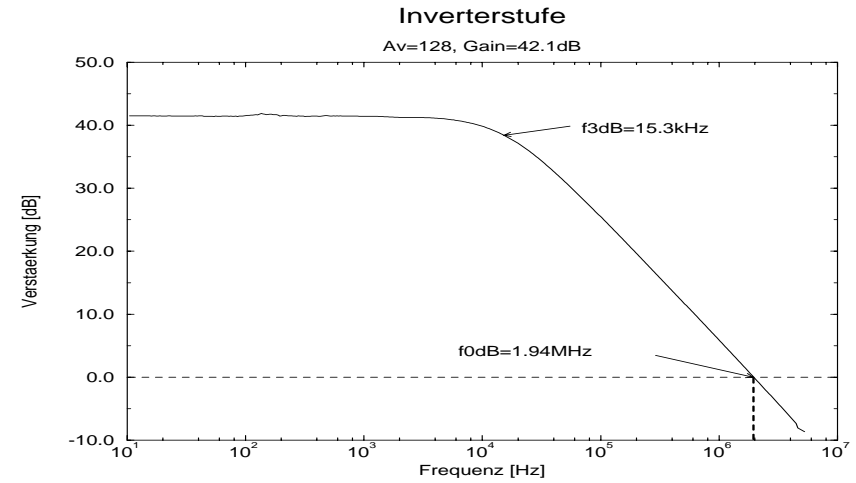
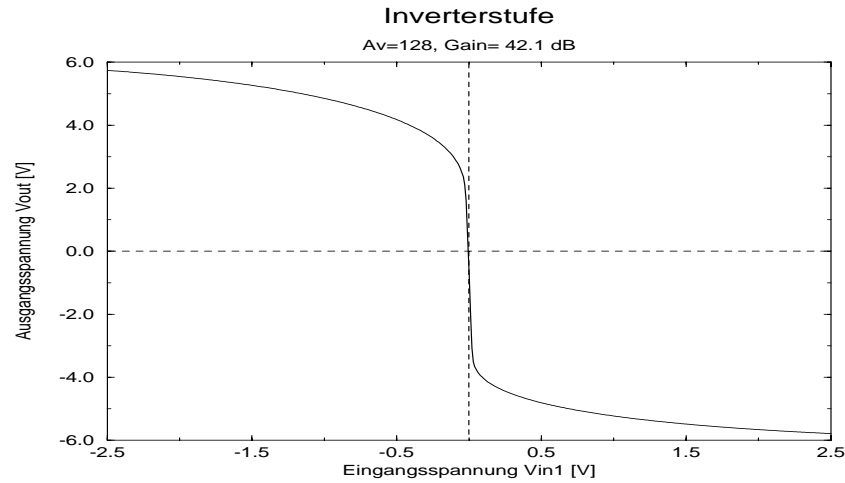


Bild A5.5: Meßkurven der Inverterstufe. Links Gleichspannungs- bzw. Großsignalverhalten, oben Kennlinie der Spannungsverstärkung, unten Verlauf der Stromaufnahme der Schaltung. Rechts Wechselspannungs- bzw. Kleinsignalverhalten: Darstellung als Bode-Plot des Frequenzgangs mit Betrag (oben) und Phase (unten).

Differenzverstärker als Eingangsstufe

Eingangsstufe fast aller Operationsverstärkerschaltungen ist die sogenannte Differenzeingangsstufe. Hauptvorteil einer solchen Auslegung ist die Unterdrückung von Störsignalen, die an beiden Eingängen in gleicher Größe anliegen. Dies können z. B. durch thermische Drift bedingte Schwankungen des Meßsignals oder auch Gleichspannungsanteile sein. Ferner wird die Verstärkung nicht massebezogener Signale ermöglicht. Hauptnachteil dieses Schaltungskonzepts ist eine Verdoppelung des Rauschens im Vergleich zu einer Schaltung mit nur einem Eingangstransistor, z.B. einer Sourcestufe [A5.6], [A5.7].

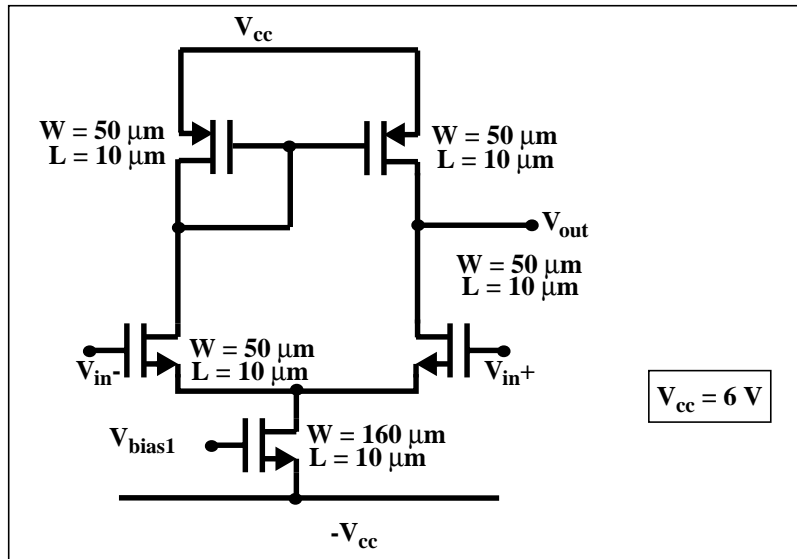


Bild A5.6: Schaltbild der Differenzeingangsstufe mit aktiver Last. Die Stromquelle dieser Stufe wird mittels eines Stromspiegels angesteuert, so daß keine externe Bereitstellung einer Biasspannung erforderlich ist.

Der weitgehend symmetrische Aufbau der in Bild A5.6 dargestellten Stufe wird durch den Einsatz einer PMOS-Stromspiegelstufe als Last durchbrochen. Der im Bild links dargestellte Eingangstransistor wird mit einer deutlich geringeren Impedanz beaufschlagt als der des rechten Zweiges. Dies hat zur Folge, daß das Ausgangssignal auf der rechten Seite als massebezogenes Spannungssignal zur Verfügung steht. Im Vergleich zu einer Anordnung mit zwei separaten PMOS-Konstantstromquellen als Last bietet diese Ausführung den Vorteil der höheren Spannungsverstärkung sowie die Freiheit neben dem zwischen beiden Drainanschlüssen der Transistoren verfügbaren Differenzsignal auch ein einzelnes massebezogenes Signal abgreifen zu können.

Der Arbeitspunkt der Stufe wird über einen ohmschen Widerstand und einen als Diode geschalteten MOS-Transistor eingestellt. Hierbei kommt eine Stromspiegelschaltung zum Einsatz. Dies erlaubt den Betrieb der Schaltung ohne externe Bereitstellung einer zusätzlichen Spannungsquelle V_{bias1} . Dennoch ist der Biasstrom I_B durch ein von außen zugängliches Kontaktfeld (engl. pad) zusätzlich extern beeinflussbar, so daß in erster Näherung im Rahmen der Beziehung $A_v \cdot f_{-3\text{dB}} = \text{GBW}$ die Niederfrequenzverstärkung A_v und die Bandbreite $f_{-3\text{dB}}$ gegeneinander austauschbar sind. Allerdings zeigt die Simulation, daß für größere Verstärkungen die Bandbreite überproportional sinkt.

TABELLE 21. NF-Verstärkung und Bandbreite

Biasstrom I_B [μA]	NF-Verstärkung A_v [1]	Bandbreite $f_{-3\text{dB}}$ [kHz]
50	50.4	8.6
100	40.7	15.5
200	33.1	26.0
400	25.0	46.2

Für eine rechnerische Analyse der Niederfrequenzverstärkung ist in Bild A5.7 das Kleinsignalersatzschaltbild dieser Stufe gezeigt. Zur Vereinfachung der Rechnung sind im folgenden die Einflüsse des endlichen Ausgangsleitwerts der Eingangstransistoren g_{o1} und g_{o2} vernachlässigt worden. Aus diesem Ersatzschaltbild kann ebenfalls eine Berechnung des Frequenzgangs erfolgen, wenngleich für eine genauere Analyse der Einsatz eines Schaltungssimulators, wie z. B. das im Rahmen dieser Arbeit benutzte SPICE 3 [A5.8], erforderlich ist.

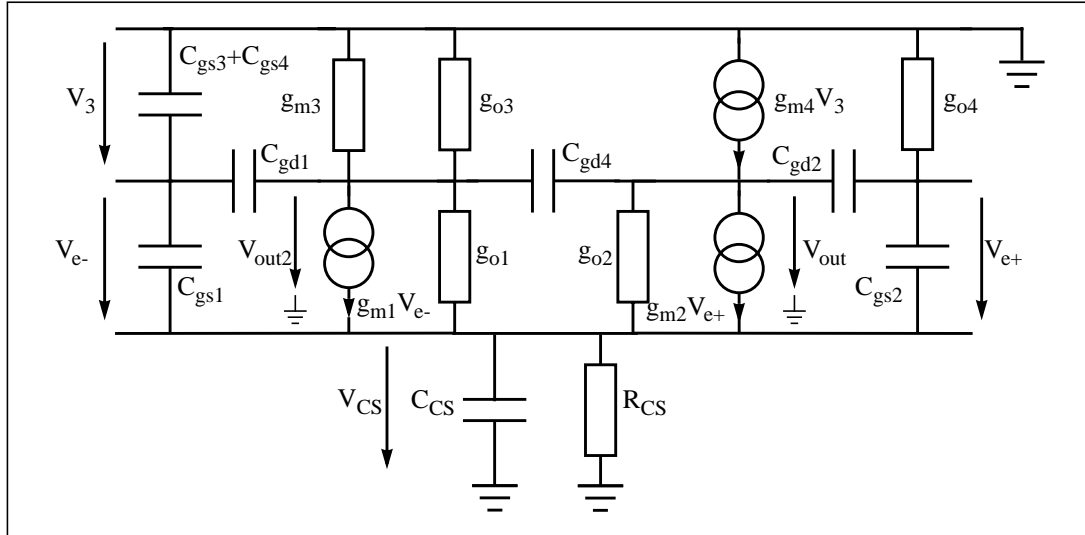


Bild A5.7: Kleinsignalersatzschaltbild der Differenzeingangsstufe. Die Eingangstransistoren sind mit 1 und 2, die Lasttransistoren mit 3 bzw. 4 sowie die Stromquelle mit CS bezeichnet.

Der Ansatz mit

$$V_{in1} = V_{e-} + V_{CS}$$

und

$$V_{in2} = V_{e+} + V_{CS}$$

sowie die Anwendung der Knotenregel führen zu folgender Spannung V_{CS} an der Stromquelle CS:

$$V_{CS} = \frac{g_{m1}V_{in1} + g_{m2}V_{in2}}{g_{m1} + g_{m2} + 1/R_{CS}}. \quad (\text{A5.10})$$

Die an den beiden Ausgängen vorhandenen Spannungssignale V_{out} und V_{out2} ergeben sich als lineare Superposition der durch die beiden Eingangssignale V_{in1} und V_{in2} bedingten Anteile. Mit der Definition

$$V_{out} = A_{11}V_{in1} + A_{12}V_{in2} \quad V_{out2} = A_{21}V_{in1} + A_{22}V_{in2} \quad (\text{A5.11})$$

ergeben sich nach Anwendung der Maschenregel die Beziehungen für die gesuchten Ausgangsspannungen. Nach Auflösung und Umstellung folgt damit für die gesuchten Verstärkungen

$$A_{11} = \frac{\frac{g_{m1}g_{m2}g_{m4}}{g_{m3} + g_{o3}} + \frac{g_{m1}g_{m4}}{g_{m3} + g_{o3}} \cdot \frac{1}{R_{CS}} + g_{m1}g_{m2}}{g_{o4} \cdot (g_{m1} + g_{m2} + 1/R_{CS})} \approx \frac{\frac{2}{g_{m1}} + \frac{g_{m1}}{R_{CS}} + g_{m1}}{g_{o4} \cdot (g_{m1} + g_{m2} + 1/R_{CS})} \approx \frac{g_{m1} + \frac{1}{2R_{CS}}}{g_{o4} \left(1 + \frac{1}{2g_{m1}R_{CS}}\right)} \approx \frac{g_{m1}}{g_{o4} \left(1 + \frac{1}{2g_{m1}R_{CS}}\right)}, \quad (\text{A5.12})$$

$$A_{12} = \frac{\frac{g_{m1}g_{m2}g_{m4}}{g_{m3}+g_{o3}} + \frac{g_{m2}}{R_{CS}} + g_{m1}g_{m2}}{g_{o4} \cdot (g_{m1}+g_{m2}+1/R_{CS})} \approx \frac{\frac{2}{g_{m1}+R_{CS}} + g_{m1}}{g_{o4} \cdot (g_{m1}+g_{m2}+1/R_{CS})} \approx \frac{g_{m1} + \frac{1}{2R_{CS}}}{g_{o4} \left(1 + \frac{1}{2g_{m1}R_{CS}}\right)} \approx \frac{g_{m1}}{g_{o4} \left(1 + \frac{1}{2g_{m1}R_{CS}}\right)}, \quad (\text{A5.13})$$

$$A_{21} = -\frac{g_{m1}}{g_{m3}+g_{o3}} \cdot \frac{g_{m2}+1/R_{CS}}{g_{m1}+g_{m2}+1/R_{CS}} \approx -\frac{g_{m1}}{g_{m3}} \cdot \frac{g_{m2} \left(1 + \frac{1}{g_{m2}R_{CS}}\right)}{g_{m2} \left(2 + \frac{1}{g_{m2}R_{CS}}\right)} \approx -\frac{g_{m1}}{2g_{m3}} \left(1 + \frac{1}{g_{m2}R_{CS}}\right) \approx -\frac{g_{m1}}{2g_{m3}} \quad (\text{A5.14})$$

und

$$A_{22} = \frac{g_{m1}}{g_{m3}+g_{o3}} \cdot \frac{g_{m2}}{g_{m1}+g_{m2}+1/R_{CS}} \approx \frac{g_{m1}}{g_{m3}} \cdot \frac{g_{m2}}{g_{m2} \left(2 + \frac{1}{g_{m2}R_{CS}}\right)} \approx \frac{g_{m1}}{2g_{m3}}. \quad (\text{A5.15})$$

Mit Hilfe der üblichen Definitionen für die Differenz- und Gleichtaktsignale [A5.9] lassen sich die Differenzverstärkung A_{dm}

$$\begin{aligned} A_{dm} &= \frac{A_{11} - A_{12} - A_{21} + A_{22}}{2} \\ &= \frac{1}{2} \left[\frac{\frac{g_{m1}g_{m2}g_{m4}}{g_{m3}+g_{o3}} + \frac{g_{m1}g_{m4}}{g_{m3}+g_{o3}} \cdot \frac{1}{R_{CS}} + g_{m1}g_{m2}}{g_{o4} \cdot (g_{m1}+g_{m2}+1/R_{CS})} + \frac{\frac{g_{m1}g_{m2}g_{m4}}{g_{m3}+g_{o3}} + \frac{g_{m2}}{R_{CS}} + g_{m1}g_{m2}}{g_{o4} \cdot (g_{m1}+g_{m2}+1/R_{CS})} + \right. \\ &\quad \left. \frac{g_{m1}}{g_{m3}+g_{o3}} \cdot \frac{g_{m2}+1/R_{CS}}{g_{m1}+g_{m2}+1/R_{CS}} + \frac{g_{m1}}{g_{m3}+g_{o3}} \cdot \frac{g_{m2}}{g_{m1}+g_{m2}+1/R_{CS}} \right] \\ &\approx \frac{1}{2} \left[\frac{g_{m1}}{g_{o4} \left(1 + \frac{1}{2g_{m1}R_{CS}}\right)} + \frac{g_{m1}}{g_{o4} \left(1 + \frac{1}{2g_{m1}R_{CS}}\right)} + \frac{g_{m1}}{2g_{m3}} \left(1 + \frac{1}{g_{m2}R_{CS}}\right) + \frac{g_{m1}}{g_{m3}} \cdot \frac{g_{m2}}{g_{m2} \left(2 + \frac{1}{g_{m2}R_{CS}}\right)} \right] \\ &\approx \frac{g_{m1}}{g_{o4}} \end{aligned} \quad (\text{A5.16})$$

und die Gleichtaktverstärkung A_{cm}

$$\begin{aligned}
 A_{cm} &= \frac{A_{11} + A_{12} + A_{21} + A_{22}}{2} \\
 &= \frac{1}{2} \left[\frac{\frac{g_{m1}g_{m2}g_{m4}}{g_{m3} + g_{o3}} + \frac{g_{m1}g_{m4}}{g_{m3} + g_{o3}} \cdot \frac{1}{R_{CS}} + g_{m1}g_{m2}}{g_{o4} \cdot (g_{m1} + g_{m2} + 1/R_{CS})} - \frac{\frac{g_{m1}g_{m2}g_{m4}}{g_{m3} + g_{o3}} + \frac{g_{m2}}{R_{CS}} + g_{m1}g_{m2}}{g_{o4} \cdot (g_{m1} + g_{m2} + 1/R_{CS})}}{g_{m3} + g_{o3}} \cdot \frac{g_{m2} + 1/R_{CS}}{g_{m1} + g_{m2} + 1/R_{CS}} + \frac{g_{m1}}{g_{m3} + g_{o3}} \cdot \frac{g_{m2}}{g_{m1} + g_{m2} + 1/R_{CS}} \right] \quad (A5.17) \\
 &\approx \frac{1}{2} \left[\frac{\frac{g_{m1}g_{m4}}{g_{m3} + g_{o3}} \cdot \frac{1}{R_{CS}} - \frac{g_{m2}}{R_{CS}}}{g_{o4} \cdot (g_{m1} + g_{m2} + 1/R_{CS})} + \frac{g_{m1}}{g_{m3} + g_{o3}} \cdot \frac{g_{m2} - (g_{m2} + 1/R_{CS})}{g_{m1} + g_{m2} + 1/R_{CS}} \right] \\
 &= \frac{\frac{1}{R_{CS}} \left(\frac{g_{m1}g_{m4}}{g_{m3} + g_{o3}} - g_{m2} \right) (g_{m3} + g_{o3}) - \frac{g_{o4}g_{m1}}{R_{CS}}}{2g_{o4} \cdot (g_{m3} + g_{o3})(g_{m1} + g_{m2} + 1/R_{CS})} \approx \frac{\frac{g_{o4}g_{m1}}{R_{CS}}}{2g_{o4} \cdot g_{m3}(g_{m1} + g_{m2} + 1/R_{CS})} \\
 &\approx \frac{\frac{g_{m1}}{R_{CS}}}{4g_{m3}g_{m1}(1 + 1/2g_{m1}R_{CS})} = \frac{-1}{4g_{m3}(R_{CS} + 1/2g_{m1})} \\
 &\approx \frac{-1}{4g_{m3}R_{CS}}
 \end{aligned}$$

finden.

TABELLE 22. Ersatzschaltbildkomponenten der Differenzeingangsstufe

	Steilheit g_m [S]	Ausgangsleitwert g_o [S]
Eingangstransistoren 1 und 2 NMOS 50/10	$1.5 \cdot 10^{-4}$	$1.5 \cdot 10^{-6}$
Lasttransistoren 3 und 4 PMOS 50/10	$1.0 \cdot 10^{-4}$	$3.5 \cdot 10^{-6}$
Stromquellentransistor CS NMOS 160/10	$6.8 \cdot 10^{-4}$	$9.6 \cdot 10^{-5}$
	bei $I_B = 200 \mu A$	bei $I_B = 200 \mu A$

Den berechneten Werten liegt ein Biasstrom von $I_B = 200 \mu A$ sowie $g_{m,NMOS} = 1.5 \cdot 10^{-2} \cdot A^{1/2} / V_{I_0}^{1/2}$, $g_{o,NMOS} = 1.5 \cdot 10^{-2} \cdot V^{-1} \cdot I_0$, $g_{m,PMOS} = 1.0 \cdot 10^{-2} \cdot A^{1/2} / V_{I_0}^{1/2}$ und $g_{o,PMOS} = 3.5 \cdot 10^{-2} \cdot V^{-1} \cdot I_0$ für einen Transistor mit den Entwurfsmaßen $W_{nom} = 50 \mu m$ und $L_{nom} = 10 \mu m$ zu Grunde. Da sich im Arbeitspunkt der Biasstrom jeweils zur Hälfte auf beide Stränge verteilt, gilt $I_0 = 0.5 \cdot I_B = 100 \mu A$.

Für die Differenzverstärkung ergibt sich $A_{dm} = -1.5 \cdot 10^{-4} / 3.5 \cdot 10^{-6} = -42.8$ ($|A_{dm}| = 32.6$ dB). Aufgrund der Nichtberücksichtigung des endlichen Ausgangsleitwerts der Eingangstransistoren 1 und 2 liegt dieser Wert

geringfügig höher als der meßtechnisch ermittelte Wert einer 36-fachen Verstärkung. Der Wert der Gleichtaktverstärkung beträgt rechnerisch $A_{cm} = -9.6 \cdot 10^{-5} / (4 \cdot 1.0 \cdot 10^{-4}) = 0.24 \Rightarrow |A_{cm}| = -12.4 \text{ dB}$. Für die Gleichtaktunterdrückung folgt damit $CMRR = A_{dm} / A_{cm} = 45 \text{ dB}$. Das Meßergebnis liegt für niedrige Frequenzen bei $CMRR = 38 \text{ dB}$, die für höhere Frequenzen abfallende Leerlaufverstärkung sorgt für eine entsprechende Absenkung des CMRR.

TABELLE 23. Kennwerte der Differenzeingangsstufe

	Berechnet	SPICE-Simulation	Messung
Differenzverstärkung A_{dm} [dB]	32.6	30.3	31.1
Stromaufnahme [μA]	400	415	424
3 dB-Grenzfrequenz $f_{-3 \text{ dB}}$ [kHz]		27	28
bei 25 pF Lastkapazität			
0 dB-Grenzfrequenz $f_{0 \text{ dB}}$ [MHz]		0.86	0.81
bei 25 pF Lastkapazität			
Transitfrequenz f_T [MHz]		0.88	
bei 25 pF Lastkapazität			

Die Grenzfrequenz $f_{-3 \text{ dB}}$ bezeichnet die Frequenz, bei der die Leerlaufspannungsverstärkung $A_{dm}(f)$ auf das $1/2^{1/2}$ -fache abgesunken ist, $f_{0 \text{ dB}}$ bezeichnet den Punkt der einfachen Verstärkung und f_T berechnet sich gemäß $f_T = A_{dm} \cdot f_{-3 \text{ dB}}$.

Im Bild A5.8 sind die Meßdaten zusammenfassend dargestellt. Die Darstellung der Spannungsverstärkung (links oben im Bild) zeigt einen Spannungshub größer als $\pm 4 \text{ V}$ um den Nullpunkt. Die Versorgungsstrommessung darunter erfolgt kombiniert für die eigentliche Schaltung und den Versorgungszweig aus Widerstand und NMOS, daher ist ein konstanter Anteil von $I_B = 200 \mu\text{A}$ der Arbeitspunkteinstellung zuzuschreiben. Die geringe Schwankung des Versorgungsstroms zeigt die gute Qualität der Stromquelle, bedingt durch den geringen Ausgangsleitwert des Langkanal-NMOS.

Der Frequenzgang in der Bode-Plot-Darstellung (rechts) zeigt wiederum das Verhalten eines Tiefpasses mit einem einzigen dominanten Pol. Dementsprechend gut ist auch die Phasenreserve der Anordnung ($\Delta\phi = 95.3^\circ$ bei $f_{0 \text{ dB}} = 810 \text{ kHz}$), die in der Literatur geforderte Werte von 45° bis 60° klar übertrifft [A5.9], [A5.10], [A5.11]. Somit ließe sich diese Stufe auch mittels einer Rückkopplung betreiben.

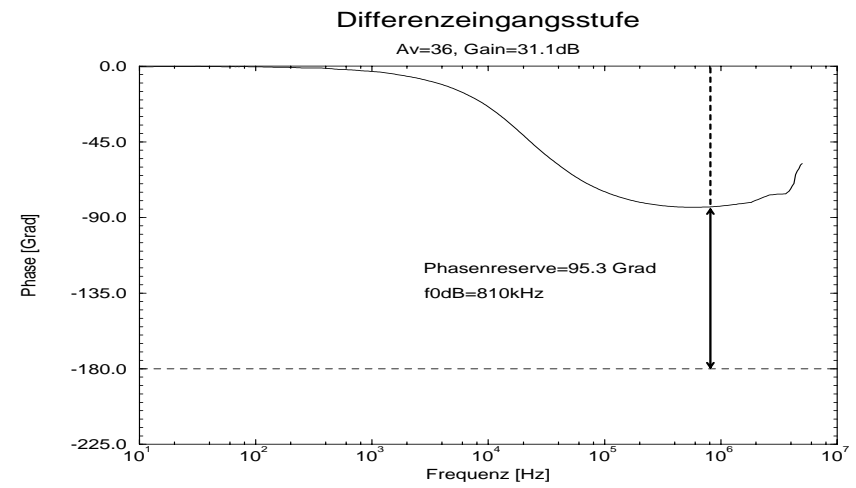
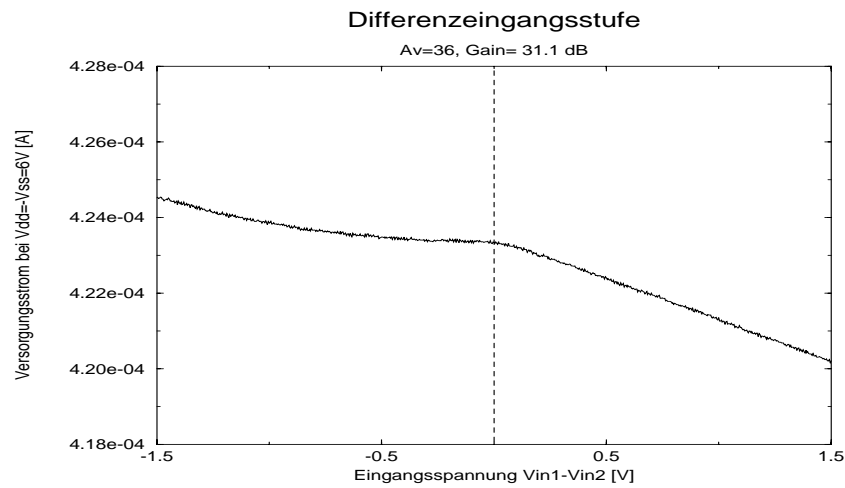
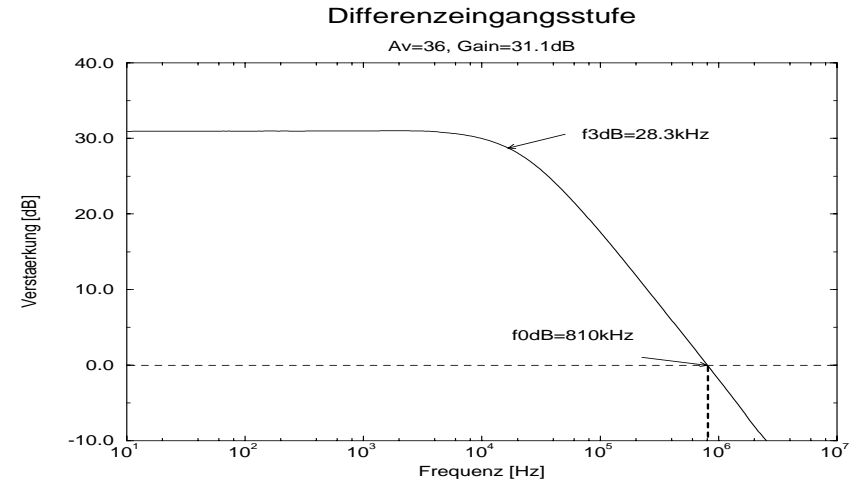
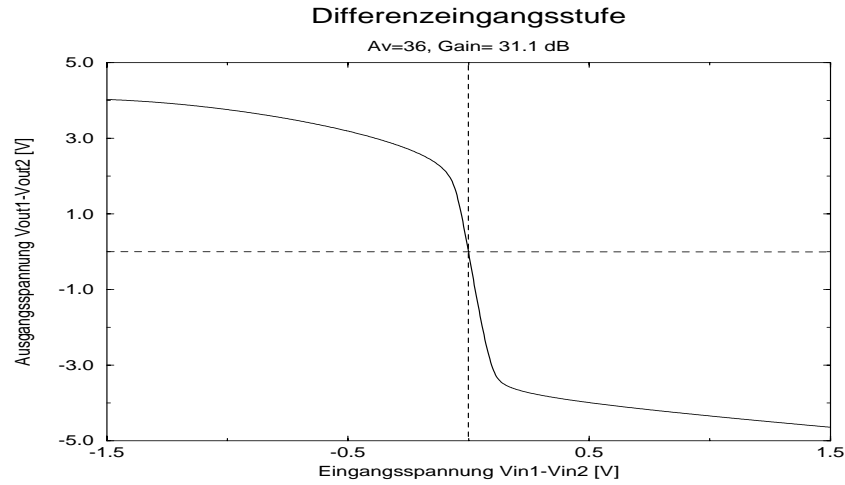


Bild A5.8: Meßkurven der Differenzeingangsstufe. Links Gleichspannungskennlinien, oben Spannungsverstärkung, unten Verlauf der Stromaufnahme der Schaltung. Rechts Bode-Plot des Frequenzgangs mit Betrag (oben) und Phase (unten).

Zweistufiger Differenzverstärker

Zur Erhöhung der Spannungsverstärkung der Differenzeingangsstufe kann eine zweite Stufe vorgesehen werden. Die Dimensionierung der Eingangsstufe ist beibehalten worden, um den Einfluß der zweiten Stufe zu verdeutlichen. Wie in Bild A5.9 dargestellt, erfolgt eine hochohmige Ankopplung der zweiten Verstärkerstufe, so daß außer einer geringen kapazitiven Belastung der ersten Stufe durch die Eingangskapazitäten der zweiten Stufe keine weitere Beeinflussung erfolgt. Die zweite Stufe ist ebenfalls differentiell aufgebaut, weist jedoch keine Konstantstromquelle auf. Der dadurch entfallende Spannungsabfall am Stromquellentransistor bedingt einen größeren Ausgangsspannungshub.

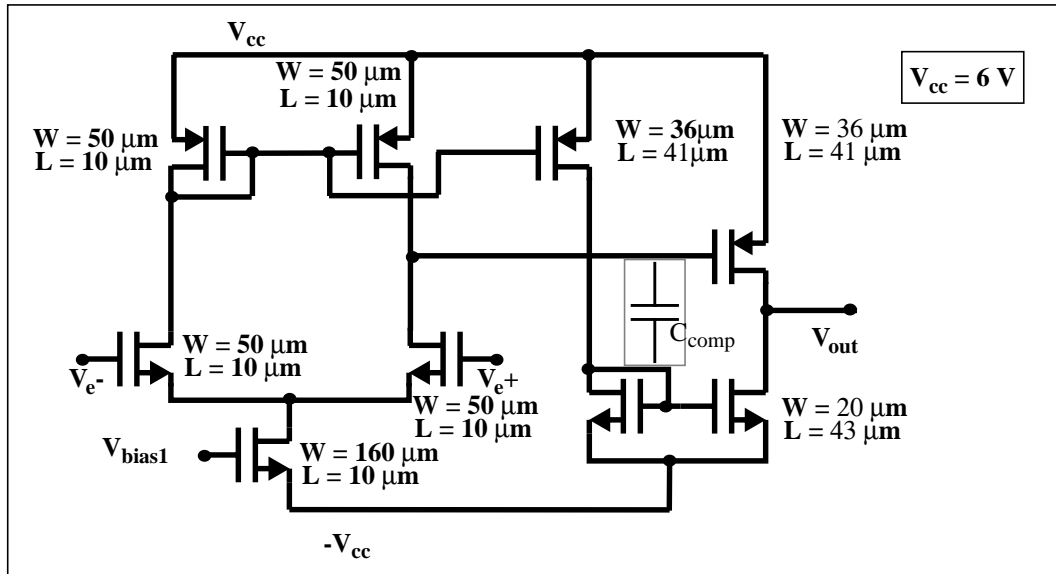


Bild A5.9: Schaltbild eines zweistufigen Verstärkers mit Differenzeingangsstufe und ebenfalls differentiell aufgebauter zweiter Verstärkerstufe.

Die Ermittlung der Spannungsverstärkung der Ausgangsstufe kann wie im Fall der Differenzeingangsstufe erfolgen, die Funktion der NMOS- und PMOS-Transistoren ist jedoch vertauscht, ferner entfällt der Beitrag der Stromquelle, d.h. der Einfluß des Widerstands R_{CS} .

Die Spannungsverstärkungen erhält man dementsprechend aus

$$[B] = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \lim_{R_{CS} \rightarrow 0} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \quad (A5.18)$$

Die Spannungsverstärkung der Gesamtanordnung ergibt sich aus dem Produkt der Einzelverstärkungen, da, wie oben dargelegt, eine gegenseitige ohmsche Beeinflussung der beiden Stufen ausgeschlossen werden kann. Damit wird

$$A_{dm} = \frac{C_{11} - C_{12} - C_{21} + C_{22}}{2} \quad \text{mit} \quad [C] = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = [B] \cdot [A]. \quad (A5.19)$$

Nach entsprechenden Vereinfachungen folgt

$$A_{dm} = \frac{g_{m, PMOS - Stufe2}}{g_{o, NMOS - Stufe2} + g_{o, PMOS - Stufe2}} \cdot \frac{g_{m, NMOS - Stufe1}}{g_{o, PMOS - Stufe1}} \quad (A5.20)$$

Aufgrund des deutlich höheren Ausgangsleitwerts des PMOS-Fet würde bei dessen Nichtberücksichtigung die Spannungsverstärkung der zweiten Stufe unrealistisch hohe Werte annehmen.

TABELLE 24. Ersatzschaltbildkomponenten der differentiellen Ausgangsstufe

	Steilheit g_m [S]	Ausgangsleitwert g_o [S]
Eingangstransistoren Stufe 2 PMOS 36/41	$7.8 \cdot 10^{-6}$	$2.9 \cdot 10^{-8}$
Lasttransistoren Stufe 2 NMOS 20/43	$6.0 \cdot 10^{-6}$	$6.0 \cdot 10^{-9}$

Den oben angegebenen Werten liegt ein Strangstrom der Ausgangsstufe von $I_o = (\text{Gesamtstrom minus Strom der Eingangsstufe})/2 = (525 \mu\text{A} - 425 \mu\text{A})/2 = 50 \mu\text{A}$ zugrunde. Für einen NMOS-Transistor mit den Entwurfsmaßen $W_{nom} = 20 \mu\text{m}$ und $L_{nom} = 43 \mu\text{m}$ gilt $g_{m, NMOS} = 8.5 \cdot 10^{-4} \cdot A^{1/2}/V_{I_o}^{1/2}$ sowie $g_{o, NMOS} = 1.2 \cdot 10^{-4} \cdot V^{-1} \cdot I_o$ und für einen PMOS mit den Maßen $W_{nom} = 36 \mu\text{m}$ und $L_{nom} = 41 \mu\text{m}$ gilt $g_{m, PMOS} = 1.1 \cdot 10^{-3} \cdot A^{1/2}/V_{I_o}^{1/2}$ und $g_{o, PMOS} = 5.7 \cdot 10^{-4} \cdot V^{-1} \cdot I_o$.

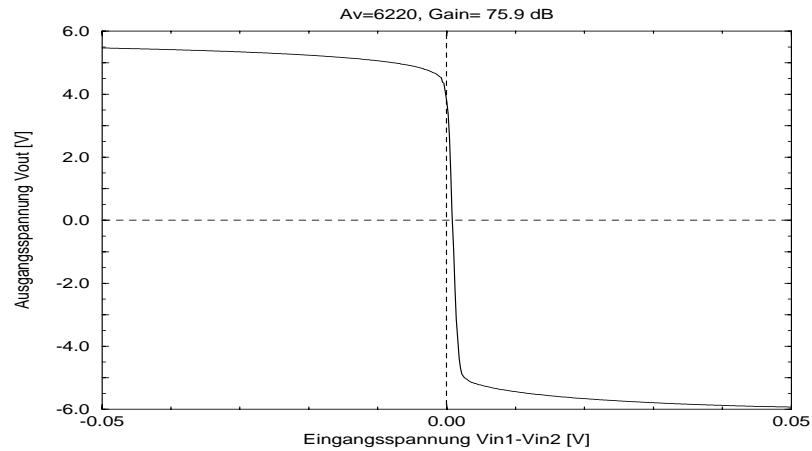
Die Spannungsverstärkung der zweiten Stufe ergibt sich zu $A_{dm2} = 222 \Rightarrow 47 \text{ dB}$. Meßtechnisch ergibt sich die Spannungsverstärkung der zweiten Stufe zu $A_{dm, Stufe2} = A_{dm, gesamt}/A_{dm, Stufe1} = 6220/36 = 172 \Rightarrow 44.7 \text{ dB}$.

Der Ausgangsspannungshub, wie in Bild A5.10 links oben dargestellt, erreicht die negative Versorgungsspannung und kommt bis auf ein halbes Volt an die positive Versorgungsspannung heran. Ein Abflachen der Ausgangskennlinie erfolgt etwa 1.5 Volt von der entsprechenden Versorgungsspannung entfernt, da jeweils ein Ausgangstransistor den Bereich der Sättigung verläßt und in den linearen Bereich übergeht. Die Stromaufnahme ist über den Aussteuerungsbereich recht konstant und liegt bei etwa $I_{gesamt, ss} = 525 \mu\text{A}$. Da die Differenzeingangsstufe für sich genommen schon eine Stromaufnahme von $I_{diff, ss} = 425 \mu\text{A}$ zeigt, verbleiben nur je etwa $50 \mu\text{A}$ für jeden Ausgangszweig. Damit ergibt sich die Notwendigkeit einer hochohmigen Ankopplung, da Impedanzen unterhalb von $R = 1 \text{ M}\Omega$ zu deutlichen Beeinträchtigungen der Spannungsverstärkung führen. Für den vorgesehenen Einsatz als Vorverstärker stellt dies allerdings keine signifikante Beschränkung dar.

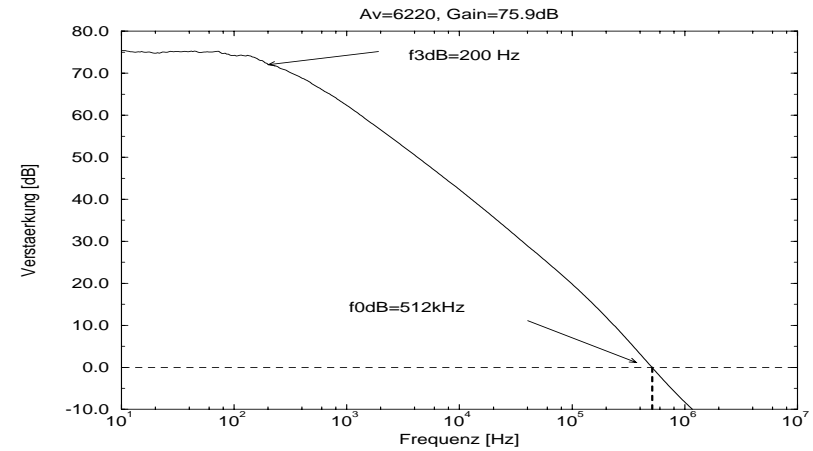
Frequenzgang und Phasenverlauf werden wie erwartet von der zweiten Stufe beeinflusst und der Pol der Ausgangsstufe sorgt für eine zusätzliche Phasendrehung schon unterhalb der 0dB-Grenzfrequenz. Bei $f_{0 \text{ dB}} = 512 \text{ kHz}$ verbleibt nur eine Phasenreserve von $\Delta\phi = 37.1^\circ$. Eine Rückkopplung ohne zusätzliche Maßnahmen wird daher leicht zu Instabilitäten führen. Aufgrund der nach außen geführten Anschlüsse ist eine kapazitive Kompensation des Verstärkers zwischen der ersten und zweiten Stufe möglich. Die im Schaltbild im C_{comp} bezeichnete Kompensationskapazität besorgt eine entgegengesetzte Phasendrehung durch Pole-Splitting [A5.11], so daß die Phasenreserve vergrößert wird. Die erforderlichen Kapazitätswerte liegen bei etwa $C_{comp} = 5 \text{ pF}$.

Die in Bild A5.11 ersichtlichen Meßkurven der Gleichtaktunterdrückung und Speisespannungsunterdrückung belegen die gute Wirksamkeit der differentiellen Auslegung der Schaltung. Das Gleichtaktunterdrückungsverhältnis von über 70 dB liegt deutlich über dem für die einfache Differenzeingangsstufe ermittelten. Mit Werten von 58 dB erreicht auch die Unterdrückung der Schwankungen der negativen Versorgungsspannung (PSRR⁻) sehr gute Werte, die von den 72 dB PSRR⁺ noch übertroffen werden.

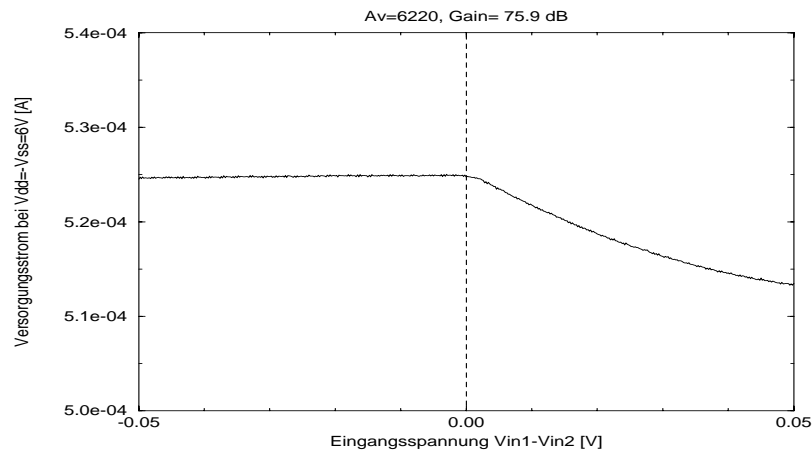
2-stufiger Verstärker mit Differenzeingang und Stromspiegel



2-stufiger Verstärker mit Differenzeingang und Stromspiegel



2-stufiger Verstärker mit Differenzeingang und Stromspiegel



2-stufiger Verstärker mit Differenzeingang und Stromspiegel

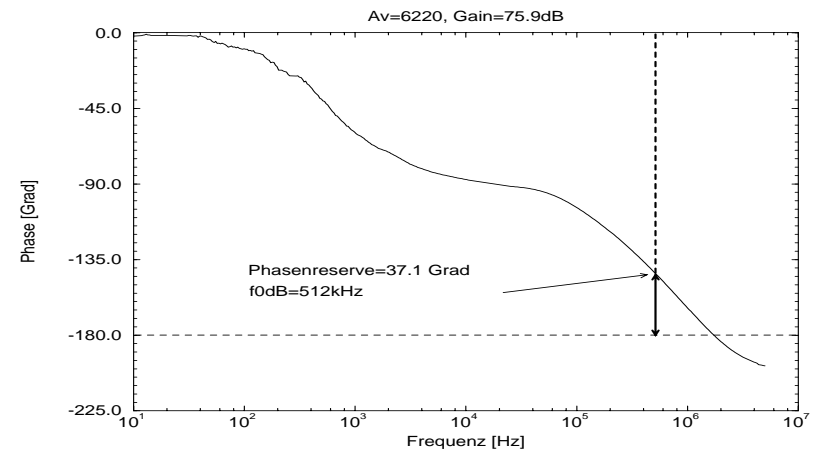


Bild A5.10: Meßkurven des zweistufigen Verstärkers. Links Gleichspannungskennlinien, oben Spannungsverstärkung, unten Verlauf der Stromaufnahme der Schaltung. Rechts Bode-Plot des Frequenzgangs mit Betrag (oben) und Phase (unten).

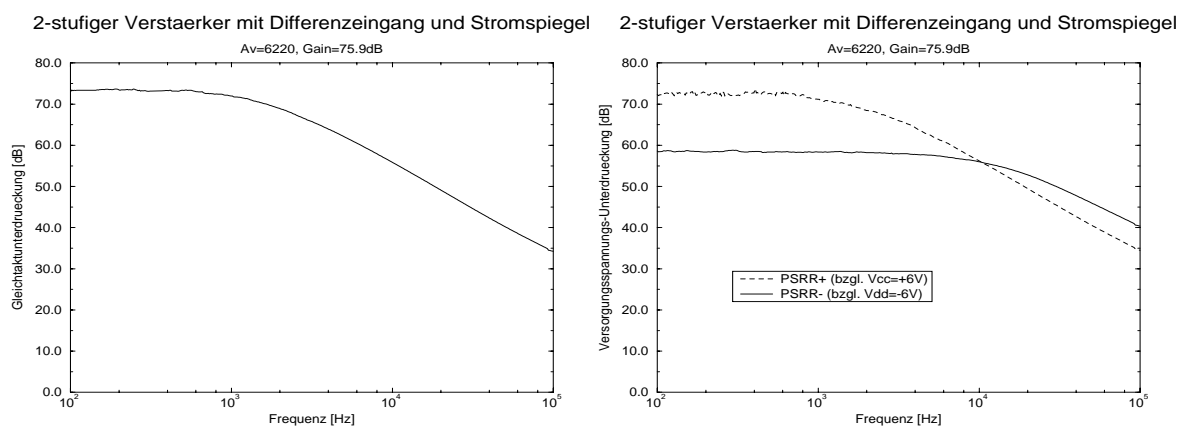


Bild A5.11: Links Gleichtaktunterdrückung (CMRR) und rechts Speisespannungsunterdrückung (PSRR) des zweistufigen Verstärkers.

In der Literatur [A5.12] werden für die Gleichtaktunterdrückung einer optimal realisierten CMOS-Differenzeingangsstufe $CMRR = 60$ dB angegeben. Werte bis 80 dB lassen sich mit aus Bipolartransistoren aufgebauten Eingangsstufen erhalten. Signifikanten Einfluß auf das maximal mögliche Unterdrückungsverhältnis haben laut Laker et al. [A5.12]

- (i) Unterschiede in den Ausgangsleitwerten der beiden Lasttransistoren Δg_o ,
- (ii) Unterschiede in den Schwellspannungen der beiden Eingangstransistoren ΔV_{th} ,
- (iii) und Unterschiede in den Steilheiten der beiden Eingangstransistoren Δg_m .

Schwankungen dieser Parameter sind durch folgende Maßnahmen zu minimieren:

- (i) Eine saubere und stabile Prozeßführung sowie
- (ii) Entwurfsregeln, die eine gegenüber Prozeßschwankungen weitgehend robuste Schaltung ermöglichen, so etwa durch eine Minimierung der geometrischen Abstände zwischen sich in beiden Zweigen entsprechenden Bauelementen.

Vergleich der Schaltungen in Standard-CMOS-Technik und in CMOS-Technik auf hochohmigem Substrat

Die auf den vorangegangenen Seiten dargestellten Verstärker-Kennlinien stammen alle von auf hochohmigem Substrat ($<100>$, Bor, float zone, $\rho = 5000 \Omega\text{cm}$) realisierten Schaltungen. Der Vergleich dieser Kennlinien mit solchen von auf Standard-CMOS-Substrat realisierten Schaltungen ($<100>$, Phosphor, Czochralski, $\rho = 4-6 \Omega\text{cm}$) zeigt keine merklichen Unterschiede in Bezug auf Niederfrequenzverstärkung, Offsetverhalten oder Frequenzgang.

A 5.3 Ansteuerschaltung für einen monolithisch integrierten pin-Schalter

In künftigen SIMMWIC-Schaltungen kommt neben einem Einsatz von CMOS-Schaltungen als signalverarbeitende Elemente auf Empfängerseite der Ansteuerung von HF-Bauelementen, wie etwa von IMPATT-Dioden [A5.13], besondere Bedeutung zu. Die CMOS-Technik stößt jedoch in diesem Anwendungsgebiet leicht an ihre Grenzen, da weder höhere Ströme noch höhere Spannungen klassische Domänen einer konventionellen CMOS-Technik sind. Die in dieser Arbeit entworfenen NMOS-Fets auf hochohmigem Substrat können hier aufgrund der Möglichkeit eines hohen negativen Substratpotentials zum Zuge kommen. Jedoch steht hier die CMOS-Technik in direkter Konkurrenz zur Bipolartechnologie, so daß insbesondere eine BiCMOS-Technik, also eine Kombination von CMOS- und Bipolartechnologie, überlegene Resultate bieten könnte. Der höhere technologische Aufwand einer solchen Technologie ist allerdings von Nachteil.

Sinnvoll ist die reine CMOS-Technik jedoch bei Anwendungen, die mit niedrigen Spannungen und mittleren Strömen auskommen, wie etwa bei einem pin-Hochfrequenz-Schalter in SIMMWIC-Technik. Mittels eines Vorwärtsstroms werden pin-Dioden in einem Zweig des Schalters in den leitfähigen Zustand gebracht und wirken für Hochfrequenzsignale kleiner Amplitude in erster Näherung als ein idealer Kurzschluß. Antiparallel dazu geschaltete Dioden im jeweils anderen Zweig des Schalters werden im Sperrbereich betrieben und beeinflussen

deshalb die Impedanz der Wellenleiterstruktur nur unmerklich. Der dynamische Hochfrequenzwiderstand einer solchen Diode ist im Durchlaßbereich invers proportional zum eingepprägten Vorwärtsstrom und bestimmt sich laut [A5.14] zu

$$R_{RF} = \rho \frac{x_l}{A} = \frac{x_l}{en_f(\mu_n + \mu_p)A} = \frac{x_l^2}{J_f \tau (\mu_n + \mu_p)A} = \frac{x_l^2}{I_f \tau (\mu_n + \mu_p)} \quad (A5.21)$$

Hierbei ist x_l die Länge der i-Zone, τ die Lebensdauer der Träger in der i-Zone, μ_n und μ_p deren Beweglichkeiten sowie I_f der Vorwärtsstrom. Die im Schalterentwurf eingesetzten pin-Dioden sind für einen Vorwärtsstrom von $I_f = 10 \text{ mA}$ ausgelegt, so daß aufgrund des Vorhandenseins von jeweils zwei pin-Dioden in beiden Zweigen ein Mindeststrom von $2 \cdot I_f = 20 \text{ mA}$ gefordert wird. Die Sperrströme der beiden antiparallel geschalteten Dioden liegen um mindestens 3-4 Größenordnungen darunter (vgl. Abs. 2.3.2).

Eine mögliche robuste Ansteuerschaltung besteht aus einem CMOS-Inverter, welcher aufgrund der hohen geforderten Ströme große Steilheiten g_m aufweisen muß. Dies erfordert große Ausgangstransistoren wie in Bild A5.12 dargestellt.

Um gleich hohe Ausgangsströme in beide Richtungen zu erreichen, ist eine Dimensionierung der Invertertransistoren erforderlich, die zu einem Schaltpunkt etwa in der Mitte der beiden Versorgungsspannungen führt. Das Interesse besteht in der Pegelumsetzung dieses Schaltpunkts auf TTL-Niveau, wie es von den üblicherweise eingesetzten Logikschaltkreis-Familien 74.., 74LS.., 74S.. usw. verfügbar ist. Die Definition der Logikpegel gemäß [A5.15]

$$\begin{aligned} 0V \leq U \leq 0,8V & \quad \text{LowPegel}(L) \\ 2,0V \leq U \leq 5,0V & \quad \text{HighPegel}(H) \end{aligned}$$

bedingt eine steile Schaltflanke oberhalb von 0.8 V und unterhalb von 2.0 V. Die Umsetzung des Inverterschaltpunkts bei $U = 0 \text{ V}$ erfolgt durch eine vorgeschaltete PMOS-Sourcestufe, die neben der Pegelumsetzung eine weitere Erhöhung der Schaltflankensteilheit bewirkt und somit einer alternativ möglichen Sourcefolgerstufe, die eine reine Pegelumsetzung und Impedanzwandlung [A5.16] vornimmt, überlegen ist. Die Sourcestufe wird durch einen Stromspiegel auf einen Arbeitspunkt von $I_B = 250 \mu\text{A}$ eingestellt.

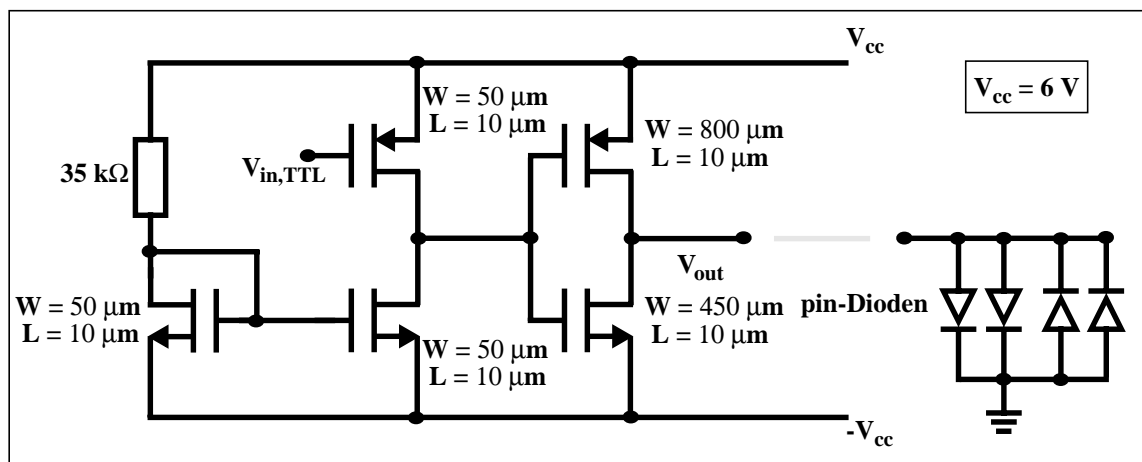


Bild A5.12: Ansteuerschaltung für den integrierten pin-Schalter. Die Inverterschaltung dient als verstärkende Endstufe zur Ansteuerung der pin-Dioden und erlaubt Diodenströme bis über 20 mA in beide Richtungen. Davor ist die Sourcestufe zur Pegelumsetzung und weiteren Verstärkung dargestellt. Der Ruhestrom der aktiven Last der Sourcestufe wird mittels eines Stromspiegels eingestellt.

Bild A5.13 zeigt die dem Entwurf zugrundeliegenden SPICE3 Simulationsergebnisse.

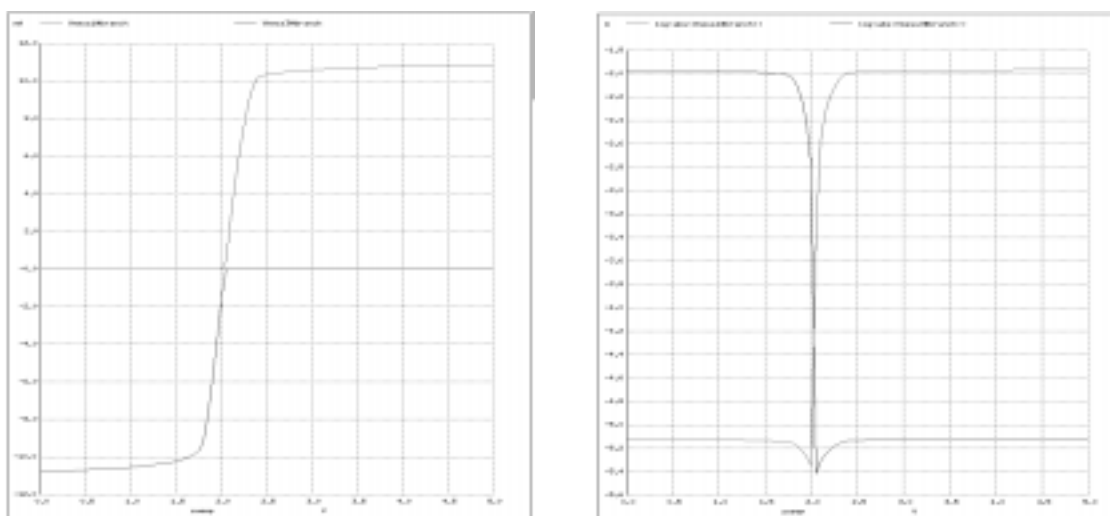
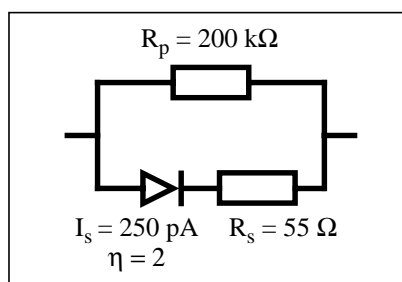


Bild A5.13: Die Simulation der Treiberstufe mittels SPICE 3 zeigt die Anforderungen an den Ausgangsstrom. In jedem Strang bestehend aus zwei parallel geschalteten pin-Dioden sollen mehr als 20 mA Vorwärtsstrom fließen können. Links: Lineare Darstellung des Stroms. Rechts: Logarithmische Darstellung des Strombetrags.

Links im Bild ist der linear aufgetragene Strom durch ein antiparalleles pin-Diodenpaar gezeigt. Es wird ein Schwellenwert bei etwa $U = 2$ V erreicht, ferner fließt durch jede Diode ein Vorwärtsstrom von mindestens $I_f = 10$ mA. Rechts in der Darstellung des Absolutbetrags des Stroms ist auch der Sperrstrom der pin-Dioden gezeigt. Als Simulationsmodell der pin-Diode, die nicht im Umfang der SPICE3-Software vorgesehen ist, wurde ein modifiziertes pn-Dioden-Modell eingesetzt.



Dieses Diodenmodell erlaubt eine weitgehende Simulation der pin-Diodencharakteristik und modelliert sowohl die Rekombinationsstrom-Anteile im Vorwärtsbereich als auch den Sperrstrom sehr genau.

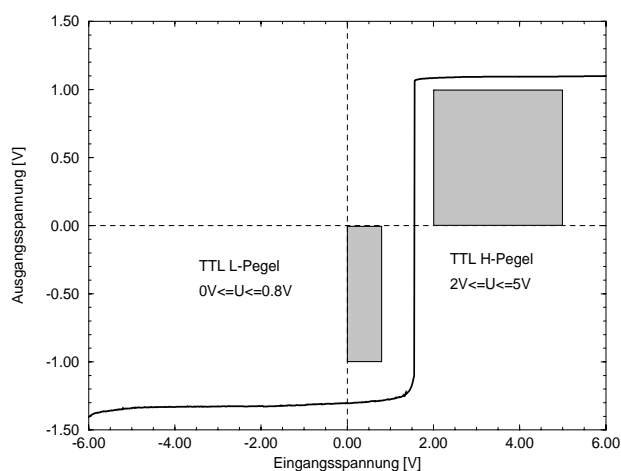


Bild A5.14: Messung der Ausgangskennlinie bei Belastung des Ausgangs mit antiparallel geschalteten pin-Dioden. Der interessierende Bereich der Eingangsspannungen zwischen 0 und 5 Volt entspricht TTL-Signalen, deren Schwellen im Diagramm markiert sind.

Das Meßergebnis der an den pin-Dioden anliegenden Ausgangsspannung ist in Bild A5.14 dargestellt. Die Dioden werden mit einer Vorwärtsspannung von etwas mehr als 1 Volt beaufschlagt, die bei den antiparallel geschalteten Dioden als Sperrspannung anliegt. Die im Bild grau dargestellten Bereiche der TTL-Logik-Pegel werden sicher eingehalten; die Schaltspannung liegt bei $U = 1.6 \text{ V}$. Während bei unbelastetem Betrieb die Ausgangsspannungen bis $U_a = -5.95 \text{ V}$ und $U_a = +5.80 \text{ V}$ reichen, wird durch die pin-Dioden eine Begrenzung auf unter $U_a = \pm 1.5 \text{ V}$ bewirkt.

Die Ausgangsströme der Schaltstufe, wie in Bild A5.15 aufgezeichnet, liegen deutlich über den geforderten 10 mA pro Diode. Insgesamt erreicht die Schaltung einen positiven Ausgangsstrom $I_A > 37 \text{ mA}$ für TTL-Low-Pegel und $I_A = -24 \text{ mA}$ bei TTL-High-Pegel. Zusätzlich dargestellt ist der Verlauf des positiven und negativen Versorgungsstroms bei $V_{cc} = 6 \text{ V}$. Aus diesen Kurven wird ersichtlich, daß die Schaltung selbst nur eine Stromaufnahme von 1 mA aufweist.

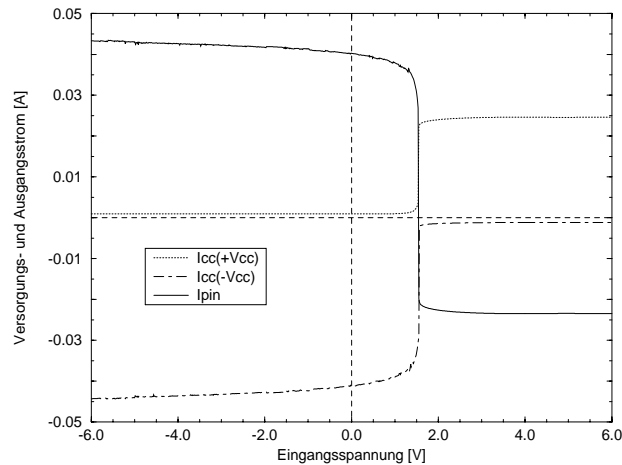


Bild A5.15: Darstellung des gemessenen positiven und negativen Versorgungsstroms, sowie des Ausgangsstroms durch die Dioden.

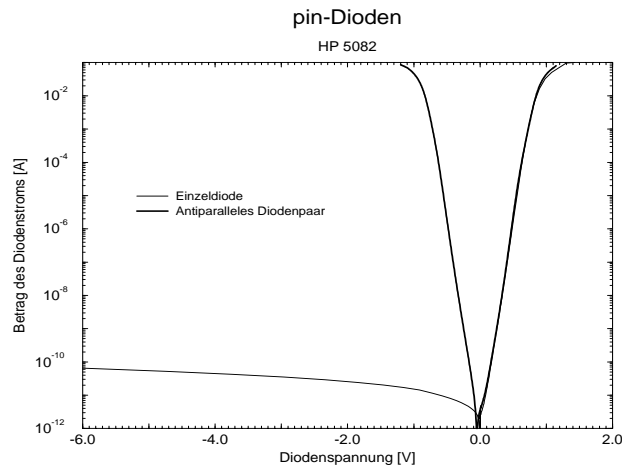


Bild A5.16: Strom-Spannungskennlinie einer einzelnen pin-Diode, bzw. zweier antiparallel geschalteter wie im Fall des pin-Schalters. Zur logarithmischen Darstellung sind die Beträge der Ströme aufgetragen worden.

Die Kennlinien der zur Messung als Last eingesetzten diskreten pin-Dioden sind in Bild A5.16 dargestellt. Im Vorwärtsbetrieb entsprechen diese Dioden dem für die Simulation eingesetzten Modell und damit den am Institut hergestellten Bauelementen, jedoch zeigen diese Dioden einen um einige Größenordnungen geringeren Sperrstrom. Dies ist bedingt durch die unterschiedliche Bauweise der Dioden. Während die diskreten Dioden Vertikalstrukturen aufweisen, die ein inhärent besseres Sperrstromverhalten bewirken, müssen die im pin-Schalter benutzten Dioden lateral aufgebaut werden und zeigen dementsprechend stärkere Oberflächengeneration und damit höhere Sperrströme.

Literaturangaben:

- [0.1] J.-F. Luy, K. M. Strohm, J. Büchler, und P. Russer, "Silicon monolithic millimetre-wave integrated circuits," *IEE Proceedings-H*, vol. 139, pp. 209-216, 1992.
- [0.2] E. Kasper, H. Kibbel, H.-J. Herzog, und A. Gruhle, "Growth of 100 GHz SiGe-heterobipolar transistor (HBT) structures," *Japan. Journ. Appl. Phys.*, vol. 33, pp. 2415-2418, 1994.
- [0.3] J.-F. Luy, K. M. Strohm, E. Sasse, A. Schüppen, J. Büchler, M. Wollitzer, A. Gruhle, F. Schäffler, U. Guettich, und A. Klaaßen, "Si/SiGe MMICs," *IEEE Trans. Microwave Theory Tech.*, vol. 43, pp. 705-714, 1995.
- [0.4] J. F. Luy, und P. Russer, *Silicon-Based Millimeter-Wave Devices*, Berlin: Springer, Kap. 2, 1994.
- [0.5] J. R. Brews in S. M. Sze, *High-speed semiconductor devices*, New York: Wiley, Kap. 3, 1990.
- [0.6] C. Y. Chang, und S. M. Sze, *ULSI technology*, New York: McGraw-Hill, 1995.
-
- [1.1] Nikkei, *Microdevices*, vol. 5, pp. 54-62, 1990.
- [1.2] S. K. Ghandhi, *VLSI fabrication principles - silicon and gallium arsenide*, 2. Aufl., New York: Wiley, 1994.
- [1.3] W. Kern, und D. A. Poutinen, "Cleaning solutions based on hydrogen peroxide for use in silicon semiconductor technology," *RCA Rev.*, vol. 31, pp. 187ff., 1970.
- [1.4] M. Watanabe, M. Harazono, Y. Hiratsuka, und T. Edamura, *Electrochem. Soc. Ext. Abstracts*, pp. 221-222, 1983.
- [1.5] D. S. Becker, W. R. Schmidt, C. A. Peterson, und D. Burkman, *Microelectronics processing, inorganic materials characterization*, Kap. 23, American Chemical Society Symp. Series No. 295, pp. 368-376, 1986.
- [1.6] E. Kasper, *Persönliche Mitteilung*, Stuttgart, 1995.
- [1.7] E. H. Nicollian, und J. R. Brews, *MOS - Physics and technology*, New York: Wiley, 1982.
- [1.8] B. E. Deal, und A. S. Grove, "General relationship for the thermal oxidation of silicon," *Jour. Appl. Phys.*, vol. 36, pp. 3770ff., 1965.
- [1.9] M. Ghezze, und D. M. Brown, "Diffusivity summary of B, Ga, P, As, and Sb in SiO₂," *Jour. Electrochem. Soc.*, vol. 120, pp. 146ff., 1973.
- [1.10] J. Mezger, "Untersuchungen für einen hochohmigen CMOS-Prozeß," *Studienarbeit am Institut für Halbleitertechnik*, 1994.
- [1.11] H. F. Wolf, *Silicon semiconductor data*, New York: Pergamon Press, 1969.
- [1.12] T. Y. Tan, E. E. Gardner, und W. K. Tice, "Intrinsic gettering by oxide precipitate induced dislocations in Czochralski Si," *Appl. Phys. Lett.*, vol. 30, pp. 175ff., 1977.
- [1.13] S. M. Sze, *VLSI technology*, 2. Aufl., New York: McGraw-Hill, 1988.
- [1.14] C. Y. Chang, und S. M. Sze, *ULSI technology*, New York: McGraw-Hill, 1995.

-
- [1.15] M. Nägele, "Prozeßdokumentation Standard-Al-Gate-CMOS", Stuttgart: Institut für Halbleitertechnik, 1994.
- [1.16] ASTM-Norm F84-93, "Standard test method for measuring resistivity of silicon wafers with an in-line four-point probe," *American Society for Testing and Materials*, 1993.
- [1.17] S. Pfüller, *Halbleitermeßtechnik*, Heidelberg: Alfred Hüthig, Kap. 4, 1977.
- [1.18] C. Bulucea, "Recalculation of Irvin's resistivity curves for diffused layers in silicon using updated resistivity data," *Solid State Electronics*, vol. 36, pp. 489ff., 1993.
- [1.19] G. Baccarani, und P. Ostoja, "Electron mobility empirically related to the phosphorus concentration in silicon," *Solid State Electr.*, vol. 18, pp. 579ff., 1975.
- [1.20] D. A. Antoniadis, A. G. Gonzales, und R. W. Dutton, "Boron in near-intrinsic <100> and <111> silicon under inert and oxidizing ambients - diffusion and segregation," *Journ. Electrochem. Soc.*, vol. 125, pp. 813ff., 1978.
- [1.21] I. Ruge, *Halbleiter-Technologie*, Kap. 6, Berlin: Springer, 1975.
- [1.22] *Semiconductor process development and monitoring using spreading resistance profiling*, Pittsburgh: SSM, 1987.
- [1.23] W. Wiedemann "Herstellung und Untersuchung von Bor-Dotierungsprofilen in Si-n-Substrat," *Diplomarbeit am Institut für Halbleitertechnik*, pp. 14 ff, 1995.
- [1.24] ASTM-Norm F674-92, "Standard practice for preparing silicon for spreading resistance measurements," *American Society for Testing and Materials*, 1992.
- [1.25] R. G. Mazur, und D. H. Dickey "A spreading resistance technique for resistivity measurements on silicon," *Journ. Electrochem. Soc.*, vol. 113, pp. 255ff., 1966.
- [1.26] S. R. Weinzierl, J. M. Heddleson, R. J. Hillard, P. Rai-Choudhury, R. G. Mazur, C. M. Osburn, und P. Potyraj, "Ultrashallow dopant profiling via spreading resistance measurements with integrated modeling," *Solid State Techn.*, 1993.
- [1.27] P. A. Schumann, und E. E. Gardner, "Application of multilayer potential distribution to spreading resistance correction factors," *Journ. Electrochem. Soc.*, vol. 116, pp. 87-91, 1969.
- [1.28] S. C. Choo, Beitrag bei *The First International Workshop on the Measurement and Characterization of Ultra-Shallow Doping Profiles in Semiconductors*, Research Triangle Park, North Carolina, 1991.
- [1.29] S. M. Hu, "Between carrier distributions and dopant atomic distribution in beveled silicon substrates," *Journ. Appl. Phys.*, vol. 53, pp. 1499-1510, 1982.
- [1.30] Analysepaket zum SSM 150 Spreading Resistance System von Solid State Measurements, Inc., USA.
- [1.31] L. C. Feldman, und J. W. Mayer, *Fundamentals of surface and thin film analysis*, New York: North Holland, 1986.

-
- [2.1] G. Vanstraelen, E. Simoen, C. Claeys, und G. J. Declerck, "Short-Channel pMOST's in a high resistivity silicon substrate - I: Analytical model," *IEEE Trans. on Electron Devices*, vol. 39, pp. 2268-2276, 1992.
- [2.2] J. R. Brews, W. Fichtner, E. H. Nicollian, und S. M. Sze, "Generalized guide for MOSFET miniaturization," *IEEE Electron Device Letters*, vol. 1, pp. 2ff, 1980.
- [2.3] C. Hu, "MOSFET scaling in the next decade and beyond," *Semiconductor International*, pp. 105-114, Juni 1994.
- [2.4] S. M. Sze, *Physics of semiconductor devices*, 2nd ed., New York: Wiley, 1981.
- [2.5] B. Eitan, und D. Frohman-Bentchkowsky, "Surface conduction in short-channel MOS devices as a limitation to VLSI scaling," *IEEE Trans. Electron Devices*, vol. 29, pp. 254-266, 1982.
- [2.6] N. F. Mott, und R. W. Gurney, *Electronic Process in Ionic Crystals*, 2nd ed., Oxford: Clarendon, Kap. 1, 1948.
- [2.7] F. C. Hsu, R. S. Muller, und C. Hu, "A simplified model of short-channel MOSFET characteristics in the breakdown mode," *IEEE Trans. on Electron Devices*, vol. 30, pp. 571-576, 1983.
- [2.8] F. C. Hsu, R. S. Muller, C. Hu, und P. K. Ko, "A simple punchthrough model for short-channel MOSFET's," *IEEE Trans. on Electron Devices*, vol. 30, pp. 1354-1359, 1983.
- [2.9] H. Wang, "A physically-based MOS transistor avalanche breakdown model," *IEEE Trans. on Electron Devices*, vol. 42, pp. 2197-2202, 1995.
- [2.10] A. Schutz, S. Selberherr, und H. W. Potzl, "A two-dimensional model of the avalanche effect in MOS transistors," *Solid-State Electronics*, vol. 25, pp. 177-183, 1982.
- [2.11] S. M. Sze, *Physics of semiconductor devices*, 2nd ed., New York: Wiley, 1981.
- [2.12] J. W. Slobboom, G. Streutker, G. J. T. Davids, und P. B. Hartog, "Surface impact ionization in silicon devices," *IEEE IEDM Technical Digest*, p. 494, 1987.
- [2.13] W. von Münch, *Skriptum zur Vorlesung - Halbleitertechnik I.*, Universität Stuttgart, Kap. 2, 1992.
- [2.14] S. E. Hansen, und M. D. Deal, *SUPREM-IV.GS User Manual*, Stanford University, 1993.
- [2.15] M. Ismail, und T. Fiez, *Analog VLSI - Signal and information processing*, New York: McGraw-Hill, Kap. 2, 1994.
- [2.16] R. Paul, *MOS-Feldeffekttransistoren*, Berlin: Springer, Kap. 2, 1994.
- [2.17] J. R. Brews, "Subthreshold behavior of uniformly and nonuniformly doped long-channel MOSFET," *IEEE Trans. Elec. Dev.*, vol. 26, no. 9, pp. 1282-1291, 1979.
- [2.18] S. M. Sze, *VLSI Technology - 2nd edition*, New York: McGraw-Hill, Kap. 9, 1988.
- [2.19] J. Mezger, "Untersuchungen für einen hochohmigen CMOS-Prozeß," *Studienarbeit am Institut für Halbleitertechnik*, pp. 55 ff, 1994.
- [2.20] B. E. Deal, E. L. MacKenna, und P. L. Castro, *Journal Electrochemical Society*, vol. 116, pp. 997ff., 1969.
- [2.21] *Semiconductor process development and monitoring using CV/IV and spreading resistance profiling*, Pittsburgh: SSM, 1995.

-
- [2.22] H. Jorke, *Persönliche Mitteilung*, Ulm: Daimler-Benz, 1996.
- [2.23] K. K. Ng, *Complete guide to semiconductor devices*, New York: McGraw-Hill, Kap. 1, 1995.
- [2.24] A. Möschwitzer, und K. Lunze, *Halbleiterelektronik*, 4. Aufl., Heidelberg: Hüthig, Kap. 1, 1980.
- [2.25] K. H. Löcherer, *Halbleiterbauelemente*, Stuttgart: Teubner, Kap. 1, 1992.
- [2.26] E. Kasper, *Persönliche Mitteilung*, Stuttgart: Institut für Halbleitertechnik, März 1996.
- [2.27] Z. Yu, D. Chen, L. So, und R. W. Dutton, *Two dimensional device simulation for silicon and hetero-structures*, Stanford University, 1994.
- [2.28] S. M. Sze, und G. Gibbons, "Avalanche breakdown voltages of abrupt and linearly graded p-n junctions in Ge, Si, GaAs, and GaP," *Appl. Phys. Letters*, vol. 8, pp. 111 ff, 1966.
- [2.29] Z. Li, und H. W. Kraner, "Gettering in high resistive float zone silicon wafers for silicon detector applications," *IEEE Trans. on Nuclear Science*, vol. 36, pp. 290-294, 1989.
- [2.30] G. Zimmer, *CMOS-Technologie*, München: Oldenbourg, Kap. 2, 1982.
- [2.31] A. Herlet, "The forward characteristic of silicon power rectifiers at high current densities," *Solid-State Electronics*, vol. 11, pp. 717ff, 1968.
- [2.32] P. D. Taylor, *Thyristor design and realization*, New York: Wiley, 1987.
- [2.33] S. C. Choo, "Effect of carrier lifetime on the forward characteristics of high-power devices," *IEEE Trans. Electron. Dev.*, vol. 17, pp. 647ff, 1970.
- [2.34] A. G. Milnes, *Semiconductor devices and integrated electronics*, New York: Van Nostrand, 1980.
- [2.35] A. Klaasen, H.-E. Sasse, "E-Band monolithic Si pin diode switches", Proc. IEEE Workshop on Silicon based high freq. dev. and circ., Günzburg, 1994.
- [3.1] M. Herrmann, E. Kasper, und D. Beck, *Abschlußbericht BMBF Vorhaben Integrationstechnik Silizium/Silizium-Germanium Millimeterwellen-IC (SiGe-SIMMWIC) - Teilprojekt „SIMMWIC Empfänger-schaltung in CMOS-Technik“*, 1996.
- [3.2] J. F. Luy, und P. Russer, *Silicon-based millimeter-wave devices*, Berlin: Springer, 1994.
- [3.3] J. Büchler, E. Kasper, P. Russer, und K. M. Strohm, "Silicon high-resistivity-substrate millimeter-wave technology," *IEEE Trans. MTT*, vol. 34, pp. 1516-1521, 1986.
- [3.4] K. M. Strohm, J. Büchler, J. F. Luy, und F. Schäffler, "A silicon technology for active high frequency circuits," *Microelectronic Engineer.*, vol. 19, pp. 717-720, 1992.
- [3.5] K. Strohm, F. Beißwanger, H. Jorke, und J. F. Luy, *3. technischer Zwischenbericht BMBF Vorhaben Integrationstechnik Silizium/Silizium-Germanium Millimeterwellen-IC (SiGe-SIMMWIC)*, 1995.
- [3.6] M. Herrmann, D. Beck, E. Kasper, J.-F. Luy, K. Strohm, und J. Büchler, "Hybrid 90 GHz rectenna chip with CMOS preamplifier," *ESSDERC 1996*, Bologna, 1996.

-
- [4.1] J.-F. Luy, K. M. Strohm, E. Sasse, A. Schüppen, J. Büchler, M. Wollitzer, A. Gruhle, F. Schäffler, U. Guettich, und A. Klaaßen, "Si/SiGe MMICs," *IEEE Trans. Microwave Theory Tech.*, vol. 43, pp. 705-714, 1995.
- [4.2] J. F. Luy, und P. Russer, *Silicon-based millimeter-wave devices*, Berlin: Springer, 1994.
- [4.3] V. Dudek, W. Appel, L. Beer, G. Digele, und B. Höfflinger, „Lithography-independent nanometer silicon MOSFET's on insulator," *IEEE Trans. Electron Devices* , vol. 43, 1996.
- [4.4] G. Wöhl, *Arbeitsanleitung Plasmaätzanlage*, interne Mitteilung, Stuttgart: IHT, 1995.
- [4.5] E. H. Nicollian, und J. R. Brews, *MOS - Physics and technology*, New York: Wiley, 1982.
- [4.6] B. E. Deal, E. H. Snow, und C. A. Mead, *J. Phys. Chem. Solids* , vol. 27, pp. 1873ff., 1966.
- [4.7] B. J. Gordon, "CV plotting: Myths and methods," *MOS - Solid State Technology*, p. 57-61, 1993.
- [4.8] *Semiconductor process development and monitoring using CV/IV and spreading resistance profiling*, Pittsburgh: SSM, 1995.
- [4.9] J. F. Luy, K. M. Strohm, H. E. Sasse, A. Schüppen, J. Büchler, M. Wollitzer, A. Gruhle, F. Schäffler, U. Guettich, und A. Klaaßen, "Si/SiGe MMIC's," *IEEE Trans. MTT*, vol. 43, pp. 705-714, 1995.
- [4.10] M. Herrmann, E. Kasper, und D. Beck, *Abschlußbericht BMBF Vorhaben Integrationstechnik Silizium/Silizium-Germanium Millimeterwellen-IC (SiGe-SIMMWIC) - Teilprojekt „SIMMWIC Empfänger-schaltung in CMOS-Technik“*, 1996.
- [4.11] K. Strohm, *Persönliche Mitteilung*, Ulm: Daimler-Benz, 1996.
- [4.12] I. J. Bahl, und R. Garg, "Simple and accurate formulas for microstrip with finite strip thickness," *Proc. IEEE*, vol. 65, pp. 1611-1612, 1977.
- [4.13] S. Ramo, J. R. Whinnery, und T. van Duzer, *Fields and waves in communication electronics*, 2. Aufl., New York: Wiley, 1984.
- [4.14] R. A. Pucel, D. J. Massé, und C. P. Hartwig, "Losses in microstrip," *IEEE Trans. MTT*, vol. 16, pp. 342-350, 1968.
- [4.15] K. C. Gupta, R. Garg, und I. J. Bahl, *Microstrip lines and slotlines*, Boston: Artech, 1979.
- [4.16] H. A. Wheeler, "Transmission-line properties of parallel wide strips by a conformal-mapping approximation," *IEEE Trans. MTT.*, vol. 12, pp. 280-289, 1964.
- [4.17] H. A. Wheeler, "Transmission-line properties of parallel strips separated by a dielectric sheet," *IEEE Trans. MTT.*, vol. 13, pp. 172-185, 1965.
- [4.18] J. D. Welch, und H. J. Pratt, "Losses in microstrip transmission systems for integrated microwave circuits," *NEREM Rec.*, vol. 8, pp. 100-101, 1966.
- [4.19] R. Horton, B. Easter, und A. Gopinath, "Variation of microstrip losses with thickness of strip," *Electron. Lett.*, vol. 17, pp. 490-491, 1971.
- [4.20] C. P. Wen, "Coplanar waveguide: A surface strip transmission line suitable for non-reciprocal gyro-magnetic device application," *IEEE Trans. MTT*, vol. 17, pp. 1087-1090, 1969.

-
- [5.1] D. Beck, M. Herrmann, und E. Kasper, "CMOS on FZ-high resistivity substrate for monolithic integration of SiGe-RF-circuitry and readout electronics," *IEEE Trans. Electron. Dev.*, vol. 44, pp. 1091-1101, 1997.
- [5.2] D. Beck, und E. Kasper, "Influence of CMOS-circuit areas on RF-damping of gold and aluminium microstripline in combined SIMMWIC-CMOS technology," *ESSDERC 1997*, Stuttgart, 1997.
- [A1.1] S. M. Sze, *VLSI Technology - 2nd edition*, New York: McGraw-Hill, Kap. 4, 1988.
- [A1.2] M. R. Haskert, und I. C. May, *Analog VLSI design*, New York: Prentice Hall, 1988.
- [A1.3] E. A. Irene, und D. W. Dong, "Silicon oxidation studies: The oxidation of heavily B- and P-doped single crystal silicon", *J. Electrochem. Soc.*, vol. 125, pp. 1146ff, 1978.
- [A1.4] C. P. Ho, J. D. Plummer, J. D. Meindl, und B. E. Deal, "Thermal oxidation of heavily phosphorus-doped silicon", *J. Electrochem. Soc.*, vol. 125, pp. 665ff, 1978.
- [A1.5] K. K. Ng, *Complete guide to semiconductor devices*, New York: McGraw-Hill, Kap. 1, 1995.
- [A1.6] G. Massobrio, und P. Antognetti, *Semiconductor device modeling with SPICE*, 2. Auflage, New York: McGraw-Hill, 1988.
- [A2.1] R. Kleinöder, *Einführung in die Netzwerkanalyse mit SPICE*, Stuttgart: Teubner, 1993.
- [A2.2] D. Ehrhardt, und J. Schulte, *Simulieren mit PSPICE - 2. Auflage*, Braunschweig: Vieweg, 1995.
- [A2.3] H. Khakzar, A. Mayer, und R. Oetinger, *Entwurf und Simulation von Halbleiterschaltungen mit SPICE*, Ehningen: Expert, 1991.
- [A2.4] G. Massobrio, und P. Antognetti, *Semiconductor device modelling with Spice*, New York: McGraw-Hill, Kap. 4 und 6, 1993.
- [A2.5] R. V. Booth, M. H. White, H. S. Wong, und T. J. Krutsick, "The effect of channel implants on MOS transistor characterization," *IEEE Trans. on Electron Devices*, vol. 34, pp. 2501-2508, 1987.
- [A2.6] C. G. Sodini, T. W. Ekstedt, und J. L. Moll, "Charge accumulation and mobility in thin dielectric MOS transistors," *Solid-State Electronics*, vol. 25, pp. 831-836, 1982.
- [A2.7] A. B. Fowler, und A. M. Hartstein, "Techniques for determining threshold," *Surface Science*, vol. 98, pp. 169-175, 1980.
- [A2.8] "Standard test method for measuring MOSFET linear threshold voltage," *ASTM Norm F 617-91*, pp. 266-271, 1991.
- [A3.1] E. H. Nicollian, und J. R. Brews, *MOS - Physics and technology*, New York: Wiley, 1982.
- [A3.2] R. Lindner, *Bell Syst. Tech. J.*, vol. 41, pp. 803ff, 1962.

-
- [A3.3] B. E. Deal, "Standardized terminology for oxide charges associated with thermally oxidized silicon," *IEEE Trans. Electron Devices*, vol. 27, pp. 606 ff, 1980.
- [A3.4] A. S. Grove, *Physics and technology of semiconductor devices*, New York: Wiley, Kap. 5, 1967.
- [A3.5] D. K. Schroder, "The concept of generation and recombination lifetimes in semiconductors," *IEEE Trans. Electron Devices*, vol. 29, pp. 1336-1338, 1982.
- [A3.6] T. J. Mego, "Improved quasistatic CV measurement method for MOS," *Solid State Technology Test & Measurement World*, pp. 107-110, 1986.
- [A3.7] E. H. Nicollian, "Advances in instrumentation for CV measurement," *Solid State Technology*, pp. 18-21, 1986.
- [A3.8] B. J. Gordon, "CV plotting: Myths and methods," *MOS - Solid State Technology*, p. 57-61, 1993.
- [A4.1] K. K. Ng, *Complete guide to semiconductor devices*, New York: McGraw-Hill, Kap. 1, 1995.
- [A4.2] W. Shockley, and W. T. Read, "Statistics of recombinations of electrons and holes," *Phys. Rev.*, vol. 87, pp. 835-842, 1952.
- [A4.3] R. N. Hall, "Electron-hole recombination in germanium," *Phys. Rev.*, vol. 87, pp. 387, 1952.
- [A4.4] D. K. Schroder in W. C. O'Mara, R. B. Herring, and L. P. Hunt, *Handbook of semiconductor technology*, Park Ridge: Noyes, Kap. 8, 1990.
- [A4.5] P. G. Wilson, "Recombination in silicon p- π -n diodes," *Solid-State Electronics*, vol. 10, pp. 145-154, 1967.
- [A4.6] M. Zerbst, "Relaxationseffekte an Halbleiter-Isolator-Grenzflächen," *Zeit. Angew. Phys.*, vol. 22, p. 30, 1966.
- [A4.7] *Semiconductor process development and monitoring using CV/IV and spreading resistance profiling*, Pittsburgh: SSM, 1995.
- [A4.8] A. Renz, "Entwicklung von Meßverfahren zur Charakterisierung eines hochohmigen CMOS-Prozesses," *2. Semesterarbeit am IHT*, Stuttgart, 1995.
- [A4.9] W. v. Münch, *Skriptum zur Vorlesung - Halbleitertechnik I*, Universität Stuttgart, 1992.
- [A5.1] M. Herrmann, D. Beck, E. Kasper, J.-F. Luy, K. Strohm, and J. Büchler, "Hybrid 90 GHz rectenna chip with CMOS preamplifier," *ESSDERC 1996*, Bologna, 1996.
- [A5.2] K. M. Strohm, J. Büchler, J. F. Luy, M. Herrmann, D. Beck, and E. Kasper, "94 GHz Schottky detector with CMOS preamplifier" *WRI 1996*, New York, 1996.
- [A5.3] M. Munk, "Erstellung eines Meßaufbaus zur Charakterisierung von CMOS-Operationsverstärkern," *1. Semesterarbeit am IHT*, Stuttgart, pp. 10-18, 1995.
- [A5.4] W. M. C. Sansen, M. Steyaert, and P. J. V. Vandelloo, "Measurement of operational amplifier characteristics in the frequency domain," *IEEE Trans. on Instrumentation and Measurement*, vol. 34, pp. 56-64, 1985.

-
- [A5.5] H. Wolf, *Lineare Systeme und Netzwerke*, Berlin: Springer, Kap. 4, 1971.
- [A5.6] Z. Y. Chong, und W. M. C. Sansen, *Low Noise Wide-Band Amplifiers in Bipolar and CMOS Technologies*, Boston: Kluwer, Kap. 2, 1990.
- [A5.7] Z. Chang, und W. M. C. Sansen, "Low-noise, low-distortion CMOS AM wide-band amplifiers matching a capacitive source," *IEEE Trans. on Solid-State Circuits*, vol. 25, pp. 833-840, 1990.
- [A5.8] B. Johnson, T. Quarles, A. R. Newton, D. O. Pederson, und A. Sangiovanni-Vincentelli, *SPICE3 Version 3f User's Manual*, Berkeley: University of California, 1992.
- [A5.9] P. R. Gray, und R. G. Meyer, *Analysis and Design of Analog Integrated Circuits*, 3rd ed., New York: Wiley, Kap. 3, 1993.
- [A5.10] F. D. Waldhauer, *Feedback*, New York: Wiley, 1982.
- [A5.11] F. Riedel, *MOS-Analogtechnik*, Berlin: Akademie, Kap. 2, 1988.
- [A5.12] K. R. Laker, und W. M. C. Sansen, *Design of Analog Integrated Circuits and Systems*, New York: Mc Graw-Hill, Kap. 4, 1994.
- [A5.13] J. F. Luy, und P. Russer, *Silicon-Based Millimeter-Wave Devices*, Berlin: Springer, Kap. 2, 1994.
- [A5.14] K. N. Ng, *Complete Guide to Semiconductor Devices*, New York: McGraw-Hill, Kap. 2, 1995.
- [A5.15] E. Gaßner, *Datenlexikon ttl 74er digital*, p. 145, 1986.
- [A5.16] U. Tietze, und C. Schenk, *Halbleiterschaltungstechnik*, 8. Aufl., Berlin: Springer, Kap. 5, 1986.

Danksagung

Die vorliegende Arbeit entstand im Rahmen eines vom Bundesministerium für Bildung und Forschung (BMBF) geförderten Verbundprojekts am Institut für Halbleitertechnik der Universität Stuttgart. Herrn Professor Dr. phil. E. Kasper übernahm dabei die Betreuung dieser Arbeit. Für seine Unterstützung und die Übernahme des Hauptberichts danke ich ihm herzlich.

Allen Mitarbeitern des Instituts, die in vielfältiger Weise zum Gelingen dieser Arbeit beigetragen haben, möchte ich an dieser Stelle meinen Dank ausdrücken. Für die Durchführung des Großteils der Wafer-Prozessierung danke ich insbesondere Frau S. Rohmer.

Für die Ausführung von Implantationen gebührt Herrn Dr. W. Appel vom Institut für Mikroelektronik, Stuttgart, mein herzlicher Dank.

Da die Durchführung der die Hochfrequenzschaltungen betreffenden Prozesse der Daimler-Benz-Forschungsgruppe von Dr. J.-F. Luy in Ulm oblag, möchte ich an dieser Stelle insbesondere Herrn Dr. K. Strohm und Herrn H. Jorke für fruchtbare wissenschaftliche Diskussionen sowie Frau S. Lindenmaier und Herrn Dr. Rössler für die prozeßtechnische Arbeit danken.

Für die Hochfrequenzcharakterisierung der Wellenleiterstrukturen bedanke ich mich ausdrücklich bei Herrn F. Beisswanger, der bis zum Abschluß des Projekts ebenfalls bei DaimlerChrysler in Ulm arbeitete.

Lebenslauf

Geburtsdatum: 13. März 1968

Geburtsort: Holzminden, Niedersachsen

Schulbildung: 08/74-07/78 Grundschule
08/78-07/84 Realschule
08/84-05/87 Allgemeinbildendes Gymnasium

Wehrdienstzeit: 10/87-12/88 Grundwehrdienst bei einer Luftwaffen-Radareinheit

Studium: 10/88-07/93 Elektrotechnik an der Technischen Hochschule Karlsruhe
davon 09/91-07/92 Studium an der University of Essex, Großbritannien
und 08/92-07/93 Studium an der ESIEE Paris, Frankreich

Berufstätigkeit: 10/93-09/96 Wissenschaftlicher Mitarbeiter am Institut für Halbleitertechnik der Universität Stuttgart
10/96-07/98 Mitarbeiter der DaimlerChrysler AG, Sindelfingen
seit 08/98 Mitarbeiter im DaimlerChrysler Projekthaus Brennstoffzelle

