

Charakterisierung von Gate-Oxiden mittels Charge-Pumping und $1/f$ -Rauschanalysen

Von der Fakultät für Elektrotechnik und Informatik
der Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des akademischen Grades
Doktor-Ingenieur (Dr.-Ing.)
genehmigte Dissertation
von

Dipl.-Phys. Gernot Krause

geboren am 03.11.1975 in Bremen

2009

1. Referent: Prof. Dr. K. R. Hofmann
2. Referent: Prof. Dr. W. Mathis
Tag der Promotion: 04.12.2008

Kurzfassung

Diese Arbeit beschäftigt sich mit der elektrischen Charakterisierung von Gate-Oxiden, die in modernen Halbleitertechnologien eingesetzt werden. Durch den Einsatz der frequenzabhängigen Charge-Pumping-Methode werden neue Erkenntnisse bezüglich der Generation von Störstellen in CMOS-Bauelementen unter Fowler-Nordheim-Stress gewonnen. Darüber hinaus kommen erstmals $1/f$ -Rauschanalysen zur Untersuchung von NROM-Speicherzellen mit ultrakurzen Kanälen zur Anwendung.

Die Stressexperimente werden an n - und p -Kanal CMOS-Transistoren mit nitridiertem Gate-Oxid durchgeführt. Durch Variation der Charge-Pumping-Frequenz ist es möglich, die Dichte von Grenzflächenzuständen getrennt von der Dichte grenzflächennaher Oxidstörstellen zu bestimmen. Dieses Verfahren wird erstmals dazu verwendet, die Generation von Störstellen unter Fowler-Nordheim-Stress in Abhängigkeit der Stresspolarisation zu beobachten. Hierbei zeigt die Grenzflächenzustandsdichte der n -MOSFETs eine geringe Polarisationsabhängigkeit, während die der p -MOSFETs keine Abhängigkeit von der Polarisierung der Stressspannung aufweist. Dagegen kann in der Dichte grenzflächennaher Oxidstörstellen für beide Transistortypen eine deutliche Polarisationsabhängigkeit beobachtet werden. Des Weiteren wird mithilfe von $C(V)$ -Messungen die Differenz der Flachbandspannung zwischen positiver und negativer Spannungsrampe bestimmt. Hierbei wird ein ungewöhnliches Verhalten im anfänglichen Stresstadium des p -MOSFETs beobachtet, das auf eine extrem hohe Dichte von Störstellen auf der Gate-Seite des Oxids hindeutet.

Neben CMOS-Bauelementen werden NROM-Speicherzellen mithilfe von $1/f$ -Rauschanalysen untersucht. Eine Degradation dieser Bauelemente tritt durch heiße Ladungsträger auf, die beim Programmieren und Löschen der Zelle entstehen. Daher wird der Stress hier experimentell durch wiederholte Programmier- und Löschoperationen erzeugt und anschließend die Spektraldichte des Drain-Stromrauschens mithilfe von Fast-Fourier-Transformationen bestimmt. Im Zuge dieser Untersuchungen werden zwei potenzielle Methoden zur Bestimmung des lateralen Schädigungsortes im Transistorkanal entwickelt. Der erste Ansatz beruht auf demselben Effekt, der zum Lesen der Speicherzellen (*reverse read mode*) verwendet wird. Im zweiten Ansatz wird die NROM-Zelle nach dem Stressen gezielt programmiert, um damit eine Erhöhung der Rauschempfindlichkeit in bestimmten Kanalbereichen herbeizuführen und gleichzeitig Rauschquellen in anderen Bereichen zu unterdrücken. Beide Experimente zeigen, dass eine erhöhte Schädigung aufgrund heißer Ladungsträger in der Nähe des Drains vorliegt.

Zur Unterstützung der experimentellen Annahmen werden zusätzlich 3D-Simulationen durchgeführt, mit dem Ziel, den Einfluss einer Elementarladung im Gate-Oxid auf den Drain-Strom zu bestimmen. Die Stärke des Einflusses in Abhängigkeit der lateralen Position der Elementarladung lässt Rückschlüsse auf die lokale Rauschempfindlichkeit zu.

Schlagnworte: Charge-Pumping, $1/f$ -Rauschen, Fowler-Nordheim-Stress, Gate-Oxid

Abstract

This publication presents the electrical characterization of gate oxides which are used in state-of-the-art semiconductor technologies. A better insight into trap generation in CMOS devices under Fowler-Nordheim stress conditions will be gained utilizing the frequency dependent charge-pumping method. Furthermore, $1/f$ noise analysis studies are applied to NROM memory cells with ultra short channels.

For the Fowler-Nordheim stress experiments n -channel as well as p -channel CMOS devices with nitrated gate oxides are used. The frequency dependent charge-pumping is able to distinguish between the density of fast interface states and the density of slower near-interface oxide traps. For the first time this method is used to study trap generation under Fowler-Nordheim stress in dependence of the polarity of stress voltage. As a result the interface state density shows no polarity dependence for p -MOS devices and only a marginal dependence in case of n -MOS devices. In contrast, the density of near-interface oxide traps shows a significant difference between positive and negative applied stress voltage. Additionally, the difference in flat-band voltage between forward and reverse voltage sweep during $C(V)$ measurements is investigated. For the first time an abnormal behaviour at the onset of stress is observed which indicates an extremely high density of pre-existing gate-side oxide traps in case of p -MOS devices.

NROM memory cells are investigated using $1/f$ noise analysis. These devices are subjected to hot carrier stress during normal write and/or erase operations. Therefore, systematic stress is applied by iterative write/erase cycles. For different stress states the drain-current is measured and analysed by Fast Fourier Transform. In line with these experiments two potential detection methods for the lateral location of stress-induced damage along the channel are developed. The first one uses the reverse read mode of the memory cell to locally influence the drain current noise sensitivity. In the second approach memory charges are regularly stored into one bit of the cell. These charges locally increase the contribution of noise sources to the power spectral density beneath the specific bit area whereas noise sources in other channel regions are suppressed. Both approaches show an increased damage near the respective drain due to initial hot carrier stress.

To support the experimental assumptions additional 3D simulations are performed to determine the influence of an elementary charge in the near-interface oxide region on the drain current. From the amplitude of the influence conclusions about local noise sensitivity can be drawn.

Keywords: charge-pumping, $1/f$ noise, Fowler-Nordheim stress, gate oxide

Inhaltsverzeichnis

Kurzfassung	i
Abstract	ii
1 Einleitung	1
2 Grundlagen hochintegrierter Bauelemente	5
2.1 Das MIS-System	5
2.1.1 Grundlagen des MOS-Kondensators	5
2.1.2 Grundlagen des MOS-Feldeffekttransistors (MOSFET)	9
2.1.3 Kurzkanaleffekte	12
2.2 Gate-Oxid und Substratgrenzfläche	17
2.2.1 Elektrische Transportmechanismen	17
2.2.2 Elektrische und strukturelle Eigenschaften von Störstellen	19
2.3 Degradationseffekte	25
2.3.1 Wasserstofffreisetzung	26
2.3.2 Anoden-Löcherinjektion	28
2.3.3 Negative Bias Temperature Instability (NBTI)	28
2.3.4 Nitrierte Oxide	29
2.4 Der MOSFET in Logikschaltungen (CMOS)	31
2.5 Der MOSFET als nichtflüchtiger Speicher (NROM)	32
2.5.1 Die NROM-Speicherzelle	34
2.5.2 Programmieren und Löschen einer NROM-Zelle	34
2.5.3 Lesen einer NROM-Zelle	35
3 Theorie der Messmethoden	39
3.1 Elektrische Charakterisierungsmethoden	39
3.2 Charge-Pumping (CP)	39
3.2.1 Die Funktionsweise der CP-Methode	40
3.2.2 Besonderheiten und Unterschiede der CP-Modi	45
3.3 Rauschen	50
3.3.1 $1/f$ -Rauschen	54
3.3.2 Analyse des $1/f$ -Rauschens mittels FFT	57
3.3.3 Bewertung der Rauschmessmethode	59

4	Charakterisierung von nitrierten Gate-Oxiden	61
4.1	Verwendete Probenstrukturen	61
4.2	Messaufbau für Charge-Pumping	62
4.2.1	Messablauf	65
4.3	Charge-Pumping-Untersuchungen	65
4.3.1	CP-Messungen bei konstanter Frequenz	65
4.3.2	Frequenzabhängige CP-Messungen	69
4.3.3	Fowler-Nordheim-Stressexperiment	74
4.3.4	Ergebnisse des Stressexperiments	76
4.3.5	Schlussfolgerung	86
4.4	Messaufbau für Rauschmessungen	86
4.5	Messungen des $1/f$ -Rauschens	88
4.5.1	Ergebnisse der Rauschmessungen	90
5	Charakterisierung von NROM-Speicherzellen	93
5.1	Verwendete Probenstrukturen	93
5.2	Stresseinprägung durch wiederholte P/L-Zyklen	94
5.3	Untersuchungen des $1/f$ -Rauschens	98
5.3.1	Rauschen in Abhängigkeit der Drain-Source-Spannungspolarität	101
5.3.2	Rauschen in Abhängigkeit des Programmierzustands	103
5.4	Simulation einer NROM-Zelle	107
5.4.1	3D-Simulationen mit ATLAS	108
5.4.2	Simulationsergebnisse	109
5.5	Schlussfolgerung	113
6	Zusammenfassung	115
A	CMOS-Charakterisierungsmessungen	119
A.1	$C(V)$ -Messungen an Kapazitäten mit nitriertem Gate-Oxid	119
A.2	Bestimmung von f_0 mittels frequenzabhängiger CP-Messung	122
A.3	Verlauf der Gate-Spannung beim FN-Stress	123
B	NROM-Charakterisierungsmessungen	125
B.1	Spannungsverläufe beim Programmieren und Löschen	125
	Symbol- und Abkürzungsverzeichnis	127
	Abbildungsverzeichnis	133
	Tabellenverzeichnis	135
	Literaturverzeichnis	137

1 Einleitung

Elektronische Kommunikationsmedien sind in den letzten Jahren fester Bestandteil unseres täglichen Lebens geworden. Im Jahr 2007 hat die weltweite Zahl von Mobiltelefonen die Grenze von drei Milliarden überschritten, die Zahl der installierten PCs liegt bei einer Milliarde weltweit und steigt jährlich mit zweistelligen Zuwachsraten. Ähnliche Prognosen werden für Internetanschlüsse gegeben [1]. Der Wunsch, an nahezu jedem Ort der Welt beliebige Daten zur Verfügung zu haben und Informationen austauschen zu können, treibt die Entwicklung dieser Technologien in rasanter Geschwindigkeit voran. Die Basis dafür bilden Halbleiterbauelemente, die mit immer kleineren Dimensionen und daher höheren Dichten auf einzelnen Halbleiterchips integriert werden können. Die Elektronik- und Halbleiterindustrie unterliegt einem enormen Wachstum und ist heute die größte Industrie weltweit [2].

Um diesem Markt und den ständig steigenden Anforderungen an die Leistungsfähigkeit seiner Produkte gerecht zu werden, müssen von der Halbleiterindustrie immer leistungsfähigere mikroelektronische Schaltungen entwickelt werden. Zum einen werden schnelle Mikroprozessoren benötigt, die große Mengen von Daten in kürzester Zeit verarbeiten können, zum anderen Speicher, die die verarbeiteten Informationen speichern. Gleichzeitig soll aber der Energiebedarf dieser mikroelektronischen Chips nicht über die Maßen steigen, um den Einsatz in mobilen Geräten zu gewährleisten.

In den letzten Jahrzehnten wurde die Steigerung der Leistungsfähigkeit von mikroelektronischen Chips hauptsächlich durch Skalierung erzielt. Das bedeutet, dass die typische Breite von Strukturen innerhalb eines Chips stetig verringert wird, um kürzere Signallaufzeiten und schnellere Schaltzeiten der Transistoren zu realisieren. Wurden vor einigen Jahren noch Chips mit Strukturbreiten von 130 nm und 90 nm hergestellt, so ist man heute bei Strukturbreiten von 45 nm angekommen. Der nächste Schritt in der Technologie sind Strukturbreiten von 32 nm. Weil die Verringerung der Strukturbreite in naher Zukunft an ihre Grenzen stoßen wird und auch heutzutage schon große Herausforderungen an den Herstellungsprozess stellt, werden seit einigen Jahren auch äquivalente Skalierungen durchgeführt. Hierbei werden zum einen neue Materialien, wie zum Beispiel High- k , eingesetzt, zum anderen werden gezielt physikalische Effekte ausgenutzt, zum Beispiel mechanische Verspannungen des Halbleiters [3]. In der Speichertechnologie wird zusätzlich durch die Entwicklung neuer Konzepte eine höhere Integrationsdichte der Speicherzellen erreicht [4]. Die zunehmende Integration mehrerer Baugruppen auf einem Chip selbst trägt ebenfalls zur Steigerung der Effizienz von elektronischen Schaltungen bei. So werden zum Beispiel in Mobilfunkgeräten immer mehr Baugruppen auf einem einzigen Chip untergebracht. Durch die Skalierungsprozesse in der Halbleitertechnologie wird aber nicht nur die Leistungsfähigkeit gesteigert, sondern auch die benötigte Chipfläche reduziert und damit

werden Produktionskosten gesenkt.

Neben den positiven Aspekten der Skalierung bringt diese aber auch eine ganze Reihe neuer Probleme mit sich. In den letzten Jahren traten diese vermehrt in den Vordergrund und beeinflussen entscheidend die Zuverlässigkeit der Produkte. Mit der Verringerung der lateralen Dimensionen der Bauelemente muss auch die Dicke des Gate-Oxids verringert werden, um ein äquivalentes Schaltverhalten der Transistoren zu gewährleisten. Dünnere Oxide weisen aber auch größere Leckströme auf, welche zu einer höheren Leistungsaufnahme der integrierten Schaltung führen. Im Falle von Speicherzellen, in denen Information in Form von Ladungen viele Jahre gespeichert werden soll, führen Leckströme zur Entladung und damit zum Verlust der gespeicherten Information. Neben den erhöhten Leckströmen treten weitere Effekte im Zuge der geometrischen Skalierung auf.

Mit der Verkleinerung der Strukturbreiten und Schichtdicken muss auch die Betriebsspannung reduziert werden. Dies kann jedoch nicht in gleichem Maße geschehen. Speicherzellen benötigen zum Beispiel bestimmte Mindestspannungen zum Lesen beziehungsweise zum Programmieren oder Löschen. Im Falle von Mikroprozessoren führt jede Reduktion der Spannung wiederum zu Einbußen in der Leistungsfähigkeit. Die Folge der nicht verhältnismäßigen Skalierung von Längen und Spannungen führt zu höheren elektrischen Feldern und damit zur Generation von so genannten heißen Ladungsträgern im Bauelement. Die Folge ist eine Degradation der elektrischen Eigenschaften der Transistoren. Hohe Spannungsabfälle über dem Gate-Oxid rufen weitere Degradationseffekte hervor, wie zum Beispiel NBTI oder einen Anstieg von Leckströmen (SILC) bis hin zum elektrischen Durchbruch des Oxids.

Die Untersuchung dieser intrinsischen Bauelementedegradation hat durch die extreme Technologieskalierung in den letzten Jahren zunehmend an Bedeutung gewonnen. Daher befasst sich diese Arbeit mit elektrischen Methoden zur Charakterisierung der Degradation in MOS-Transistoren, die in 130 nm und 75 nm Halbleitertechnologien gefertigt wurden. Im Vordergrund steht hierbei die Untersuchung des Gate-Oxids, das eine gewisse Festigkeit gegenüber Degradationseffekten und elektrischen Durchbrüchen aufweisen muss.

Die Arbeit ist in zwei Teile untergliedert. Im ersten Teil werden Degradationseffekte unter Fowler-Nordheim-Stressbedingungen mit sowohl positiver als auch negativer Gate-Spannungspolarität untersucht. Es standen dazu n - und p -MOS-Transistoren zur Verfügung, die in digitalen Logikschaltungen (CMOS-Technologien) zum Einsatz kommen. Zur Analyse der Schädigung werden Kapazitäts-Spannungsmessungen sowie frequenzabhängige Charge-Pumping-Methoden angewendet, mit denen zwischen Schädigungen an der Grenzfläche und im grenzflächennahen Oxidbereich unterschieden werden kann. Der zweite Teil der Arbeit konzentriert sich auf die Untersuchung von NROM-Speicherzellen, die als nichtflüchtige Speicher eingesetzt werden. Hierbei wird das $1/f$ -Rauschen des Drain-Stroms analysiert, wodurch Aussagen über die Degradation des MOS-Transistors möglich sind. Die Rauschanalysen erfolgen in Abhängigkeit der Zahl wiederholter Programmier- und Löszyklen. Im Vordergrund dieser Stressexperimente steht hierbei die Entwicklung einer Methode, mit welcher der Ort der Schädigung im Transistorkanal aufgelöst werden kann. Unterstützend kommen 3D-Simulationen zum Einsatz.

In den nachfolgenden Kapiteln werden zunächst die Grundlagen integrierter Bauelemente beschrieben und das MIS-System vorgestellt, das die Basis für den MOS-Feldeffekttransistor bildet. Es wird die Funktionsweise des MOS-Feldeffekttransistors (MOSFET) erläutert und die für diese Arbeit relevanten Degradationsmechanismen vorgestellt, die typischerweise im Gate-Oxid oder an dessen Grenzflächen unter hohen Spannungen auftreten. Zur Veranschaulichung der Auswirkungen der Bauelementedegradation auf mikroelektronische Schaltungen werden mögliche Einsatzgebiete von MOS-Transistoren in Logikschaltungen dargestellt. Ebenso wird die Funktionsweise einer NROM-Speicherzelle beschrieben sowie die Effekte, die die Arbeitsweise der Zelle bei einer Degradation des Gate-Oxids negativ beeinflussen können.

Darüber hinaus werden die in dieser Arbeit zur Anwendung kommenden elektrischen Messmethoden vorgestellt. Zunächst wird die Charge-Pumping-Methode erläutert, die für die stressabhängigen Gate-Oxid-Untersuchungen an Logiktransistoren verwendet wurde. Diese Methode zeichnet sich durch eine Vielzahl von unterschiedlichen Modi aus, von denen die wichtigsten kurz dargestellt werden. Danach werden die Messungen und Analyseverfahren vorgestellt, die zur Ermittlung des $1/f$ -Rauschens in NROM-Speicherzellen verwendet wurden. Das Augenmerk richtet sich hierbei auf die Fast-Fourier-Transformation (FFT), die in dieser Arbeit mittels einer Software durchgeführt wurde.

Der erste experimentelle Teil der Arbeit befasst sich mit den Untersuchungen von nitridierten Gate-Oxiden unter Fowler-Nordheim-Stressbedingungen. Die hier verwendeten Teststrukturen sind typische n - und p -MOS-Transistoren, welche in CMOS-Logikschaltungen zum Einsatz kommen. Die Analyse der Degradation wird mit frequenzabhängigen Charge-Pumping-Messungen durchgeführt. Zunächst wird der dazu verwendete Messaufbau beschrieben und danach die frequenzabhängigen Charge-Pumping-Messungen sowie die Details des Stressexperiments vorgestellt. Es werden die Ergebnisse der Untersuchungen dargestellt, sowie die daraus getroffenen Schlussfolgerungen. Darüber hinaus werden zur Evaluation der Methode Messungen des $1/f$ -Rauschens durchgeführt und bewertet.

Der zweite experimentelle Teil befasst sich mit $1/f$ -Rauschanalysen sowie 3D-Simulationen an NROM-Speicherzellen. Zunächst wird beschrieben, wie die Speicherzellen durch wiederholte Programmier- und Löschkzyklen gestresst wurden. Danach werden zwei unterschiedliche Ansätze vorgestellt, mit denen die Schädigung im Transistorkanal lokal bestimmt werden kann. Die Ergebnisse der beiden Experimente werden dargestellt und bewertet. Zur Verifizierung der Ansätze werden dreidimensionale Simulationen der NROM-Zelle mit scharf definierten Ladungspaketen im Gate-Oxid durchgeführt.

Am Schluss der Arbeit wird eine Zusammenfassung der Ergebnisse und ein Ausblick auf mögliche Folgeuntersuchungen und zukünftige Forschungsarbeiten gegeben.

2 Grundlagen hochintegrierter Bauelemente

2.1 Das MIS-System

Der MIS-Kondensator (*engl.*: Metal-Insulator-Semiconductor) stellt das grundlegende Bauelement in siliziumbasierten integrierten Schaltungen dar. Er ist Bestandteil von Feldeffekttransistoren in CMOS-Schaltungen und wird in strukturell modifizierter Form auch in nichtflüchtigen Speichern eingesetzt. Daher lassen sich alle wichtigen physikalischen Vorgänge und Probleme von Halbleiterbauelementen anhand dieses Systems studieren. Die Zuverlässigkeit (*engl.*: reliability) und Langlebigkeit (*engl.*: endurance) der Bauelemente hängt dabei besonders stark von den Eigenschaften des Isolators und der Grenzfläche zwischen Isolator und Halbleiter ab [2]. Der Isolator besteht in CMOS-Schaltungen aus einem Oxid. Daher wird ein MIS-Kondensator mit einer isolierenden Oxidschicht auch als MOS-Kondensator (*engl.*: Metal-Oxide-Semiconductor) bezeichnet. Für siliziumbasierte Halbleitertechnologien wird seit vielen Jahren erfolgreich Siliziumdioxid (SiO_2) als Isolator verwendet [5]. In modernen Prozessen kommt aber auch immer häufiger Hafniumoxid (HfO_2) zum Einsatz, das aufgrund seiner hohen Dielektrizitätszahl einige Vorteile gegenüber Siliziumdioxid aufweist [6]. Der Metallkontakt besteht in modernen Schaltungen in der Regel aus hochdotiertem Poly-*Si*, und als Halbleiter wird Silizium eingesetzt.

2.1.1 Grundlagen des MOS-Kondensators

In Abbildung 2.1 ist der Aufbau eines MOS-Kondensators skizziert. Zwischen dem Siliziumsubstrat und dem Poly-*Si*-Gate befindet sich eine Siliziumdioxidschicht der Dicke d_{ox} . Die Rückseite des Substrats ist mit einem ohmschen Kontakt versehen, und die am Gate angelegte Spannung wird mit V_G bezeichnet. Bei einem idealen MOS-Kondensator wird angenommen, dass Ladungen nur im Halbleiter und im Gate-Kontakt existieren und das Oxid frei von Ladungen ist. Außerdem wird angenommen, dass das Oxid einen idealen Isolator darstellt, durch den unter Gleichstrombedingungen kein Strom fließt.

Es gibt grundsätzlich zwei Arten von MOS-Kondensatoren, den *n*-Typ und den *p*-Typ. Beim *n*-MOS-Kondensator ist das Siliziumsubstrat *p*-dotiert mit einer Akzeptordichte N_A . Das Metall-Gate besteht in der Regel aus hochdotiertem n^+ -Poly-*Si*. Der *p*-MOS-Kondensator besitzt im

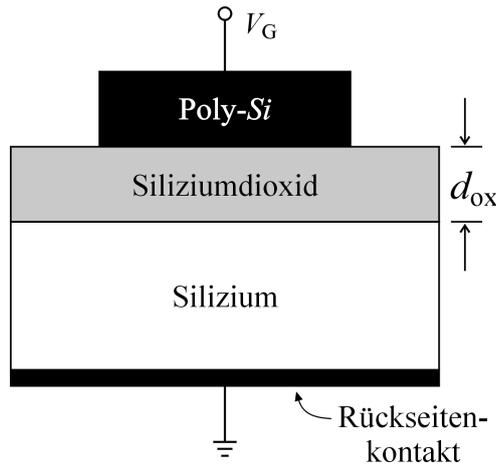


Abbildung 2.1: Aufbau eines MOS-Kondensators. Die Dicke des Siliziumdioxids wird mit d_{ox} angegeben, die angelegte Spannung am Poly-*Si* (Gate) mit V_G .

Gegensatz dazu ein n -dotiertes Siliziumsubstrat mit einer Donatordichte N_D und einem hochdotierten p^+ -Poly-*Si* als Gate-Kontakt. Da sich beide Bauelemente analog zueinander verhalten, beziehen sich die nachfolgenden Betrachtungen ausschließlich auf den n -MOS-Kondensator.

Jedes der verwendeten Materialien besitzt eine charakteristische Energiebandstruktur [7]. Bringt man diese Materialien zusammen, so ändert sich der Verlauf der Energiebänder. Abbildung 2.2a zeigt das Banddiagramm eines MOS-Kondensators mit einem p -dotiertem Substrat für den Fall, dass keine äußere Spannung am Gate angelegt ist. Aufgrund der unterschiedlichen Austrittsarbeiten $q\phi_{\text{Poly}}$ des Poly-*Si* und $q\phi_{\text{Si}}$ des Halbleiters verbiegen sich die Bänder im Silizium nahe der Grenzfläche zum Oxid. Dies führt zu einer Ansammlung von Ladungen im Halbleiter, die durch Ladungen mit entgegengesetztem Vorzeichen am Gate-Kontakt kompensiert werden. Daraus resultiert ein Spannungsabfall V_{ox} über dem Oxid, der ein elektrisches Feld $\mathcal{E}_{\text{ox}} = \frac{V_{\text{ox}}}{d_{\text{ox}}}$ hervorruft.

Um das MOS-System feldfrei zu bekommen - man spricht in diesem Fall auch von flachen Bändern -, ist eine äußere Spannung V_{fb} nötig, die als Flachbandspannung bezeichnet wird [5]. Das Banddiagramm für den so genannten Flachbandfall ist in Abbildung 2.2b dargestellt. Die Flachbandspannung ist im idealen Fall durch die Differenz der Austrittsarbeiten gegeben:

$$V_{\text{fb}} = \phi_{\text{Poly}} - \phi_{\text{Si}} = \phi_{\text{Poly}} - \left(\chi + \frac{E_g}{2q} + \psi_F \right) . \quad (2.1)$$

Dabei bezeichnet $q\chi$ die Elektronenaffinität im Silizium, E_g die Energiebandlücke im Silizium und ψ_F das Fermi-Potenzial. Das Fermi-Potenzial kann mithilfe des Oberflächen-Fermi-Potenzials ψ_o mit dem Oberflächenpotenzial ψ_s über

$$\psi_F = \psi_s + \psi_o \quad (2.2)$$

in Verbindung gesetzt werden (vergleiche Abbildung 2.2a). Im Flachbandfall ist der Kondensator entladen und die Nettoladungsdichte an der Grenzfläche, sowie das elektrische Feld im Oxid

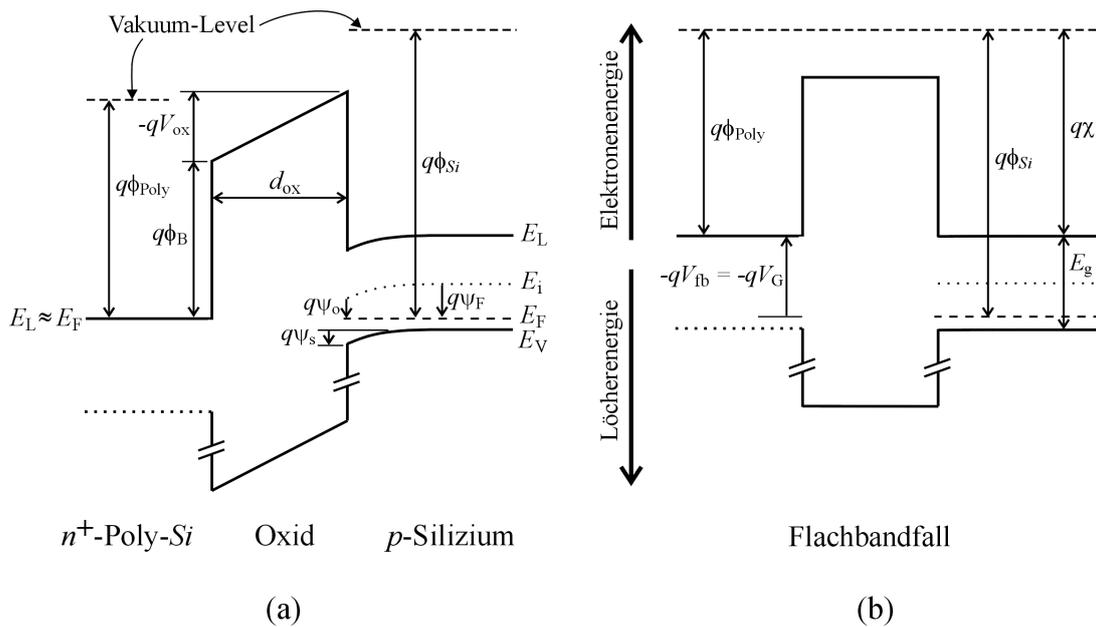


Abbildung 2.2: Bandstruktur eines MOS-Kondensators. (a) Ohne äußere Spannung. (b) Bei angelegter Flachbandspannung V_{fb} .

betragen Null.

Durch Anlegen einer äußeren Spannung kann das Oberflächenpotenzial ψ_s und damit die Ladungsträgerdichte an der Grenzfläche beeinflusst werden. Grundlegend werden je nach Polarität der Gate-Spannung drei Zustände unterschieden, in denen sich das System befinden kann: Akkumulation, Verarmung und Inversion. Diese Zustände sind in der oberen Hälfte von Abbildung 2.3 skizziert. In der unteren Hälfte ist die Ladungsverteilung in Abhängigkeit des Ortes für die jeweiligen Zustände aufgetragen.

Wenn eine negative Spannung am Gate-Kontakt anliegt, biegen sich die Bänder nahe der Grenzfläche nach oben. Dadurch liegt die Valenzbandkante E_V hier näher am Fermi-Niveau E_F , welches flach ist, da in der idealen MOS-Struktur kein Strom fließt. Die Ladungsträgerdichte hängt exponentiell von der Energiedifferenz $E_F - E_V$ ab. An der Oberfläche des Halbleiters sammeln sich Majoritätsladungsträger, in diesem Fall Löcher, an. Dieser Zustand wird als Akkumulation bezeichnet (Abbildung 2.3a). Die Ladung Q_s an der Halbleiteroberfläche wird durch eine gleichgroße Ladungsmenge Q_P mit entgegengesetztem Vorzeichen im Poly-Si kompensiert. Ist die Gate-Spannung dagegen leicht positiv, biegen sich die Bänder an der Siliziumoberfläche nach unten. Die Majoritätsladungsträger werden in diesem Fall verdrängt und die Struktur befindet sich in Verarmung (Abbildung 2.3b). Es bildet sich dabei eine Raumladungszone der Weite W_R im Silizium aus. In dieser Zone sind alle Akzeptoren ionisiert und die Ladungsmenge im Halbleiter berechnet sich aus $Q_B = -qN_A W_R$. Wird die positive Spannung am Gate weiter erhöht, verbiegen sich die Bänder an der Oberfläche des Halbleiter so stark, dass das Fermi-Niveau das intrinsische Energieniveau E_i kreuzt und sich oberhalb davon befindet. In diesem Fall ist die Konzentration an Minoritätsladungsträgern, hier Elektronen, größer als die der Ma-

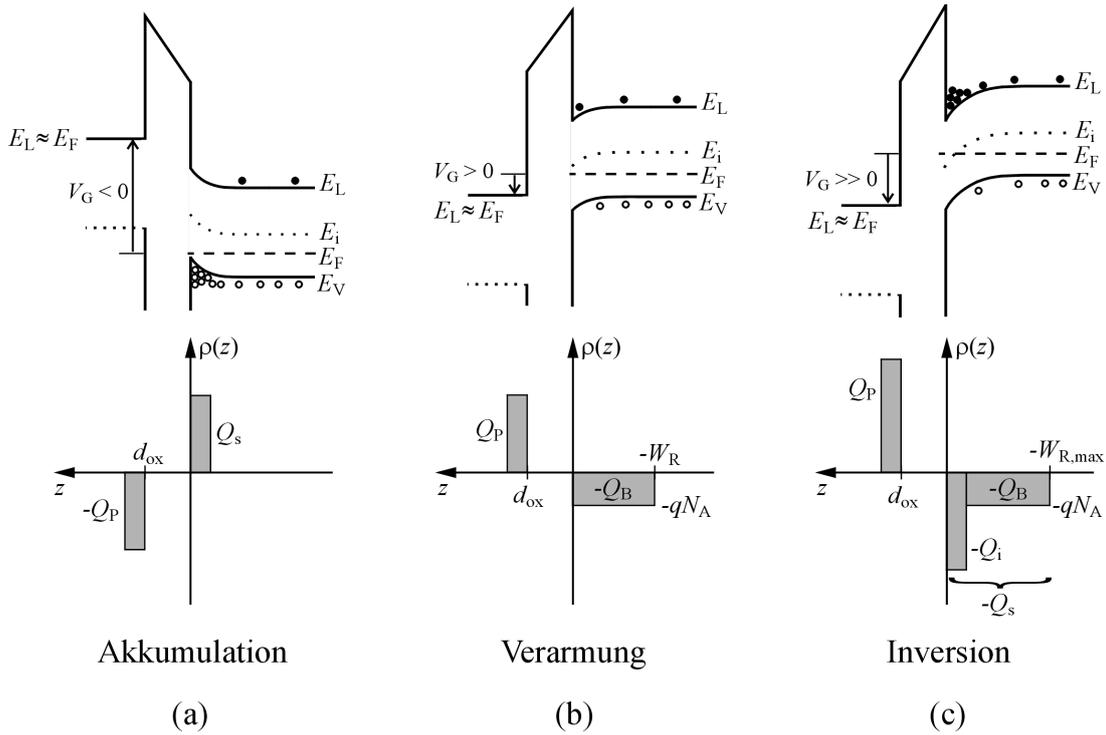


Abbildung 2.3: Energiebandverläufe (oben) eines idealen MOS-Kondensators bei angelegter Gate-Spannung. Je nach Vorzeichen der Gate-Spannung werden drei verschiedene Zustände des Systems unterschieden. (a) Akkumulation: $V_G < 0$, (b) Verarmung: $V_G > 0$, (c) Inversion: $V_G \gg 0$. Die gefüllten Kreise symbolisieren Elektronen, die leeren Löcher. Darunter ist jeweils die Ladungsverteilung in Abhängigkeit des Ortes aufgetragen.

poritäten. Dieser Zustand wird als Inversion bezeichnet (Abbildung 2.3c). Die Raumladungszone besitzt in diesem Zustand eine maximale Ausdehnung $W_{R,max}$ und es bildet sich eine zusätzliche, aus Elektronen bestehende Inversionsladung Q_i an der Oberfläche. Die Summe der Ladungen in der Raumladungszone und der Inversionsschicht bilden die gesamte Halbleiterladung $Q_s = Q_B + Q_i$.

Die Oberflächenladung Q_s ist gegeben durch

$$Q_s = \mp \frac{\sqrt{2}\epsilon_s}{\beta L_D} F\left(\beta\psi_s, \frac{n_{p,0}}{p_{p,0}}\right), \quad (2.3)$$

wobei $n_{p,0}$ und $p_{p,0}$ die Gleichgewichtsladungsträgerdichten der Elektronen beziehungsweise Löcher bezeichnen, ϵ_s die Dielektrizitätskonstante im Silizium und $\beta \equiv q/k_B T$ ist. Das Vorzeichen ergibt sich je nach Zustand der Oberfläche. In Akkumulation ($\psi_s < 0$) gilt für n -MOS-Kondensatoren das positive Vorzeichen. L_D ist hier die Debye-Länge für Löcher und die Funktion $F(\beta\psi_s, \frac{n_{p,0}}{p_{p,0}})$ ist gegeben durch

$$F\left(\beta\psi_s, \frac{n_{p,0}}{p_{p,0}}\right) \equiv \sqrt{(e^{-\beta\psi_s} + \beta\psi_s - 1) + \frac{n_{p,0}}{p_{p,0}} (e^{-\beta\psi_s} - \beta\psi_s - 1)} \geq 0. \quad (2.4)$$

Im Verarmungszustand steigt die Oberflächenladung proportional zur Quadratwurzel des Oberflächenpotenzials, $Q_s \propto \sqrt{\psi_s}$. Der Zustand der starken Inversion setzt in etwa ein wenn für das Oberflächenpotenzial gilt

$$\psi_s \simeq 2\psi_F \approx \frac{2}{\beta} \ln \left(\frac{N_A}{n_i} \right) \quad . \quad (2.5)$$

Für die Charge-Pumping-Untersuchungen in Abschnitt 3.2.1 werden außerdem die Ladungsträgerdichten der Majoritäten und Minoritäten an der Oberfläche benötigt. Diese sind gegeben durch

$$p_s = n_i e^{\beta(\psi_F - \psi_s)} \quad \text{und} \quad (2.6)$$

$$n_s = n_i e^{\beta(\psi_s - \psi_F)} \quad . \quad (2.7)$$

Dabei bezeichnet $n_i = 9,65 \cdot 10^9 \text{ cm}^{-3}$ die intrinsische Ladungsträgerkonzentration im Halbleiter [2].

2.1.2 Grundlagen des MOS-Feldeffekttransistors (MOSFET)

Ein MOSFET ist ein MOS-Kondensator mit zusätzlichen Source- und Drain-Kontakten zwischen denen ein Strom fließen kann, wenn sich der MOSFET im Zustand der Inversion befindet. Da zum Stromfluss entweder nur Elektronen oder Löcher beitragen, ist der MOSFET ein unipolares Bauelement. In Abbildung 2.4 ist der Aufbau eines MOSFETs dargestellt. Der MOS-Kondensator besteht aus einer Schicht Siliziumdioxid der Dicke d_{ox} mit einem Poly-Si-Gate darüber. Dieser Schichtstapel ist durch einen Strukturierungsprozess auf eine Länge L_G und eine Weite W_G begrenzt. Links und rechts davon befinden sich im Siliziumsubstrat hochdotierte n^+ -Gebiete, die mittels Implantation erzeugt wurden. Durch Diffusionsprozesse dehnen sich diese Gebiete während der Herstellung leicht unter das Oxid aus, so dass sie einen tatsächlichen Abstand $L < L_G$ voneinander einnehmen. Die Isolation zu benachbarten Bauelementen wird durch einen Oxidwall erreicht.

Bringt man den zentralen MOS-Kondensator durch Anlegen einer positiven Gate-Spannung in Inversion, so bildet die Inversionsschicht einen leitenden Kanal zwischen dem Source- und dem Drain-Gebiet. Inversion setzt ein, wenn die Gate-Spannung eine charakteristische Schwellspannung V_t übersteigt. Wird zusätzlich eine Drain-Spannung $V_D > 0$ am Drain-Kontakt angelegt, so befindet sich die Halbleiteroberfläche im Nichtgleichgewicht und ein Drain-Strom I_D fließt zwischen Source und Drain. Die Abhängigkeit des Drain-Stroms von der Gate-Spannung wird als Transferkennlinie wiedergegeben und ist in Abbildung 2.5 sowohl halblogarithmisch als auch linear dargestellt. In der halblogarithmischen Darstellung können drei Bereiche definiert werden, der Sperrbereich, der Unterschwellbereich und der Bereich, in dem der Transistor eingeschaltet ist. Im Sperrbereich fließt kein Strom durch den Kanal, sondern nur ein Leckstrom $I_{D,leak}$, der durch den pn -Übergang am Drain verursacht wird. Im Unterschwellbereich steigt der Drain-Strom exponentiell mit der Gate-Spannung an. Der Transistorkanal befindet sich in

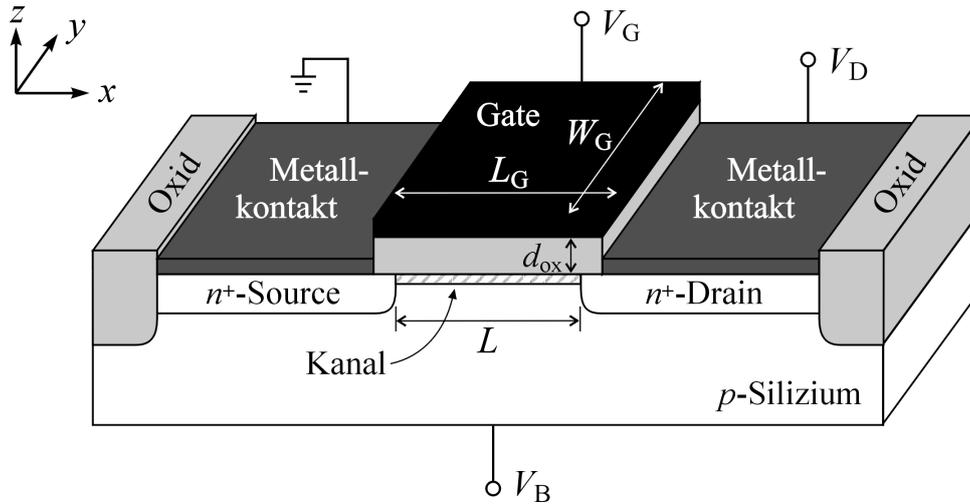


Abbildung 2.4: Aufbau eines MOSFETs. Zentraler Bestandteil ist ein MOS-Kondensator mit einer Oxiddicke d_{ox} und einem Gate der Länge L_G und Weite W_G . Links und rechts daneben befinden sich hochdotierte Source- und Drain-Gebiete, zwischen denen sich bei entsprechender Gate-Source-Spannung V_G ein leitender Kanal bildet. Der MOSFET ist durch einen Oxidwall elektrisch von benachbarten Bauelementen getrennt.

schwacher Inversion und der Strom durch den Transistorkanal wird in diesem Bereich durch Diffusion bestimmt. Oberhalb der Schwellspannung V_t befindet sich die Oberfläche in mittlerer bis starker Inversion und der Drain-Strom steigt linear mit der Gate-Spannung an (siehe lineare Darstellung der Transferkennlinie in Abbildung 2.5).

Im Folgenden soll der Transistor in Inversion näher betrachtet werden. Die Gate-Spannung ist hierbei größer als die Schwellspannung, und der Transistor befindet sich daher im eingeschalteten Zustand. Die Schwellspannung ist gegeben durch

$$V_t = V_{\text{fb}} + 2\psi_F - \frac{Q_B}{C_{\text{ox}}} = V_{\text{fb}} + 2\psi_F + \sqrt{2} \frac{\epsilon_s}{L_D C_{\text{ox}}} \sqrt{\frac{2\psi_F - V_B}{\beta}} \quad (2.8)$$

Dabei gibt $C_{\text{ox}} = \frac{\epsilon_{\text{ox}}}{d_{\text{ox}}}$ die Oxidkapazität an und V_B bezeichnet die Substratspannung, mit der die Verarmungszone im Silizium und damit die Schwellspannung des Transistors beeinflusst werden kann. Aufgrund der positiven Drain-Spannung setzt die Inversion am Drain ein. Das Oberflächenpotenzial in der Nähe des Drain-Kontaktes ist dabei in guter Näherung gegeben durch [5]

$$\psi_s \simeq V_D + 2\psi_F \quad (2.9)$$

Abhängig von einer Sättigungsspannung $V_{D,\text{sat}} \equiv V_G - V_t$ werden drei Bereiche unterschieden, in denen ein MOSFET betrieben werden kann:

$$\begin{aligned} V_D &\ll V_{D,\text{sat}} && \text{linearer Bereich} \\ V_D &\lesssim V_{D,\text{sat}} && \text{nichtlinearer Bereich} \\ V_D &\gtrsim V_{D,\text{sat}} && \text{Sättigungsbereich} \end{aligned} \quad (2.10)$$

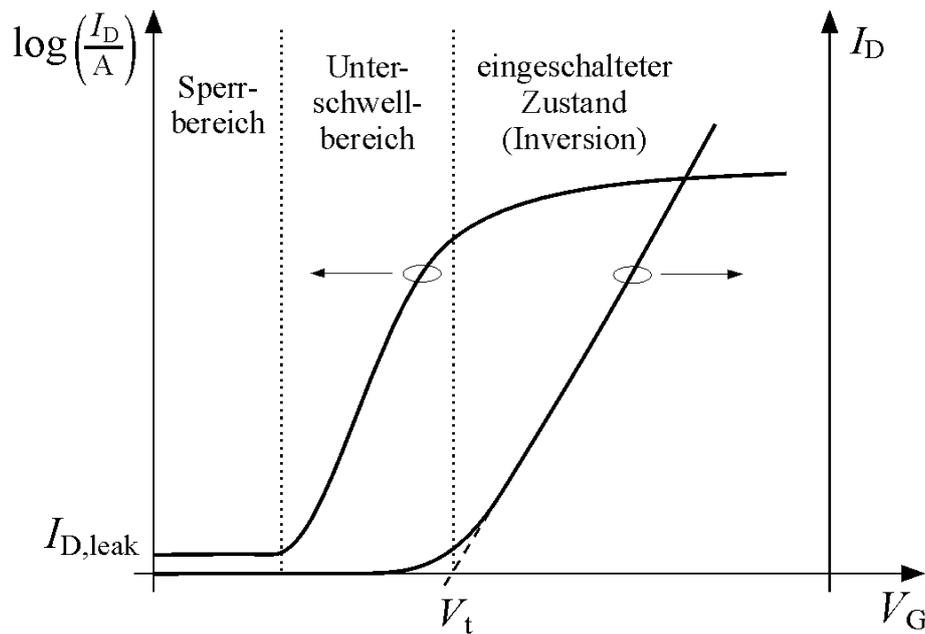


Abbildung 2.5: Transferkennlinie eines MOSFETs. Der Drain-Strom ist in Abhängigkeit der Gate-Spannung sowohl logarithmisch als auch linear aufgetragen. Der Sperrbereich, der Unterschwellbereich und der Bereich, in dem der Transistor eingeschaltet ist (Inversion), sind gekennzeichnet.

In Abbildung 2.7 sind der lineare Bereich und der Sättigungsbereich zusammen mit dem Übergangszustand dargestellt. In (a) ist die Drain-Spannung größer Null aber kleiner als die Differenz zwischen Gate- und Schwellspannung. Der Transistor operiert im linearen Bereich. Die Inversionsladung an der Oberfläche des Halbleiters bildet einen nahezu homogenen leitfähigen Kanal, über den Elektronen vom Source zum Drain fließen. Der Widerstand des Kanals verhält sich in diesem Bereich linear zur angelegten Drain-Spannung und der Strom ist gegeben durch

$$I_D = \frac{W}{L} \mu_n C_{\text{ox}} \left[(V_G - V_t) V_D - \frac{1}{2} V_D^2 \right] , \quad (2.11)$$

wobei $W = W_G$ die Weite des Kanalbereichs angibt und hier mit der physikalischen Gate-Weite gleichgesetzt wurde. Mit μ_n ist die Beweglichkeit der Elektronen im Kanal angegeben. Bei weiter ansteigender Drain-Spannung wird die Inversionsladung in der Nähe des Drains verdrängt. Der Widerstand des Kanals steigt an. In Abbildung 2.7b ist der Zustand skizziert, wenn $V_D = V_{D,\text{sat}}$ ist. Die Inversionsladung am Drain ist vollständig verdrängt und der Kanal an dieser Stelle abgeschnürt (*engl.*: Pinch-off).

Abbildung 2.7c zeigt schließlich den Zustand, in dem die Drain-Spannung die Drain-Sättigungsspannung $V_{D,\text{sat}}$ überschritten hat. Der Pinch-off-Punkt hat sich weiter in Richtung Source bewegt und verkürzt den Kanal dadurch auf die Länge $L' < L$. Dabei hinterlässt er eine Verarmungszone an der Halbleiteroberfläche. Das Potenzial an diesem Punkt entspricht der Drain-Sättigungsspannung und der Widerstand zwischen Pinch-off-Punkt und Drain ist näherungsweise Null. Das bedeutet, dass der Gesamtwiderstand zwischen Source und Drain nicht weiter

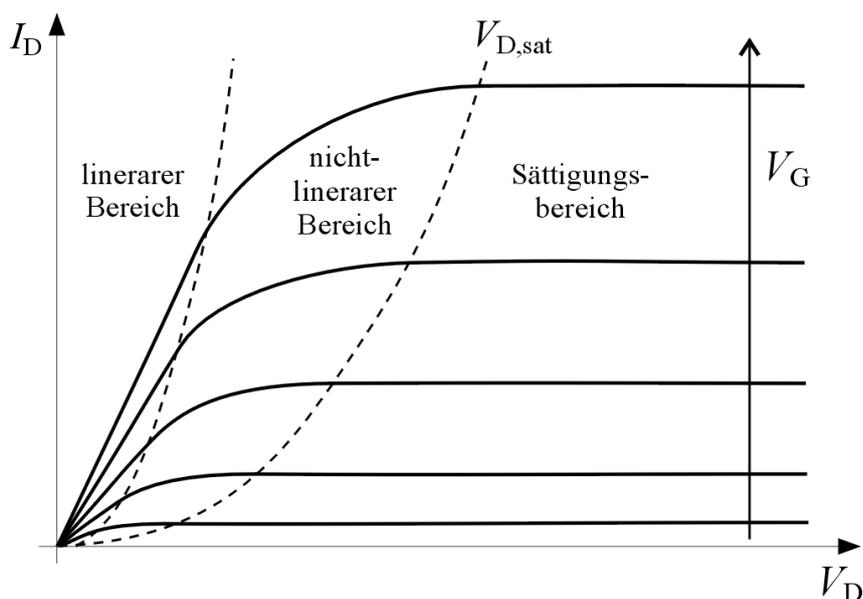


Abbildung 2.6: Ausgangskennlinienfeld eines idealen MOSFETs. Der Drain-Strom ist über der Drain-Spannung für verschiedene Gate-Spannungen aufgetragen. Der lineare Bereich, der nichtlineare Bereich und der Sättigungsbereich sind gekennzeichnet.

steigt und somit der Strom sättigt. Der Drain-Strom in diesem Zustand ist im Idealfall gegeben durch

$$I_{D,\text{sat}} = \frac{W}{L} \mu_n C_{\text{ox}} \frac{(V_G - V_t)^2}{2} \quad (2.12)$$

Ein typisches Ausgangskennlinienfeld eines idealen MOSFETs in Inversion ist in Abbildung 2.6 gezeigt. Hierbei ist der Drain-Strom über der Drain-Spannung für verschiedene Gate-Spannungen aufgetragen. Die drei Bereiche, in denen der Transistor betrieben werden kann, sind gekennzeichnet. Zunächst steigt der Strom linear mit V_D an (linearer Bereich). Aufgrund der kleinen Drain-Spannung ist der Term $\frac{1}{2}V_D^2$ aus Gleichung 2.11 hier vernachlässigbar klein. Im nichtlinearen Bereich nimmt der Einfluss des quadratischen Terms zu und die Steigung des Drain-Stroms mit zunehmender Drain-Spannung verringert sich. Bei Erreichen der Sättigungsspannung $V_{D,\text{sat}}$ befindet sich der Transistor im Sättigungsbereich und der Strom steigt mit zunehmender Drain-Spannung nicht weiter an.

2.1.3 Kurzkanaleffekte

Zur Steigerung der Leistungsfähigkeit der Bauelemente wurden diese in den vergangenen Jahren schrittweise verkleinert. Dies führte unter anderem zur Verkürzung der Schaltzeiten, zur Verringerung von Leistungsverlusten und zu einer Erhöhung der Integrationsdichte. In Folge konnte die benötigte Chipfläche verkleinert und somit eine Kostenreduktion erzielt werden. Die Skalierung der Bauelemente bringt aber auch Probleme mit sich, die im Folgenden näher erläu-

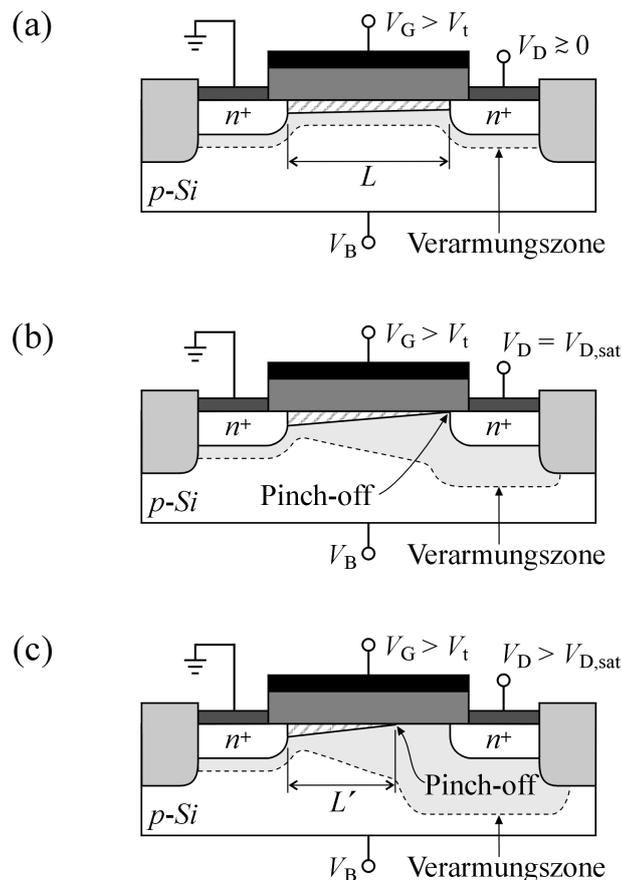


Abbildung 2.7: Betriebsmodi eines eingeschalteten ($V_G > V_t$) MOSFETs. (a) Linearer Bereich ($V_D \gtrsim 0$): Der Kanal ist weitgehend homogen, jedoch nimmt sein Querschnitt nahe Drain mit steigender Drain-Spannung ab. (b) Einsetzen der Sättigung ($V_D = V_{D,sat}$): Der Kanalquerschnitt am Drain schrumpft auf Null und es entsteht ein Abschnürpunkt (Pinch-off). Die Verarmungszone am Drain dehnt sich weiter ins Substrat aus. (c) Sättigung ($V_D > V_{D,sat}$): Der Pinch-off-Punkt bewegt sich in Richtung Source und der Kanal verkürzt sich auf die Länge L' .

tert werden.

Die Betrachtungen aus Abschnitt 2.1.2 beziehen sich auf so genannte Langkanaltransistoren. Das bedeutet, dass die Ausdehnung der Verarmungszonen am Source und Drain klein ist im Vergleich zur Länge L des Transistorkanals. Diese Annahme ist für Transistoren mit kleinen Gate-Längen nicht mehr gültig. Insbesondere die aufgrund der Drain-Spannung vom Drain ausgehenden elektrischen Felder beeinflussen den Kanal zusätzlich zur Gate- beziehungsweise Substratspannung. Damit wird die Steuerwirkung des MOSFETs herabgesetzt, welches sich direkt auf den Drain-Strom auswirkt, der in Sättigung weiter stark ansteigt. Die Feldlinien vom Gate enden in einem Kurzkanaltransistor nicht mehr nur auf der Verarmungsladung Q_B , sondern auch in den pn -Übergängen der Source- und Drain-Gebiete. Dadurch ist für die Inversion weniger Ladung auf dem Gate erforderlich und somit auch weniger Verarmungsladung $Q'_B < Q_B$.

Die effektive Schwellspannung aus Gleichung 2.8 verschiebt sich daher bei einem n -Kanal-MOSFET um den Betrag

$$\Delta V_t = -\frac{1}{C_{\text{ox}}}(Q'_B - Q_B) \quad . \quad (2.13)$$

zu kleineren Werten [8].

Zwei weitere Effekte, die auf die Verkürzung der Kanallänge zurückzuführen sind, sollen in den folgenden Abschnitten näher erläutert werden. Die Generation von heißen Ladungsträgern (HC, *engl.*: hot carriers) führt zur Degradation der Transistoreigenschaften. In nichtflüchtigen Speichern dagegen wird der Effekt der heißen Ladungsträger zum Programmieren der Zellen ausgenutzt. Dies trifft insbesondere für die in dieser Arbeit untersuchten NROM-Speicherzellen zu (siehe Kapitel 2.5).

Ein weiterer Effekt tritt bei hohen Drain-Spannungen auf, wodurch es zu einer elektrostatischen Rückkopplung des Drains auf die Source-Region kommt. Dieser Effekt wird als Drain-Induced Barrier Lowering (DIBL) bezeichnet. Ein an DIBL angelehnter Effekt spielt bei der Auswirkung von grenzflächennahen Oxidstörstellen auf die Rauscheigenschaften eines MOSFETs eine Rolle (siehe Abschnitt 5.4.2).

Heiße Ladungsträger

Bei der Skalierung der Bauelemente wird die Kanallänge überproportional zur Drain-Spannung verringert. Dies erhöht das zum Kanal parallele elektrische Feld \mathcal{E}_{\parallel} an der Halbleiteroberfläche. Im unteren Teil von Abbildung 2.8 ist der typische Verlauf des elektrischen Feldes entlang des Kanals in einem Kurzkanal-MOSFET als gestrichelte Linie wiedergegeben.

Auf ihrem Weg vom Source zum Drain werden die Elektronen durch das Feld extrem beschleunigt. Dabei nehmen sie zusätzlich Energie auf, ohne diese an das Kristallgitter durch Stöße wieder abzugeben. Diese erhöhte Energie, die deutlich oberhalb der Leitungsbandkante vom Silizium liegt, kann auch mit einer erhöhten effektiven Temperatur gleichgesetzt werden, woraus die Bezeichnung heiße Ladungsträger abgeleitet ist. Ist die Energiedifferenz zur Leitungsbandkante höher als die Energiebarriere des Siliziumdioxids ($\phi_B \simeq 3,15$ eV), können die Elektronen ins Oxid injiziert werden, welches zu einem zusätzlichen Gate-Strom führt [2, 9].

Zusätzlich zur Injektion der Elektronen ins Oxid kann ein weiterer Effekt in der Drain-Region auftreten. Die Elektronen verlieren dabei ihre Energie durch Stoßionisation und erzeugen ein Elektron-Loch-Paar. Das erzeugte Elektron fließt über den Drain-Kontakt ab, während das Loch entweder über das Substrat oder über den Source-Kontakt abfließen kann. In Transistoren mit sehr kurzen Kanallängen entsteht hierdurch ein parasitärer $nnpn$ -Bipolartransistor. Ein kleiner Teil der erzeugten heißen Löcher jedoch wird auch Richtung Oxid beschleunigt, welches unerwünschte Nebeneffekte hervorrufen kann [2].

Die generierten heißen Ladungsträger (Elektronen und Löcher) verursachen eine Degradation des Bauelements [10, 11], die auf einen schmalen Bereich in der Nähe des Drains begrenzt ist. Die ins Oxid injizierten Ladungen werden in dort vorhandene Störstellen eingefangen und

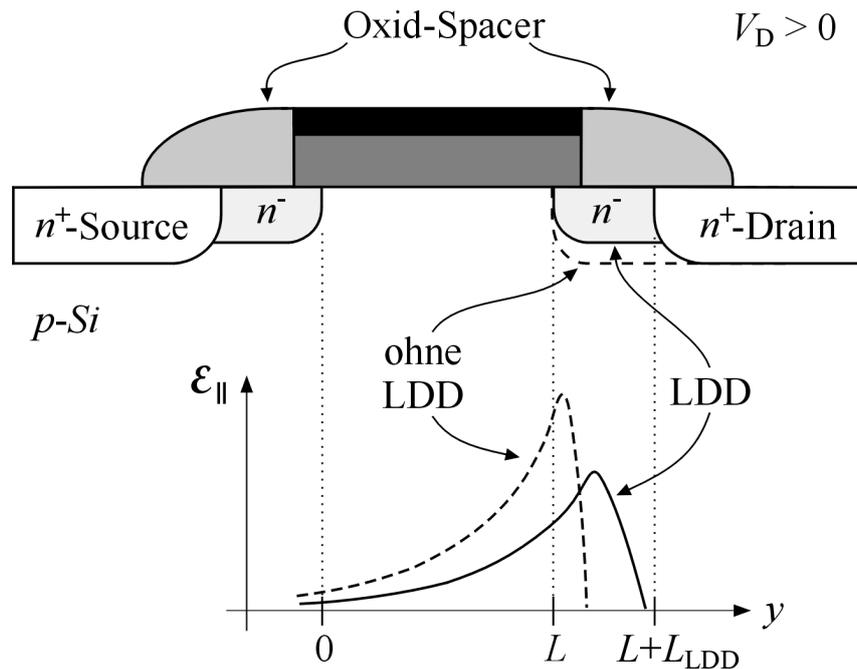


Abbildung 2.8: Verlauf der parallelen Komponente des elektrischen Feldes an der Halbleiteroberfläche in einem Kurzkanal-MOSFET (gestrichelte Linie). Durch die Einführung einer geringer dotierten Zone zwischen dem Kanal und dem Drain-Gebiet (LDD) wird die elektrische Feldstärke am Drain reduziert (durchgezogene Linie).

erzeugen somit als feste Ladungen eine lokale Verschiebung der Schwellspannung [12]. Gleichzeitig generieren die heißen Ladungsträger Grenzflächenzustände an der $Si-SiO_2$ -Grenzfläche [13]. Die Degradation des Bauelements macht sich hauptsächlich in einer Verschiebung der Schwellspannung und einer Verringerung des Drain-Stroms bemerkbar [5].

Um diese Effekte zu minimieren wurden verschiedene technologische Modifikationen eingeführt. Eine Möglichkeit, die Entstehung von heißen Ladungsträgern zu reduzieren bezieht sich auf das Design des Transistors. Bei der so genannten LDD-Struktur (*engl.*: lightly doped drain) wird das Drain-Gebiete am Übergang zum Kanal schwächer dotiert [14]. Dadurch entsteht eine schwach dotierte Region zwischen dem Kanal und dem übrigen hoch dotierten Drain-Gebiet, wie im oberen Teil von Abbildung 2.8 dargestellt. Das elektrische Feld entlang des Kanal wird durch diese Modifikation in der Drain-Region abgeschwächt (durchgezogene Linie in Abbildung 2.8 unten).

Technologisch wird die unterschiedliche Dotierung mittels eines Oxid-Spacers aus Siliziumdioxid realisiert. Dieser befindet sich auf beiden Seiten des Gates. Die Dotieratome werden durch den Spacer ins Si -Substrat implantiert. Aufgrund der variierenden Dicke des Spacers entsteht eine gradueller Verlauf der Dotierkonzentration im darunter liegenden Substrat. Dies geschieht auf beiden Seiten des Gates, also auch in der Source-Region. Nachteilig wirkt sich hierbei der höhere Serienwiderstand aufgrund der geringeren Dotierung aus. Um den Widerstand möglichst gering zu halten, ist es wünschenswert, die geringere Dotierung ausschließlich auf der

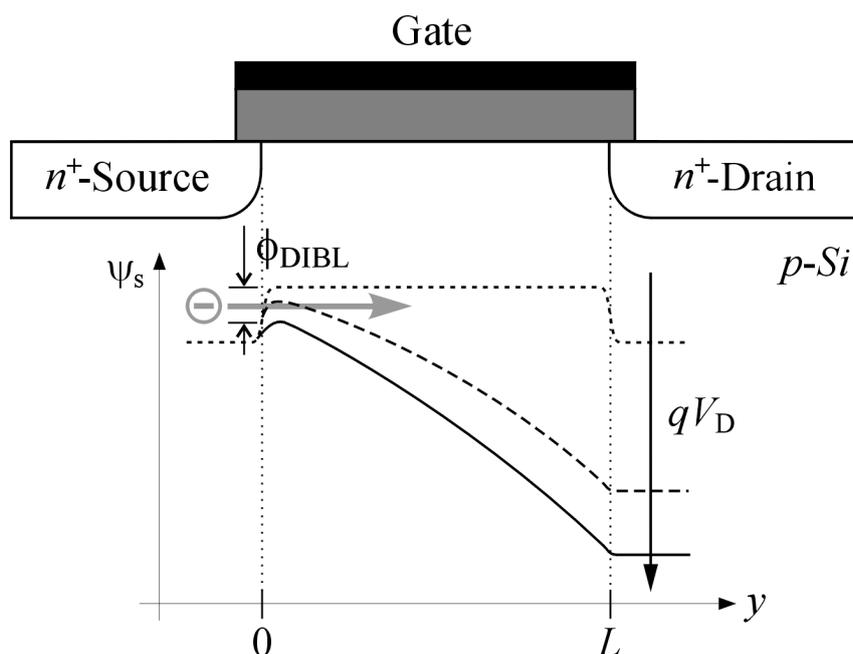


Abbildung 2.9: Darstellung des Oberflächenpotenzials entlang des Transistorkanals. Durch Erhöhung der Drain-Spannung kommt es zum Übergreifen des elektrischen Feldes am Drain auf die Source-Region und damit zur Absenkung der Potenzialbarriere (DIBL) um den Betrag ϕ_{DIBL} .

Drain-Seite anzuwenden. Dies ist jedoch nur über die Einführung weiterer Prozessschritte möglich, die wiederum die Herstellungskosten ansteigen lassen. Wird der Transistor allerdings auch in der entgegengesetzten Richtung betrieben, das heißt mit vertauschten Source- und Drain-Spannungen (vergleiche Kapitel 2.5), wäre die beidseitige LDD-Struktur von Vorteil.

Eine weitere Möglichkeit, die schädigende Wirkung von heißen Ladungsträgern abzuschwächen, besteht darin, das Gate-Oxid resistenter zu machen. Dies geschieht zum Beispiel durch eine Nitridierung des Oxids, welche in Abschnitt 2.3.4 näher beschrieben wird.

Drain-Induced Barrier Lowering (DIBL)

Nehmen die Ausdehnungen der Verarmungszonen am Source und Drain einen Großteil des Transistorkanals ein, kommt es zum Übergreifen des elektrischen Feldes am Drain auf die Source-Region. Dabei wird die Potenzialbarriere am Source in Abhängigkeit der angelegten Drain-Spannung um einen Betrag $\phi_{\text{DIBL}}(V_{\text{D}})$ abgesenkt. Dies wird als drain-induzierte Barrierenabsenkung (*engl.*: drain-induced barrier lowering) bezeichnet. Das Oberflächenpotenzial entlang des Kanals ist im unteren Teil von Abbildung 2.9 dargestellt. Im Sättigungsbetrieb, also bei ausreichend hoher Drain-Spannung und bei einer Gate-Spannung nahe der Schwellspannung, wird der Drain-Strom im Wesentlichen durch die Emission von Ladungsträgern über die Potenzialbarriere am Source bestimmt [5].

Der Extremfall des DIBL tritt ein, wenn sich die Verarmungszonen von Source und Drain berühren. Das elektrische Feld des Drains wirkt dann direkt auf das Source-Gebiet. Dadurch ist der Transistorstrom nur noch durch die Verarmungszone bestimmt und kann nicht mehr über die Gate-Spannung gesteuert werden [2]. Dieser Effekt bekommt bei der Verkürzung der Kanallängen zunehmend Gewicht und stellt große Herausforderungen für das Design zukünftiger Bauelemente dar [15].

2.2 Gate-Oxid und Substratgrenzfläche

Das Gate-Oxid isoliert das Gate elektrisch vom Substrat und im Falle eines MOSFETs vom leitenden Kanal. Es bildet somit das Dielektrikum des MOS-Kondensators. Die typische Dicke dieser Oxidschicht variiert je nach Verwendung des Bauelements. Für den Einsatz in Logikschaltungen (CMOS), wie zum Beispiel Mikroprozessoren, werden Dicken im Bereich von 1,5 nm verwendet [16]. Unterhalb dieser Dimension verliert das Siliziumdioxid rasch seine isolierende Wirkung und setzt somit die Ausfallsicherheit des Bauelements deutlich herab [17, 18]. Für nichtflüchtige Speicher werden dickere Oxide verwendet, da es hier auf eine gute Isolierung zwischen Transistorkanal und Speicherschicht ankommt, damit keine Speicherladungen verloren gehen. Außerdem müssen die Oxide resistent gegen hohe elektrische Feldstärken sein, die beim Programmieren und Löschen der Speicherzellen auftreten [5]. Typische Gate-Oxiddicken für nichtflüchtige Speicher liegen im Bereich von 6 bis 8 nm [19]. Dünnere Oxide sind für Speicheranwendungen problematisch, da die Leckströme durch das Oxid extrem zunehmen [20, 21]. Als Isolatormaterial wird Siliziumdioxid (SiO_2) aufgrund seiner zahlreichen Vorteile bis heute in der Mikroelektronik eingesetzt [5]. SiO_2 besitzt eine große Energiebandlücke von 9 eV, einen hohen Schichtwiderstand von $10^{15} \Omega\text{cm}$ und eine hohe elektrische Durchschlagsfestigkeit von 10^7 V/cm . Mit Silizium bildet es darüber hinaus eine sehr defektarme Grenzfläche [22, 23]. Neuste Generationen von CMOS-Schaltungen verwenden Isolatoren mit hoher Dielektrizitätszahl, so genannte High- k -Materialien, in Verbindung mit Gates aus Metall. Die Firma Intel setzt diese bereits in ihrer Volumenproduktion ein [24].

In den folgenden Abschnitten sollen aber die elektrischen und strukturellen Eigenschaften von Siliziumdioxid näher betrachtet werden.

2.2.1 Elektrische Transportmechanismen

Im Idealfall ist die Oxidschicht isolierend. Dennoch kann ein elektrischer Transport durch das Oxid zum Beispiel auf Basis von quantenmechanischem Tunneln stattfinden. Dabei werden zwei Arten von Tunneln unterschieden, direktes Tunneln und Fowler-Nordheim-Tunneln (FN-Tunneln). Diese Mechanismen sind in Abbildung 2.10 skizziert.

Beim direkten Tunneln durchquert ein Ladungsträger das gesamte Oxid. Dieser Prozess findet

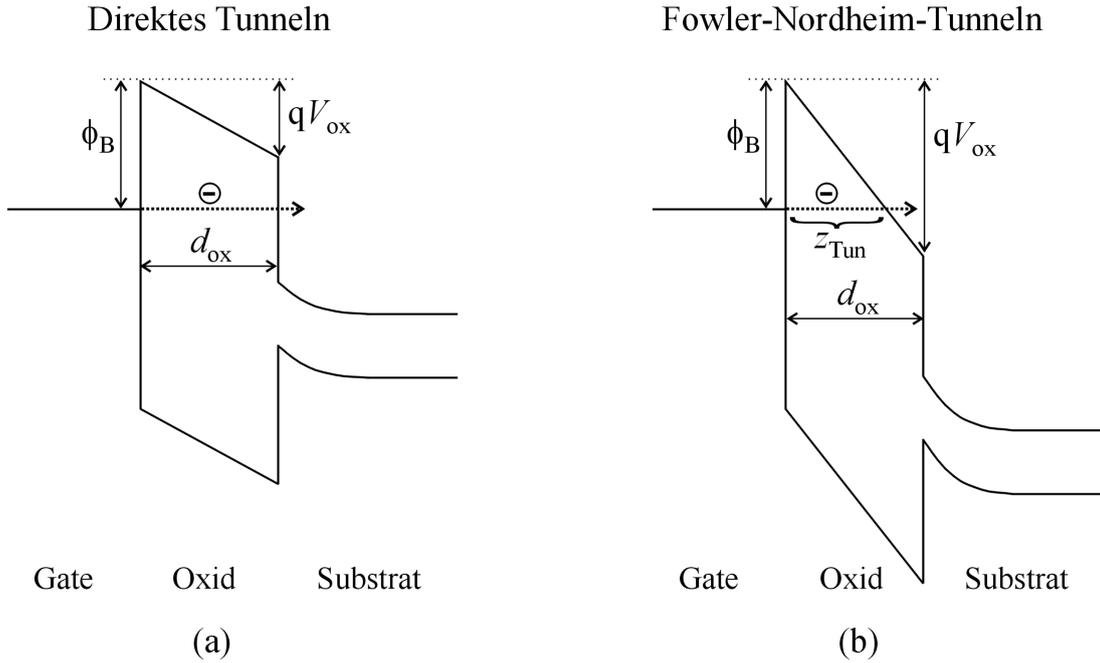


Abbildung 2.10: Tunneln von Ladungsträgern durch Gate-Isolatoren. (a) Beim direkten Tunneln wird der Isolator komplett durchquert. (b) Beim Fowler-Nordheim-Tunneln wird aufgrund der starken Bandverkipfung nur eine Distanz $z_{\text{Tun}} < d_{\text{ox}}$ überwunden.

bei dünnen Isolatoren mit einer Dicke $d_{\text{ox}} \lesssim 4 \text{ nm}$ statt [25].

Die elektrische Tunnelstromdichte für direktes Tunneln in Abhängigkeit des elektrischen Feldes \mathcal{E}_{ox} ist gegeben durch [26]

$$j_{\text{D}}(\mathcal{E}_{\text{ox}}) = \frac{A_{\text{T}} \mathcal{E}_{\text{ox}}}{(1 - K(\mathcal{E}_{\text{ox}}))^2} \cdot e^{-\frac{B_{\text{T}}}{\mathcal{E}_{\text{ox}}} (1 - K^3(\mathcal{E}_{\text{ox}}))} \cdot \left[1 - e^{-\frac{3}{2} \frac{B_{\text{T}} E_{\text{F}}}{\mathcal{E}_{\text{ox}} \phi_{\text{B}}} (1 - K(\mathcal{E}_{\text{ox}}))} \right], \quad (2.14)$$

wobei die vom elektrischen Feld abhängige Konstante $K(\mathcal{E}_{\text{ox}})$ wie folgt definiert ist:

$$K(\mathcal{E}_{\text{ox}}) := \sqrt{1 - \frac{q \mathcal{E}_{\text{ox}} d_{\text{ox}}}{\phi_{\text{B}}}}. \quad (2.15)$$

Die Tunnelkonstanten A_{T} und B_{T} sind gegeben durch

$$A_{\text{T}} := \frac{q^3}{16\pi^2 \hbar \phi_{\text{B}}} \quad (2.16)$$

$$B_{\text{T}} := \frac{4}{3} \frac{\sqrt{2m^* \phi_{\text{B}}^3}}{q \hbar} \quad (2.17)$$

und hängen im Wesentlichen von der Barrierenhöhe ϕ_{B} und der effektiven Elektronenmasse im Oxid m^* ab.

Bei dickeren Oxiden mit $d_{\text{ox}} \gtrsim 4 \text{ nm}$ kann kein direktes Tunneln mehr stattfinden [25]. Stattdessen setzt aber bei ausreichender Bandverbiegung Fowler-Nordheim-Tunneln ein [27, 28],

wie in Abbildung 2.10b dargestellt ist. Im Unterschied zum direkten Tunneln durchqueren die Ladungsträger beim FN-Tunneln eine dreieckige Potenzialbarriere. Die effektive Tunneldistanz z_{Tun} hängt dabei stark vom elektrischen Feld \mathcal{E}_{ox} ab. Für das $Si-SiO_2$ -System mit einer Barrierrhöhe von $\phi_B = 3,15$ eV setzt der FN-Strom etwa bei einer Oxidfeldstärke von 6 MV/cm ein [5]. Die Stromdichte des FN-Tunnelns ist gegeben durch

$$j_{\text{FN}}(\mathcal{E}_{\text{ox}}) = A_T \mathcal{E}_{\text{ox}}^2 \cdot \exp\left(-\frac{B_T}{\mathcal{E}_{\text{ox}}}\right) \quad . \quad (2.18)$$

Die Konstanten A_T und B_T sind bereits durch die Gleichungen 2.16 und 2.17 definiert. Der FN-Strom hängt im Gegensatz zum direkten Tunnel stark vom äußeren elektrischen Feld ab, da sich die Tunneldistanz mit zunehmender Erhöhung des Feldes weiter verkürzt.

Neben dem direkten Tunneln und dem FN-Tunneln treten in Oxiden weitere Ladungstransportmechanismen auf. Ladungsträger können über eine thermische Anregung in das Leitungsband des Isolators angehoben werden (Schottky-Emission). Dieser Mechanismus ist bei Raumtemperatur aber aufgrund der hohen Barriere zwischen Silizium und Siliziumdioxid ($\phi_B = 3,15$ eV) vernachlässigbar.

Weitaus wahrscheinlicher ist das Auftreten von störstellenunterstütztem Tunneln (*engl.*: trap-assisted tunneling) [29]. Hierfür ist das Vorhandensein von Störstellen im Oxid notwendig, die als Zwischenzustände für mehrere aufeinanderfolgende Tunnelvorgänge fungieren. Dabei liegen die Distanzen der einzelnen Tunnelvorgänge im Bereich, in dem direktes Tunneln möglich ist. Störstellen im Oxid können beim Herstellungsprozess entstehen, werden aber in der Regel durch elektrischen Stress des Bauelements erzeugt. Derartige Ströme durch den Isolator werden daher auch als stressinduzierte Leckströme (SILC, *engl.*: stress-induced leakage current) bezeichnet [30, 31].

Findet beim störstellenunterstützten Tunneln zusätzlich eine thermische Anregung statt, so liegt Poole-Frenkel-Emission vor. Hierbei kommt es zum Tunneln eines Ladungsträgers über das Störstellenniveau mit anschließender, durch das elektrische Feld beschleunigter thermischer Anregung ins Oxidleitungsband [32, 33]. Dieser Leitungsmechanismus tritt typischerweise in Nitriden (zum Beispiel Si_3N_4) oder in stark geschädigtem SiO_2 auf [5].

Die elektrischen und strukturellen Eigenschaften von Störstellen werden im folgenden Kapitel näher beschrieben.

2.2.2 Elektrische und strukturelle Eigenschaften von Störstellen

Siliziumdioxid als Gate-Isolator liegt als amorphe Schicht vor. Anders als kristallin hergestellte Gate-Isolatoren aus High- k -Materialien, wie zum Beispiel Gadoliniumoxid (Gd_2O_3) [34], haben die SiO_2 -Moleküle keine langreichweitige Ordnung. In der SiO_2 -Matrix, die auch als α -Quarz bezeichnet wird, sind die Siliziumatome in der Regel über ein Sauerstoffatom miteinander chemisch gebunden. Störungen dieses Bindungsgefüges treten auf, wenn Sauerstoffatome fehlen oder durch Wasserstoffatome ersetzt worden sind. Zusätzlich treten diese beiden Fehler

in unterschiedlichen Konfigurationen auf, die somit eine ganze Reihe an elektrisch aktiven Störstellen bilden.

Darüber hinaus treten weitere Ladungen im Oxid auf, deren Klassifizierung nach Art und Position auf B. E. Deal zurückgeht [35]. In Abbildung 2.11 sind die unterschiedlichen Arten von Oxidladungen dargestellt. Innerhalb des Isolators befinden sich geladene Störstellen (Q_{ot}), die auf fehlenden Sauerstoffatomen basieren. Nahe der Grenzfläche zum Substrat liegen feste positive Ladungen (Q_f) vor, die durch ionisierte Si -Atome hervorgerufen werden. Direkt an den $Si-SiO_2$ -Grenzflächen zum Substrat und zum Poly- Si existieren Grenzflächenladungen (Q_{it}), die intrinsisch vorhanden sind oder durch Stress erzeugt wurden. Des Weiteren können sich bewegliche Alkali-Ionen (Q_m) im Oxid befinden. Dazu gehören in der Regel Li^+ , Na^+ oder K^+ , die allerdings in modernen Herstellungsprozessen keine Rolle mehr spielen.

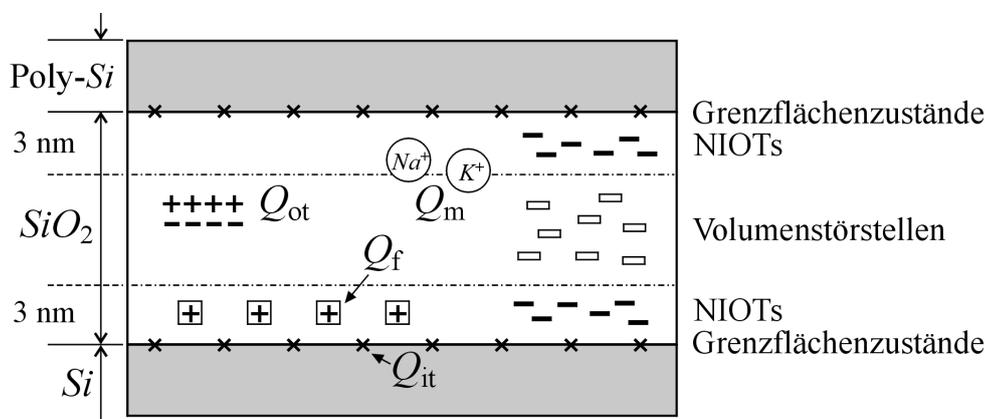


Abbildung 2.11: Klassifizierung von Ladungen nach Art und Position im Oxid [35]. Zusätzlich wird eine räumliche Einteilung in Grenzflächenzustände, grenzflächen-nahe Zustände (NIOts) und Volumenstörstellen vorgenommen [36].

Im Jahre 1992 wurde die Einteilung von Oxidladungen neu vorgenommen und der Begriff der „border-traps“ [36] eingeführt. Darüber hinaus sind auch die Begriffe „grenzflächennahe Oxidstörstellen“ (NIOt, *engl.*: Near-Interface Oxide Traps) oder „slow interface traps“ gebräuchlich [37]. Unter all diesen Bezeichnungen werden Störstellen zusammengefasst, die von den Grenzflächen des Isolators aus beladen werden können. Die Zone der NIOts erstreckt sich dabei bis etwa 3 nm in das Oxid hinein. In der Literatur werden die NIOts im Zusammenhang mit Untersuchungen des $1/f$ -Rauschens auf einen schmalen Bereich begrenzt [38]. Diese Grenze ist jedoch fließend und beruht auf der Annahme, dass die Zeitskalen für Tunnelvorgänge von Ladungsträgern im Bereich von Minuten liegen [36].

Strukturell unterscheiden sich die NIOts nicht von Störstellen im Oxidvolumen [39, 40]. Die Unterscheidung zwischen Störstellen im Oxidvolumen und grenzflächennahen Oxidstörstellen wird in aktuellen Technologiegenerationen, in denen Gate-Oxide in der Regel wesentlich dünner sind als 6 nm, hinfällig. Hier wären quasi alle Volumenstörstellen auch gleichzeitig NIOts.

In dieser Arbeit wird für alle Störstellen, die nicht direkt an der Grenzfläche liegen (schnelle Störstellen) und daher über Tunnelvorgänge kommunizieren, die Bezeichnung NIOTs oder langsame Störstellen verwendet.

Oxidstörstellen durch Sauerstofffehlstellen

In Abbildung 2.12 sind die möglichen Konfigurationen von Sauerstofffehlstellen skizziert, die in der Oxidmatrix vorkommen können. Die Sauerstofffehlstelle (Abbildung 2.12a) wird durch eine direkte Bindung der beiden benachbarten Si -Atome überbrückt (Abbildung 2.12b) [41]. Diese direkte $Si-Si$ -Bindung ist aufgrund der räumlichen Lage der Atome um etwa 4% gegenüber den Bindungsabständen im reinen Si -Gitter gestreckt [42]. Durch Abgabe eines Bindungselektrons oder durch Einfangen eines Lochs entsteht ein E'_δ -Defekt (Abbildung 2.12c). Es bleibt ein einfach positiv geladenes Si -Rumpfatom zurück. Durch eine zusätzliche Umkonfiguration der Bindung, bei der sich das verbleibende Elektron mit einem rückwärtigen Sauerstoffatom bindet, entsteht der metastabile E'_γ -Defekt. Dieser Zustand kann wiederum ein Elektron einfangen (Abbildung 2.12d).

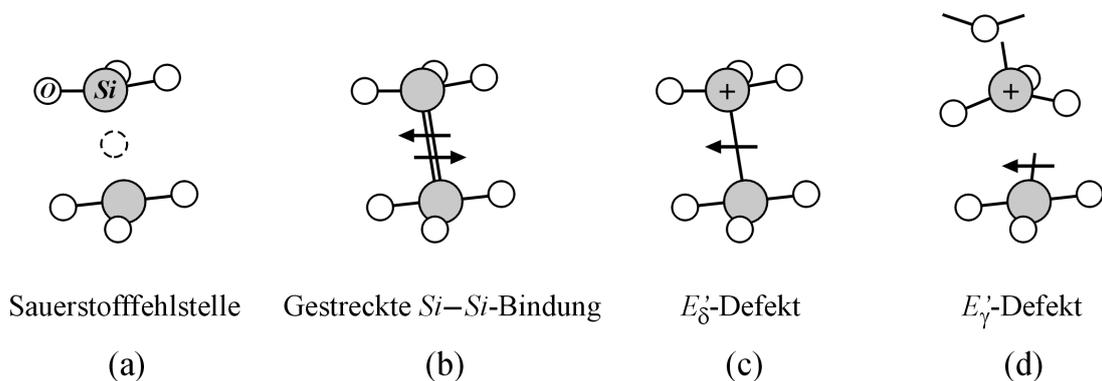


Abbildung 2.12: Sauerstofffehlstellen und mögliche Konfigurationen in einer SiO_2 -Matrix
 (a) Ein Sauerstoffatom fehlt zwischen zwei Si -Atomen. (b) Über der Fehlstelle bildet sich eine gestreckte $Si-Si$ -Bindung. (c) Bei Abgabe eines Bindungselektrons (oder Einfang eines Lochs) bildet sich ein E'_δ -Defekt. (d) Konfiguriert sich die Bindung zusätzlich um, entsteht ein E'_γ -Defekt [41].

Die Bezeichnungen der Defekte gehen auf Messungen der Elektronenspinresonanz (ESR) zurück [43, 44]. Alternative Bezeichnungen der Defekte orientieren sich an der chemischen Struktur oder sind aus unabhängigen Untersuchungen sowohl an der SiO_2 -Matrix als auch an α -Quarz hervorgegangen.

Die in Abbildung 2.12 gezeigten Konfigurationen der Defekte können durch Abgabe beziehungsweise Einfang eines Elektrons ineinander übergehen. So wechselt der neutrale Zustand der gestreckten $Si-Si$ -Bindung durch Abgabe eines Elektrons in den einfach positiv geladenen

E'_δ -Zustand. Dabei ändert sich die energetische Lage der Störstelle relativ zum intrinsischen Energieniveau E_i des Halbleiters. Im neutralen Zustand mit zwei Elektronen beträgt die Störstellenenergie $E_T - E_i = -3,70$ eV, nach Abgabe eines Elektrons $-1,38$ eV [45]. Der E'_γ -Zustand (Abbildung 2.12d) kann durch Einfang eines Elektrons vom einfach positiv geladenen Zustand in den neutralen Ladungszustand wechseln. Dabei wird seine energetische Lage von $E_T - E_i = 0,68$ eV auf $-1,02$ eV abgesenkt [45].

Da die Bindungswinkel einer statistischen Verteilung unterliegen, ergibt sich im Volumen ein Quasi-Kontinuum von energetischen Zuständen im Oxid. Bei der elektrischen Charakterisierung der Störstellen wird daher deren Zahl pro Volumeneinheit und Energieintervall angegeben. Diese Dichte wird in dieser Arbeit mit D_{ot} [$\text{cm}^{-3}\text{eV}^{-1}$] bezeichnet.

Durch Wasserstoff induzierte Oxidstörstellen

Oxidstörstellen entstehen auch mit Beteiligung eines Wasserstoffatoms. Ausgangspunkt ist wiederum die Sauerstofffehlstelle, die im vorigen in Abschnitt beschrieben wurde. Ein Wasserstoffatom ersetzt den Platz des Sauerstoffs und bildet somit eine Wasserstoffbrücke. Diese Wasserstoffbrückenbindung (Abbildung 2.13a) kann wiederum verschiedene Konfigurationen einnehmen. In den Abbildungen 2.13b und 2.13c sind der E'_4 -Defekt und der E'_β -Defekt gezeigt, die sich durch die Lage des Si -Atoms relativ zu den rückgebundenen Sauerstoffatomen unterscheiden [46]. Neben dem neutralen und dem positiv geladenen Zustand sind bei Störstellen, die durch Wasserstoffbrücken erzeugt werden, auch negative Ladungszustände der Defekte möglich. Eine Besonderheit tritt beim neutralen Zustand auf, der je nachdem, ob er mit einem Elektron besetzt ist oder nicht, verschiedene energetische Lagen relativ zum intrinsischen Niveau einnimmt [47].

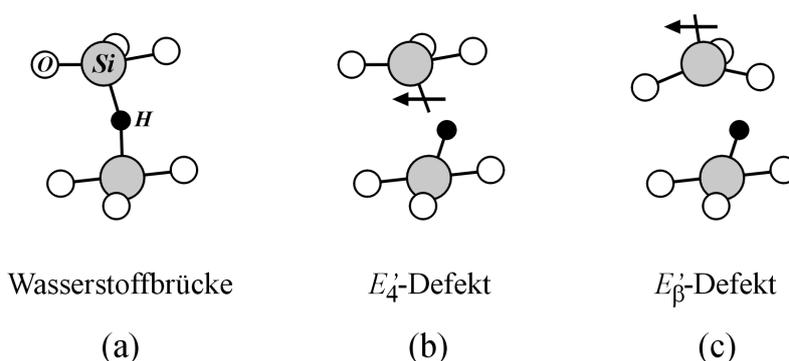


Abbildung 2.13: Wasserstoffbrückenbindungen und mögliche Konfigurationen in einer SiO_2 -Matrix (a) Eine Sauerstofffehlstelle wird durch ein Wasserstoffatom überbrückt. (b) Das Aufbrechen der Bindung eines Si -Atoms zu dem Wasserstoffatom erzeugt einen E'_4 -Defekt. (c) Durch Veränderung der Lage des Si -Atoms relativ zu den rückgebundenen Sauerstoffatomen entsteht ein E'_β -Defekt [46].

Defekte, die durch Wasserstoffbrücken entstehen, spielen eine besondere Rolle bei der Entstehung von stressinduzierten Leckströmen (SILC). Der E'_4 -Defekt erfüllt aufgrund seiner geringen Energielaxation bei Umladung ($\Delta E \leq 2,2 \text{ eV}$) alle Voraussetzungen für stöstellenunterstütztes Tunneln. Begünstigt wird der SILC zusätzlich durch die absolute energetische Lage des Defektes nahe der Si -Bandlücke. Bei höheren elektrischen Feldern im Oxid trägt auch der E'_β -Defekt zum SILC bei [45]. Daher können diese Störstellen schon bei relativ geringen Gate-Spannungen be- oder entladen werden. Die Auswirkungen dieser Störstellen auf die Funktion von MOS-Bauelementen wird näher in Kapitel 2.3 erläutert.

Grenzflächenzustände

In den vorangegangenen Abschnitten wurden Störstellen im Volumen des Siliziumdioxids betrachtet. Des Weiteren treten Defekte aber auch an der Grenzfläche zum Siliziumsubstrat beziehungsweise zum Poly- Si -Gate auf. Letztere sind allerdings messtechnisch schwer zugänglich, so dass unter dem Begriff Grenzflächenzustände meist Störstellen an der Si - SiO_2 -Grenzfläche behandelt werden.

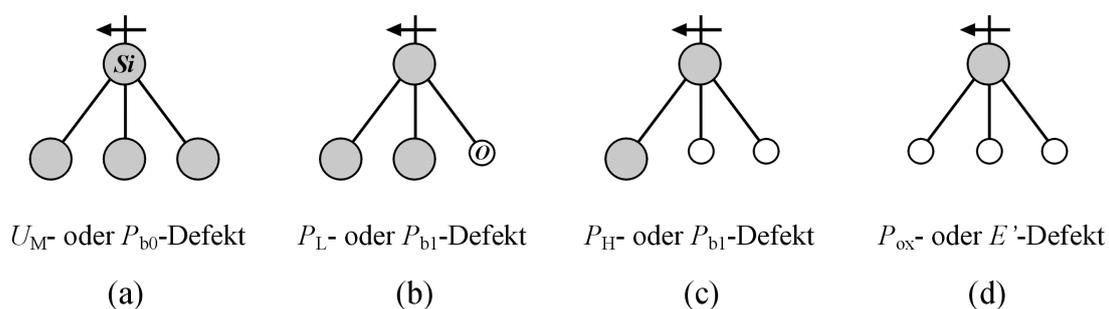


Abbildung 2.14: Chemische Struktur von Zuständen an der Si - SiO_2 -Grenzfläche. (a)-(d) Alle Zustände besitzen ein dangling bond und unterscheiden sich durch die Zahl der rückgebundenen Sauerstoff- beziehungsweise Siliziumatome [48].

Die Defekte im Übergangsbereich vom Silizium zum Siliziumdioxid, der etwa aus zwei Atomlagen besteht [49], unterscheiden sich aus mikroskopischer Sicht von den Störstellen im Oxidvolumen. Jedes Si -Atom ist mit einer unterschiedlichen Zahl von Sauerstoff- beziehungsweise Siliziumatomen gebunden, wie in Abbildung 2.14 dargestellt ist [48]. Alle Grenzflächenzustände bestehen aus einer aufgebrochenen Siliziumbindung (*engl.*: dangling bond). In Abbildung 2.14a ist ein Si -Atom mit drei weiteren Si -Atomen gebunden. Dieser Zustand wird als U_M -Defekt [48] oder aus ESR-Messungen als P_{b0} -Zentrum bezeichnet [50]. Ist das Si -Atom mit einem (Abbildung 2.14b) oder zwei (Abbildung 2.14c) Sauerstoffatomen gebunden, entstehen P_L - beziehungsweise P_H -Defekte, die im Zusammenhang mit ESR-Messungen auch als P_{b1} -Defekte bezeichnet werden [51]. Die vierte Möglichkeit, eine Rückbindung des Si -

Atoms mit drei Sauerstoffatomen, entspricht der SiO_2 -Stöchiometrie (Abbildung 2.14d). Dieser Defekt wird hier als P_{ox} -Defekt bezeichnet und ist bereits aus Abschnitt 2.2.2 als Gruppe der E' -Defekte behandelt worden. Seine energetische Lage befindet sich innerhalb des Si -Leitungsbandes, so dass er mit klassischen Messmethoden, die die Dichte der Grenzflächenzustände innerhalb der Si -Bandlücke bestimmen, nicht zugänglich ist. Zu diesen Messmethoden gehören unter anderem Charge-Pumping, DLTS (*engl.*: Deep Level Transient Spectroscopy) und quasistatische $C(V)$ -Messungen.

Die energetischen Zustände der Störstellen sind auch hier aufgrund von statistischen Schwankungen in den Bindungswinkeln verbreitert und führen zu einem Kontinuum von Zuständen innerhalb der Si -Bandlücke. In der Praxis werden diese Grenzflächenzustände analog zu den Störstellen im Oxidvolumen mit einer flächen- und energiebezogenen Dichte D_{it} [$cm^{-2}eV^{-1}$] ausgedrückt. Abbildung 2.15 zeigt den Verlauf der Dichte der verschiedenen Defekte über der Energie.

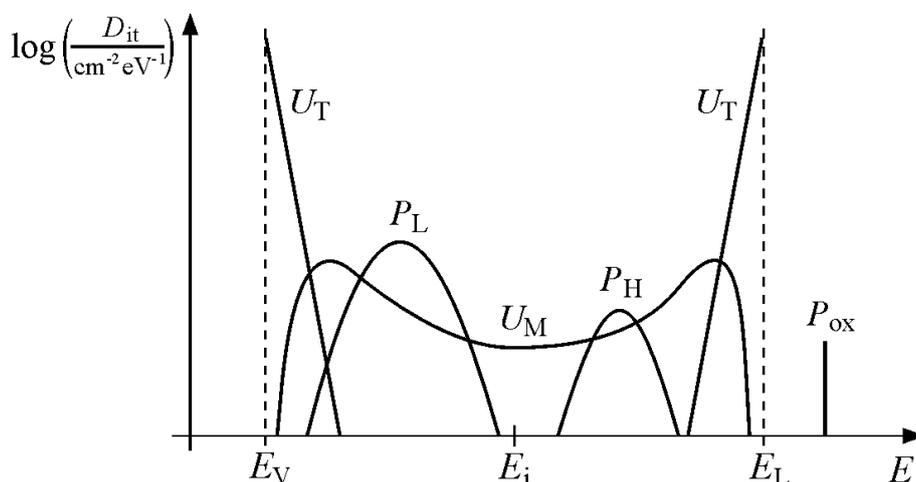


Abbildung 2.15: Zustandsdichten verschiedener Defekte an der Si - SiO_2 -Grenzfläche über der Energie. Die U_M -, P_L - und P_H -Defekte verteilen sich innerhalb der Si -Bandlücke, während der P_{ox} -Zustand und die Tail-Zustände (U_T) innerhalb des Leitungsbandes beziehungsweise nahe der Bandkanten liegen [48].

Zusätzlich zu den bereits behandelten dangling-bond-Defekten existieren so genannte Tail-Zustände, die als U_T -Defekte bezeichnet werden. Sie entstehen durch gestreckte $Si-Si$ -Bindungen und liegen energetisch nahe der Valenz- beziehungsweise Leitungsbandkante des Halbleiters.

Mit messtechnischen Verfahren wie zum Beispiel $C(V)$ -Messungen wird stets die Summe der Zustandsdichten aller Grenzflächendefekte gemessen [52]. Beim Charge-Pumping kann darüber hinaus nur eine mittlere Dichte der Zustände innerhalb eines bestimmten Energieintervalls ΔE (siehe Abschnitt 3.2.1) bestimmt werden [53].

Die Dichte von Grenzflächenzuständen kann in der heutigen Bauelementefertigung auf Werte in

der Größenordnung von $10^{10} \text{ cm}^{-2} \text{ eV}^{-1}$ reduziert werden. Eine wesentliche Verbesserung der Grenzflächenqualität wird dabei durch eine Temperung (POA, *engl.*: Post Oxidation Anneal) erzielt, die nach der Oxidierung durchgeführt wird. Während der Temperung, die in Wasserstoffatmosphäre durchgeführt wird, werden offene *Si*-Bindungen an der Grenzfläche mit Wasserstoffatomen abgesättigt und somit elektrisch inaktiv [54]. Generell ist die Zustandsdichte in Oxiden auf (100)-orientierten *Si*-Wafers geringer als auf (111)-Wafers [55]. Dies liegt an der geringeren Zahl von *Si*-Atomen an der (100)-Oberfläche und der damit verbundenen geringeren Zahl von Bindungspartnern. In der Folge haben sich (100)-orientierte Wafer als Substrat in der Halbleiterindustrie durchgesetzt.

2.3 Degradationseffekte

Im elektrischen Betrieb der Bauelemente treten Effekte auf, die das Bauelement und insbesondere dessen Isolator schädigen. Alle diese Schädigungsmechanismen, die in den folgenden Abschnitten näher erläutert werden, erzeugen im Oxid und an dessen Grenzflächen Störstellen [56, 57], wie sie im vorangegangenen Abschnitt 2.2.2 beschrieben wurden. Mithilfe von störstellenunterstützten Tunneln tragen diese zum Auftreten von Leckströmen durch das Oxid bei [58] und können letztendlich zum elektrischen Durchbruch des Isolators führen [59].

Bei CMOS-Bauelementen ist insbesondere auch eine Verschiebung der Schwellspannung als kritisch zu betrachten. Hervorgerufen wird dies durch den Einbau von festen Ladungen in der Oxidmatrix unter Stressbedingungen. Darüber hinaus werden durch Stress Grenzflächenzustände generiert, die ebenfalls zur Verschlechterung der Eigenschaften des Bauelements führen. In MOSFETs, die als Speicherzellen arbeiten (zum Beispiel NROM), treten oftmals Degradationseffekte durch heiße Ladungsträger auf. Diese entstehen bei der Programmierung und beim Löschen der Speicherzelle. In MOSFETs treten heiße Ladungsträger aber auch im Zuge der Verkürzung der Kanallänge auf (siehe Abschnitt 2.1.3).

Abbildung 2.16 zeigt einige mögliche Schädigungsmechanismen. Das Auftreten dieser Effekte hängt von der zur Verfügung stehenden Energie ab [23]. Bei geringen Spannungen, bei denen moderne MOSFETs arbeiten, findet eine Generation von Störstellen durch die Freisetzung von Wasserstoff (HR, *engl.*: Hydrogen Release) statt (a). Hierfür ist eine Mindestenergie von 2 eV relativ zur unteren Kante des Oxidleitungsbandes nötig [60]. Bei höheren Spannungen, wenn eine Schwellenenergie von etwa 5 eV zur Verfügung steht, tritt Anoden-Löcherinjektion (AHI, *engl.*: Anode Hole Injection) auf (b) [61] und ab 9 eV, welches der Energiebandlücke von SiO_2 entspricht, setzt die Generation von Elektronen-Loch-Paaren durch Stoßionisation (II, *engl.*: Impact Ionization) ein (c) [62]. Dieser Mechanismus ist nur in Oxiden präsent, deren Dicke im Bereich von 20 nm und darüber liegt, da die Elektronen in dünneren Oxiden nicht genügend Energie aus dem elektrischen Feld beziehen können [63].

Die Effekte der Wasserstofffreisetzung und der Anoden-Löcherinjektion werden in den folgenden Kapiteln eingehender beschrieben. Der Mechanismus der Stoßionisation wird dagegen

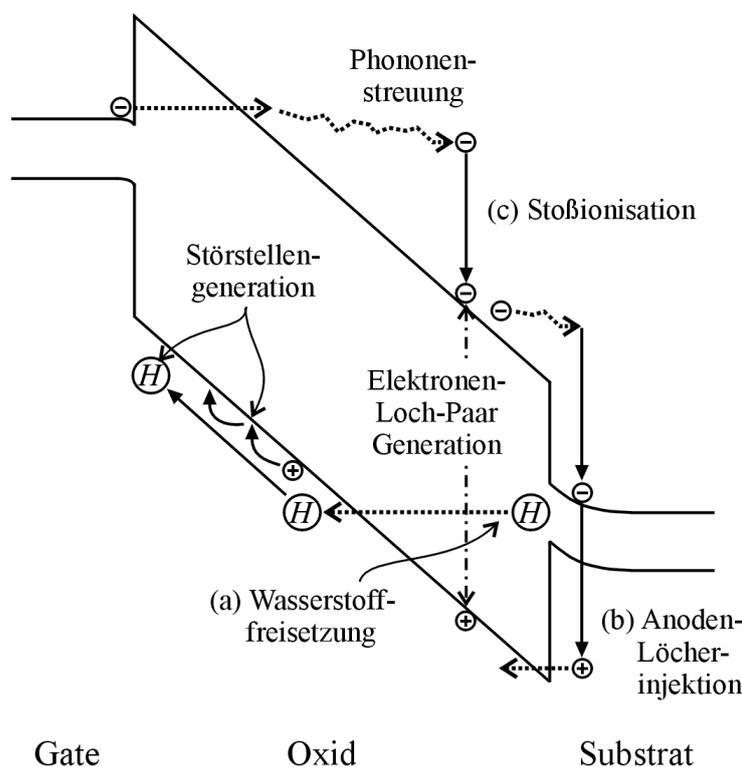


Abbildung 2.16: Mögliche Schädigungsmechanismen, die im SiO_2 in Abhängigkeit der Elektronenenergie auftreten können. (a) Wasserstofffreisetzung (HR). (b) Anoden-Löcherinjektion (AHI). (c) Elektronen-Loch-Paar-Generierung durch Stoßionisation.

nicht näher erläutert, da die in dieser Arbeit untersuchten Transistoren wesentlich kleinere Oxiddicken aufweisen.

2.3.1 Wasserstofffreisetzung

Wie bereits in Abschnitt 2.2.2 erwähnt, sind die Grenzflächen des Isolators zum Si -Substrat beziehungsweise zum Poly- Si durch einen Nachbehandlungsprozess in Wasserstoffatmosphäre mit H -Atomen abgesättigt (passiviert). Die Reaktionsgleichung für die Passivierung in molekularem Wasserstoff lautet



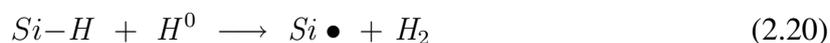
Dabei symbolisiert $Si \bullet$ eine ungesättigte Siliziumbindung.

Der gebundene Wasserstoff kann unter elektrischen Stressbedingungen auf der jeweiligen Anodenseite freigesetzt werden und in Richtung Kathode diffundieren (siehe Abbildung 2.16a) [60, 62, 64]. Die Grenzfläche wird dabei depassiviert. Auf dem Weg durch das Oxid bildet der Wasserstoff vorwiegend in der Nähe der Grenzflächen des Isolators $Si-O-H$ -Defekte,

die sehr effiziente Einfangzentren für Elektron darstellen [65]. Diese wiederum formen SiO^- -Defekte, die zu einer Anhäufung von negativen Ladungen nahe der Grenzfläche führen können [66].

Direkt an der Grenzfläche bleiben ungesättigte Bindungen (dangling bonds) zurück und bewirken einen Anstieg der Grenzflächenzustandsdichte. In ESR-Messungen wurden diese als P_b -Defekte identifiziert [67]. Für diesen Prozess sind heiße Elektronen mit einer Energie von etwa 2 eV relativ zum Oxidleitungsband notwendig [60]. Ob der Wasserstoff direkt durch die heißen Elektronen [68, 69] oder durch Löcher [70, 71] freigesetzt wird, ist in der Literatur nicht eindeutig geklärt. Der Generationsprozess über Löcher deutet auf einen Zusammenhang des HR-Modells mit dem im nächsten Abschnitt behandelten Modell der Anoden-Löcherinjektion hin [16]. Neuere Untersuchungen zeigen, dass nicht nur heiße, sondern auch kalte Löcher an der Freisetzung von Wasserstoff beteiligt sind [72].

Der Wasserstoff an der Grenzfläche kann auch unter Beteiligung von atomarem Wasserstoff H^0 freigesetzt werden. Dies lässt sich über die Reaktionsgleichung



beschreiben [50]. In diesem System kann der atomare Wasserstoff aber auch eine Passivierung über die Reaktion



bewirken. Die Reaktionen nach Gleichungen 2.20 und 2.21 treten bei Raumtemperatur gleichzeitig auf, wobei die Passivierung mit atomarem Wasserstoff bevorzugt wird [73].

Um dieses Modell zu überprüfen, wurden in der Literatur Experimente durchgeführt, in denen das Oxid mit Deuterium anstatt mit Wasserstoff behandelt wurde. Ein Deuteriumatom besitzt eine höhere Masse als Wasserstoff. Daher wurde erwartet, dass die Depassivierung der Grenzfläche mit einer geringeren Rate vonstatten geht als bei Wasserstoff passivierten Grenzflächen [74]. Eine Verbesserung der Oxideigenschaften in Bezug auf Durchbruchfestigkeit und Grenzflächenzustände konnte nicht in allen Arbeiten beobachtet werden [75]. In späteren Arbeiten wurde der so genannte Isotopeneffekt allerdings nachgewiesen: In Proben, die mit Deuterium behandelt wurden, konnte eine Verringerung der durch Stress verursachten Leckströme (SILC) und der Flachbandspannungsverschiebung festgestellt werden [76–78]. Ein Grund für die scheinbar widersprüchlichen Aussagen könnte sein, dass der Isotopeneffekt bei zu geringen Gate-Spannungen nicht auftritt [79, 80].

Das Modell der Wasserstofffreisetzung ist heute weitgehend etabliert, welches durch zahlreiche Untersuchungen bestärkt wird. Die Beteiligung von Wasserstoff ist dabei signifikant. So konnte gezeigt werden, dass durch Wasserstoff gesättigte Oxidschichten eine beschleunigte Degradation des Isolators bewirken [60, 81, 82]. Des Weiteren zeigten reine SiO_2 -Schichten, die einer Atmosphäre aus atomarem Wasserstoff ausgesetzt waren, eine deutliche Degradation der Grenzfläche [50, 54, 73]. Diese Degradation verhielt sich ähnlich zu Schädigungen, die unter elektrischem Stress oder bei Bestrahlung auftraten, allerdings ohne dass ein elektrisches Feld beteiligt war [23].

2.3.2 Anoden-Löcherinjektion

Verschiedene Arbeiten zeigen, dass der Transport von Löchern durch das Oxid Schädigungen hervor ruft [83–87]. Diese Löcher werden an der jeweiligen Anodenseite des Isolators erzeugt, indem heiße Elektronen ihre Energie an ein Elektron im Valenzband der Anode abgeben. Dieses wird dadurch ins Leitungsband gehoben und erzeugt dabei ein heißes Loch, das genügend Energie besitzt, um die Barriere zum Oxid in Richtung Kathode zu überwinden (siehe Abbildung 2.16b) [59]. Beim Passieren des Oxids erzeugen diese heißen Löcher Störstellen. Der Mechanismus wird über die Präsenz von gestreckten $Si-O$ -Bindungen [88] erklärt, die durch Einfangen eines Lochs aufbrechen können [89]. Dabei bewegt sich das Sauerstoffatom auf eine Position, auf der es seinen mechanischen Stress relaxieren kann und hinterlässt eine offene Si -Bindung [84]. Letztere stellt nach dem Einfang eines Loches eine positiv geladene Störstelle dar, die in ESR-Messungen nachgewiesen wurde [90]. Das Sauerstoffatom kann als neutrale Störstelle für Elektronen dienen. Gestreckte $Si-O$ -Bindungen befinden sich vorwiegend in der Nähe der Grenzflächen in einem Bereich von 1 – 3 nm, welches durch XPS-Messungen (*engl.*: X-ray Photoelectron Spectroscopy) nachgewiesen wurde [91]. Allerdings wurde in Experimenten auch eine homogene Verteilung von neutralen Störstellen im Oxid gefunden [86]. Dies deutet darauf hin, dass $Si-O$ -Bindungen generell aufgebrochen werden und nicht nur die gestreckten [92].

Die Energie, die für die Erzeugung der heißen Löcher an der Anode notwendig ist, beträgt etwa 5 eV relativ zur Leitungsbandkante des Oxids. Es konnte in der Literatur allerdings gezeigt werden, dass eine Löcherinjektion auch unterhalb dieser Energieschwelle stattfindet [93]. Dies liefert eine mögliche Erklärung für die Generation von Störstellen bei kleinen Gate-Spannungen [94, 95].

2.3.3 Negative Bias Temperature Instability (NBTI)

In p -Kanal-MOSFETs wurde bei Stressexperimenten unter negativen Gate-Spannungen und bei Temperaturen oberhalb der Raumtemperatur ein weiterer Degradationsmechanismus identifiziert [96]. Dieser Effekt, für den sich in der Literatur der Begriff „Negative Bias Temperature Instability“ (NBTI) durchgesetzt hat, wurde bereits 1967 von Deal *et al.* beschrieben [97]. NBTI äußert sich durch eine Zunahme der Grenzflächenzustandsdichte und einer negativen Verschiebung der Flachbandspannung [96, 98, 99], welches auf eine Anhäufung positiver Ladungen im Oxid hindeutet. In n -Kanal-MOSFETs ist dieser Effekt für beide Spannungspolaritäten vernachlässigbar, ebenso wie für positive Gate-Spannungen in p -MOSFETs [100, 101]. In neusten CMOS-Schaltungen mit Gate-Isolatoren aus einem High- k -Material mit einem Metall-Gate gewinnt jedoch ein analoger Effekt an Bedeutung: Unter positiver Gate-Spannung degradieren die n -Kanal-MOSFETs ähnlich stark oder sogar stärker als die p -MOSFETs unter negativer Spannung. Dieser Effekt wird daher als PBTI (*engl.*: Positive Bias Temperature Instability) bezeichnet [102, 103].

Der NBTI-Effekt wurde mit dem Freisetzen von Wasserstoff analog zum HR-Mechanismus (siehe Abschnitt 2.3.1) erklärt, bei dem Grenzflächenzustände (vorwiegend P_b -Defekte) erzeugt werden [99, 104, 105]. In der Literatur wird dieser Vorgang mit dem „Reaction-Diffusion“-Modell beschrieben [106]. Die Beobachtung einer Verringerung der Grenzflächenzustandsdichte nach Abschalten des Stresses (Ausheilung, *engl.*: recovery) bestärkt diese Theorie. Über eine Rückdiffusion von Wasserstoff zur Grenzfläche und die erneute Passivierung der dangling bonds kann dieses Ausheilen der Grenzfläche erklärt werden [107, 108]. Des Weiteren findet der Effekt Unterstützung im Auftreten eines Isotopeneffekts nach der Temperaturbehandlung der Proben in einer Deuteriumatmosphäre [101]. Deuterium hat aufgrund seiner höheren Masse im Vergleich zu Wasserstoff eine geringere Diffusionskonstante. Der NBTI-Effekt ist damit durch Diffusionsprozesse limitiert, welches die starke Temperaturabhängigkeit erklärt [109, 110].

Die Anhäufung positiver Ladung im Oxid durch NBTI-Stress kann ebenfalls durch das Aufbrechen einer $Si-H$ -Bindung erklärt werden. Im Unterschied zur Generation von P_b -Defekten an der Grenzfläche ist das Si -Atom hierbei zu drei Sauerstoffatomen rückgebunden [111]. Dieser Defekt entspricht dem P_{ox} - beziehungsweise E' -Zentrum (siehe Abbildung 2.14d) und liegt nahe der $Si-SiO_2$ -Grenzfläche. Durch Einfang eines Lochs wird dieser Defekt positiv geladen und erzeugt die beobachtete negative Verschiebung der Schwellspannung [101, 112].

Allerdings tritt der Effekt auch bei Stress unter geringen Gate-Spannungen auf, bei dem die Konzentration von heißen Löchern auf der Anodenseite vernachlässigbar klein ist [96]. In neuesten Untersuchungen des NBTI-Effekts wurde während des elektrischen Stresses auch die Substratspannung V_B der Transistoren variiert. Darin wurde gezeigt, dass $Si-H$ -Bindungen auch durch Löcher des Inversionskanals aufgebrochen werden [108]. Dies erklärt insbesondere die erhöhte Degradation der p -Kanal-MOSFETs unter Stress schon bei kleinen negativen Gate-Spannungen im Vergleich zur Degradation bei positiven Gate-Spannungen.

In anderen Arbeiten dagegen wird die Degradation der Grenzfläche ausschließlich auf die Depassivierung der P_b -Defekte durch atomaren Wasserstoff zurück geführt. Dieser wird demnach im Gegensatz zum HR-Modell nicht an der $Si-SiO_2$ -Grenzfläche freigesetzt, sondern an der gate-seitigen Grenzfläche [113]. Begründet wird dies durch die nicht zur Verfügung stehende Aktivierungsenergie von etwa 2,8 eV für Löcher [114], um die wasserstoffgesättigte Bindung des P_b -Zentrums aufzubrechen. Für das Aufbrechen der Bindung mittels atomaren Wasserstoff dagegen reicht eine Aktivierungsenergie von 0,3 – 0,4 eV aus [115].

2.3.4 Nitrierte Oxide

Die Erzeugungsrate von Störstellen im Volumen oder an der Grenzfläche des Isolators kann durch eine Nitridierung des Oxids herabgesetzt werden [116–118]. Insbesondere wird der Einfang von negativer Ladung reduziert [119]. Oxinitride (SiO_xN_y) verringern die Diffusion von Dotieratomen und anderen Verunreinigungen [120]. Ein wesentliches Ziel bei der Nitridierung des Oxids ist die Verringerung von Gate-Leckströmen (SILC) [121–123] und die Unterdrückung der Bor-Diffusion in p -MOSFETs [124–126]. Des Weiteren weisen nitrierte Gate-

Oxide eine höhere Beständigkeit gegen Strahlung [127, 128] und hohe elektrische Feldstärken auf [89]. Auf das Degradationsverhalten von MOSFETs wurde ebenfalls eine positive Wirkung beobachtet. Eine mäßige Nitridierung verringert den NBTI-Effekt und senkt die Degradation durch heiße Ladungsträger [126, 129]. Bei zu starker Nitridierung nimmt die NBTI-Degradation allerdings wieder zu.

Die Herstellung von Oxinitriden kann über einen thermischen Oxinitridierungsprozess oder über chemische oder physikalische Abscheidungsverfahren erfolgen [126]. Bei den thermischen Prozessen wird das Silizium direkt unter Verwendung von N_2O oder NO oxidiert. Der Vorteil dieser Methode liegt darin, dass kein Wasserstoff in die Schicht eingebaut wird. Eine weitere Möglichkeit, ein Oxinitrid herzustellen, ist die Temperung einer SiO_2 -Schicht in Ammoniak (NH_3). Danach ist allerdings ein Reoxidationsschritt notwendig, um den beteiligten Wasserstoff wieder zu entfernen, da sonst eine Verschlechterung der elektrischen Eigenschaften eintritt [130, 131]. Eine ausführlichere Beschreibung der Herstellungsmethoden findet sich in [126]. Die Stickstoffkonzentration im Oxid ist nicht homogen verteilt. Nahe der Grenzfläche zum Silizium findet eine Anhäufung von Stickstoffatomen statt [132, 133]. Dies beruht auf der Tatsache, dass N -Atome hauptsächlich mit $Si-Si$ -Bindungen reagieren und seltener mit $Si-O$ -Bindungen [126].

In Abbildung 2.17 sind die möglichen Konfigurationen eines Oxinitrids an der Grenzfläche zum Silizium dargestellt. Durch die um Eins höhere Valenz des Stickstoffatoms im Vergleich zum Sauerstoff, gibt es eine Vielzahl von Einbaumöglichkeiten des Stickstoffs [5]. In (a) und (c) ist das N -Atom an drei benachbarte Si -Atome gebunden, in (b) nur zu zwei, wobei die dritte

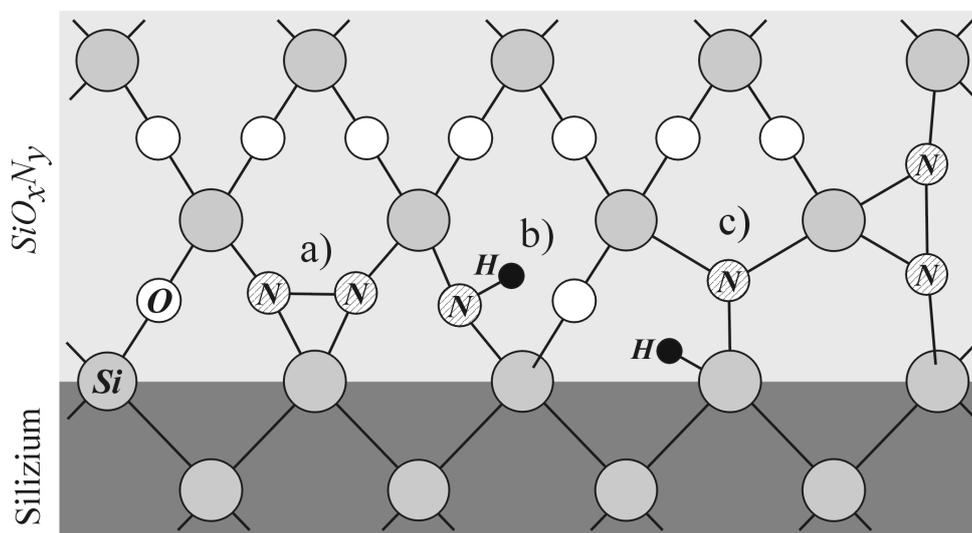


Abbildung 2.17: Mögliche Konfiguration eines Oxinitrids nahe der $Si-SiO_2$ -Grenzfläche [5]. (a) Der Stickstoff bildet entsprechend seiner Valenzen drei Bindungen zu Nachbaratomen. (b) Die Sauerstoffbrücke wurde durch Stickstoff ersetzt und die offene Bindung durch ein Wasserstoffatom gesättigt. (c) Ähnliche Situation wie (a), jedoch bleibt an der Grenzfläche zum Silizium ein dangle bond zurück, das durch ein H -Atom abgesättigt sein kann.

offene Bindung durch ein H -Atom abgesättigt ist. Der Einbau von Stickstoff führt aufgrund der stärkeren Bindung zu einer größeren Stabilität des Oxinitrids im Hinblick auf die Generation von Störstellen [126]. Ähnlich wie im Siliziumdioxid wird die Defektentstehung im Oxinitrid ebenfalls auf die Erzeugung von dangling bonds zurückgeführt [134].

Bezüglich des NBTI-Effekts wurde in nitridierten Oxiden jedoch eine erhöhte Degradation beobachtet [129, 135, 136]. Zurückgeführt wird dies auf eine höhere Zahl von Oxidstörstellen, die durch den Einbau von Stickstoff entstehen [137, 138]. Der Stickstoff bildet Si_3-N -Bindungen ähnlich den Sauerstoffbrücken in einem reinen Siliziumdioxid (siehe Abbildung 2.17). Beide Bindungsarten agieren als Einfangzentren für Löcher, jedoch ist die Reaktionsenergie bei den $Si-N$ -Bindungen geringer als bei den $Si-O$ -Bindungen [139]. Die Stickstoffzentren fangen somit schneller Löcher ein, woraus die für NBTI typische Anhäufung positiver Ladungen im Oxid resultiert. Des Weiteren werden durch den Löchereinfang Zentren für Wasserstoff gebildet [140], die den diffusiven Transport des Wasserstoffs beeinflussen [141, 142]. Es hat sich gezeigt, dass die Degradation des Bauelements durch NBTI stärker ist, wenn die Stickstoffkonzentration an der $Si-SiO_2$ -Grenzfläche hoch ist. Eine geringere Degradation ergab sich in nitridierten Oxiden, in denen sich die größte Stickstoffanhäufung in einem größeren Abstand von der Grenzfläche befand [143].

2.4 Der MOSFET in Logikschaltungen (CMOS)

Je nach Anwendung von aktiven MOSFETs in modernen hochintegrierten Schaltungen werden unterschiedliche Anforderungen an den Gate-Isolator gestellt. In dieser Arbeit sollen hauptsächlich die Anforderungen behandelt werden, die die Leistungsfähigkeit des Bauelements und die Zuverlässigkeit (*engl.*: reliability) des Isolators betreffen. MOSFETs werden als aktives Bauelement überwiegend in Logikschaltungen (CMOS, *engl.*: Complementary MOS) und in nichtflüchtigen Speichern (NVSM, *engl.*: Non-Volatile Semiconductor Memory) eingesetzt. Als passives Bauelement, zum Beispiel als Auswahltransistor, finden sie in DRAM- und FLOTOX-Speichern Anwendung [5].

Die speziellen Anforderungen an den Gate-Isolator eines MOSFETs in nichtflüchtigen Speichern wird im nächsten Kapitel behandelt. Dieses Kapitel beschreibt zunächst die Auswirkungen von Degradationseffekten in Logikschaltungen, welche überwiegend in Mikroprozessoren und als periphere Schaltungen von Speichern verwendet werden. Das Grundelement einer CMOS-Schaltung ist der Inverter, der aus einem p - und einem n -MOSFET besteht [5].

Abbildung 2.18 zeigt ein vereinfachtes Schaltbild eines Inverters, der mit einer Versorgungsspannung V_{DD} gegenüber einem Massepotenzial betrieben wird. Das Ausgangspotenzial V_{out} ist stets dem Eingangspotenzial V_{in} entgegengesetzt, das heißt eine Spannung V_{DD} am Eingang zieht den Ausgang auf Massepotenzial und umgekehrt. Die Eingangsspannung wird auf die beiden Gates der Transistoren gegeben, wobei im Falle von $V_{in} = V_{DD}$ der p -Kanal-MOSFET sperrt und der n -Kanal-MOSFET einschaltet. Die Spannung am Ausgang wird dabei mit dem

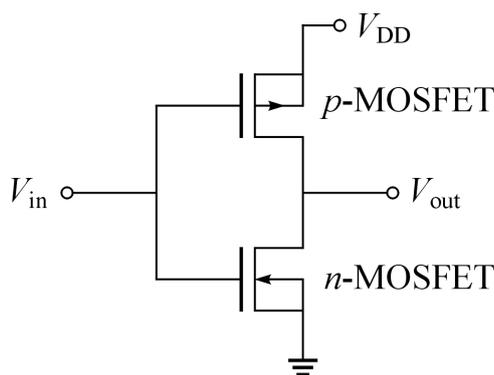


Abbildung 2.18: Schaltbild eines CMOS-Inverters. In Abhängigkeit der Spannung (V_{DD} oder Massepotenzial) am Eingang V_{in} schaltet einer der beiden Transistoren durch, während der andere sperrt. Am Ausgang V_{out} liegt jeweils das entgegengesetzte Potenzial im Vergleich zum Eingang an.

Massepotenzial der Schaltung verbunden. Im Falle von $V_{in} = 0$ V sperrt der n -Kanal-MOSFET, während der p -Kanal-MOSFET einschaltet. Die Ausgangsspannung entspricht dann V_{DD} . Strom fließt nur kurzzeitig während der Schaltvorgänge, die durch ihre charakteristischen Zeiten τ_n und τ_p gekennzeichnet sind. Um die Funktion des Inverters zu gewährleisten, sind die Werte der Schwellspannung der beiden Transistoren (p - und n -Kanal) beziehungsweise die effektiven Transistorströme extrem wichtig. Verschiebt sich allerdings die Schwellspannung eines oder beider Transistoren aufgrund von Degradation, zum Beispiel durch NBTI beim p -MOSFET [99, 101, 144], so sinkt der Transistorstrom, welches zu einer Verlängerung der Schaltzeiten führt. Außerdem kann es passieren, dass die Transistoren unter den jeweils anliegenden Eingangsspannungen entweder nicht sauber sperren oder nicht eingeschaltet werden können. Die Funktion des Schaltkreises wird dadurch erheblich beeinträchtigt und die maximale Frequenz, mit der die Schaltung betrieben werden kann, sinkt.

An einem Inverter können eine ganze Reihe von Eigenschaften einer CMOS-Schaltung studiert werden. So werden zum Beispiel durch eine geschlossene Kette von hintereinandergeschalteten Invertiern so genannte Ringoszillatoren gebildet. Diese bestehen aus einer ungeraden Anzahl von Invertiern und schwingen bei Anlegen einer Versorgungsspannung. Die Frequenz verhält sich invers zu den Schaltzeiten der MOSFETs und erlaubt Rückschlüsse auf die maximale Frequenz der CMOS-Schaltung. Die hauptsächlichen Ursachen der Degradation sind heiße Ladungsträger und NBTI. Über den zeitlichen Verlauf der maximalen Frequenz können Rückschlüsse auf das Degradationsverhalten gezogen werden [145].

2.5 Der MOSFET als nichtflüchtiger Speicher (NROM)

Nichtflüchtige Speicher sind Halbleiterbauelemente, die ihre gespeicherte Information nach Abschalten der Versorgungsspannung nicht verlieren. Sie werden anhand ihrer Programmier- und

Löschverfahren in zwei Gruppen unterteilt. Die erste Gruppe beinhaltet die EPROMs (*engl.*: Electrically Programmable Read Only Memory), die Speicher mit überwiegend lesendem Zugriff (ROM, *engl.*: Read Only Memory) darstellen. Diese werden elektrisch zum Beispiel mittels heißer Elektronen programmiert und können nur durch eine Bestrahlung mit UV-Licht wieder gelöscht werden. Der Nachteil hierbei ist, dass immer der gesamte Speicher gelöscht wird und anschließend wieder neu geschrieben werden muss. Die zweite Gruppe bilden die EEPROMs (*engl.*: Electrically Erasable Programmable Read Only Memory), die elektrisch sowohl programmiert als auch gelöscht werden können. Das Löschen geschieht im Allgemeinen durch FN-Tunneln von Ladungsträgern [5].

Für den Einsatz von Speichern in modernen, überwiegend mobilen elektronischen Geräten hat sich in den letzten Jahrzehnten eine weitere Gruppe von nichtflüchtigen Speichern entwickelt, so genannte Flash-Speicher, die im Sinne ihrer Programmierereigenschaften zwischen die EPROMs und die EEPROMs einzuordnen ist [5]. Diese Speicher werden ähnlich den EPROMs über heiße Ladungsträger programmiert, aber analog zu EEPROMs über FN-Tunneln gelöscht. Auch hierbei werden wie bei den EEPROMs stets komplette Zellenfelder gleichzeitig gelöscht und anschließend neu geschrieben.

Ein großer Teil der Flash-Speicher basiert auf dem Konzept eines Floating-Gates, das in den Gate-Oxid-Stapel eines MOSFETs eingefügt ist. In Abbildung 2.19a ist ein solcher Floating-Gate-Speicher dargestellt. Das Floating-Gate ist elektrisch leitend (meist Poly-Si), aber nach außen elektrisch isoliert. Zum Substrat wird es durch einen Tunnelisolator, zum Control-Gate durch einen Zwischenisolator begrenzt. Dadurch ist es möglich, auf dem Floating-Gate dauerhaft Ladungen zu speichern. Diese Speicherladungen werden zum Beispiel durch eine Injektion von heißen Elektronen aus dem Kanal (vergleiche Abschnitt 2.1.3) oder über FN-Tunneln

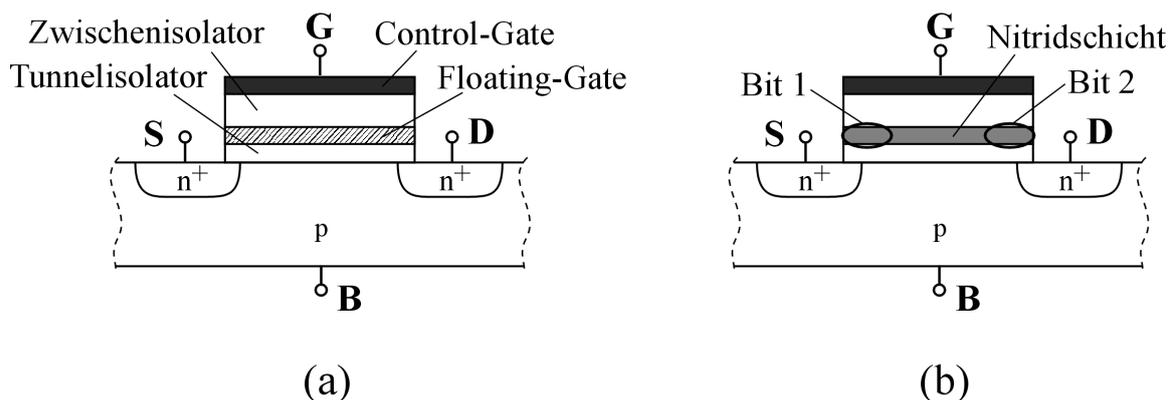


Abbildung 2.19: (a) Aufbau eines Floating-Gate-Speichers. In den Gate-Oxid-Stapel ist ein elektrisch leitendes, aber nach außen isoliertes Floating-Gate eingebettet, auf dem Ladungen gespeichert werden können. Das Oxid zwischen Floating-Gate und Siliziumsubstrat wird als Tunnelisolator und das Oxid oberhalb des Floating-Gates als Zwischenisolator bezeichnet. (b) In einer NROM-Speicherzelle ist das Floating-Gate durch eine Nitridschicht ersetzt, die eine lokalisierte Speicherung von Ladungen ermöglicht (Bit 1 und Bit 2 nahe der beiden n^+p -Übergänge).

(siehe Abschnitt 2.2.1) durch den Tunnelisolator eingebracht beziehungsweise wieder entfernt. Gespeicherte Ladungen auf dem Floating-Gate verschieben die Flachbandspannung und damit die Schwellspannung des Transistors. Bei einer festen Lesespannung V_G am Control-Gate und $V_D > 0$ variiert der Drain-Strom dabei um mehrere Größenordnungen, wodurch zwischen dem programmierten und dem gelöschten Zustand unterschieden werden kann.

2.5.1 Die NROM-Speicherzelle

Ein NROM-Speicher gehört ebenfalls zu den Flash-Speichern, basiert aber auf dem Konzept eines SONOS-Speichers (*engl.*: Silicon-Oxide-Nitrid-Oxide-Semiconductor) [146]. Im Gegensatz zum Floating-Gate-Speicher ist die ladungsspeichernde Schicht hierbei durch ein Siliziumnitrid (Si_3N_4) ersetzt worden, wie in Abbildung 2.19b dargestellt. Man spricht in diesem Fall von einem ONO-Dielektrikum (*engl.*: Oxide-Nitride-Oxide). Die Ladungen können sich in der Nitridschicht nicht frei verteilen, sondern werden in energetisch tiefen Störstellen eingefangen und sind dort lokalisiert [147]. Aufgrund dieser Lokalisierung der Speicherladungen kann eine Speicherzelle mit zwei Bits realisiert werden. Dazu werden Ladungen unabhängig voneinander nahe des Source- und Drain-Kontakts eingebracht (Bit 1 und Bit 2 in Abbildung 2.19b). Mithilfe eines geeigneten Leseverfahrens, das in Abschnitt 2.5.3 näher erläutert wird, können die beiden Bits unabhängig ausgelesen werden [146, 148]. Die Möglichkeit, zwei Bits pro Zelle speichern zu können ist ein wesentlicher Vorteil der NROM-Zelle gegenüber dem SONOS-Speicher, bei dem die Nitridschicht über dem gesamten Kanalbereich durch Tunneln von Ladungsträgern be- und entladen wird. Somit kann dort nur ein Bit gespeichert werden, wohingegen mit der NROM-Zelle bei vergleichbarer Transistorgröße die doppelte Speicherdichte erzielt werden kann.

2.5.2 Programmieren und Löschen einer NROM-Zelle

Das Setzen eines Bits des NROM-Speichers (Programmieren) erfolgt durch die Injektion von heißen Elektronen aus dem Transistorkanal. Durch eine ausreichend große Spannung am Drain und eine in etwa doppelt so hohe Spannung am Gate werden die Elektronen entlang des Kanals beschleunigt und in die ONO-Schicht injiziert [149]. Der Source-Kontakt bleibt dabei geerdet. Typische Schreibspannungen sind $V_D = 4,5$ V und $V_G = 9$ V, die für wenige Mikrosekunden anliegen [146, 150]. Die Elektronen werden von tiefen Störstellen innerhalb der Nitridschicht nahe des Drains eingefangen. Das Maximum der lateralen Verteilung der Elektronen liegt dabei in etwa auf Höhe des pn -Übergangs des Drain-Kontakts. Die Ausläufer der Verteilung erstrecken sich aber deutlich über den drain-nahen Kanalbereich [151–153].

Zum Löschen eines Bits wird die Gate-Spannung auf einen hohen negativen Wert gesetzt (typischerweise zwischen -5 und -7 V), während die Drain-Spannung in etwa der beim Programmieren entspricht [146, 150]. Bei diesen Spannungen entsteht im n^+ -Gebiet des Drains unterhalb der ONO-Schicht eine Zone tiefer Verarmung, in der durch Band-zu-Band-Tunneln (BBT)

Löcher generiert werden [154]. Diese Löcher können im lateralen Feld der Drain-Region genügend Energie aufnehmen und werden anschließend durch das Tunneloxid ins Nitrid injiziert. Dort rekombinieren die heißen Löcher entweder direkt mit den vorhandenen Elektronen oder das Profil der injizierten Löcher kompensiert das der Elektronen [146, 155]. Da der Transistor symmetrisch bezüglich Source und Drain aufgebaut ist, kann durch Vertauschen von Source- und Drain-Spannung auf die gleiche Weise das Bit nahe des Source-Kontakts programmiert beziehungsweise gelöscht werden.

Im Allgemeinen decken sich die Profile der injizierten Löcher und Elektronen jedoch nicht exakt, sondern liegen leicht verschoben [151]. Dies führt nach iterativem Programmieren und Löschen zu einer Anhäufung von Ladungen im Nitrid und damit zu einer Beeinträchtigung nachfolgender Programmier- und Löschoptionen. Eine bessere Übereinstimmung der Profile kann jedoch durch eine geeignete Wahl der Spannungen erzielt werden [150]. Darüber hinaus ist die Fähigkeit der NROM-Zelle, die gespeicherten Ladungen zu halten (*engl.*: retention) von entscheidender Bedeutung. Ein Verlust von Ladungen ins Substrat durch direktes Tunneln konnte in experimentellen Untersuchungen aufgrund der hohen Tunneloxidstärke als vernachlässigbar eingestuft werden [152, 156]. Jedoch ist ein Ladungsverlust durch störstellenunterstütztes Tunneln denkbar, bei dem die Ladungsträger in einem zweistufigen Tunnelprozess über vorhandene Oxidstörstellen entweichen (vergleiche Abschnitt 2.2.1). Der Vollständigkeit halber sei an dieser Stelle erwähnt, dass sich gespeicherte Ladungsträger auch lateral in der Nitridschicht umverteilen können und somit das Lesen der Zelle erschweren oder sogar unmöglich machen [156, 157]. Ein Verlust der gespeicherten Information wäre die Folge. Dieser Effekt soll hier aber nicht näher betrachtet werden.

Eine weit größere nachteilige Auswirkung auf die Lebensdauer (*engl.*: endurance) der NROM-Zelle haben Grenzflächenzustände und grenzflächenennahe Oxidstörstellen (NIOTs). Sie werden durch heiße Ladungsträger beim Programmieren und Löschen generiert und bewirken eine permanente Verschiebung der Schwellspannung, wodurch das Lesen der Zelle erschwert wird [4, 158]. Charge-Pumping-Untersuchungen an SONOS- und NROM-Speichern zeigten, dass die Programmierung mit heißen Elektronen einen deutlich größeren Anstieg der Grenzflächenzustandsdichte verursacht, als eine Programmierung über Fowler-Nordheim-Tunneln [159].

2.5.3 Lesen einer NROM-Zelle

Zum Lesen einer NROM-Zelle wird der Transistor in umgekehrter Richtung (*engl.*: reverse read mode) betrieben und der Drain-Strom als Indikator für den Speicherzustand gemessen. Das bedeutet, dass die Rollen von Source und Drain im Vergleich zum Programmieren vertauscht sind. Um das in Abbildung 2.19b dargestellte Bit 2 zu lesen, wird der Source-Kontakt mit typischerweise 1,5 V belegt, während der Drain geerdet ist. Am Gate wird eine Spannung von etwa 3 V angelegt. Der Transistor arbeitet unter dieser Spannungsconfiguration im Sättigungsbereich, wie in Abbildung 2.20a dargestellt.

In der Pinch-off Region nahe des Source kommt es aufgrund der hohen Spannung zu einem

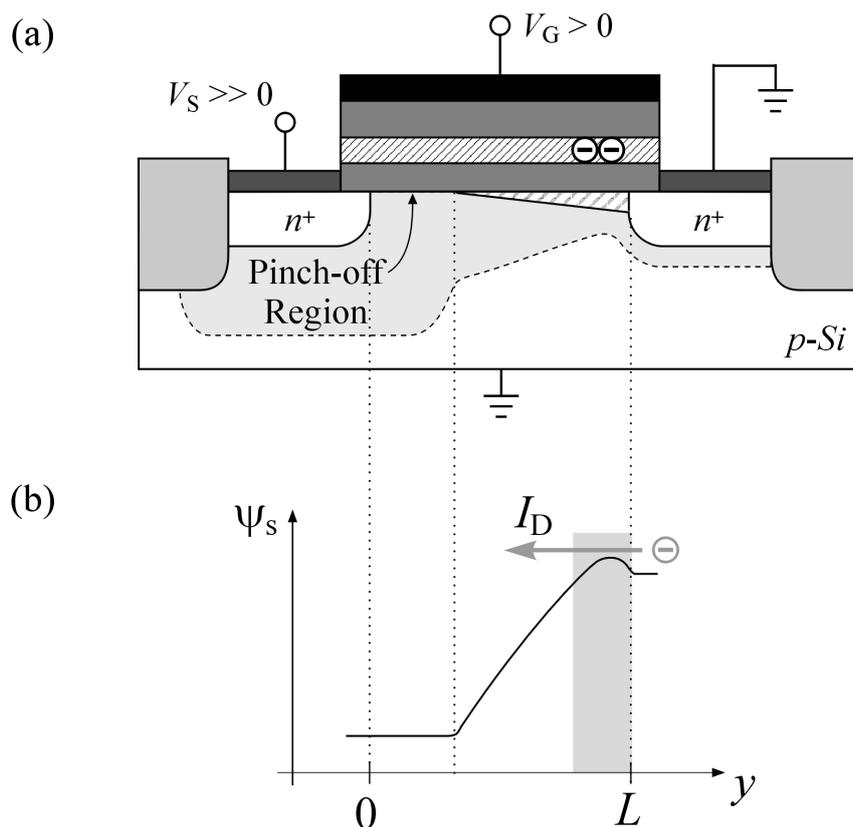


Abbildung 2.20: (a) NROM-Speicherzelle beim Lesen des drain-seitigen Bits. Der MOSFET befindet sich aufgrund der hohen Source-Spannung in Sättigung, wodurch am Source eine Pinch-off Region entsteht. Der Transistorstrom wird fast ausschließlich durch das elektrostatische Potenzial von Nitridladungen auf der Drain-Seite beeinflusst.
 (b) Oberflächenpotenzial entlang des Kanals. Der Drain-Strom I_D wird durch das Potenzialmaximum am Drain bestimmt, dessen Höhe empfindlich durch die gespeicherten Nitridladungen beeinflusst wird.

Durchschlagen (*engl.*: punch through) [157]. Falls Ladungen in der Nitridschicht oberhalb dieser Region vorhanden sind, hat deren elektrostatische Wirkung nur einen sehr geringen Einfluss auf die Kanalleitfähigkeit [160]. Der Drain-Strom I_D wird in dieser Situation überwiegend durch das Potenzial von Ladungen auf der entgegengesetzten Kanalseite beeinflusst (graue Fläche in Abbildung 2.20b). Die Folge ist eine Reduktion des Stroms im Kanal, welches sich in eine Verschiebung der Schwellspannung übertragen lässt.

Abbildung 2.21 zeigt die Transferkennlinien einer unprogrammierten NROM-Zelle ($V_t = V_{t,g}$) und einer programmierten Zelle ($V_t = V_{t,p} > V_{t,g}$) in beiden Source-Drain-Polaritäten. Im Falle der programmierten Zelle ist das Bit nahe des Drains mit Elektronen beladen, während das Bit auf der Source-Seite unbeladen ist. Die Kennlinie unter $V_{DS} < 0$ ist im programmierten Zustand deutlich gegenüber der unprogrammierten zu höheren Spannungen verschoben. Der bei einer festen Gate-Spannung $V_G = V_{G,lesen}$ gemessene Drain-Strom unterscheidet sich zwischen den beiden Zuständen um mehrere Größenordnungen und ermöglicht somit die Detektion der

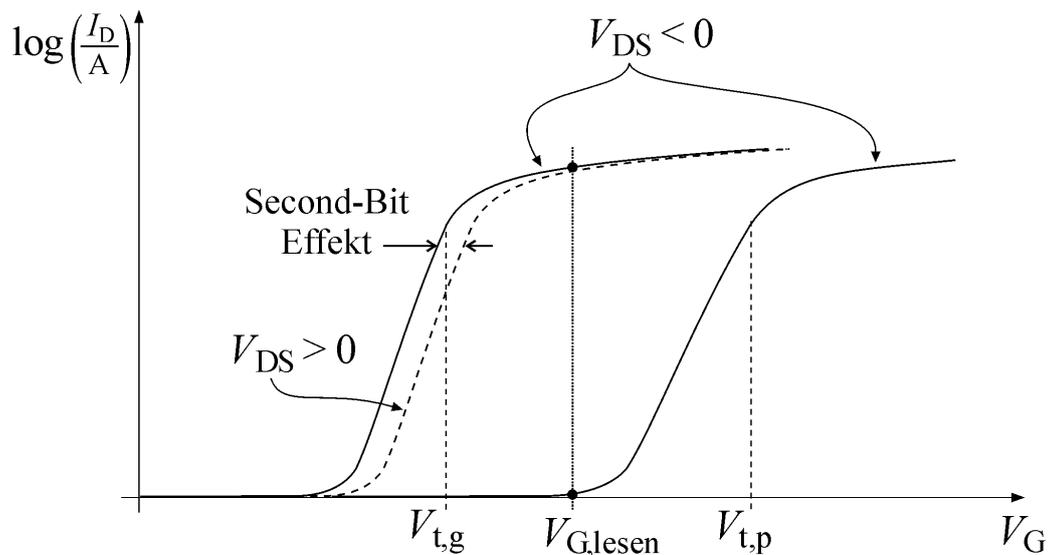


Abbildung 2.21: Transferkennlinien einer drain-seitig programmierten ($V_t = V_{t,p}$) und einer gelöschten ($V_t = V_{t,g}$) NRROM-Zelle bei $V_{DS} < 0$ (durchgezogene Kurven). Die Schwellspannung verschiebt sich durch die Programmierung von $V_{t,g}$ nach $V_{t,p}$. Der Strom bei der Lesespannung $V_{G,lesen}$ unterscheidet sich zwischen beiden Zuständen um mehrere Größenordnungen. Durch Vertauschen der Drain-Source-Polarität ($V_{DS} > 0$, gestrichelt) wird im gleichen Zustand das unprogrammierte source-seitige Bit ausgelesen. Die Transferkennlinie entspricht in etwa der der gelöschten Zelle und ist aufgrund des Second-Bit-Effekts nur leicht dagegen verschoben.

gespeicherten Ladungen. Durch Vertauschen der Source-Drain-Polarität (*engl.*: forward read mode) kann das Bit auf der Source-Seite unabhängig vom drain-seitigen Bit gelesen werden. In Abbildung 2.21 ist zusätzlich die Kennlinie für den Fall $V_{DS} > 0$ (gestrichelt) dargestellt. Obwohl in dem source-seitigen Bit keine Ladungen gespeichert sind, ist die Kennlinie gegenüber der der unprogrammierten Zelle leicht verschoben. Die Ursache hierfür ist ein geringfügiges Übersprechen der drain-seitigen Ladungen. In der Literatur wird dies als Second-Bit-Effekt bezeichnet [155].

3 Theorie der Messmethoden

3.1 Elektrische Charakterisierungsmethoden

Zur kontinuierlichen Weiterentwicklung und Optimierung von MOS-Bauelementen und deren Herstellungsprozessen ist die elektrische Charakterisierung der Bauelemente von entscheidender Bedeutung. Hierzu zählt besonders die Bestimmung der elektrischen Eigenschaften des Gate-Isolators und dessen Grenzfläche zum Siliziumsubstrat.

Für die elektrische Charakterisierung stehen eine große Zahl von Messmethoden zur Verfügung. Zu den am weitesten verbreiteten zählen die $C(V)$ -Methode [161], die Charge-Pumping-Methode [53, 162], Rauschmessungen [163], sowie die DLTS-Methode [164, 165]. Diese haben hauptsächlich die Charakterisierung der $Si-SiO_2$ -Grenzfläche und des grenzflächennahen Oxidbereichs zum Ziel. Darüber hinaus existieren weitere Methoden, wie zum Beispiel SILC-Messungen [30, 58, 166], die im Wesentlichen das Siliziumdioxid charakterisieren.

Mit steigender Integrationsdichte von mikroelektronischen Schaltungen werden immer wieder neue Herausforderungen an die Messmethoden gestellt. Diese wurden daher in der Vergangenheit ständig erweitert und angepasst, um zum Beispiel steigenden Leckströmen gerecht zu werden [167]. Darüber hinaus nimmt neben der Bestimmung der Grenzflächenzustände auch die Charakterisierung von Volumenstörstellen im Oxid an Bedeutung zu.

Dieses Kapitel stellt die in dieser Arbeit zum Einsatz kommenden Messmethoden und deren theoretischen Hintergrund näher vor. Hierzu zählt insbesondere die Charge-Pumping-Methode, die in der Lage ist, zwischen Störstellen an der Grenzfläche und Störstellen im grenzflächennahen Oxidbereich zu unterscheiden. Eine weitere Methode, die hier vorgestellt wird, ist die Messung und Analyse von $1/f$ -Rauschen in MOSFETs.

3.2 Charge-Pumping (CP)

Der dem Charge-Pumping zugrunde liegende Effekt wurde erstmals 1969 von Brugler und Jaspers beschrieben [53]. Beim Anlegen von Gate-Pulsen an MOS-Transistoren wurde ein kleiner Nettostrom von Ladungen beobachtet, der über den Substratanschluss des Bauelements floss. Nähere Untersuchungen zeigten, dass sich dieser Substratstrom proportional zur Frequenz des Gate-Pulses verhält und proportional zur Gate-Fläche ist. Dieser Linearität verdankt der Effekt, der offensichtlich eine Art Ladungspumpe darstellt, seine Bezeichnung Charge-Pumping (CP).

Die Interpretation des Effekts erfolgt mittels eines Rekombinationsprozesses an der $Si-SiO_2$ -Grenzfläche, bei dem Grenzflächenzustände involviert sind. Bei jedem Zyklus des Gate-Pulses, bei dem der MOSFET zwischen Inversion und Akkumulation umgeschaltet wird, bleibt ein Teil der Inversionsladung in Grenzflächenzuständen gespeichert. Am Ende eines Gate-Pulses rekombinieren diese Ladungen mit Majoritätsladungsträgern aus dem Substrat. Daraus resultiert, dass der Substratstrom direkt mit der Zahl der Grenzflächenzustände in Verbindung steht und somit eine Bestimmung der Grenzflächenzustandsdichte möglich ist.

3.2.1 Die Funktionsweise der CP-Methode

Im Folgenden sollen die Vorgänge an der Grenzfläche beim Umschalten von Inversion zu Akkumulation und umgekehrt näher betrachtet werden. Dafür wird ein n -Kanal-MOSFET herangezogen, der wie in Abbildung 3.1a beschaltet wurde. Die Pulsform am Gate ist in Abbildung 3.1b dargestellt. Dabei bezeichnet T_P die Periodendauer des Pulses, t_h und t_l die Dauern der konstanten Phasen, t_r und t_f die Dauern der Anstiegs- beziehungsweise Abfallflanke und $V_{Ampl} = V_{G,max} - V_{G,min}$ die Amplitude des Pulses. Bei jedem Wechsel der Gate-Spannung von $V_{G,min}$ nach $V_{G,max}$ oder umgekehrt durchläuft die Grenzfläche drei verschiedene Stadien, die als Akkumulation, Verarmung und Inversion bezeichnet werden. Jede Phase kann durch charakteristische Zeitkonstanten beschrieben werden.

Abbildung 3.2 zeigt die Prozesse an der Grenzfläche, die während eines Umschaltvorgangs von Akkumulation zu Inversion in einem n -MOSFET auftreten [162]. Zunächst befindet sich die Grenzfläche in Akkumulation und alle Grenzflächenzustände unterhalb des Quasi-Fermi-

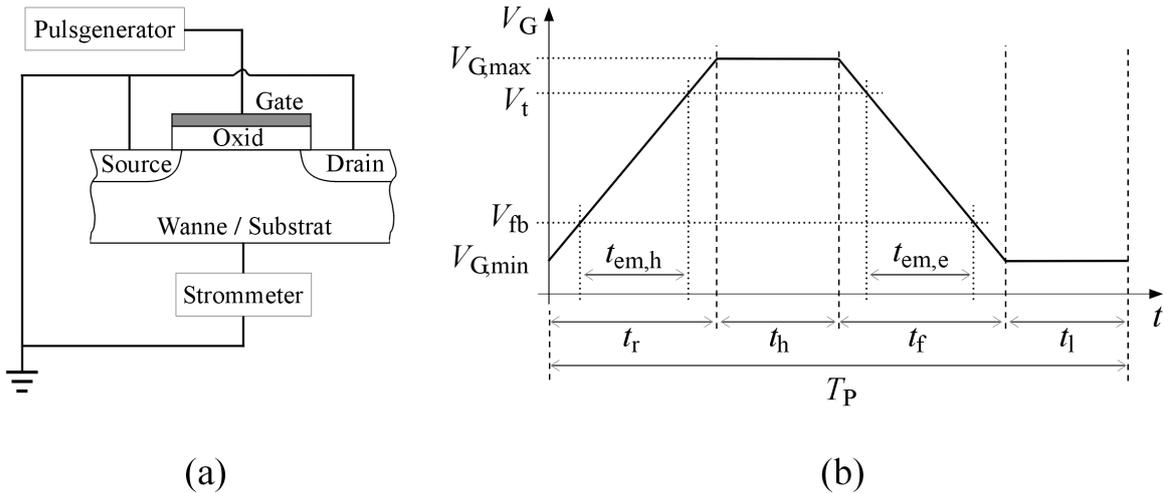


Abbildung 3.1: (a) Beschaltung eines MOSFETs bei Charge-Pumping-Messungen. (b) Typischer zeitlicher Verlauf der Gate-Spannung während eines CP-Zyklus mit Kennzeichnung der relevanten Zeitphasen. Die Phasen, in denen sich die Gate-Spannung zwischen der Schwellenspannung und der Flachbandspannung befindet, werden mit $t_{em,h}$ und $t_{em,e}$ bezeichnet.

Niveaus $E_{F,akk}$ sind mit Elektronen gefüllt. Alle Zustände oberhalb dieser Energie sind leer. In diesem Zustand befinden sich alle Energiezustände im Gleichgewicht mit den Energiebändern, deren Grenzen mit E_V (Valenzbandkante) und E_L (Leitungsbandkante) bezeichnet sind.

Wenn die Gate-Spannung mit einer gegebenen Rate ansteigt, ändert sich das Oberflächenpotential ebenfalls mit einer gewissen Rate $\frac{d\psi_s}{dt}$. Daraus resultiert die Rate $\frac{dQ_t}{dt}$, mit der sich die Ladungsträgerdichte Q_t der mit Löchern besetzten Grenzflächenzustände (ss) ändern muss, um den Gleichgewichtszustand zu erhalten. Sie ist gegeben durch

$$\left. \frac{dQ_t}{dt} \right|_{ss} = -q^2 D_{it} \frac{d\psi_s}{dt} \quad , \quad (3.1)$$

wobei q die Elementarladung und D_{it} die mittlere Grenzflächenzustandsdichte bezeichnen. Die Änderung der Besetzung der Zustände an der Grenzfläche geschieht durch die Emission von Löchern ins Valenzband, die dann über das Substrat abfließen. Die Emissionsrate (em) der Löcher ist gegeben durch

$$\left. \frac{dQ_t}{dt} \right|_{em} = -q \frac{dn_t(t)}{dt} \quad , \quad (3.2)$$

wobei $n_t(t)$ die Dichte der mit Löchern besetzten Grenzflächenzustände ist. Der Gleichgewichtszustand an der Grenzfläche ist solange gegeben, wie gilt

$$\left. \frac{dQ_t}{dt} \right|_{em} > \left. \frac{dQ_t}{dt} \right|_{ss} \quad . \quad (3.3)$$

Dieser Fall ist in Abbildung 3.2 mit Phase 1 bezeichnet. Sobald diese Bedingung nicht mehr erfüllt ist und die Emissionsrate der Löcher kleiner wird als die durch die zeitliche Änderung der Gate-Spannung vorgegebene Rate, befindet sich die Grenzfläche in einem Nichtgleichgewichtszustand. Dieser ist in Abbildung 3.2 als Phase 2 gekennzeichnet und wird vollständig durch die Emission der Löcher bestimmt. Der Übergang von der Gleichgewichtsphase zur Nichtgleichgewichtsphase ist erreicht, wenn die Gate-Spannung leicht oberhalb der Flachbandspannung V_{fb} liegt [168]. In guter Näherung kann hierfür jedoch der Wert für die Flachbandspannung angenommen werden [162].

Zu Beginn von Phase 3 in Abbildung 3.2 erreicht die Gate-Spannung die Schwellspannung des Kanals. Unter dieser Bedingung sinkt die Zeitkonstante für den Einfang von Elektronen in Grenzflächenzustände soweit ab, dass dieser Prozess dominiert. Die Zeitkonstante ist gegeben durch [53]

$$\tau_t = \frac{1}{\nu_{th} \sigma_n n_s} \quad , \quad (3.4)$$

wobei $\nu_{th} = 1 \cdot 10^7 \text{ cm s}^{-1}$ die thermische Geschwindigkeit der Ladungsträger bezeichnet, σ_n den Einfangquerschnitt für Elektronen (für Löcher σ_p) und n_s die Volumendichte der Elektronen an der Grenzfläche (für Löcher p_s). Die nicht emittierten Löcher in den Grenzflächenzuständen rekombinieren dabei mit den Elektronen, die aus den Source- und Drain-Gebieten in den Kanal fließen. Wenn die Gate-Spannung die Schwellspannung übersteigt, werden alle Grenzflächenzustände mit Elektronen gefüllt und der Gleichgewichtszustand im Kanal ist wieder hergestellt.

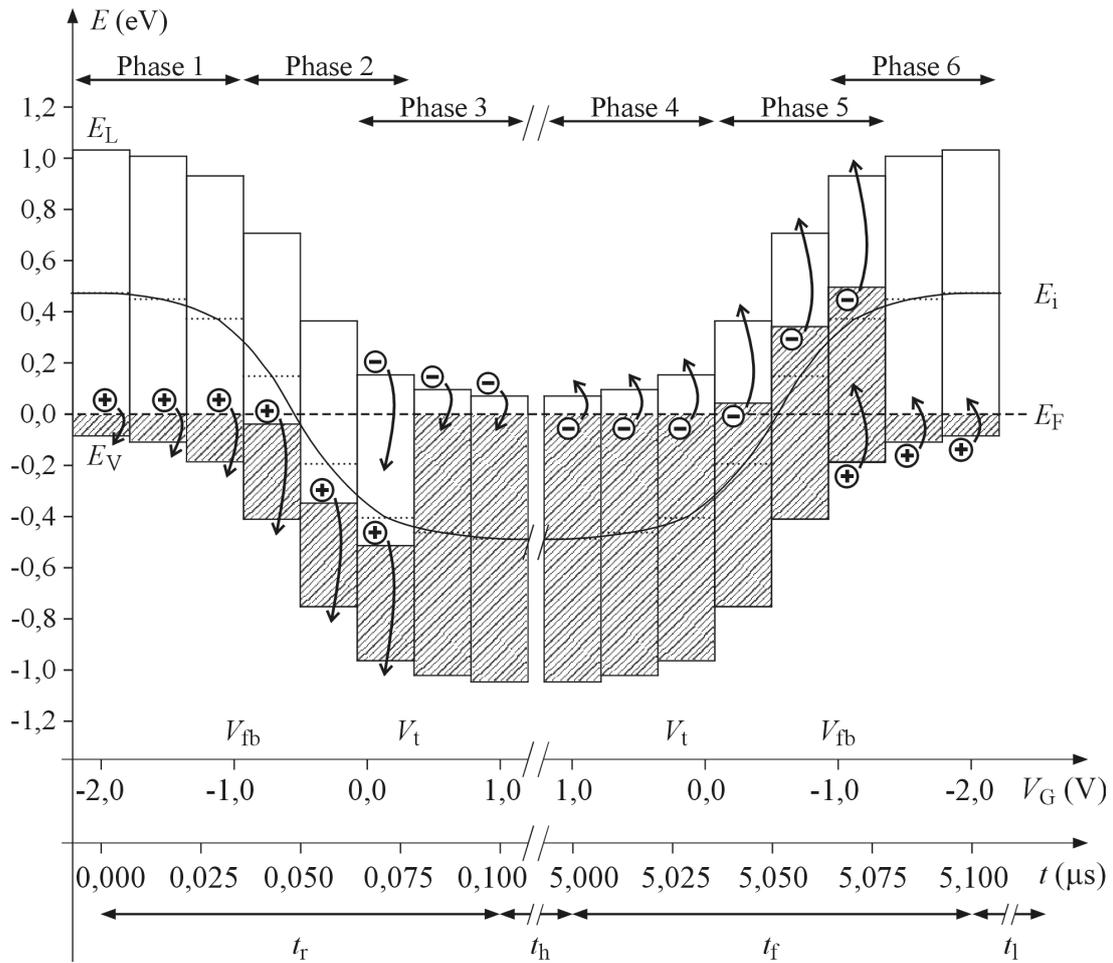


Abbildung 3.2: Verschiebung der Energiebandlücke relativ zum Quasi-Fermi-Niveau an der Grenzfläche während eines CP-Zyklus [162]. Beispiel eines n -Kanal-MOSFETs bei einer Messfrequenz von $f = 100$ kHz. Grenzflächenzustände, deren Energieniveaus innerhalb des schraffierten Bereiches liegen, sind mit Elektronen besetzt.
 Phase 1: Gleichgewichtsemission von Löchern ins Valenzband des Substrats
 Phase 2: Wie Phase 1, aber Nichtgleichgewichtsemission
 Phase 3: Einfang von Elektronen, die aus den Source- und Drain-Gebieten in den Kanal fließen
 Phase 4: Gleichgewichtsemission von Elektronen ins Valenzband des Substrats
 Phase 5: Wie Phase 4, aber Nichtgleichgewichtsemission
 Phase 6: Einfang von Löchern aus dem Substrat

Beim Zurückschalten der Gate-Spannung von Inversion nach Akkumulation spielen sich ähnliche Prozesse an der Grenzfläche ab. Zunächst werden Elektronen aus den Grenzflächenzuständen in einem Gleichgewichtszustand in das Leitungsband emittiert und fließen über die Source- und Drain-Gebiete ab (Phase 4). Wenn die Schwellenspannung unterschritten wird, tritt erneut ein Nichtgleichgewichtszustand ein, in dem weiterhin Elektronen in das Leitungsband emittiert

werden und über Source und Drain abfließen (Phase 5). Anschließend, nach Unterschreiten der Flachbandspannung, sinkt die Zeitkonstante für den Einfang von Löchern in Grenzflächenzustände stark ab und der Löchereinfang wird zum dominierenden Prozess. Die aus dem Substrat an die Grenzfläche strömenden Löcher rekombinieren mit den in Grenzflächenzuständen verbliebenen Elektronen (Phase 6).

Aus den oben beschriebenen sechs Phasen während eines Zyklus der Gate-Spannung resultieren vier Stromkomponenten im Transistor, die in Abbildung 3.3a skizziert sind. Sie werden durch

$$I_1 = q^2 D_{it} \Delta\psi_{he} f A_G \quad , \quad (3.5)$$

$$I_2 = -q^2 D_{it} \Delta\psi_e f A_G \quad , \quad (3.6)$$

$$I_3 = q^2 D_{it} \Delta\psi_{ee} f A_G \quad \text{und} \quad (3.7)$$

$$I_4 = -q^2 D_{it} \Delta\psi_h f A_G \quad (3.8)$$

ausgedrückt, wobei f die Frequenz des Gate-Spannungspulses, A_G die Gate-Fläche und $\Delta\psi_{he}$, $\Delta\psi_e$, $\Delta\psi_{ee}$ und $\Delta\psi_h$ die beteiligten Energieintervalle bezeichnen. Die Energieintervalle sind in Abbildung 3.3b dargestellt.

Der am Substrat gemessene Nettostrom ist damit gegeben durch

$$I_B = I_{CP} = I_1 + I_4 = q^2 D_{it} (\Delta\psi_{he} - \Delta\psi_h) f A_G \quad , \quad (3.9)$$

während die Summe der am Source und Drain gemessenen Ströme durch

$$I_S + I_D = -I_{CP} = I_2 + I_3 = q^2 D_{it} (\Delta\psi_{ee} - \Delta\psi_e) f A_G \quad (3.10)$$

ausgedrückt wird. Daraus ergibt sich, dass das zum Charge-Pumping beitragenden Energieintervall gegeben ist durch

$$\Delta E = \Delta\psi_{he} - \Delta\psi_h \equiv \Delta\psi_{ee} - \Delta\psi_e \quad . \quad (3.11)$$

In diesem Energieintervall, das in Abbildung 3.3b durch den schraffierten Bereich dargestellt ist, findet die Rekombination von Elektronen und Löchern beim Charge-Pumping statt. Die Grenzen dieses Energieintervalls sind gegeben durch [162]

$$E_{em,e} \simeq E_i - \frac{k_B T}{q} \ln \left(\nu_{th} \sigma_n n_i t_{em,e} + e^{\frac{(E_i - E_{F,inv})}{k_B T}} \right) \quad \text{und} \quad (3.12)$$

$$E_{em,h} \simeq E_i + \frac{k_B T}{q} \ln \left(\nu_{th} \sigma_p n_i t_{em,h} + e^{\frac{(E_{F,akk} - E_i)}{k_B T}} \right) \quad , \quad (3.13)$$

wobei n_i die intrinsische Ladungsträgerkonzentration im Silizium darstellt, E_i das intrinsische Energieniveau, $E_{F,inv}$ und $E_{F,akk}$ die Quasi-Fermi-Niveaus in Inversion beziehungsweise Akkumulation und $t_{em,e}$ und $t_{em,h}$ die Zeitdauern der Nichtgleichgewichtszustände während der Emission von Elektronen beziehungsweise Löchern. Die Ausdrücke sind durch eine Taylor-Reihenentwicklung der exakten Ausdrücke aus [169] hervorgegangen. Das zum CP-Strom I_{CP} beitragende Energieintervall ist damit gegeben durch

$$\Delta E = E_{em,e} - E_{em,h} = -2 \frac{k_B T}{q} \ln \left(\nu_{th} n_i \sqrt{\sigma_n \sigma_p} \sqrt{t_{em,e} t_{em,h}} \right) \quad . \quad (3.14)$$

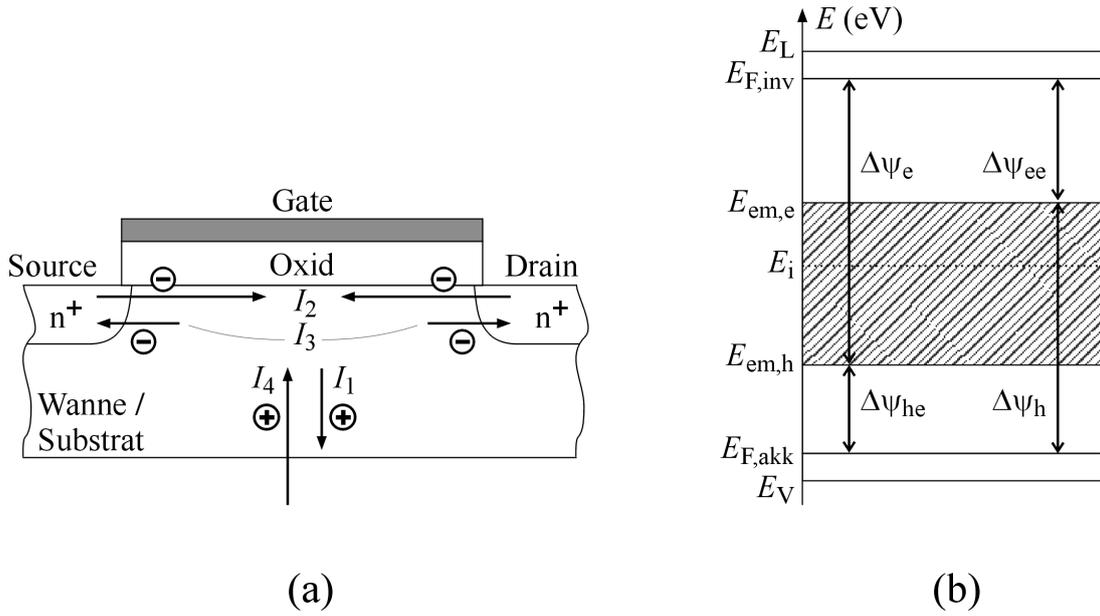


Abbildung 3.3: Stromkomponenten und beteiligte Energieintervalle beim Charge-Pumping am Beispiel eines n -Kanal-MOSFETs. (a) Elektronen fließen über die Source- und Drain-Gebiete in den Kanal, Löcher über das Substrat. (b) Relevante Energiebereiche und deren Bezeichnung. Grenzflächenzustände im schraffierten Bereich tragen zum CP-Strom bei.

Dabei wurden die Exponentialterme in den Gleichungen 3.12 und 3.13 vernachlässigt unter der Annahme, dass die Energieniveaus bei der Emission von Elektronen und Löchern näher zur Bandmitte liegen als die Quasi-Fermi-Niveaus $E_{F,inv}$ und $E_{F,akk}$. Dies ist der Fall, sobald die Anstiegs- und Abfallflanken des Gate-Spannungspulses wenige Nanosekunden oder mehr betragen [162]. Die Zeiten, die den Elektronen und Löchern für die Emission zur Verfügung stehen, sind aufgrund der Geometrie des Gate-Spannungspulses (vergleiche Abbildung 3.1b) gegeben durch

$$t_{em,e} = \frac{|V_{fb} - V_t|}{V_{ampl}} t_f \quad \text{und} \quad (3.15)$$

$$t_{em,h} = \frac{|V_{fb} - V_t|}{V_{ampl}} t_r \quad . \quad (3.16)$$

Schließlich ergibt sich aus den obigen Betrachtungen der Ausdruck für den CP-Strom zu

$$I_{CP} = qD_{it}fA_G\Delta E \quad . \quad (3.17)$$

Die Gleichung 3.17 zeigt, dass der CP-Strom direkt proportional zur Grenzflächenzustandsdichte D_{it} ist.

3.2.2 Besonderheiten und Unterschiede der CP-Modi

Charge-Pumping kann in mehreren Variationen betrieben werden, im Folgenden Modi genannt. Sie unterscheiden sich im Wesentlichen in der Form des Spannungspulses am Gate. Hier kann zum einen die Amplitude des Pulses variiert werden, zum anderen der zeitliche Anteil der ansteigenden beziehungsweise abfallenden Flanke zur Periodendauer. Im Allgemeinen wird Charge-Pumping nicht unter stationären Bedingungen gemessen, sondern in Abhängigkeit der Pulsfrequenz oder eines charakteristischen Gate-Spannungswertes, zum Beispiel $V_{G,\min}$. Die Anwendung der unterschiedlichen CP-Modi ermöglicht es, die Degradation von MOSFETs detailliert zu untersuchen [170].

CP mit konstanter Amplitude

Im CP-Modus mit konstanter Amplitude wird die Spannungsdifferenz $V_{G,\max} - V_{G,\min}$ konstant gehalten und der CP-Strom für verschiedene Werte von $V_{G,\min}$ gemessen. Der Gate-Spannungspuls überfährt dabei den gesamten Bereich von Akkumulation bis Inversion, wie in Abbildung 3.4a dargestellt. Trägt man den CP-Strom in Abhängigkeit von $V_{G,\min}$ auf, erhält man eine typische glockenförmige Kurve wie sie in Abbildung 3.4b dargestellt ist. Dieser charakteristische Kurvenverlauf wird auch als Elliot-Kurve bezeichnet [171].

Die Vorgänge in den einzelnen Bereichen der Elliot-Kurve (A bis E) werden im Folgenden kurz beschrieben. Im Bereich A befindet sich der Transistor zu jeder Zeit während des Gate-Spannungspulses in Akkumulation. Alle Grenzflächenzustände sind mit Löchern gefüllt, so dass keine Rekombination stattfindet und kein CP-Strom fließt. Im Übergangsbereich B nähert sich die maximale Gate-Spannung der Schwellspannung und der Transistor befindet sich zu diesen Zeiten in schwacher Inversion. Dabei erhöht sich drastisch die Konzentration der Elektronen an der Grenzfläche und die Rekombination mit den Löchern in den Grenzflächenzuständen setzt ein. Dies führt zu einem rapiden Anstieg des am Substrat gemessenen CP-Stroms. Dieser Übergang liegt daher in der Nähe von $V_t - V_{\text{Ampl}}$. Im Bereich C der Elliot-Kurve findet der im vorigen Abschnitt beschriebene konventionelle CP-Effekt statt. Hier erreicht der CP-Strom sein Maximum $I_{\text{CP,max}}$. Die minimale Gate-Spannung befindet sich unterhalb der Flachbandspannung und die maximale Gate-Spannung oberhalb der Schwellspannung. In diesem Zustand wird bei jedem Wechsel von Akkumulation zu Inversion und zurück eine bestimmte Ladungsmenge von den Source- und Drain-Gebieten ins Substrat gepumpt. Diese ist proportional zu den in den Grenzflächenzuständen gespeicherten Ladungen und bewirkt den CP-Strom am Substrat. Bereich D verhält sich analog zu Bereich B, nur dass hier die Löcherkonzentration ausschlaggebend ist. Da die minimale Gate-Spannung den Transistor nicht mehr ausreichend in Akkumulation bringt, verringert sich die Zahl der Rekombinationsprozesse aufgrund der sinkenden Oberflächenkonzentration von Löchern drastisch und der CP-Strom fällt rapide ab. Dieser Übergang liegt damit in der Nähe der Flachbandspannung V_{fb} des Transistors. Im Bereich E befindet sich der Transistor schließlich zu jeder Zeit in Inversion und die Grenzflächenzustände sind

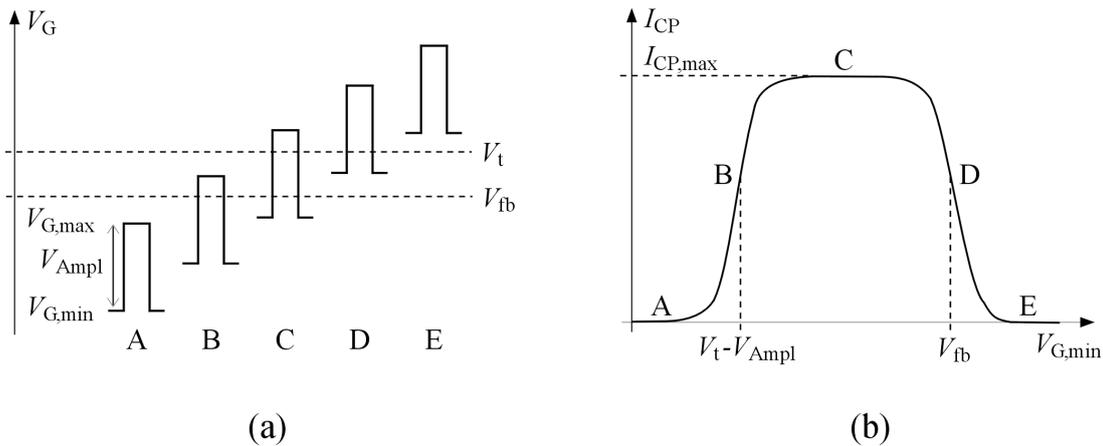


Abbildung 3.4: Charge-Pumping mit konstanter Amplitude. (a) Lage des Gate-Spannungspulses bezüglich Flachbandspannung V_{fb} und Schwellspannung V_t des Transistors für unterschiedliche Spannungsbedingungen (Bereiche A bis E). (b) CP-Strom in Abhängigkeit der minimalen Gate-Spannung $V_{G,\text{min}}$.
 Bereich A: Der Transistor befindet sich während des Spannungspulses zu jeder Zeit in Akkumulation. Alle Grenzflächenzustände sind mit Löchern gefüllt, so dass keine Rekombination stattfindet und kein CP-Strom fließt.
 Bereich B: Übergangsbereich mit einsetzenden Rekombinationsprozessen und ansteigendem CP-Strom.
 Bereich C: Konventioneller CP-Effekt mit stetigem Wechsel zwischen Akkumulation und Inversion. Der CP-Strom erreicht sein Maximum $I_{\text{CP,max}}$.
 Bereich D: Übergangsbereich mit aussetzender Rekombination und fallendem CP-Strom.
 Bereich E: Der Transistor befindet sich zu jeder Zeit in Inversion. Löcher gelangen nicht mehr zur Grenzfläche und der gemessene Substratstrom fällt wieder auf Null ab.

vollständig mit Elektronen besetzt. Die für die Rekombination benötigten Löcher können nicht mehr an die Grenzfläche gelangen, so dass der gemessene Substratstrom wieder auf Null abfällt. Tatsächlich besteht der CP-Strom in den Bereichen A und E allerdings aus den Leckströmen der Source- und Drain-Dioden.

Dieser CP-Modus mit konstanter Amplitude wird in der Literatur oft als Standard-Modus betrachtet [172] und eignet sich sehr gut, um einen ersten Eindruck über die mittlere Grenzflächenzustandsdichte zu erhalten [173–176]. Darüber hinaus wurden mit dieser Methode auch einzelne Störstellen in MOSFETs mit Gate-Flächen zwischen $1 \cdot 10^{-9} \text{ cm}^2$ und $1 \cdot 10^{-8} \text{ cm}^2$ nachgewiesen [177, 178].

CP mit variabler Amplitude

Ein weiterer CP-Modus besteht darin, die minimale Gate-Spannung $V_{G,\text{min}}$ konstant zu halten und die maximale Gate-Spannung $V_{G,\text{max}}$ und damit die Amplitude des Pulses sukzessive ansteigen zu lassen. Dieser Modus ist in Abbildung 3.5 dargestellt.

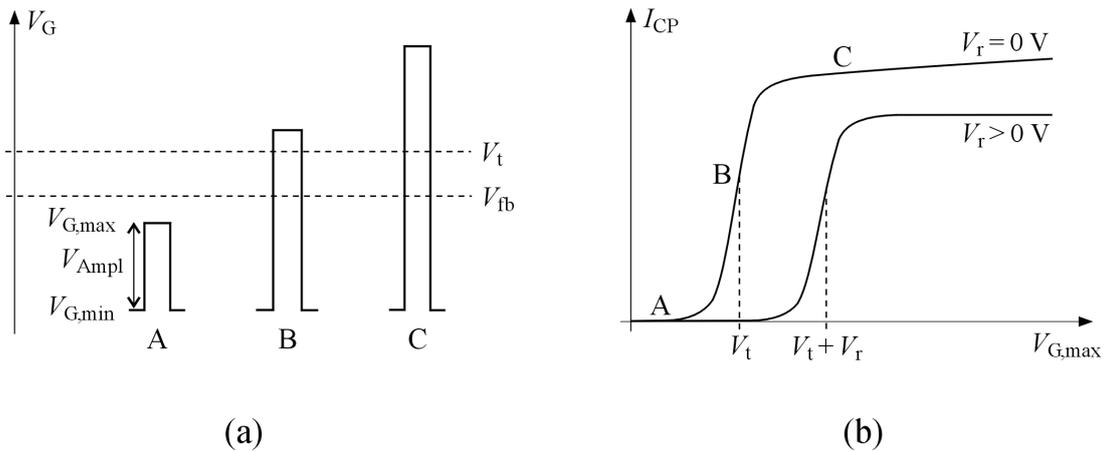


Abbildung 3.5: Charge-Pumping mit variabler Amplitude. (a) Lage des Gate-Spannungspulses bezüglich Flachbandspannung V_{fb} und Schwellspannung V_t des Transistors für unterschiedliche Spannungsbedingungen (Bereiche A bis C). (b) CP-Strom in Abhängigkeit der maximalen Gate-Spannung $V_{G,max}$. Bereich A und B: Analog zu Abbildung 3.4. Bereich C: Konventioneller CP-Effekt mit stetigem Wechsel zwischen Akkumulation und Inversion. Der CP-Strom steigt aufgrund der geometrischen Komponente weiter leicht an. Durch Anlegen einer Sperrspannung V_r an die Source- und Drain-Kontakte wird dieser Effekt unterdrückt.

Hierbei unterteilt sich die gemessene CP-Kurve in Abhängigkeit von $V_{G,max}$ in drei Bereiche A bis C, wovon der Bereich C im Folgenden näher erläutert wird. Die physikalischen Vorgänge in den Bereichen A und B sind dieselben wie beim zuvor beschriebenen CP-Modus mit konstanter Amplitude. Im Bereich C erreicht der CP-Strom eine Sättigung, da die Lage des Fermi-Niveaus an der Grenzfläche mit steigender Gate-Spannung nahezu konstant bleibt. In Experimenten wurde allerdings ein weiterer Anstieg des CP-Stroms mit steigendem $V_{G,max}$ beobachtet [53, 162, 171]. Dieser Anstieg wird als geometrische Komponente bezeichnet und resultiert aus einem geringen Teil von Inversionsladungen, der beim Umschalten des Gate-Spannungspulses den Kanal nicht schnell genug verlassen kann und dort verbleibt. Diese Ladungen rekombinieren mit einströmenden Majoritätsladungsträgern aus dem Substrat und bilden eine zusätzliche Komponente im CP-Strom. Die geometrische Komponente erweitert die Gleichung 3.17 für den CP-Strom zu

$$I_{CP} = f A_G [q D_{it} \Delta E - \alpha_{geo} C_{ox} (V_{G,max} - V_t)] \quad . \quad (3.18)$$

Dabei bezeichnet α_{geo} den Bruchteil der Inversionsladung, der nicht in die Source- und Drain-Gebiete abfließt und C_{ox} die Oxidkapazität. Experimentell kann die geometrische Komponente unterdrückt werden, indem eine Sperrspannung V_r von wenigen Volt an die Source- und Drain-Dioden angelegt wird [53]. Allerdings führt diese Sperrspannung zu einer Absenkung des CP-Sättigungsstroms, weil sich die Raumladungszonen der Source- und Drain-Gebiete ausdehnen und den Transistorkanal verkürzen [171]. Damit verringert sich auch die zum Charge-Pumping beitragende Gate-Fläche. Die Auswirkungen einer Sperrspannung an den Source- und Drain-Kontakten sind schematisch in Abbildung 3.5b dargestellt.

Dieser CP-Modus mit variabler Amplitude eignet sich besonders gut zur Detektion von Störstellen in der Nähe der Source- und Drain-Gebiete und damit zur Bestimmung von Schädigungen durch heiße Ladungsträger [179–182]. Betrachtet man den tatsächlichen Verlauf der Schwellspannung und der Flachbandspannung entlang des Kanals, so ist dieser nicht homogen aufgrund der Dotierübergänge nahe Source und Drain. Am Beispiel des n -Kanal-MOSFETs sind V_t und V_{fb} in diesen Regionen zu negativen Spannungen verschoben. Dies bewirkt, dass hier der CP-Strom von Akkumulation kommend eher einsetzt als in der Kanalmitte und somit eine leichte Schulter in der CP-Kurve aus Abbildung 3.5 zwischen den Bereichen A und B entsteht [183, 184]. Unter der Annahme einer homogen verteilten Grenzflächenzustandsdichte bei einem ungestressten Transistor und bei bekanntem Verlauf der Schwellspannung $V_t(x)$ entlang des Kanals kann der lokale Anstieg der Störstellendichte unter Stress bestimmt werden [185]. Auf diese Weise lässt sich neben der Grenzflächenzustandsdichte auch die Dichte von festen Oxidladungen getrennt davon bestimmen [186]. Rosmeulen *et al.* nutzten diese Methode dazu, um die Verteilung von gespeicherten Ladungen in nichtflüchtigen Speichern zu detektieren [187].

CP mit variablen Pulsflanken

Neben der Variation der Spannungen, wie in den beiden vorangegangenen Anschnitten beschrieben, können auch die zeitlichen Parameter des Gate-Spannungspulses beim CP variiert werden. Mit diesem CP-Modus ist es möglich, die Energieverteilung der Grenzflächenzustände über einen weiten Bereich der Energiebandlücke zu bestimmen [162, 188]. Dazu wird die Zeitdauer der ansteigenden Flanke des Spannungspulses sukzessive verlängert und jeweils der CP-Strom gemessen. Die Frequenz, das Tastverhältnis¹ und die Zeitdauer der fallenden Flanke werden dabei konstant gehalten. Am Beispiel des n -Kanal-MOSFETs ändert sich dadurch die energetische Grenze der Löcheremission $E_{em,h}$ (Gleichung 3.13). Somit dehnt sich bei kürzer werdender Anstiegsflanke das beitragende Energieintervall ΔE in der unteren Hälfte der Energiebandlücke immer weiter in Richtung Valenzbandkante aus. Durch die Differentiation des CP-Stroms nach der Flankendauer t_r ergibt sich die Grenzflächenzustandsdichte in Abhängigkeit der unteren energetischen Grenze zu

$$D_{it}(E_{em,h}) = -\frac{t_r}{qA_G k_B T f} \frac{dI_{CP}}{dt_r} . \quad (3.19)$$

Dies liefert unmittelbar die Energieverteilung von Grenzflächenzuständen in der unteren Hälfte der Bandlücke. Analog dazu kann das Experiment mit einer variablen fallenden Pulsflanke und konstant gehaltener steigender Flanke durchgeführt werden. Hierbei verschiebt sich die obere energetische Grenze $E_{em,e}$ in Abhängigkeit von t_f und die Energieverteilung der Grenzflächen-

¹Das Tastverhältnis gibt den Bruchteil der Pulsperiode an, der aus ansteigender Flanke und anschließendem konstanten Verlauf besteht. Im Englischen wird dieser auch als *duty cycle* bezeichnet.

zustände in der oberen Hälfte der Bandlücke ergibt sich analog zu

$$D_{it}(E_{em,e}) = -\frac{t_f}{qA_G k_B T f} \frac{dI_{CP}}{dt_f} \quad (3.20)$$

In der Literatur fand dieser CP-Modus häufig Anwendung, um die energetische Verteilung der unter verschiedenen Stressbedingungen generierten Grenzflächenzustände zu beobachten [189–191].

CP mit variabler Frequenz

Ein weiterer sehr wichtiger Modus der CP-Methode besteht darin, die Frequenz des CP-Pulses zu variieren. Da dieser Modus hauptsächlich in dieser Arbeit zur Anwendung kam, wird das Verfahren näher in Abschnitt 4.3.2 beschrieben. An dieser Stelle soll kurz auf die Möglichkeiten eingegangen werden, die dieser CP-Modus für die Charakterisierung von MOS-Bauelementen bietet.

Die Idee hierbei ist, durch Absenkung der Messfrequenz den Ladungsträgern an der Grenzfläche mehr Zeit zu geben, um über Tunnelprozesse mit grenzflächennahen Oxid-Störstellen zu kommunizieren [192]. Dazu werden die Zeitdauern der Pulsflanken t_r und t_f konstant gehalten, damit das beitragende Energieintervall ΔE ebenfalls konstant bleibt. Die Zeiten t_h und t_l , in der die Gate-Spannung konstant ist, verlängern sich aufgrund des gleichbleibenden Tastverhältnisses gleichmäßig mit abnehmender Frequenz, wie in Abbildung 3.6a dargestellt ist. Dadurch können Ladungsträger über Grenzflächenzustände in tiefer im Oxid liegende Störstellen tunneln, wenn deren Tunnelzeitkonstante kleiner ist als die Zeit, in der sich der Transistor in Inversion beziehungsweise Akkumulation befindet. Dies führt zu einer zusätzlichen CP-Komponente im gemessenen Strom, die von der Messfrequenz abhängt und mit fallender Frequenz zunimmt.

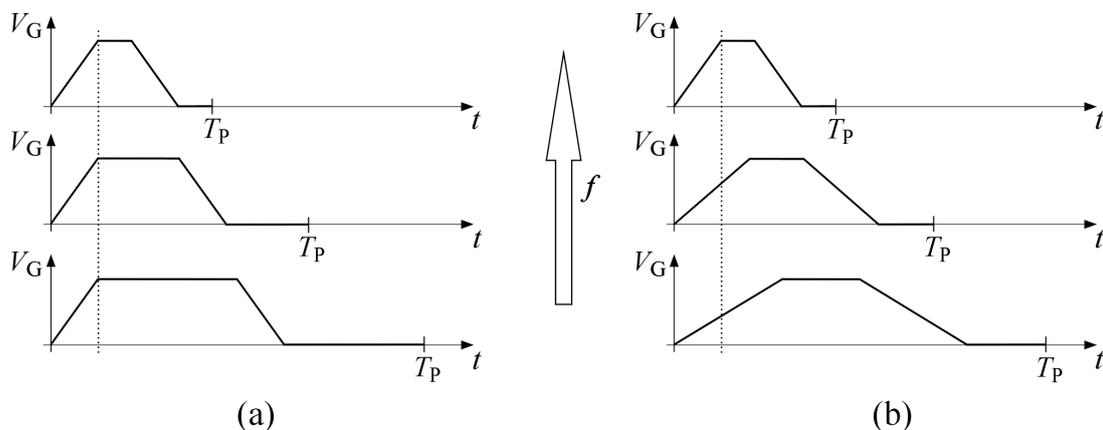


Abbildung 3.6: Charge-Pumping mit variabler Frequenz. (a) Die Zeitdauern der Flanken des Gate-Spannungspulses werden bei Erhöhung der Frequenz konstant gehalten. (b) Die gesamte Pulsform wird mit Erhöhung der Frequenz skaliert.

Der Vollständigkeit halber sei an dieser Stelle erwähnt, dass das frequenzabhängige CP auch so betrieben werden kann, dass die Pulsflanken stets einen festen prozentualen Anteil an der gesamten Periodendauer T_P des Pulses einnehmen, wie in Abbildung 3.6b dargestellt. Dies führt zu einer anderen Abhängigkeit des CP-Stroms von der Frequenz. Hierüber kann das geometrische Mittel der Einfangquerschnitte von Störstellen bestimmt werden [162]. Der Extremfall dieser Methode liegt vor, wenn die Pulsflanken zusammengenommen 100 % der Pulsperiode einnehmen. In diesem Fall liegen Dreieckspulse vor, wenn $\frac{t_r}{T_P} = 0,5$, oder Sägezahnpulse, wenn die Anteile asymmetrisch auf die steigende und fallende Pulsflanke verteilt werden. Diese Methode ist eingehender in [193] dargestellt.

Mit dem oben beschriebenen CP-Modus werden also neben den Grenzflächenzuständen auch die grenzflächennahen Oxidstörstellen erfasst. Paulsen *et al.* nutzten diese Methode, um Oxidstörstellen in strahlungsgeschädigten MOSFETs und in SONOS-Speicherezellen zu detektieren [194]. Des Weiteren wurde diese Methode auch an Transistoren mit dünnen High- k -Dielektrika durchgeführt, indem der dabei einsetzende störende Einfluss von direkten Tunnelströmen eliminiert wurde [195]. D. Bauza und Y. Maneglia entwickelten die Methode weiter und extrahierten Profile der Störstellendichte in einem grenzflächennahen Oxidbereich bis etwa 1 nm Abstand zur Grenzfläche [196].

3.3 Rauschen

Bei jeder Messung von kleinen elektrischen Signalen stößt man mit zunehmender Verstärkung der Signale an eine untere Grenze der Auflösung, die durch spontane Fluktuationen des Signals hervorgerufen wird. Diese Fluktuationen werden als Rauschen bezeichnet. Der Begriff des Rauschens stammt aus der Gebiet der Akustik, wo Fluktuationen der Spannung oder des Stroms bei hinreichend großer Verstärkung ein Rauschen in einem Lautsprecher hervorrufen [163].

Fluktuationen einer physikalischen Größe werden durch Zufallsvariablen beschrieben, zum Beispiel ein Strom durch $I(t)$, wobei t die Zeit ist. Diese Größen können durch ihren Mittelwert \bar{I} und Varianz $\overline{\Delta I^2}$ charakterisiert werden. Eine besonders wichtige und leistungsfähige Methode um fluktuierende Größen zu analysieren ist die Fourier-Methode. Sie beschreibt die zeitlich fluktuierende Größe durch die spektrale Leistungsdichte in einem infinitesimal kleinen Frequenzintervall, im Falle eines Stroms also durch $S_I(f)$ mit der Einheit A^2/Hz . Die Spektraldichte² hängt nicht mehr von der Zeit t ab, sondern von deren Fourier-Transformierten, der Frequenz f . Der Vorteil dieser Methode ist, dass das Rauschen über nur eine Größe quantitativ beschrieben wird.

Die wichtigsten und am häufigsten in Halbleitern auftretenden Rauscharten sind thermisches Rauschen, Schrotrauschen (*engl.*: shot noise), Generations-Rekombinationsrauschen (G-R-Rauschen), RTS (*engl.*: Random Telegraph Signal) und $1/f$ -Rauschen (auch Funkelrauschen, *engl.*:

²In dieser Arbeit werden die Begriffe Spektraldichte sowie spektrale Dichte synonym für die spektrale Leistungsdichte verwendet. Aus dem Englischen ist auch der Begriff *power spectrum* bekannt.

flicker noise). Die verschiedenen Rauscharten unterscheiden sich in der Frequenzabhängigkeit ihrer spektralen Dichte. Die Spektraldichte von thermischem Rauschen und Schrotrauschen ist frequenzunabhängig, während sich die des G-R-Rauschens und RTS proportional zu $1/f^2$ verhält. Der Name des $1/f$ -Rauschens rührt von der zur Frequenz invers verlaufenden Spektraldichte her.

In dieser Arbeit wird speziell das $1/f$ -Rauschen behandelt und für elektrische Charakterisierungen von Halbleiterbauelementen herangezogen. Die Ursachen und die in der Literatur existierenden Modelle dieser Rauschart werden ausführlich in Abschnitt 3.3.1 erläutert. Der Vollständigkeit halber werden in den folgenden Abschnitten zunächst weitere mögliche Rauscharten und deren Quellen kurz vorgestellt.

Thermisches Rauschen

Jeder elektrische Widerstand R mit einer Temperatur T unterliegt thermischem Rauschen, auch bekannt als Johnson-Nyquist-Rauschen [197]. Dies wird verstanden durch zufällige Bewegungen von Ladungsträgern, die Fluktuationen der messbaren Größen, zum Beispiel I oder V , hervorrufen. Die Spektraldichte des thermischen Stromrauschens ist gegeben durch

$$S_I = \frac{4k_B T}{R} \quad (3.21)$$

und somit, wie bereits erwähnt, unabhängig von der Frequenz. Es wird daher oft auch als weißes Rauschen bezeichnet und ist fundamental [198].

Thermisches Rauschen tritt auch im Kanal eines MOSFETs auf. Arbeitet der Transistor im ohmschen Bereich, so kann der Kanal als homogener Widerstand aufgefasst werden. Bei nicht homogenem Kanal müsste dieser für Rauschbetrachtungen in mehrere kleine Teile Δx mit Widerstand R_x unterteilt werden. Die spektrale Dichte des Rauschens eines MOSFETs ist dann durch

$$S_I = 4k_B T \frac{\mu W}{L^2} \int_0^L Q_i dx \quad (3.22)$$

gegeben, wobei Q_i die Flächenladungsdichte der Inversionsschicht angibt und μ die Beweglichkeit der Kanalladungsträger [197].

Schrotrauschen

Schrotrauschen in Halbleitern entsteht, wenn Ladungsträger über eine Barriere hinweg emittiert werden. Dies geschieht beispielsweise an pn -Übergängen, an Schottky-Barrieren oder in Bipolartransistoren. Unter Vernachlässigung von Rekombinationsprozessen während der Emission passiert ein Ladungsträger die Barriere in einer mittleren Zeit τ und ruft dabei einen Stromimpuls hervor. Der Gesamtstrom ist die Summe der einzelnen statistisch unkorrelierten Stromimpulse. Die spektrale Dichte des Stroms ist beim Schrotrauschen ebenso wie beim thermischen

Rauschen frequenzunabhängig und gegeben durch

$$S_I = 2qI \quad . \quad (3.23)$$

Diese Formel gilt für Frequenzen $f \ll (2\pi\tau)^{-1}$, die durch die mittlere Transmissionszeit der Barriere gegeben ist.

G-R-Rauschen und RTS

Bei den Betrachtungen zum Schrotrauschen im vorigen Abschnitt wurden Rekombinationsprozesse vernachlässigt. In Halbleitern, die Dotierstoffe, Störstellen oder Rekombinationszentren aufweisen, treten Generations- und Rekombinationsprozesse von Elektronen und Löchern auf. Bei diesen Vorgängen ändert sich die Zahl N der Elektronen beziehungsweise je nach Betrachtungsweise die der Löcher. Diese Fluktuation ΔN der Teilchenzahl erzeugt ein Rauschen, das als Generations-Rekombinationsrauschen (G-R-Rauschen) bezeichnet wird. Die Spektraldichte ist gegeben durch

$$S_N = 4\overline{\Delta N^2} \frac{\tau}{1 + (2\pi f\tau)^2} \quad . \quad (3.24)$$

Hierbei gibt τ die mittlere Lebensdauer der betrachteten Teilchen an. Die Spektraldichte stellt ein Lorentzspektrum dar und ist in Abbildung 3.7 skizziert.

Für kleine Frequenzen, wenn $2\pi f\tau \ll 1$, ist die spektrale Dichte annähernd konstant. Bei großen Frequenzen, wenn $2\pi f\tau \gg 1$, fällt die spektrale Dichte dagegen mit $1/f^2$ ab. Der Übergang ist durch die charakteristische Frequenz $f_c = (2\pi\tau)^{-1}$ gekennzeichnet.

Der bisher betrachtete Fall gilt für Prozesse mit einer einheitlichen Zeitkonstante τ . Treten mehrere Prozesse mit unterschiedlichen Zeitkonstanten τ_i auf, so setzt sich die Spektraldichte aus der gewichteten Summe der einzelnen Spektraldichten zusammen:

$$S_N = 4\overline{\Delta N^2} \sum_{i=1}^M g(\tau_i) \frac{\tau_i}{1 + (2\pi f\tau_i)^2} \quad , \quad (3.25)$$

wobei M die Anzahl der Prozesse angibt und $g(\tau_i)$ die normierte Gewichtungsfunktion ist, für die gilt

$$\sum_{i=1}^M g(\tau_i) = 1 \quad . \quad (3.26)$$

Für den Fall kontinuierlich verteilter Zeitkonstanten geht die Summe in Gleichung 3.25 in ein Integral über:

$$S_N = 4\overline{\Delta N^2} \int_0^\infty \frac{\tau g(\tau) d\tau}{1 + (2\pi f\tau)^2} \quad . \quad (3.27)$$

Die Funktion $g(\tau)$ ist hier eine Wahrscheinlichkeitsdichte. Der Ausdruck $g(\tau)d\tau$ gibt die normierte Wahrscheinlichkeit an, eine Zeitkonstante zwischen τ und $\tau + d\tau$ zu finden, wobei die Normierungsbedingung

$$\int_0^\infty g(\tau) d\tau = 1 \quad (3.28)$$

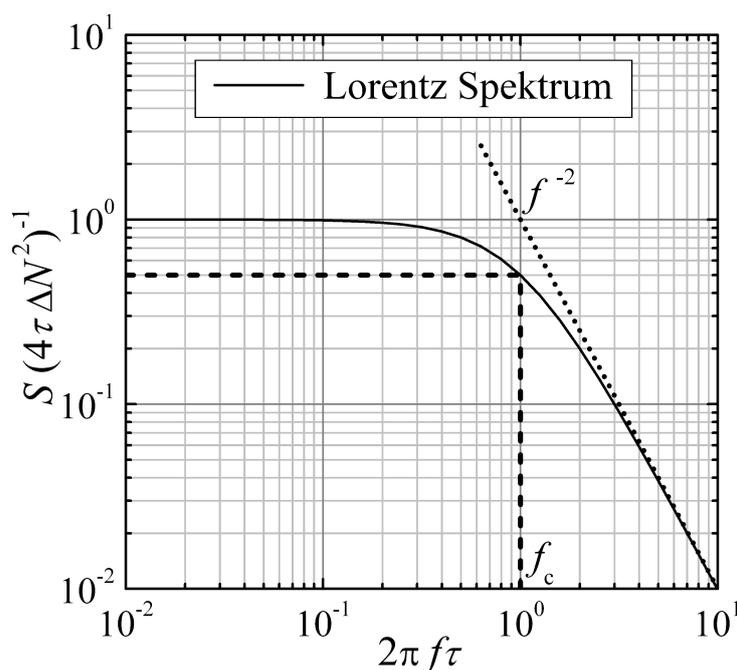


Abbildung 3.7: Verlauf eines normierten Lorentzspektrums. Für kleine Frequenzen zeigt sich ein Plateau, welches zu großen Frequenzen mit $1/f^2$ abfällt. Der Übergangsbereich ist durch die Grenzfrequenz f_c gekennzeichnet, bei der die Spektraldichte auf 50 % des Plateauwertes abgefallen ist.

gilt [163].

Ein besonderer Fall des G-R-Rauschens liegt vor, wenn sich die Teilchenzahl um den Wert $\Delta N = 1$ ändert und die Gesamtteilchenzahl N klein ist. Dieses Rauschen wird als RTS (Random Telegraph Signal) bezeichnet und hat ebenfalls ein Lorentzspektrum. Es tritt in Halbleiterbauelementen auf und ist besonders gut in MOSFETs mit sehr kleinen Gate-Flächen zu beobachten. Bei üblichen Störstellendichten von $4 \cdot 10^{10} \text{ cm}^{-2}$ an der $Si-SiO_2$ Grenzfläche und typischen Gate-Längen und -Weiten von je 50 nm ergibt sich etwa eine Störstelle pro Transistor. Mit der mittleren Emissionszeit τ_e und der mittleren Einfangzeit τ_c der Störstelle ist die Varianz eines reinen RTS gegeben durch

$$\overline{\Delta N^2} = \frac{1}{2 + \frac{\tau_e}{\tau_c} + \frac{\tau_c}{\tau_e}} \quad (3.29)$$

Durch Einsetzen der Varianz in Gleichung 3.24 erhält man die Spektraldichte des RTS, die einem Lorentzspektrum mit der charakteristischen Zeitkonstante $\tau = (\tau_e^{-1} + \tau_c^{-1})^{-1}$ folgt [163].

In Abbildung 3.8 ist ein Gesamtspektrum des Rauschens eines Halbleiters über einen sehr großen Frequenzbereich wiedergegeben. Im mittleren Frequenzbereich ist das G-R-Rauschen zu erkennen, das gemäß Gleichung 3.24 zunächst einen konstanten Verlauf zeigt und oberhalb von f_c proportional zu $1/f^2$ abfällt. Bei Frequenzen oberhalb von f_c tritt thermisches Rauschen zu Tage, das entsprechend Gleichung 3.21 ebenfalls einen konstanten Verlauf zeigt [198]. Am

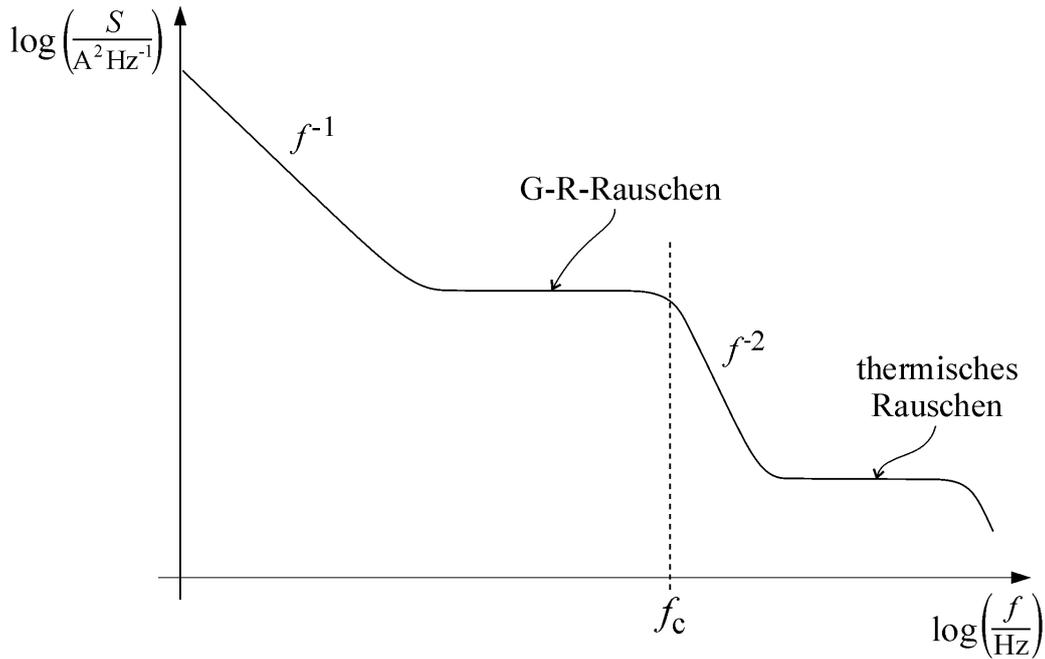


Abbildung 3.8: Generelles Rauschspektrum eines Halbleiters. Für kleine Frequenzen zeigt sich ein Verlauf proportional zu $1/f$, im mittleren Frequenzbereich dominiert G-R-Rauschen und im oberen thermisches Rauschen.

unteren Ende des Frequenzbereiches dominiert das so genannte $1/f$ -Rauschen, welches im folgenden Abschnitt näher erläutert wird.

3.3.1 $1/f$ -Rauschen

Einen besonderen Platz unter allen Arten von Rauschen nimmt das so genannte $1/f$ -Rauschen oder Funkelrauschen ein. Es tritt in nahezu allen elektronischen Bauelementen auf und besitzt eine Spektraldichte S , die eine Frequenzabhängigkeit von $1/f^\gamma$ aufweist. Der Exponent γ liegt nahe um Eins, in der Regel gilt $0,8 \lesssim \gamma \lesssim 1,2$. Eine besondere Eigenschaft des $1/f$ -Rauschens ist die Skaleninvarianz. Das bedeutet, dass es keine charakteristische Frequenz oder Zeitkonstante gibt. Das $1/f$ -Rauschen existiert auf kleinen Zeitskalen ebenso wie auf großen Skalen und überdeckt demnach fast den gesamten Frequenzbereich [199]. Berechnet man jedoch die gesamte Rauschleistung über

$$P = \int_0^\infty S df \quad , \quad (3.30)$$

so divergiert das Integral an den Grenzen und die Leistung strebt gegen Unendlich. Genauer betrachtet kann die Spektraldichte daher nicht über den gesamten Frequenzbereich mit der exakten $1/f$ -Abhängigkeit verlaufen, sondern muss für Frequenzen $f < f_0$ langsamer und für $f > f_{\max}$ schneller als $1/f$ variieren [163]. Caloyannides hat die Gültigkeit des $1/f$ -Verlaufs jedoch bis hinunter zu Frequenzen von 10^{-6} Hz nachgewiesen [200].

Die Ursachen des $1/f$ -Rauschens in MOSFETs werden in der Literatur vielfach diskutiert [201–206]. In einigen Studien wird die Fluktuation der Ladungsträgerdichte im Kanal basierend auf dem McWorther-Modell [207] für das Rauschen verantwortlich gemacht [208–212]. Diese entstehen durch den Ladungsträgeraustausch zwischen den Grenzflächenzuständen und dem Kanal. Das McWorther-Modell beruht ursprünglich auf Betrachtungen des Rauschens im Halbleitervolumen und wurde von den Autoren für die Betrachtungen an der Halbleiteroberfläche angepasst. Unterstützung findet das Modell in der vielfach beobachteten Korrelation zwischen der Rauschleistung und der Grenzflächenzustandsdichte [204, 212–214].

Mathematisch kann gezeigt werden, dass sich ein $1/f$ -Spektrum ergibt, wenn Ladungsträger aus dem Kanal über Tunnelprozesse mit Oxidstörstellen kommunizieren. Dieser Ladungsträgeraustausch führt zu Fluktuationen der Ladungsträgerzahl im Kanal. Zur Bestimmung der Spektraldichte kann daher die Gleichung 3.27 aus dem vorangegangenen Abschnitt herangezogen werden. Die Zeitkonstante für das Tunneln eines Ladungsträgers in eine Störstelle im Abstand z ist gegeben durch

$$\tau(x) = \tau_0 e^{\frac{z}{\lambda}} \quad , \quad (3.31)$$

wobei τ_0 die Zeitkonstante an der Grenzfläche und λ den Abschwächungskoeffizienten der Wellenfunktion des tunnelnden Ladungsträgers bezeichnet [208]. Unter der Annahme einer konstanten Störstellendichte in einem Bereich $0 < z \leq z_{\max}$ ist die Wahrscheinlichkeit, einen Ladungsträger im Intervall zwischen z und $z + dz$ einzufangen, gegeben durch

$$g(\tau)d\tau = \begin{cases} \frac{dz}{z_{\max}} & : 0 < z \leq z_{\max} \\ 0 & : \text{sonst} \end{cases} \quad . \quad (3.32)$$

Mit $\tau_{\max} = \tau(z_{\max})$ ergibt sich die Wahrscheinlichkeitsverteilung zu

$$g(\tau)d\tau = \frac{dz}{z_{\max}} = \frac{d\tau}{\tau} \frac{\lambda}{z_{\max}} = \frac{d\tau}{\tau} \frac{1}{\ln\left(\frac{\tau_{\max}}{\tau_0}\right)} \quad , \quad (3.33)$$

wenn $\tau_0 < \tau \leq \tau_{\max}$, ansonsten Null. Die Einschränkung von τ ist notwendig, damit die Wahrscheinlichkeit über Gleichung 3.28 normierbar bleibt. Eingesetzt in Gleichung 3.27 ergibt sich dann das $1/f$ -Spektrum

$$S_N = \frac{\overline{\Delta N^2}}{f \ln\left(\frac{\tau_{\max}}{\tau_0}\right)} \quad , \quad (3.34)$$

wenn $f_{\max} \ll f \ll f_0$, wobei $f_{\max} = \tau_{\max}/2\pi$ und $f_0 = \tau_0/2\pi$ [163].

Andere Studien gehen von einer Fluktuation der Beweglichkeit $\Delta\mu$ von Ladungsträgern als Ursache für das $1/f$ -Rauschen aus [215–217]. Dieses Modell wird als Hooge-Modell [218] bezeichnet und legt Streuprozesse von freien Ladungsträgern am Atomgitter eines Festkörpers zu Grunde. Für homogene Halbleiter und Metalle lässt sich für die spektrale Dichte des Stromrauschens die empirische Formel

$$S_I = \frac{\alpha_H}{fN} I^2 \quad (3.35)$$

aufstellen, in der N die Anzahl freier Ladungsträger darstellt und α_H als Hooge-Parameter bezeichnet wird. Der Wert für den Hooge-Parameter ergibt sich aus experimentellen Daten zu

etwa $2 \cdot 10^{-3}$ [215].

Um dieses Modell auf die Oberfläche von Halbleitern anzuwenden, wie es für die Beschreibung des $1/f$ -Rauschens in MOSFETs notwendig ist, wurde eine Modifikation der Gleichung 3.35 eingeführt [216, 219]. Dabei wurde für den Hooge-Parameter der Ausdruck

$$\alpha'_H = \alpha_H \left(\frac{\mu}{\mu_1} \right)^2 \quad (3.36)$$

verwendet [216]. Hierbei ist μ_1 die Beweglichkeit, wenn nur Gitterstreuungen im Halbleitervolumen auftreten, und μ die reduzierte Beweglichkeit unter Berücksichtigung weiterer, gleichzeitig auftretender Streuprozesse, zum Beispiel Streuungen an der Grenzfläche eines MOSFETs. Hung *et al.* vereinheitlichten 1990 die beiden existierenden Modelle zum $1/f$ -Rauschen in MOSFETs [220]. Dieses neue Modell basiert auf Beobachtungen von RTS an MOSFETs mit Gate-Flächen kleiner als $1 \mu\text{m}^2$. Das Einfangen eines Kanalladungsträgers in eine nahe Störstelle im Siliziumdioxid führt zu einer Fluktuation ΔN der Ladungsträgerzahl. Gleichzeitig wirkt der eingefangene Ladungsträger aber auch als Coulomb-Streuzentrum für die anderen Kanalladungsträger und ruft somit Fluktuationen $\Delta\mu$ der Kanalbeweglichkeit hervor. Auf diese Weise sind die beiden Rauschquellen korreliert und können in dem vereinheitlichten Modell zusammengefasst werden. Das Modell kann die meisten experimentellen Daten in der Literatur erklären [220].

Die Spektraldichte des Drain-Stroms eines MOSFETs ist in dem vereinheitlichten Modell allgemein gegeben durch

$$S_{I_D} = \frac{k_B T I_D^2}{\lambda q f W L^2} \int_0^L D_T(E_F) \left[\frac{1}{n(x)} \pm \alpha \mu \right]^2 dx \quad , \quad (3.37)$$

wobei $D_T(E_F)$ die Volumendichte der Störstellen an der Fermi-Energie E_F bezeichnet. D_T wird dabei als räumlich konstant angenommen. $n(x)$ ist die Flächendichte der Ladungsträger entlang des Kanals zwischen Source ($x = 0$) und Drain ($x = L$). Der Beitrag der Beweglichkeitsfluktuationen zur gesamten Spektraldichte wird über den Streukoeffizienten α bestimmt. Das Vorzeichen dieses Beitrags muss durch experimentelle Daten bestimmt werden und ist in den meisten Fällen positiv [220].

Für sehr kleine Drain-Spannungen kann eine konstante Ladungsträgerdichte

$$n(x) = \frac{C_{\text{ox}}}{q A_G} (V_G - V_t) \quad (3.38)$$

im Kanal angenommen werden. $A_G = WL$ bezeichnet hierbei die Gate-Fläche. Im Allgemeinen wird statt der Spektraldichte des Stroms die auf das Quadrat des Stroms normierte Spektraldichte $S'_{I_D} := S_{I_D}/I_D^2$ angegeben. Mit diesen Konventionen wird Gleichung 3.37 zu

$$S'_{I_D} = \frac{k_B T}{\lambda q f A_G} \left(\frac{1}{n} \pm \alpha \mu \right)^2 D_T(E_F) \quad . \quad (3.39)$$

Die Spektraldichte ist demnach direkt proportional zu Störstellendichte $D_T(E_F)$ an der Fermi-Energie, welches auch experimentell bestätigt wurde [212].

3.3.2 Analyse des $1/f$ -Rauschens mittels FFT

Das im vorigen Kapitel beschriebene $1/f$ -Rauschen eignet sich besonders gut zur elektrischen Charakterisierung von Gate-Oxiden in MOSFETs. Die spektrale Dichte ist direkt proportional zur Störstellendichte $D_T(E_F)$ im Oxid nahe des Fermi-Niveaus (siehe Gleichung 3.39). Zur Bestimmung der Spektraldichte wird in dieser Arbeit der Drain-Strom bei konstanter Drain- und Gate-Spannung gemessen. Der Drain-Strom rauscht dabei um seinen Mittelwert \bar{I} . Die Spektraldichte des Stromrauschens ergibt sich über eine Fourier-Transformation der Abweichungen vom Mittelwert $\Delta I(t) = I(t) - \bar{I}$ zu

$$S_{I_D} = |\widetilde{\Delta I}(f)|^2 + |\widetilde{\Delta I}(-f)|^2 \quad , \quad (3.40)$$

wobei $\widetilde{\Delta I}(f)$ die Fourier-Transformierte der Schwankungen des Drain-Stroms $\Delta I(t)$ darstellt und gegeben ist durch

$$\widetilde{\Delta I}(f) = \int_{-\infty}^{\infty} \Delta I(t) e^{2\pi i f t} dt \quad . \quad (3.41)$$

Da $\Delta I(t)$ eine reelle Funktion ist, sind die beiden Summanden in Gleichung 3.40 gleich und die Spektraldichte kann auf der positiven Frequenzachse ($f > 0$) dargestellt werden als

$$S_{I_D} = 2|\widetilde{\Delta I}(f)|^2 \quad . \quad (3.42)$$

Dies wird auch als einseitige Spektraldichte bezeichnet und in dieser Arbeit für alle Rauschanalysen herangezogen.

In der Praxis liegt die kontinuierliche Funktion $\Delta I(t)$ allerdings nur als Satz diskreter Werte vor, die mit einem konstanten Zeitabstand Δt über einen endlichen Zeitraum T aufgezeichnet wurden. Daraus ergibt sich über das Nyquist-Theorem eine maximale Frequenz

$$f_c = \frac{1}{2\Delta t} \quad , \quad (3.43)$$

die bei einer diskreten Messung erfasst werden kann und durch die Abtastrate $1/\Delta t$ bestimmt ist. In Abbildung 3.9a ist dieser Zusammenhang dargestellt. Von dem Signal der durchgezogenen Kurve, das nur Frequenzen kleiner als f_c enthält, werden in konstanten Zeitabständen Δt Messwerte erfasst. Das Signal der gestrichelten Kurve enthält höhere Frequenzanteile als f_c , liefert aber bei der gleichen Abtastrate dieselben Messwerte. Anhand der Messwerte sind die beiden Signale nicht zu unterscheiden.

Dies hat Auswirkungen auf die Fourier-Transformation des Signals. Für den Fall, dass das Signal $\Delta I(t)$ keine höheren Frequenzen als f_c aufweist, gibt es keine Komplikationen bei der anschließenden Fourier-Transformation. Im Allgemeinen enthält das Messsignal aber höhere Frequenzen, die durch das diskrete Abtasten fälschlicherweise in den Frequenzbereich $0 < f < f_c$ gespiegelt werden. Dies bezeichnet man als Aliasing-Effekt [221]. In Abbildung 3.9b ist dieser Effekt dargestellt. Die Anteile der Fourier-Transformierten mit Frequenzen $f > f_c$ werden in den Bereich $0 < f < f_c$ zurück gespiegelt und verfälschen somit das erzeugte Spektrum.

Der Aliasing-Effekt kann auf verschiedene Weisen unterdrückt werden. Wenn die maximale

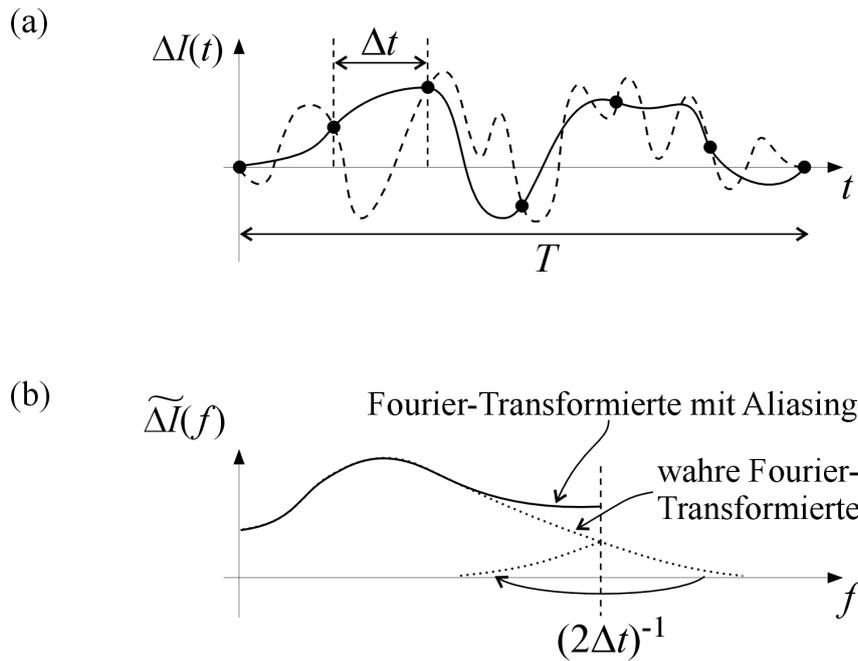


Abbildung 3.9: (a) Abtasttheorem nach Nyquist: Das Signal $\Delta I(t)$ (durchgezogene Kurve) wird zu diskreten Zeitpunkten im Abstand Δt erfasst (Punkte). Ein weiteres Signal mit Frequenzanteilen $f > f_c$ (gestrichelte Kurve) liefert bei gleicher Abtastrate dieselben Messwerte. Beide Signale sind anhand der Messwerte nicht zu unterscheiden. (b) Aliasing-Effekt: Enthält ein Messsignal höhere Frequenzanteile als die kritische Frequenz f_c , so addiert sich zur wahren Fourier-Transformierten im Bereich $0 < f < f_c$ der gespiegelte Anteil der höheren Frequenzen.

Frequenz f_{\max} des zu messenden Systems bekannt ist, kann die Abtastrate entsprechend hoch gewählt werden, so dass $f_c > f_{\max}$. Im Allgemeinen ist f_{\max} nicht bekannt beziehungsweise nicht bestimmbar, da das System über einen extrem weiten Frequenzbereich rauscht. Ebenso ist die maximale Abtastrate durch die Leistungsfähigkeit des Messgeräts bestimmt. In diesem Fall kann der Aliasing-Effekt durch einen vorgeschalteten analogen Tiefpass mit $f_{3\text{dB}} \simeq f_c$ unterdrückt werden [221].

Eine weitere Möglichkeit zur Unterdrückung des Aliasing ist die geeignete Wahl des Messmodus am Messgerät. Dazu wird die Integrationszeit t_{int} bei der Erfassung eines Messpunktes so gewählt, dass sie möglichst genau dem Zeitabstand Δt zwischen zwei Messpunkten entspricht. Für die Aufnahme des n -ten Messpunktes zum Zeitpunkt t_n integriert das Messgerät das Eingangssignal zwischen t_n und $t_n + t_{\text{int}} \leq t_{n+1}$ auf und gibt den Mittelwert für diesen Zeitraum an. Im Idealfall, wenn $t_{\text{int}} = \Delta t$, werden dadurch alle Schwankungen mit Frequenzen größer als f_c herausgemittelt. Praktisch ist dieser Fall nicht zu realisieren, da jedes Messgerät endliche Totzeiten zur internen Verarbeitung der Daten besitzt. Jedoch kann durch die Wahl einer möglichst nahe an Δt liegenden Integrationszeit $t_{\text{int}} \lesssim \Delta t$ eine sehr gute Unterdrückung des Aliasing-Effekts erzielt werden [222]. Dieses Verfahren wurde in dieser Arbeit zur Aufnahme der Messdaten angewendet.

3.3.3 Bewertung der Rauschmessmethode

In der Literatur werden neben der hier vorgestellten Drain-Strommessung und Auswertung auch andere Ansätze zur Messung und Analyse des Transistorrauschens verwendet. Eine Möglichkeit besteht darin, über einen Lastwiderstand die Spannungsschwankungen am Drain zu messen [223–225]. Diese werden über einen Spannungsverstärker mit geringem Eigenrauschen verstärkt und einem Signalanalysator zugeführt. Signalanalysatoren haben meist den Vorteil, dass sie Fourier-Transformationen in Echtzeit durchführen und die erzeugten Spektren mitteln können.

Eine weitere Variante der Rauschmessungen besteht darin, das Drain-Stromrauschen direkt über einen Transimpedanzverstärker zu messen, dessen Ausgangssignal dann wiederum von einem Signalanalysator ausgewertet wird [225, 226]. In manchen Publikationen wird das Rauschen des Gate-Leckstroms untersucht. Hierzu werden die Spannungsschwankungen am Gate-Kontakt ebenfalls mit einem Transimpedanzverstärker erfasst und analysiert [227].

Alle diese Ansätze benötigen einen sehr aufwendigen experimentellen Aufbau. Die verwendeten Vorverstärker und Signalanalysatoren dürfen kein zu hohes Eigenrauschen aufweisen und müssen aufeinander abgestimmt sein, um saubere Ergebnisse zu liefern. Bei der Verwendung von Transimpedanzverstärkern ist ebenfalls die Kenntnis der Verstärkungskurve (*engl.*: gain) wichtig, damit die Messwerte aufgrund des frequenzabhängigen Verstärkungsfaktors korrigiert werden können [224]. Der Verstärkungsfaktor fällt besonders am Rande des genutzten Frequenzbandes ab.

Aus diesen Gründen ist in dieser Arbeit, wie bereits oben beschrieben, ein sehr viel einfacher Ansatz gewählt worden. Der Nachteil der hier verwendeten Methode ist allerdings, dass die Messungen und die anschließende Auswertung mittels Software um ein Vielfaches länger dauern als mit einer entsprechenden FFT-fähigen Hardware. Für die Untersuchung von zum Beispiel Relaxationsphänomenen eignet sich dieser Messansatz daher nicht. Die Relaxation eines Stresszustands oder ein zeitlich driftender Drain-Strom wurden in dieser Arbeit nicht in dem Maße beobachtet, dass sie einen Einfluss auf die Ergebnisse hätten.

4 Charakterisierung von nitridierten Gate-Oxiden

Dieses Kapitel beschreibt die elektrischen Untersuchungen von MOS-Transistoren mit nitridierten Gate-Oxiden. Dabei steht die Charakterisierung der Degradation des Gate-Oxids unter Fowler-Nordheim-Stress (FN-Stress) im Vordergrund. Für die elektrische Untersuchung kommt zum einen die Charge-Pumping-Methode (Kapitel 4.3) und zum anderen die Messung des $1/f$ -Rauschens (Kapitel 4.5) zur Anwendung. Die untersuchten Bauelemente sind MOSFETs, die in CMOS-Logikschaltungen eingesetzt werden. Diese Schaltungen setzen sich sowohl aus p -MOSFETs als auch n -MOSFETs zusammen. Eine mögliche Degradation eines Transistors durch Stresseinwirkung, die sich zum Beispiel in einer Verschiebung der Schwellspannung äußert, kann zum Ausfall einer solchen Schaltung führen. Dies ist besonders gravierend, wenn vorwiegend einer der beiden Transistortypen (p -MOSFET oder n -MOSFET) degradiert. Daher wird das Degradationsverhalten im Folgenden für beide Transistortypen genauer untersucht. Die Polarität der Gate-Spannung spielt dabei eine wesentliche Rolle.

4.1 Verwendete Probenstrukturen

Die für die folgenden elektrischen Untersuchungen verwendeten n - und p -Kanal-MOSFETs wurden von der Firma LSI zur Verfügung gestellt. Beide Transistortypen sind gemeinsam auf einem Wafer vorhanden, der über eine n -Grunddotierung verfügt. Der n -Kanal-MOSFET ist in eine p -dotierte Substratwanne (ein so genannter p -Well) eingebettet. Die effektiven Kanaldotierungen wurden für beide Leitungstypen jeweils aus $C(V)$ -Messungen an MOS-Kapazitäten, die ebenfalls auf dem Wafer vorhanden waren, ermittelt (siehe Anhang A.1). Der n -Kanal-MOSFET besitzt demnach eine p -Wanne mit einer Dotierung von $N_A = 3,4 \cdot 10^{17} \text{ cm}^{-3}$, der p -Kanal-MOSFET eine n -Dotierung von $N_D = 6,1 \cdot 10^{16} \text{ cm}^{-3}$. Die nominelle Weite W_G des aktiven Kanalbereichs beträgt bei allen Transistoren $50 \mu\text{m}$, die nominelle Poly-Gate-Länge $L_G = 0,24 \mu\text{m}$. Beide Transistortypen sind mit einem LDD (Lightly Doped Drain) prozessiert worden, um die Generation von heißen Ladungsträgern gering zu halten.

Das Gate-Oxid besteht aus $6,4 \text{ nm}$ nitridiertem SiO_2 . Die Nitridierung erfolgte auf dem gesamten Wafer mittels einer NO -Temperung und reicht vom Gate bis zum Substrat. SIMS-

Messungen¹ zeigen, dass sich die größte Konzentration an Stickstoff auf der Gate-Seite des Oxid-Stapels befindet. Das Gate wurde beim n -Kanal-MOSFET aus Phosphor dotiertem n^+ -Poly- Si hergestellt, beim p -Kanal-MOSFET aus Bor dotiertem p^+ -Poly- Si . Für beide Leitungstypen lag die Dotierkonzentration in der Größenordnung von $1 \cdot 10^{20} \text{ cm}^{-3}$. Alle diese Angaben zum Gate-Oxid und zum Poly- Si beruhen auf persönlicher Korrespondenz mit dem Hersteller [228]. Die Berechnung des Oberflächenpotentials über das Berglund-Integral (siehe Anhang A.1) ergibt eine Flachbandspannung für den n -MOSFET von $V_{fb} = -1,0 \text{ V}$ und für den p -MOSFET von $V_{fb} = 1,0 \text{ V}$. Die wichtigsten Probenparameter sind in Tabelle 4.1 zusammengefasst.

	n -MOSFET	p -MOSFET
Kanaldotierung	$N_A = 3,4 \cdot 10^{17} \text{ cm}^{-3}$	$N_D = 6,1 \cdot 10^{16} \text{ cm}^{-3}$
Kanalweite W_G	$50 \mu\text{m}$	$50 \mu\text{m}$
Kanallänge L_G	$0,24 \mu\text{m}$	$0,24 \mu\text{m}$
Oxiddicke d_{ox}	$6,4 \text{ nm}$	$6,4 \text{ nm}$
Maximum der Oxidnitridierung	gate-seitig	gate-seitig
Flachbandspannung V_{fb}	$-1,0 \text{ V}$	$1,0 \text{ V}$
Schwelspannung V_t	$0,75 \text{ V}$	$-0,50 \text{ V}$
Dotierung des Poly- Si -Gates	$n^+ (N_D \sim 1 \cdot 10^{20} \text{ cm}^{-3})$	$p^+ (N_A \sim 1 \cdot 10^{20} \text{ cm}^{-3})$

Tabelle 4.1: Die wichtigsten Probenparameter der untersuchten MOS-Transistoren [228].

Jeder der untersuchten MOSFETs verfügt über getrennte elektrische Anschlüsse. Neben dem Source-, Drain- und Gate-Kontakt kann auch das Substrat beziehungsweise die Wanne von der Waferoberfläche aus kontaktiert werden. Die Schwelspannung wurde jeweils bei einem konstanten Drain-Strom von $I_{D,t} = 10 \mu\text{A}$ bestimmt.

4.2 Messaufbau für Charge-Pumping

Alle elektrischen Untersuchungen fanden in einem Waferprober der Firma Süss statt (Abbildung 4.1). Der Waferprober besteht aus einem heizbaren Messing-Chuck mit 10 cm Durchmesser. Auf diesem Chuck werden die Wafer mittels Vakuum angesaugt und somit fixiert. Zum Anfahren einer bestimmten Position auf dem Wafer befinden sich Stellschrauben an der Halterung des Chucks. Der Chuck stellt gleichzeitig einen elektrischen Rückkontakt dar, der gesondert belegt

¹SIMS (Secondary Ion Mass Spectroscopy) ist eine Methode, bei der die Oberfläche des zu untersuchenden Materials durch einen Ionenstrahl abgetragen wird. Die dabei entstehenden Sekundärionen werden in einem Massenspektrometer analysiert und geben Aufschluss über die atomare Zusammensetzung der Probe. SIMS ist ein zerstörendes Messverfahren und wird im Ultrahochvakuum durchgeführt.



Abbildung 4.1: Verwendeter Waferprober der Firma Süss, bestehend aus einem heizbaren Messing-Chuck, der über Stellschrauben verfahrbar ist (Mitte, hier mit Wafer belegt), vier Messnadeln (jeweils zwei links und rechts) und einem Lichtmikroskop.

werden kann oder einfach auf Masse des Messaufbaus gelegt wird. Die elektrische Kontaktierung der Bauelemente auf dem Wafer erfolgt von oben mittels Messnadeln, die mithilfe eines Lichtmikroskops über Stellschrauben mikrometergenau positioniert werden können. Der komplette Waferprober befindet sich in einer geerdeten Metallbox zur elektrischen Abschirmung. Abbildung 4.2 zeigt den schematischen Aufbau für Charge-Pumping-Messungen wie sie in dieser Arbeit durchgeführt wurden. Da das Substrat beziehungsweise die Wanne (einheitlich mit B bezeichnet) der verwendeten Proben von der Waferoberfläche aus kontaktiert werden konnte, wurde der Rückkontakt über den Chuck nicht benötigt. Um mögliche störende elektrische Einflüsse des Chucks auf die Probe zu unterdrücken, wurde eine isolierende Folie zwischen dem

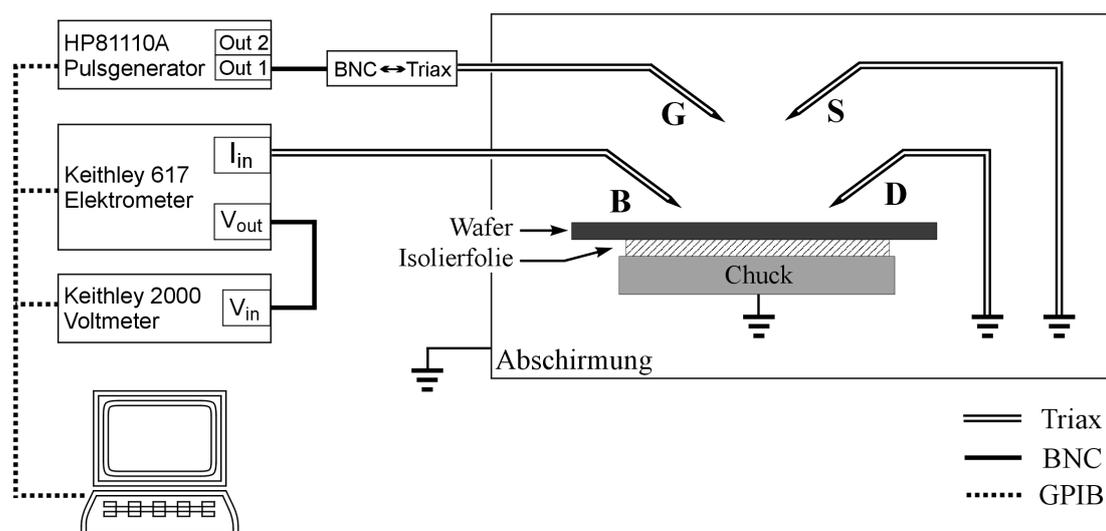


Abbildung 4.2: Schematische Darstellung des Messaufbaus für Charge-Pumping-Messungen. G: Gate-Kontakt, S: Source-Kontakt, D: Drain-Kontakt, B: Wannenananschluss

Wafer und dem Chuck eingeführt.

Die Messnadeln sind über dreiadrige Koaxialkabel (so genannte Triaxkabel) mit den Messgeräten außerhalb der Metallbox verbunden. Das Besondere bei den Triaxkabeln ist, dass der Innenleiter mit zwei koaxialen und elektrisch voneinander getrennten Abschirmungen versehen ist. Die mittlere Abschirmung (der so genannte Guard-Ring) wird von dem jeweiligen Messgerät durch eine aktive Steuerung (im Allgemeinen Operationsverstärker) auf demselben Potenzial gehalten, wie der Innenleiter. Somit wird der Leckstrom und die kapazitive Kopplung des Innenleiters zur Umgebung auf ein Mindestmaß reduziert. Dies ermöglicht extrem empfindliche Messungen von Strömen bis in den Subfemto-Ampère-Bereich.

Der Gate-Kontakt wurde mit einem Spannungspulsgenerator verbunden. Da dieser am Ausgang nur einen Anschluss für zweiadrige Koaxialkabel (so genannte BNC-Kabel) bereithält, wurde das Triaxkabel mittels eines Adapters in ein BNC-Kabel überführt, wobei der Guard-Ring innerhalb des Adapters endete. Der Substrat- beziehungsweise Wannenananschluss (B) wurde mit einem Elektrometer verbunden, das Ströme mit einer Auflösung von 0,1 fA messen kann. Am Ausgang V_{out} des Elektrometers liegt stets ein Spannungssignal an, das dem gemessenen Stromsignal proportional ist. Dieses Spannungssignal wird von einem Voltmeter aufgezeichnet. Alle Messgeräte werden über die GPIB-Schnittstelle von einem PC gesteuert. Als Steuerungssoftware wurde LabVIEW von der Firma National Instruments verwendet. Die Source- und Drain-Kontakte waren während der Charge-Pumping-Messungen stets geerdet.

Das Elektrometer misst nur die Gleichstromanteile des Substratstroms der Probe, da der Eingangverstärker eine sehr hohe Zeitkonstante besitzt. Der am Substrat- beziehungsweise Wannenananschluss der Probe zu erwartende Strom ist ein Wechselstrom mit im Gegensatz zum Charge-Pumping-Strom relativ großer Amplitude. Der Charge-Pumping-Strom selber ist diesem als Gleichstromsignal überlagert. Durch die hohe Zeitkonstante des Eingangskreises des

Elektrometers werden die Wechselstromanteile heraus gefiltert. Dieser Effekt wurde genutzt, um den Charge-Pumping-Strom, der einen Gleichstrom darstellt, zu messen.

4.2.1 Messablauf

Der Ablauf der elektrischen Messungen folgte einem stets gleichen Prinzip. Zunächst wurden alle Parameter und Geräteeinstellungen mittels der Steuerungssoftware LabVIEW über die GPIB-Schnittstelle an die Messgeräte übertragen. Danach wurde eine Messung gestartet. Nach Beendigung der Messung wurden die Daten wiederum mit LabVIEW aus dem Speicher der Messgeräte ausgelesen und gespeichert.

Bei den Charge-Pumping-Messungen wurde ein spezielles Verfahren angewendet, um mögliche parasitäre Ströme herauszufiltern. Der eingesetzte Pulsgenerator bietet die Möglichkeit, den Puls am Ausgang ein- und auszuschalten, ohne die übrigen Parameter neu übertragen zu müssen. Daher wurden bei jedem Messpunkt zunächst alle Pulsparameter an das Gerät übertragen, der Puls aber noch nicht eingeschaltet. In diesem Zustand befindet sich der Ausgang des Pulsgenerators und somit der Gate-Kontakt der Probe auf Massepotenzial und es fließt kein CP-Strom. Jeder Strom, der hierbei am Substrat gemessen wird, ist parasitär und rührt von Leckströmen an der Probe oder am Messaufbau her.

Aus diesem Grund wurde bei jedem Messpunkt der Substratstrom der Probe sowohl mit ausgeschaltetem als auch mit eingeschaltetem Gate-Spannungspuls gemessen. Die Differenz der beiden Ströme ergibt den von allen parasitären Strömen befreiten CP-Strom. Bei beiden Messungen wurden jeweils 1000 Messwerte mit dem Voltmeter aufgezeichnet und anschließend gemittelt.

4.3 Charge-Pumping-Untersuchungen

Dieser Abschnitt beschreibt detailliert die in dieser Arbeit angewendeten Charge-Pumping-Methoden. Zunächst wird die CP-Methode bei konstanter Frequenz vorgestellt und der Einfluss der Pulsform auf die Messergebnisse erläutert. Danach wird die Methode durch Variation der Frequenz erweitert. Hierdurch ist es möglich, eine getrennte Bestimmung der Dichte von Grenzflächenzuständen und der Dichte von grenzflächennahen Oxidstörstellen vorzunehmen.

4.3.1 CP-Messungen bei konstanter Frequenz

Bei einer typischen Charge-Pumping-Messung wird am Gate des MOSFETs eine kontinuierliche Spannungspulsfolge angelegt und der mittlere Substratstrom gemessen. Zur Interpretation des gemessenen CP-Stroms I_{CP} und insbesondere zur absoluten Bestimmung experimenteller

Daten, wie zum Beispiel der Grenzflächenzustandsdichte D_{it} oder der Dichte von grenzflächen-nahen Oxidstörstellen (NIOTs), ist die exakte Pulsform am Gate entscheidend. Die Form des Spannungspulses am Gate ist durch fünf Parameter bestimmt:

- Pulsamplitude V_{Ampl}
- Basisspannung $V_{G,\text{min}}$ ($V_{G,\text{max}} = V_{G,\text{min}} + V_{\text{Ampl}}$)
- Periode T_P beziehungsweise Frequenz $f = 1/T_P$
- Dauer der Anstiegs- und Abfallflanke, t_r beziehungsweise t_f
- Tastverhältnis α_{Tast}

Die einzelnen Größen sind in Abbildung 4.3 veranschaulicht. Für die CP-Messungen in dieser Arbeit wurden die Dauer der Anstiegs- und Abfallflanke gleich eingestellt, $t_{\text{Flanke}} := t_r = t_f$, und das Tastverhältnis betrug stets $\alpha_{\text{Tast}} = 0,5$.

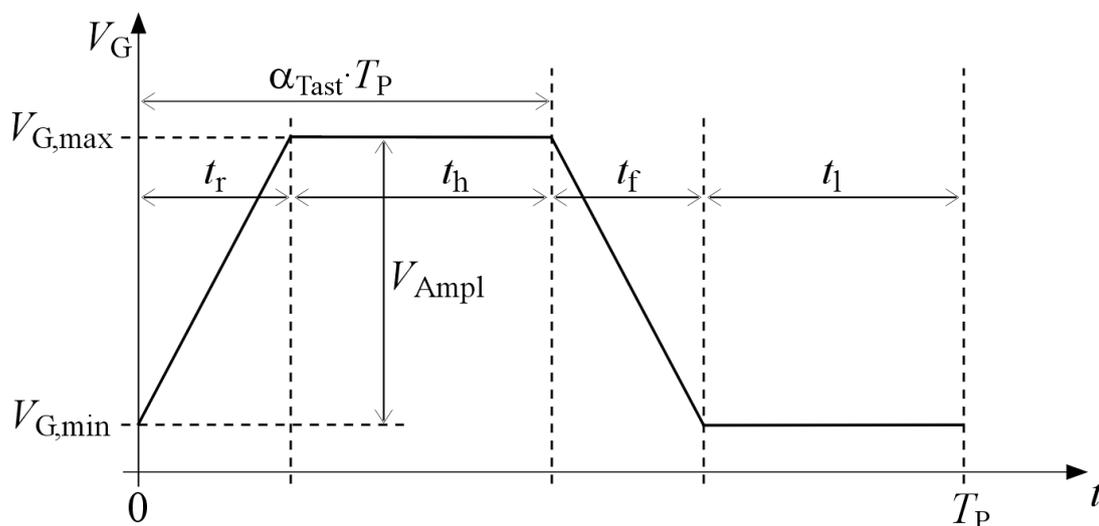


Abbildung 4.3: Pulsparameter beim Charge-Pumping, mit denen die Form des Spannungspulses am Gate definiert ist.

Abbildung 4.4 zeigt jeweils eine CP-Kurve sowohl für den n - als auch den p -MOS-Transistor bei einer Frequenz von $f = 100 \text{ kHz}$ und einer Amplitude von $V_{\text{Ampl}} = 3,0 \text{ V}$. Die Flanken des Gate-Spannungspulses betrugen $t_{\text{Flanke}} = 100 \text{ ns}$. Aufgetragen ist der gemessene CP-Strom I_{CP} in Abhängigkeit der Basisspannung $V_{G,\text{min}}$ im Bereich von $-4,0$ bis $+1,0 \text{ V}$.

Der Kurvenverlauf spiegelt die charakteristische Form einer CP-Kurve mit einem ausgeprägten Maximum wider. Dieses liegt bei dem n -Kanal-MOSFET bei $V_{G,\text{min}} = -2,0 \text{ V}$ und bei dem p -Kanal-MOSFET bei $V_{G,\text{min}} = -0,4 \text{ V}$. Beide MOSFETs zeigen einen nicht verschwindenden CP-Strom an einem bestimmten Ende der dargestellten Kurve. Das ist bei dem n -Kanal-MOSFET der Spannungsbereich $V_{G,\text{min}} < -3,0 \text{ V}$ und bei dem p -Kanal-MOSFET der Bereich

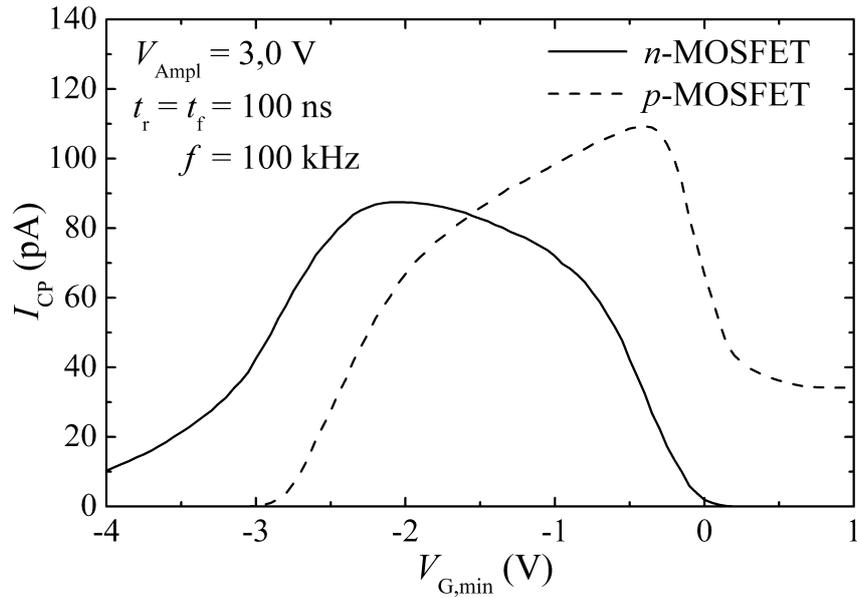


Abbildung 4.4: CP-Kurven sowohl für den n - als auch den p -Kanal-MOSFET bei einer Frequenz von $f = 100$ kHz und einer Amplitude von $V_{\text{Ampl}} = 3,0$ V. Probenparameter: siehe Tabelle 4.1.

$V_{\text{G,min}} > 0,2$ V. Hier befinden sich die Transistoren aufgrund des angelegten Spannungspulses am Gate jeweils im Akkumulationsbereich, in dem alle Grenzflächenzustände mit Löchern (n -Kanal-MOSFET) beziehungsweise Elektronen (p -Kanal-MOSFET) gefüllt sind. Aufgrund des Leckstroms der pn -Übergänge zu den Source- beziehungsweise Drain-Gebieten findet hier eine Rekombination mit deren Majoritätsladungsträgern statt [189].

Eine weitere zusätzliche Komponente zum CP-Strom, die als geometrische Komponente bezeichnet wird [229], kann an den in dieser Arbeit verwendeten Bauelementen ausgeschlossen werden. Diese CP-Komponente tritt auf, wenn das Gate von Inversion nach Akkumulation umgeschaltet wird und nicht alle freien Minoritätsladungsträger in die Source- beziehungsweise Drain-Gebiete abfließen. Diese rekombinieren dann mit Majoritätsladungsträgern aus dem Substrat [162]. Die geometrische Komponente hängt damit von der Gate-Länge und der Umschaltzeit des Spannungspulses am Gate ab. Für die hier verwendeten Geometrien der Gate-Fläche von $W/L \gg 1$, Kanallängen unterhalb von $1 \mu\text{m}$ und den Dauern der Flanken des Gate-Pulses von $t_{\text{Flanke}} = 100$ ns spielt die geometrische Komponente keine Rolle [53, 230].

Die mittlere Grenzflächenzustandsdichte D_{it} wird aus dem Maximum der CP-Kurve bestimmt. Unter diesen Spannungsbedingungen am Gate sind die Eindringtiefen z_e und z_h der Elektronen beziehungsweise Löcher von der Grenzfläche ins Oxid gleich groß, welches detaillierter in Abschnitt 4.3.2 gezeigt wird. Dies führt zu einer maximalen Anzahl von Rekombinationen und somit zu einem Maximum im CP-Strom. Für D_{it} folgt aus Gleichung 3.17, wenn für I_{CP} der maximale CP-Strom eingesetzt wird,

$$D_{\text{it}} = \frac{I_{\text{CP,max}}}{qfA_G\Delta E} \quad , \quad (4.1)$$

wobei f die Frequenz des Gate-Spannungspulses ist und A_G die Gate-Fläche bezeichnet. Der zum CP-Strom beitragende Energiebereich ΔE ist nach Gleichung 3.14 gegeben durch

$$\Delta E = -2 \frac{k_B T}{q} \ln \left(\nu_{th} n_i \sqrt{\sigma_n \sigma_p} \sqrt{t_{em,e} t_{em,h}} \right) \quad (4.2)$$

Dabei bezeichnen ν_{th} die mittlere thermische Geschwindigkeit der Ladungsträger, n_i die intrinsische Ladungsträgerkonzentration im Silizium und σ_n und σ_p die Einfangquerschnitte der Störstellen für Elektronen beziehungsweise Löcher. Die Zeitdauern der nicht stationären Emission von Elektronen beziehungsweise Löchern während des Umschaltvorgangs zwischen Inversion und Akkumulation $t_{em,e}$ und $t_{em,h}$ wurden bereits in Abschnitt 3.2.1 definiert (Gleichungen 3.15 und 3.16).

Da in dieser Arbeit stets Gate-Spannungspulse mit gleicher Anstiegs- und Abfallflanke $t_r = t_f$ angelegt wurden, gilt entsprechend $t_{em} := t_{em,e} = t_{em,h}$. Das geometrische Mittel der Einfangquerschnitte $\sqrt{\sigma_n \sigma_p}$ kann über

$$\sqrt{\sigma_n \sigma_p} = \frac{1}{\nu_{th} n_i} \frac{V_{Ampl}}{|V_{fb} - V_t|} \frac{f_0}{\sqrt{\alpha_{Tast}(1 - \alpha_{Tast})}} \quad (4.3)$$

berechnet werden [162]. Die Frequenz f_0 wird in Anhang A.2 bestimmt. Hierzu ist eine frequenzabhängige CP-Messung mit einem dreieckförmigen Spannungspuls am Gate notwendig. Frequenzabhängige CP-Messungen werden in Abschnitt 4.3.2 näher erläutert. Damit sind alle wichtigen Größen der hier verwendeten Transistoren ermittelt und die mittlere Grenzflächenzustandsdichte kann bestimmt werden.

Die Ergebnisse für den n -MOSFET sowie den p -MOSFET sind in Tabelle 4.2 zusammengefasst. Die ungestressten MOSFETs weisen eine mittlere Grenzflächenzustandsdichte von $6,9 \cdot 10^{10} \text{ cm}^{-2} \text{ eV}^{-1}$ für den n -Kanal-Transistor und $9,6 \cdot 10^{10} \text{ cm}^{-2} \text{ eV}^{-1}$ für den p -Kanal-Transistor auf.

	n -MOSFET	p -MOSFET
$I_{CP,max}$	87,44 pA	109,30 pA
$V_{G,min} _{I_{CP,max}}$	-2,0 V	-0,4 V
A_G	$1,2 \cdot 10^{-7} \text{ cm}^2$	$1,2 \cdot 10^{-7} \text{ cm}^2$
f_0	14,96 Hz	54,28 Hz
$\sqrt{\sigma_n \sigma_p}$	$5,3 \cdot 10^{-16} \text{ cm}^2$	$2,3 \cdot 10^{-15} \text{ cm}^2$
t_{em}	58,3 ns	50,0 ns
ΔE	0,66 eV	0,59 eV
D_{it}	$6,9 \cdot 10^{10} \text{ cm}^{-2} \text{ eV}^{-1}$	$9,6 \cdot 10^{10} \text{ cm}^{-2} \text{ eV}^{-1}$

Tabelle 4.2: Probenparameter aus CP-Messungen für die Bestimmung der mittleren Grenzflächenzustandsdichte

Die beiden geometrischen Mittel der Einfangquerschnitte für den n - und den p -Kanal-MOSFET unterschieden sich etwa um einen Faktor 4. In dem hier beschriebenen einfachen Modell werden die Einfangquerschnitte als unabhängig von der Energie angenommen [231, 232]. Dies ist im Allgemeinen nicht der Fall und könnte die Ursache für die unterschiedlichen Werte der experimentell ermittelten Einfangquerschnitte sein.

4.3.2 Frequenzabhängige CP-Messungen

Bisher wurden CP-Charakterisierungsmessungen bei einer konstanten Frequenz von 100 kHz gezeigt. Mit dieser Methode kann die mittlere Dichte von Grenzflächenzuständen, also schnell kommunizierenden Störstellen, bestimmt werden. Darüber hinaus ist man aber auch an der Dichte von langsamen Störstellen im Oxid (NIOTs) interessiert, die sich in einem etwas größeren Abstand (typischerweise ~ 1 nm) von der $Si-SiO_2$ -Grenzfläche befinden. Für die getrennte Bestimmung von Grenzflächenzustandsdichte und der Dichte der grenzflächennahen Oxidstörstellen reicht die Messungen des CP-Stroms bei einer einzelnen Frequenz nicht aus. Im Folgenden wird die Methode der frequenzabhängigen CP-Messung beschrieben, die bereits 1974 von M. Declercq und P. Jespers vorgestellt [233] und von Y. Maneglia und D. Bauza zur Extraktion von Störstellenprofilen weiterentwickelt wurde [234, 235].

Bei dieser Methode wird der maximale CP-Strom bei verschiedenen Frequenzen gemessen. Die Flanken des Spannungspulses am Gate sowie die Amplitude werden dabei konstant gehalten. In dieser Arbeit wurde für die Flanke $t_{\text{Flanke}} = 100$ ns und für die Amplitude $V_{\text{Ampl}} = 3,0$ V gewählt. Diese Wahl der Pulsform führt mit fallender Frequenz f zu einer steigenden Dauer der Bereiche mit konstanter Gate-Spannung ($V_G = V_{G,\text{min}}$ beziehungsweise $V_G = V_{G,\text{max}}$) während eines Pulses. Auf diese Weise steigt auch die Verweildauer des Bauelements in Inversion beziehungsweise Akkumulation und damit die Wahrscheinlichkeit, dass Störstellen in einem immer tieferen Abstand von der Grenzfläche besetzt werden. Aus der Shockley-Read-Hall Statistik lässt sich eine Füllfunktion $\Delta F(z)$ für die Besetzungswahrscheinlichkeit ableiten. Sie gibt an, mit welcher Wahrscheinlichkeit eine Störstelle im Abstand z_T von der Grenzfläche während der Dauer eines CP-Pulses zunächst ein Elektron und anschließend ein Loch einfängt. Sie ist näherungsweise gegeben durch

$$\Delta F(z) \simeq \frac{[1 - e^{-c_n(z)/2f}] \cdot [1 - e^{-c_p(z)/2f}]}{1 - e^{-(c_n(z)+c_p(z))/2f}} =: \frac{F_n(z) \cdot F_p(z)}{1 - e^{-(c_n(z)+c_p(z))/2f}}, \quad (4.4)$$

wobei c_n und c_p die Einfangraten der Störstellen für Elektronen beziehungsweise Löcher bezeichnen. Die Funktionen $F_n(z)$ und $F_p(z)$ sind über

$$F_n(z) := 1 - e^{-c_n(z)/2f} \quad (4.5)$$

$$F_p(z) := 1 - e^{-c_p(z)/2f} \quad (4.6)$$

definiert. Hierbei wurde angenommen, dass die Emission von Ladungsträgern aus Störstellen in die Energiebänder vernachlässigbar ist. Dies ist bei den in dieser Arbeit gewählten kurzen

Übergangszeiten von Akkumulation nach Inversion ($t_{\text{Flanke}} = 100 \text{ ns}$) und den hohen Ladungsträgerdichten an der Grenzfläche während der beiden Pulsphasen mit konstanter Gate-Spannung gewährleistet. Die Einfangraten sind gegeben durch

$$c_n(z) = n_s \sigma_n(z) \nu_{\text{th}} \quad (4.7)$$

$$c_p(z) = p_s \sigma_p(z) \nu_{\text{th}} \quad (4.8)$$

Dabei bezeichnen n_s und p_s die Ladungsdichten von Elektronen beziehungsweise Löchern an der Grenzfläche während der Pulsphasen mit konstanter Gate-Spannung. Die Einfangquerschnitte $\sigma_n(z)$ und $\sigma_p(z)$ für Elektronen beziehungsweise Löcher können über das Modell von F. P. Heiman und G. Warfield [236] bestimmt werden. Sie sind gegeben durch

$$\sigma_n(z) = \sigma_n e^{-z/\lambda_e} \quad (4.9)$$

$$\sigma_p(z) = \sigma_p e^{-z/\lambda_h} \quad (4.10)$$

wobei σ_n und σ_p die Einfangquerschnitte an der Grenzfläche bezeichnen. Die Abschwächungskoeffizienten λ_e und λ_h sind im Wesentlichen durch die Energiebarriere für Elektronen beziehungsweise Löcher bestimmt und beschreiben die Abnahme der Amplituden der Wellenfunktionen innerhalb des Oxids. Trotz der unterschiedlichen Energiebarrieren für Elektronen und Löcher wurden die Abschwächungskoeffizienten in dieser Arbeit näherungsweise als gleich angenommen ($\lambda_e = \lambda_h = 0,7 \cdot 10^{-8} \text{ cm}$ [234]).

Abbildung 4.5 zeigt als durchgezogene Linie die Füllfunktion $\Delta F(z)$, sowie die Verläufe der in den Gleichungen 4.5 und 4.6 definierten Funktionen $F_n(z)$ beziehungsweise $F_p(z)$ [235]. Die Einfangrate c_p für die Löcher ist in diesem Beispiel größer gewählt, als die der Elektronen ($c_p > c_n$). Dadurch ergeben sich für Elektronen und Löcher unterschiedliche Eindringtiefen z_e beziehungsweise z_h . Wie sich aus Gleichung 4.4 und Abbildung 4.5 ablesen lässt, wird die Füllfunktion $\Delta F(z)$ im Wesentlichen durch die kleinere Einfangrate (hier c_n) bestimmt und kann näherungsweise durch die Funktion $F_n(z) = 1 - e^{-c_n(z)/2f}$ beschrieben werden.

Die Eindringtiefen z_e und z_h sind definiert durch die Tiefe \bar{z} , an der $\Delta F(\bar{z}) = 0,5$ gilt:

$$z_e = \lambda_e \ln \left[\frac{c_n(0)}{2f \cdot \ln(2)} \right] \quad (4.11)$$

$$z_h = \lambda_h \ln \left[\frac{c_p(0)}{2f \cdot \ln(2)} \right] \quad (4.12)$$

Im Allgemeinen ist also der für CP relevante Oxidbereich durch das Minimum

$$z_m = \min(z_e, z_h) \quad (4.13)$$

der Eindringtiefen bestimmt. In diesem Bereich beträgt die Wahrscheinlichkeit, dass eine Störstelle während der entsprechenden Phasen des CP-Pulses sowohl ein Elektronen als auch ein Loch einfängt, Eins. Durch Rekombination der beiden Ladungsträger trägt diese Störstelle dann

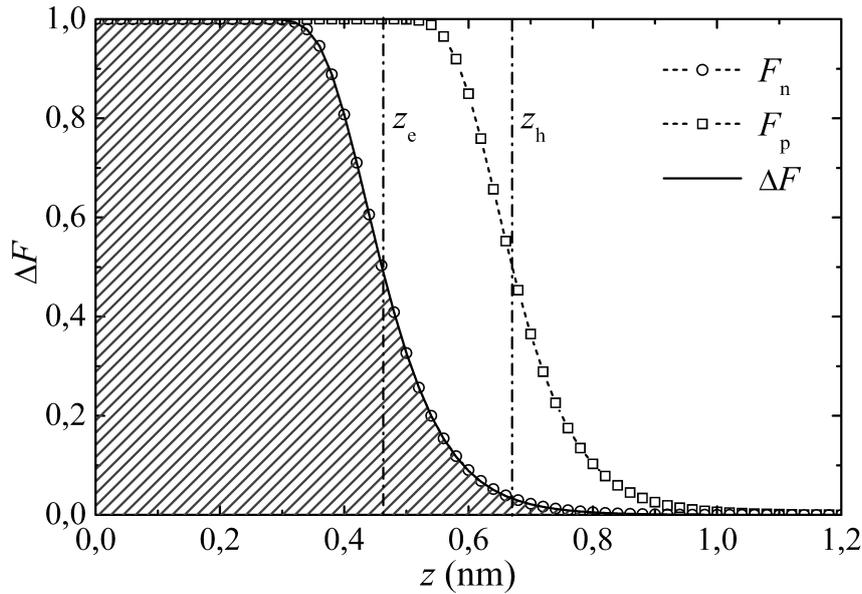


Abbildung 4.5: Beispiel einer Füllfunktion $\Delta F(z)$ (durchgezogene Linie) für den Fall $c_n < c_p$. Die offenen Symbole geben die Besetzungswahrscheinlichkeiten für die Elektronen beziehungsweise Löcher an. Der zum CP beitragende Oxidbereich (schraffierte Fläche) wird durch den Ladungsträgertyp mit der geringeren Eindringtiefe (hier $z_e < z_h$) bestimmt [235].

zum CP-Strom bei. Die größte Eindringtiefe wird demnach dann erzielt, wenn $z_e = z_h$ gilt, welches gleichbedeutend mit $c_n = c_p$ ist. Diese Situation ist am Maximum der CP-Kurve gegeben [237], hier gilt die Beziehung

$$\frac{\sigma_p}{\sigma_n} = \frac{n_s}{p_s} \quad (4.14)$$

In dieser Arbeit wurde für die frequenzabhängigen CP-Messungen ein Frequenzbereich von $f_{\min} = 1$ kHz bis $f_{\max} = 2$ MHz gewählt. Hieraus ergeben sich für die vorliegenden MOS-Transistoren Eindringtiefen von $z_m(f_{\max}) \simeq 0,8$ nm am oberen Ende des Frequenzbereiches und $z_m(f_{\min}) \simeq 1,3$ nm am unteren. Für die Berechnung wurden die Parameter aus Tabelle 4.3 herangezogen. Die Ladungsträgerkonzentrationen n_s und p_s an der Grenzfläche in Akkumulation beziehungsweise Inversion sind über den Zusammenhang $\psi_s(V_G)$ zwischen Oberflächenpotential und Gate-Spannung und mithilfe der bekannten Gleichungen für MOS-Transistoren errechnet worden [2]. Die Abhängigkeit des Oberflächenpotentials von der Gate-Spannung wurde über das Berglund-Integral (Gleichung A.3) [238] aus Messungen der quasistatischen $C(V)$ -Kurven ermittelt (siehe Anhang A.1). Die Werte für σ_n und σ_p ergeben sich aus Gleichung 4.14 und dem Wert für $\sqrt{\sigma_n \sigma_p}$ aus Tabelle 4.2.

Durch Variation der Messfrequenz lässt sich also mit dieser CP-Methode der Oxidbereich einstellen, aus dem Störstellen zum CP-Strom beitragen, einstellen. Unter der Annahme, dass die Störstellen im Oxid einer gewissen Verteilung $D_{ot}(z) > 0$ gehorchen, ist ein Anstieg des CP-Strom zu kleinen Frequenzen zu erwarten. In Abbildung 4.6 ist die maximale rekombinierte

	<i>n</i> -MOSFET	<i>p</i> -MOSFET
$V_{G,\min}$	-2,0 V	-0,4 V
$V_{G,\max}$	1,0 V	2,6 V
n_s	$2,24 \cdot 10^{19} \text{ cm}^{-3}$	$6,60 \cdot 10^{19} \text{ cm}^{-3}$
p_s	$2,52 \cdot 10^{19} \text{ cm}^{-3}$	$5,36 \cdot 10^{18} \text{ cm}^{-3}$
σ_n	$5,6 \cdot 10^{-16} \text{ cm}^2$	$6,5 \cdot 10^{-16} \text{ cm}^2$
σ_p	$5,0 \cdot 10^{-16} \text{ cm}^2$	$8,1 \cdot 10^{-15} \text{ cm}^2$
$z_m(f_{\min})$	1,28 nm	1,37 nm
$z_m(f_{\max})$	0,75 nm	0,84 nm

Tabelle 4.3: Auflistung der Probenparameter, die zur Bestimmung der minimalen Eindringtiefe z_m nach Gleichung 4.13 verwendet wurden.

Ladung $Q_{CP,\max}$, die über

$$Q_{CP,\max} = \frac{I_{CP,\max}}{f} \quad (4.15)$$

definiert ist, je einer Messreihe am *n*-MOSFET und am *p*-MOSFET über der Frequenz aufgetragen.

Erwartungsgemäß lässt sich bei beiden Kurven ein Anstieg zu kleinen Frequenzen beobachten. Die Volumendichte der Störstellen im Oxid kann über

$$D_{ot}(z_m) = -\frac{\log e}{q\lambda_e A_G \Delta E} \frac{dQ_{CP,\max}}{d \log f} \Big|_{f(z_m)} \quad (4.16)$$

bestimmt werden [234].

Dieses Verfahren wird in der Literatur benutzt, um Störstellenprofile im Gate-Oxid zu bestimmen [234]. Bei den in dieser Arbeit untersuchten MOSFETs ergab sich stets ein linearer Verlauf der maximalen rekombinierten CP-Ladung über der logarithmierten Frequenz. Dies galt insbesondere auch bei höheren Stresszuständen (siehe Abschnitt 4.3.3). Aufgrund dieser Tatsache kann die Ableitung in Gleichung 4.16 als konstant betrachtet werden,

$$\frac{dQ_{CP,\max}}{d \log f} \simeq \text{const} \quad (4.17)$$

Die Störstellendichte D_{ot} im untersuchten Tiefenbereich des Oxids (0,8 bis 1,3 nm) ist damit ebenfalls näherungsweise konstant.

Die mittlere Dichte D_{it} der schnellen Störstellen nahe der Grenzfläche wird nach Gleichung 4.1 aus dem maximalen CP-Strom bei der größten Messfrequenz von $f = 2$ MHz bestimmt. Sie erfasst damit alle schnellen Störstellen mit einer räumlichen Lage zwischen der *Si-SiO₂*-Grenzfläche ($z = 0$) und $z = z_m = 0,8$ nm. Die beiden Messreihen aus Abbildung 4.6 ergeben die Störstellendichten $D_{it,0}$ und $D_{ot,0}$ für die ungestressten MOSFETs, welche in Tabelle 4.4 aufgeführt sind.

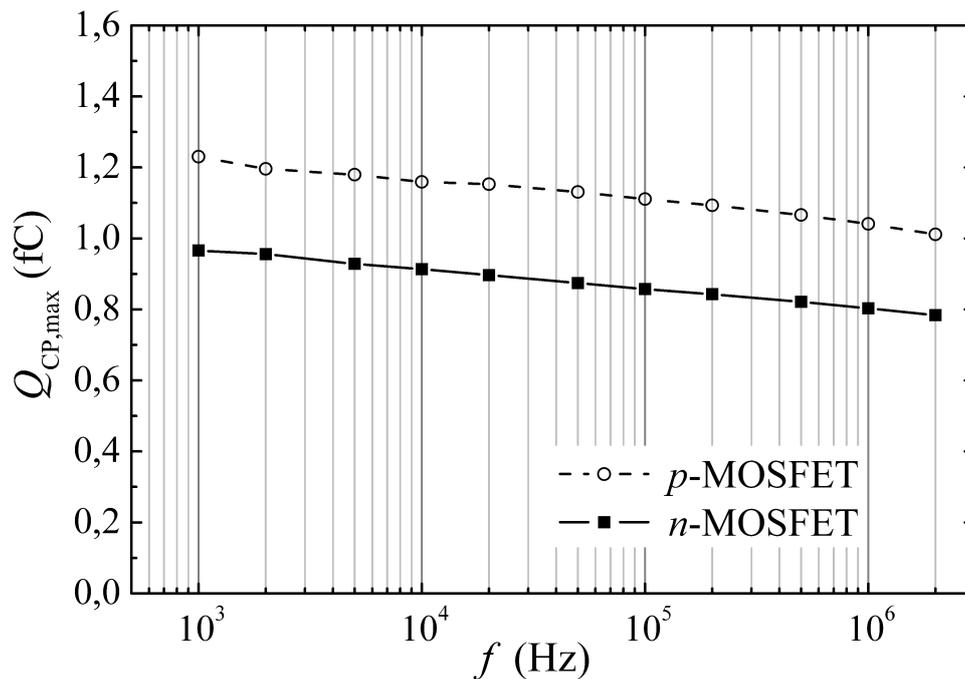


Abbildung 4.6: Maximale rekombinierte Ladung $Q_{CP,max}$ über der Frequenz für den n -MOSFET und den p -MOSFET. Aufgrund der mit fallender Frequenz größer werdenden Eindringtiefe der Ladungsträger ins Oxid tragen mehr Störstellen zum CP-Strom bei und $Q_{CP,max}$ steigt zu kleinen Frequenzen an. Aus der Steigung der linearen Anpassung der Kurven kann die Dichte der Oxidstörstellen berechnet werden. Probenparameter: siehe Tabelle 4.1.

Die bestimmten Grenzflächenzustandsdichten $D_{it,0}$ für n - und p -MOSFET liegen hier etwas niedriger als in Tabelle 4.2. Dies ist darauf zurückzuführen, dass im vorangegangenen Kapitel mit einer Frequenz von 100 kHz gemessen wurde und hier die Bestimmung bei der maximalen Frequenz von 2 MHz stattgefunden hat. Bei der höheren Messfrequenz ist die Eindringtiefe und somit die Zahl der erfassten Störstellen geringer.

Im folgenden Abschnitt wird das Fowler-Nordheim-Stressexperiment beschrieben und anschließend werden in Abschnitt 4.3.4 die Ergebnisse der frequenzabhängigen CP-Untersuchungen in Abhängigkeit des eingepprägten Stresses dargestellt.

	n -MOSFET	p -MOSFET
$D_{it,0}$	$6,2 \cdot 10^{10} \text{ cm}^{-2} \text{ eV}^{-1}$	$9,0 \cdot 10^{10} \text{ cm}^{-2} \text{ eV}^{-1}$
$D_{ot,0}$	$2,7 \cdot 10^{17} \text{ cm}^{-3} \text{ eV}^{-1}$	$3,3 \cdot 10^{17} \text{ cm}^{-3} \text{ eV}^{-1}$

Tabelle 4.4: Grenzflächenzustandsdichte $D_{it,0}$ und Volumendichte der Oxidstörstellen $D_{ot,0}$ für die ungestressten MOSFETs, bestimmt aus frequenzabhängigen CP-Messungen.

4.3.3 Fowler-Nordheim-Stressexperiment

Im Vordergrund dieser Arbeit steht die Untersuchung der Degradation von Bauelementen unter elektrischer Stresseinwirkung. An den vorliegenden n - und p -MOSFETs soll speziell die Generation von Störstellen unter Fowler-Nordheim-Stresseinwirkung betrachtet werden. Dazu wurde eine hohe elektrische Stromdichte von $j_{inj} = 0,1 \text{ mA cm}^{-2}$ am Gate-Kontakt des Bauelements eingepreßt, die zu einem Fowler-Nordheim-Tunneln von Ladungsträgern durch das Gate-Oxid führt. Die Gate-Stromdichte wurde während der Stressphase konstant gehalten. Dieses Verfahren wird auch als Konstantstromstress bezeichnet (CCS, engl.: Constant Current Stress).

Der Messaufbau aus Abbildung 4.2 wurde für die Stressexperimente erweitert. Für das Einprägen des elektrischen Stresses wurde ein Parameter-Analyzer eingefügt. Um ein Umschalten zwischen Stressexperiment und Charge-Pumping-Messungen zu ermöglichen, wurde zusätzlich eine Schaltmatrix in den Messaufbau integriert. Der erweiterte Messaufbau ist in Abbildung 4.7 skizziert.

Der Ablauf eines Stressexperiments war zyklisch: Zunächst wurde das ungestresste Bauelement mit den im vorigen Kapitel beschriebenen Charge-Pumping-Methoden untersucht. Dabei wurden die Dichte der schnellen Grenzflächenzustände und die Dichte der langsameren Oxid-

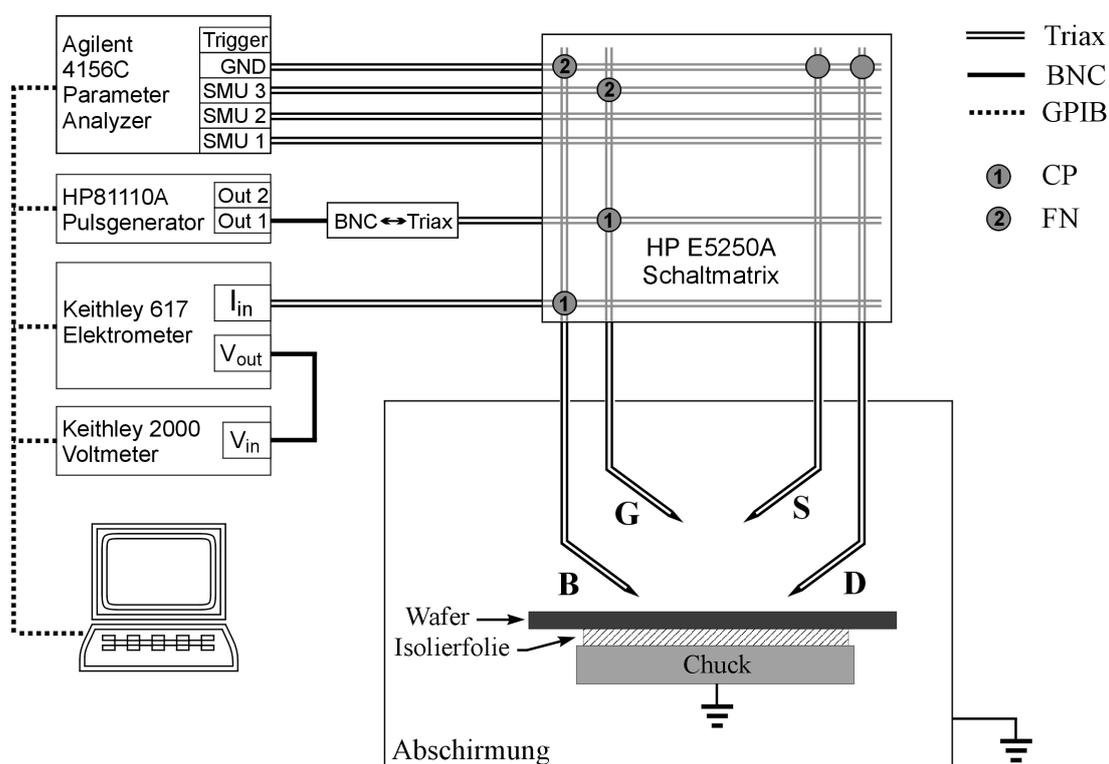


Abbildung 4.7: Skizze des erweiterten Messaufbaus für die Kombination von Charge-Pumping-Messungen mit dem Fowler-Nordheim-Stressexperiment. In der Schaltmatrix sind die beiden Konfiguration für CP (1) und FN-Stress (2) durch Punkte gekennzeichnet. Source und Drain des Bauelements sind stets mit der Ground-Unit (GND) des Parameter-Analyzers verbunden.

störstellen aus den experimentellen Daten bestimmt. Danach wurde das Bauelement für eine bestimmte Zeit t_{inj} elektrisch gestresst. Aus der Stresszeit und der konstanten Stresstromdichte j_{inj} kann über

$$Q_{inj} = j_{inj} t_{inj} \quad (4.18)$$

die Ladungsmenge pro Flächeneinheit bestimmt werden, die durch das Gate-Oxid geflossen ist. Sie ist ein Maß für den eingepprägten Stress und wird kumulativ angegeben. Die Stresseinwirkung wurde in regelmäßigen Abständen unterbrochen, wobei die Zeitabstände auf der logarithmischen Skala in etwa gleich gewählt wurden. Während jeder Stressunterbrechung wurde das Bauelement mit den zuvor beschriebenen Charge-Pumping-Methoden untersucht.

Beim Fowler-Nordheim-Stressexperiment sind grundsätzlich zwei Modi zu unterscheiden. An den Gate-Kontakten kann zum Einprägten des Stresstroms sowohl positive als auch negative Spannung angelegt werden. Bei positiver Spannung spricht man von PBS (*engl.*: Positive Bias Stress) und bei negativer von NBS (*engl.*: Negative Bias Stress). Die Auswirkungen der beiden Stresspolaritäten auf die Bandstruktur sind in Abbildung 4.8 dargestellt. Bei der PBS-Bedingung tunneln die Elektronen vom *Si*-Substrat durch eine dreieckförmige Potenzialbarriere zum Gate. Man spricht in diesem Fall auch von Substratmission. Anders verhält es sich unter der NBS-Bedingung. Hierbei tunneln die Elektronen vom Gate durch die Potenzialbarriere im Gate-Oxid zum Substrat. Dieser Fall wird als Gate-Emission bezeichnet.

Wie bereits erwähnt wurde für alle in dieser Arbeit durchgeführten Stressexperimente je nach Polarität der Gate-Spannung eine konstante Stresstromdichte von $j_{inj} = \pm 0,1 \text{ mA cm}^{-2}$ ver-

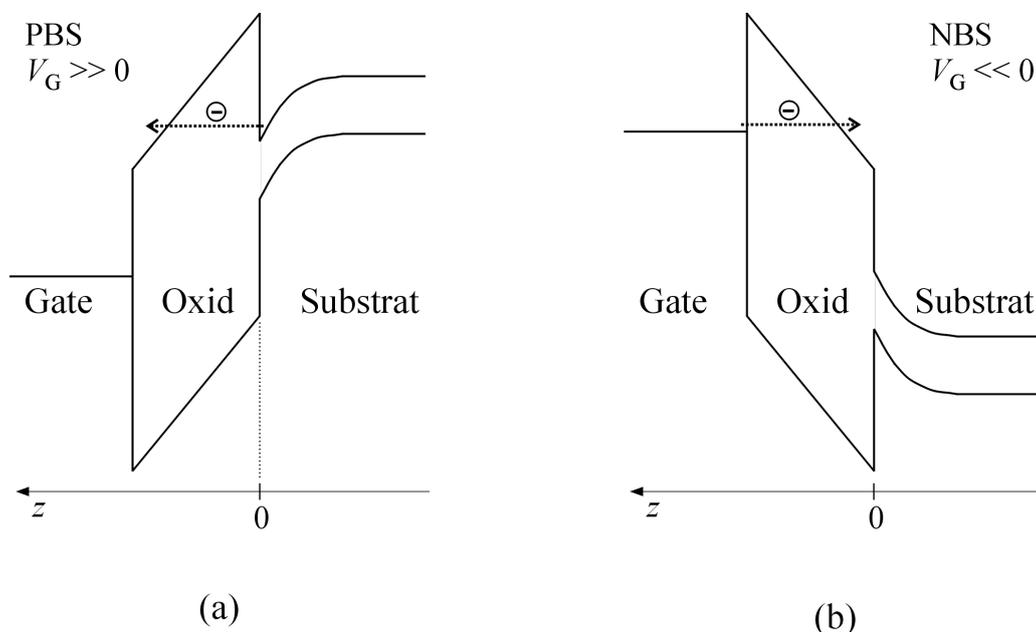


Abbildung 4.8: Darstellung der Bandstruktur für die beiden möglichen Stresspolaritäten beim Fowler-Nordheim-Stressexperiment. Die Elektronen tunneln unter PBS-Bedingung (a) vom Substrat zum Gate und unter NBS-Bedingung (b) vom Gate zum Substrat.

wendet. Die während der Stressphasen anliegenden Gate-Spannungen und korrespondierenden elektrischen Feldstärken im Oxid sind für die jeweiligen MOSFETs in Tabelle 4.5 angegeben. Die Vorzeichen der Felder ergeben sich aus der Definition der z -Achse in Abbildung 4.8. Für die Berechnung der Feldstärken wurde die Bandverbiegung an der $Si-SiO_2$ -Grenzfläche nach Abbildung A.3 aus Anhang A.1 berücksichtigt. Die Bandverbiegung im Poly- Si -Gate wurde wegen der hohen Dotierung von $\sim 1 \cdot 10^{20} \text{ cm}^{-3}$ vernachlässigt.

	n -MOSFET	p -MOSFET
$V_{G,PBS}$	6,1 V	6,9 V
$V_{G,NBS}$	-6,8 V	-7,45 \rightarrow -7,9 V
$\mathcal{E}_{ox,PBS}$	$-8,9 \cdot 10^6 \text{ V cm}^{-1}$	$-9,2 \cdot 10^6 \text{ V cm}^{-1}$
$\mathcal{E}_{ox,NBS}$	$9,1 \cdot 10^6 \text{ V cm}^{-1}$	11,3 \rightarrow $12,0 \cdot 10^6 \text{ V cm}^{-1}$

Tabelle 4.5: Gate-Spannungen und elektrische Feldstärken im Oxid während der FN-Stressphasen. Beim p -MOSFET driftete die Spannung während des Experiments von -7,45 V nach -7,9 V (siehe Anhang A.3).

Bemerkenswert bei den Werten für den p -Kanal-MOSFET unter NBS-Bedingung in Tabelle 4.5 ist die starke Drift der Gate-Spannung während der Stressphase (siehe Abbildung A.5 in Anhang A.3) und die damit verbundene hohe elektrische Feldstärke von mehr als $11 \cdot 10^6 \text{ V cm}^{-1}$. Dieses Verhalten wird zusammen mit den Ergebnissen der Stressexperimente im nächsten Abschnitt eingehender diskutiert.

4.3.4 Ergebnisse des Stressexperiments

In Abbildung 4.9 sind zwei CP-Messreihen dargestellt, die an einem n -Kanal-MOSFET durchgeführt wurden. Die untere Kurve zeigt die maximale rekombinierte Ladung $Q_{CP,max}$ eines ungestressten Transistors in Abhängigkeit der Frequenz, die obere dagegen eine Messung am selben Transistor nach dem Einprägen von $Q_{inj} = -2,0 \text{ C cm}^{-2}$.

Die Kurve nach dem Stressen des Transistors zeigt eine deutlich höhere rekombinierte Ladung im Vergleich zum ungestressten Transistor. Bei allen gemessenen Frequenzen wurde eine höhere Ladung $Q_{CP,max}$ beobachtet. Dies zeigt, dass aufgrund der Stresseinwirkung zusätzliche Störstellen entstanden sind. Gleichzeitig hat die Steigung der Kurve zugenommen. Daher ist offensichtlich auch die Zahl der langsamen Oxidstörstellen aufgrund des Stresses größer geworden.

Im Folgenden werden die Beiträge der schnellen Grenzflächenzustände und der langsameren Oxidstörstellen getrennt voneinander aus den CP-Messungen extrahiert und in Abhängigkeit der injizierten Stressladungsdichte dargestellt.

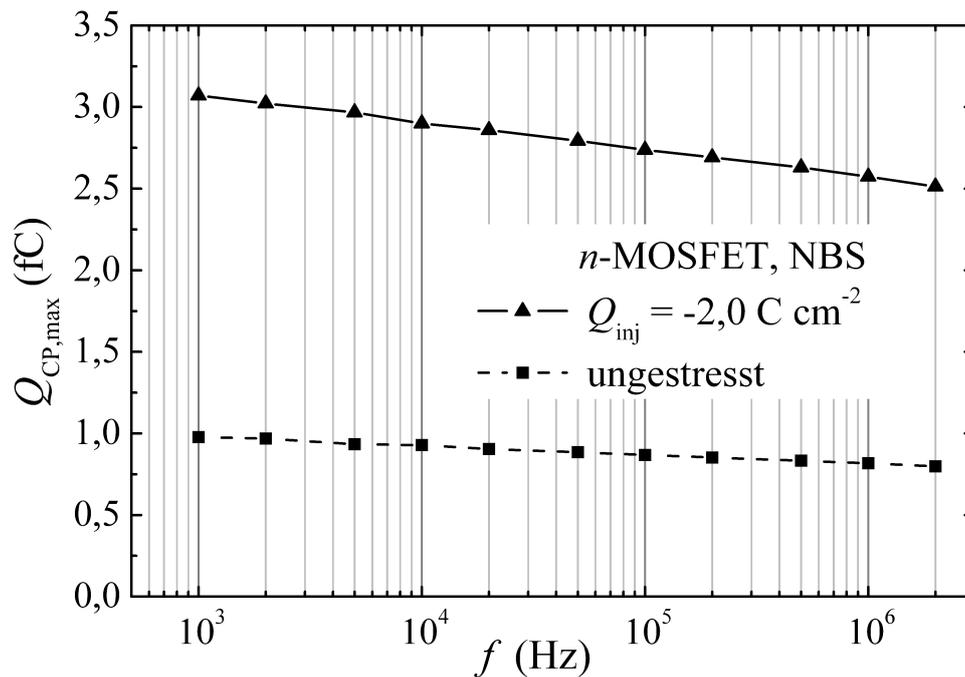


Abbildung 4.9: Vergleich zweier CP-Messreihen. Die untere Kurve zeigt die rekombinierte Ladung in Abhängigkeit der Frequenz eines ungestressten n -Kanal-Transistors. Die obere Kurve zeigt die rekombinierte Ladung an demselben Transistor nach dem Einprägen einer Stressladungsdichte von $-2,0 \text{ C cm}^{-2}$ unter NBS. Probenparameter: siehe Tabelle 4.1.

Generation von Grenzflächenzuständen und Oxidstörstellen

Abbildung 4.10 zeigt die aus CP-Messungen bestimmte mittlere Dichte der schnellen Grenzflächenzustände D_{it} für den n -Kanal-MOSFET (volle Symbole) und den p -Kanal-MOSFET (offene Symbole) über dem Betrag der injizierten Stressladungsdichte Q_{inj} . Für beide Transistortypen wurde das Stressexperiment mit beiden Stresspolaritäten (NBS und PBS) bis zu einer maximalen absoluten Stressladungsdichte von 15 C cm^{-2} , also 150.000 s ($\simeq 41,7 \text{ h}$) bei $0,1 \text{ mA cm}^{-2}$ Stressstromdichte, durchgeführt.

Die Zustandsdichte des p -MOSFETs zeigt für alle Stresszustände keinen signifikanten Unterschied zwischen den beiden Stresspolarisationen NBS und PBS. Eine schwache Abhängigkeit von der Stresspolarisation ist hier nur für den n -MOSFET zu beobachten. In der Literatur dagegen wurde eine deutliche Polarisationsabhängigkeit mithilfe von $C(V)$ -Untersuchungen vielfach beobachtet [239–241]. Diese Asymmetrie in der Degradation kann über das Modell der Wasserstofffreisetzung (siehe Abschnitt 2.3.1) erklärt werden. Im Falle negativer Gate-Spannung (NBS) während des Stresses erreichen die Elektronen auf der Substratseite (Anode) genügend Energie, um durch Relaxation heiße Löcher zu erzeugen. Diese können an der Grenzfläche durch Aufbrechen von Bindungen Wasserstoff freisetzen. Auf diese Weise werden Grenzflächenzustände generiert, die unmittelbar in der CP-Messung beobachtet werden. Im Falle der entgegengesetzten Stresspolarität (PBS) befindet sich die Anode auf der Gate-Seite des Isola-

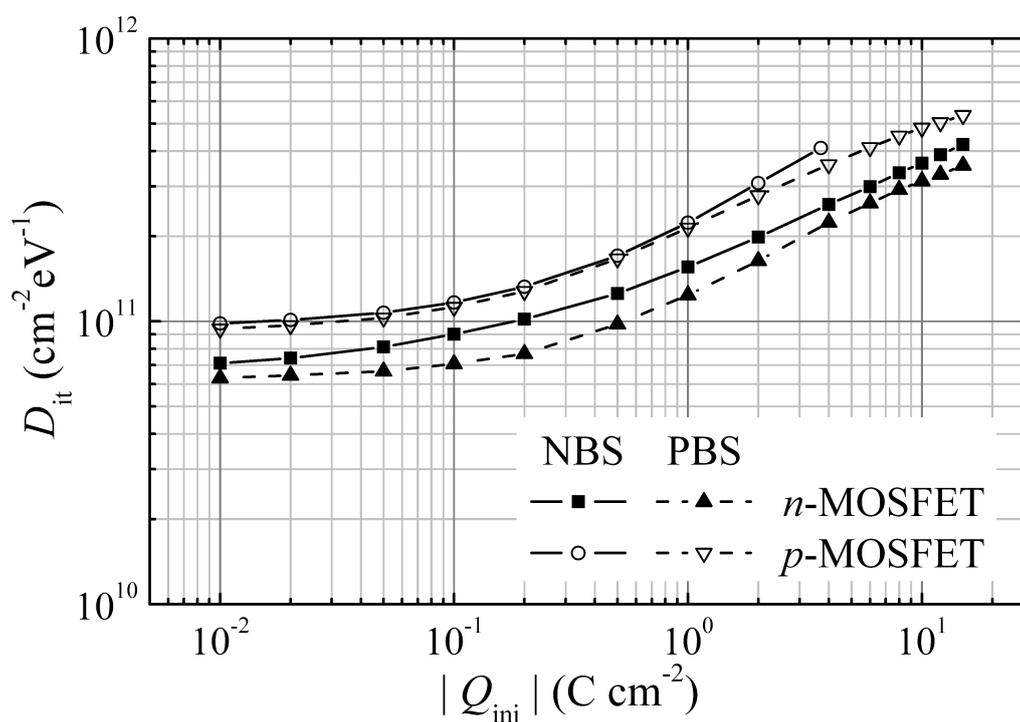


Abbildung 4.10: Dichte der schnellen Grenzflächenzustände D_{it} unter FN-Stress mit $|j_{inj}| = 0,1 \text{ mA cm}^{-2}$ in Abhängigkeit der injizierten Stressladungsdichte. Für den n -MOSFET zeigt sich ein deutlicher Unterschied im Verlauf der Kurven zwischen NBS und PBS, die Kurven des p -MOSFETs unterscheiden sich dagegen kaum. Probenparameter: siehe Tabelle 4.1.

tors. Hier erzeugte Grenzflächenzustände sind mittels CP nicht messbar. Jedoch wird auch hier Wasserstoff an der Grenzfläche freigesetzt, der wiederum zur Substratgrenzfläche diffundieren kann. Dort entstehen Grenzflächenzustände durch Depassivierung von wasserstoffgesättigten Bindungen. Dieser Vorgang wurde experimentell durch eine Erhöhung der Stickstoffkonzentration nahe der Si - SiO_2 -Grenzfläche verifiziert [240]. Stickstoff verlangsamt die Diffusion von Wasserstoff und unterdrückt somit die Generation von Grenzflächenzuständen durch gateseitig freigesetzten Wasserstoff [242]. Für dieses Modell spricht ebenso die Tatsache, dass die Kurve des n -MOSFETs unter PBS (gefüllte Dreiecke) bei den hier gezeigten Ergebnissen im Gegensatz zu den Werten für NBS einen leicht verzögerten Anstieg der Grenzflächenzustandsdichte zeigt. Bis zu einer Stressladungsdichte von $|Q_{inj}| \simeq 0,1 \text{ C cm}^{-2}$ verläuft die Kurve sehr flach, erst danach ist ein Anstieg zu beobachten, der sich für höhere Stressladungsdichten ($|Q_{inj}| > 2,0 \text{ C cm}^{-2}$) wieder der Kurve für NBS annähert.

Anhand dieser Ergebnisse liegt der Schluss nahe, dass die Polarisation der Gate-Spannung beim FN-Stress nur einen geringen Einfluss auf das Generationsverhalten von Grenzflächenzuständen hat.

Darüber hinaus konnte das Experiment am p -MOSFET unter NBS hier nur bis zu einer Stressladungsdichte von $|Q_{inj}| = 3,7 \text{ C cm}^{-2}$ durchgeführt werden, da an dieser Stelle ein elektrischer

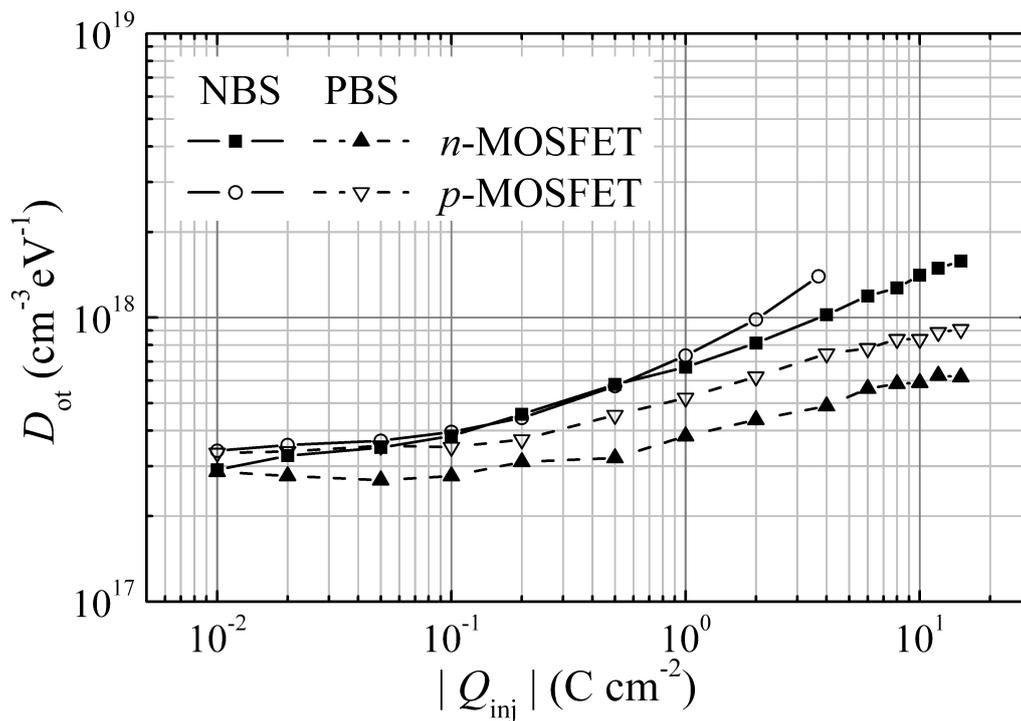


Abbildung 4.11: Dichte der langsamen Oxidstörstellen D_{ot} unter FN-Stress mit $|j_{inj}| = 0,1 \text{ mA cm}^{-2}$ in Abhängigkeit der injizierten Stressladungsdichte. Für beide Transistortypen zeigt sich eine deutliche Polaritätsabhängigkeit im Verlauf der Kurven. Probenparameter: siehe Tabelle 4.1.

Durchbruch des Isolators auftrat. Kurz vor dem Durchbruch zeigt die Grenzflächenzustandsdichte zudem einen größeren Anstieg. Offensichtlich tritt hier ein weiterer Degradationsmechanismus zu Tage. In vorangegangenen Untersuchungen an den gleichen Proben zeigte sich ein starker NBTI-Effekt der p -Kanal-Transistoren [243]. Der NBTI-Effekt bewirkt eine beschleunigte Generation von Grenzflächenzuständen, die in nitrierten Gate-Oxiden zusätzlich verstärkt auftritt (vergleiche Abschnitt 2.3.4). Dieser Mechanismus könnte hier als Ursache für den starken Anstieg von D_{it} vorliegen.

In Abbildung 4.11 ist analog die Volumendichte der Oxidstörstellen D_{ot} aus denselben Stressexperimenten über dem Betrag der Stressladungsdichte dargestellt. Hier zeigt sich in der Generation von Störstellen ein etwas anderes Bild. Den stärksten Anstieg weisen jeweils die unter NBS gestressten MOSFETs (n - und p -Kanal) auf. Beide Kurven zeigen dabei in etwa den gleichen Verlauf, jedoch steigt die Kurve des p -Kanal-MOSFETs kurz vor dem elektrischen Durchbruch ab $|Q_{inj}| \simeq 1 \text{ C cm}^{-2}$ wiederum deutlich stärker an. Unter PBS fällt die Generation von Störstellen deutlich geringer aus und zeigt für beide Transistortypen eine einsetzende Sättigung bei hohen Stressladungen.

Im Gegensatz zum Verhalten von D_{it} zeigt sich die erwartete Asymmetrie bezüglich der Stresspolarität hier für beide Transistortypen. Darüber hinaus ist sie viel stärker ausgeprägt und die ermittelten Dichten D_{ot} nähern sich für hohe injizierte Stressladungsdichten nicht einander an.

Insbesondere nimmt der Anstieg der Kurve für den p -Kanal-MOSFET unter NBS mit steigender Stressladung deutlich zu.

Eine Erklärung liefert ebenfalls das HR-Modell im Zusammenhang mit dem AHI-Modell. Hierbei spielen sich die Degradationsprozesse allerdings nicht direkt an der Grenzfläche, sondern im grenzflächennahen Bereich ab. Unter NBS werden auf der Substratseite (Anode) wiederum heiße Löcher erzeugt und ins Oxid injiziert. Diese generieren Störstellen im Oxidvolumen, vorwiegend über das Aufbrechen von $Si-O$ -Bindungen [108]. Besonders effektiv ist dieser Prozess nahe der Oxid-Grenzflächen, da hier gestreckte Bindungen vorliegen, die einfacher aufgebrochen werden können (vergleiche Abschnitt 2.3.2). Gleichzeitig diffundiert der an der Substratgrenzfläche freigesetzte Wasserstoff in Richtung Gate. Im Oxidvolumen und insbesondere nahe der Grenzfläche kann dieser Störstellen erzeugen [66, 244, 245]. Beide Prozesse, die direkte Schädigung durch heiße Löcher und die Schädigung durch diffundierenden Wasserstoff könnte die ausgeprägte Degradation unter NBS erklären. Unter PBS fehlen die heißen Löcher auf der Substratseite und eine Erzeugung von Störstellen nahe der Grenzfläche ist nur durch den am Gate freigesetzten Wasserstoff möglich. Dieses wird dadurch bestärkt, dass der Verlauf von D_{ot} in Abbildung 4.11 für PBS erst nach einer Stressladungsdichte von $|Q_{inj}| \simeq 0,5 \text{ C cm}^{-2}$ signifikant ansteigt und im späteren Verlauf ($|Q_{inj}| > 8 \text{ C cm}^{-2}$) zu sättigen beginnt.

Analog zu der Betrachtung von D_{it} konnte auch hier für den p -Kanal-MOSFET unter NBS ein größerer Anstieg der Störstellendichte kurz vor dem Durchbruch beobachtet werden. NBTI verursacht eine vermehrte Generation von Störstellen im grenzflächennahen Bereich [72, 245]. Daher steht dieser Effekt auch hier im Verdacht, für die verstärkte Degradation verantwortlich zu sein. Charakteristisch für den NBTI-Effekt ist auch die Anhäufung positiver Ladung nahe der Substratgrenzfläche. Um zu überprüfen, ob unter diesen Stressbedingungen vermehrt positive Ladung nahe der Substratgrenzfläche angehäuft wird, soll im folgenden Abschnitt die Verschiebung der Flachbandspannung und insbesondere deren Hysterese unter Stresseinwirkung betrachtet werden. Die Lage der Flachbandspannung gibt Aufschluss über die im Oxid befindliche Nettoladung.

Betrachtung der Flachbandspannung

Zur Bestimmung der Flachbandspannung wurden quasistatische $C(V)$ -Messungen an MOS-Kapazitäten durchgeführt, die auf demselben Wafer vorhanden waren (siehe Kapitel 4.1). Die Gate-Oxide der Kapazitäten wurden dementsprechend mit demselben Oxidationsprozess hergestellt. Es wird daher davon ausgegangen, dass das Verhalten unter Stresseinwirkung vergleichbar zu den Ergebnissen an den MOSFETs ist. Die Kapazitätsmessungen wurden an einem $C(V)$ -Messplatz (CV82) mit Messgeräten der Firma Keithley durchgeführt. Eine nähere Beschreibung des Messaufbaus findet sich in [47].

An den Kapazitäten wurde das gleiche Stressexperiment durchgeführt wie an den MOSFETs zuvor. Über den Gate-Kontakt wurde eine konstante FN-Stromdichte von $j_{inj} = \pm 0,1 \text{ mA cm}^{-2}$ eingepreßt bis eine maximale Stressladungsdichte von $|Q_{inj}| = 5,0 \text{ C cm}^{-2}$ durch den Iso-

lator geflossen war. Die Stresseinwirkung wurde auch hier in regelmäßigen Abständen unterbrochen, um eine quasistatische $C(V)$ -Charakteristik aufzunehmen. Die $C(V)$ -Kurven wurden dabei jeweils in Richtung steigender ($\delta V +$) und fallender ($\delta V -$) Gate-Spannung gemessen. Die Gate-Spannung wurde dabei mit einer Rate von etwa $0,2 \text{ V s}^{-1}$ zwischen $-3,0 \text{ V}$ und $3,0 \text{ V}$ variiert (*engl.*: sweep). In beiden Sweep-Richtungen wurde jeweils die Flachbandspannung V_{fb} bestimmt. Eine Verschiebung der Flachbandspannung während des Stresses ist ein deutliches Indiz für die Entstehung von festen Ladungen im Oxid beziehungsweise an der Grenzfläche. Für die ungestressten Proben ergibt sich eine Flachbandspannung von $V_{\text{fb},0} = -0,57 \text{ V}$ für die n -MOS-Kapazität und $V_{\text{fb},0} = 0,75 \text{ V}$ für die p -MOS-Kapazität. Diese Werte wurden bei einer konstanten Flächenkapazität von $0,25 \mu\text{F cm}^{-2}$ aus den gemessenen $C(V)$ -Kurven abgelesen. Daher unterscheiden sie sich etwas von den Werten, die in Tabelle 4.1 für die MOSFETs angegeben sind.

In Abbildung 4.12 ist die Flachbandspannung der n -MOS-Kapazität über dem Betrag der injizierten Stressladungsdichte für beide Stresspolaritäten aufgetragen. Dabei wird zusätzlich zwischen den beiden Sweep-Richtungen der Gate-Spannung unterschieden. Unter beiden Stresspolaritäten ergibt sich eine Verschiebung von V_{fb} zu kleinen Spannungen, die für NBS wesentlich stärker ausgeprägt ist als für PBS. Dies deutet auf eine Anhäufung positiver Ladung nahe der Si-SiO_2 -Grenzfläche hin. Hierbei gilt zu beachten, dass die Wirkung von Oxidladungen auf die Flachbandspannung linear mit ihrem Abstand zur Siliziumgrenzfläche gewichtet ist. Die Flachbandspannung in Anwesenheit von Oxidladungen ist gegeben durch

$$V_{\text{fb}} = V_{\text{fb}}(Q_{\text{f}} = 0) - \frac{Q_{\text{f}}}{C_{\text{ox}}} \frac{d_{\text{ox}} - z}{d_{\text{ox}}}, \quad (4.19)$$

wobei Q_{f} die Menge der festen Ladungen symbolisiert, z deren Abstand von der Substratgrenzfläche und $V_{\text{fb}}(Q_{\text{f}} = 0)$ die Flachbandspannung ohne Ladungen im Oxid. Ladungen nahe der Si-SiO_2 -Grenzfläche bewirken daher eine stärkere Verschiebung der Flachbandspannung als Ladungen, die weiter entfernt liegen.

Für die Anhäufung positiver Ladung wird atomarer Wasserstoff verantwortlich gemacht, der an der Grenzfläche freigesetzt wird und in den grenzflächennahen Bereich diffundiert [244, 246]. Dies spricht für das Modell der Wasserstofffreisetzung. Bestärkt wird es zusätzlich durch die Tatsache, dass eine Verschiebung der Flachbandspannung unter PBS wiederum verzögert einsetzt. Dieses Verhalten wurde zuvor schon bei der Betrachtung der Störstellendichten beobachtet. Unter PBS wird der Wasserstoff an der gate-seitigen Grenzfläche freigesetzt und muss erst zum Substrat diffundieren. Die Anhäufung positiver Ladung setzt dadurch verzögert ein.

Ein etwas anderes Bild zeigt sich für die p -MOS-Kapazität in Abbildung 4.13. Zwar verschiebt sich die Flachbandspannung auch hier zu negativeren Werten, jedoch ist die Verschiebung wesentlich stärker ausgeprägt. Die Anhäufung positiver Ladungen läuft offensichtlich mit einer höheren Rate ab als bei der n -MOS-Kapazität. Unter NBS wird wiederum eine größere Verschiebung beobachtet als unter PBS. Dies spricht erneut für das Modell der Wasserstofffreisetzung, bei dem die Schädigung auf der Substratseite höher ausfällt, wenn dort die Anode liegt (NBS).

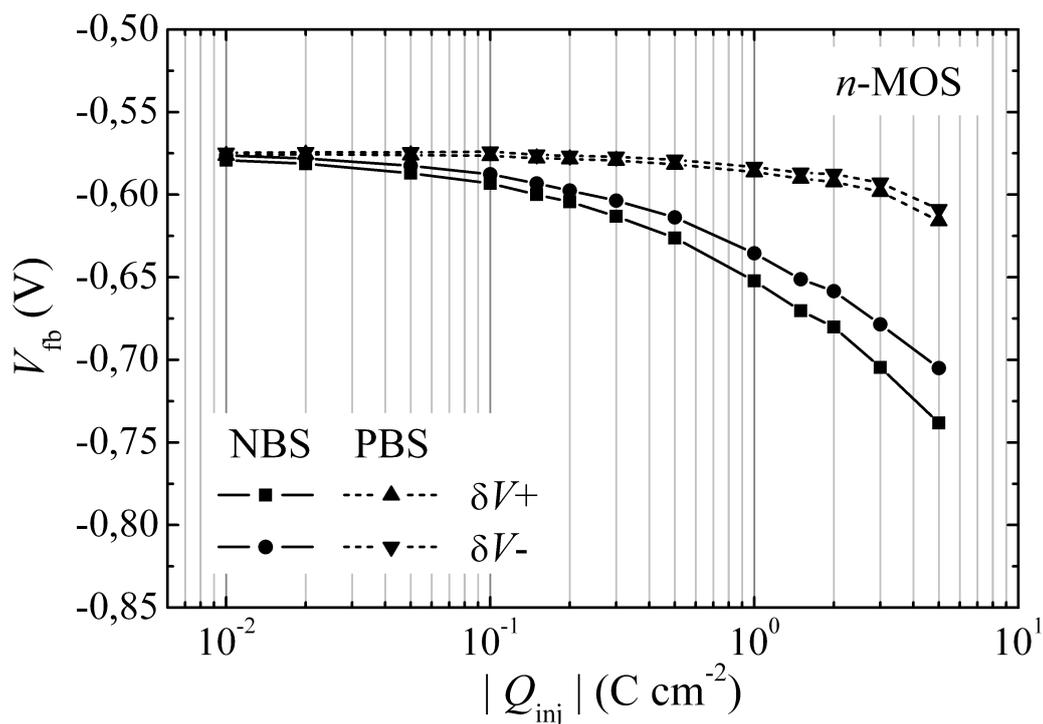


Abbildung 4.12: Verlauf der Flachbandspannung der n -MOS-Kapazität während des FN-Stressexperiments, bestimmt aus Messungen der $C(V)$ -Charakteristik. Die Flachbandspannung der ungestressten Kapazität beträgt $V_{fb,0} = -0,57$ V. Für beide Stresspolaritäten (NBS und PBS) ist die Flachbandspannung jeweils für positive ($\delta V+$) und negative ($\delta V-$) Sweep-Richtung der Gate-Spannung über der injizierten Stressladungsdichte dargestellt. Unter NBS zeigt sich eine deutlich größere Verschiebung von V_{fb} zu negativeren Spannungen. Außerdem ist die Flachbandspannung bei positiver Sweep-Richtung stets kleiner als bei negativer. Probenparameter: $d_{ox} = 6,4$ nm, $L_G = W_G = 200$ μm .

Wie bereits zuvor vermutet scheint unter diesen Stressbedingungen eine vermehrte Anhäufung positiver Ladung auf der Substratseite des Gate-Oxids stattzufinden. Allerdings kann dieser Degradationsmechanismus nicht allein für die Schädigung verantwortlich sein. Die Verschiebung der Flachbandspannung setzt schon bei geringen injizierten Stressladungen deutlich ein. Eine Verzögerung aufgrund von Diffusionseffekten wie bei der n -MOS-Kapazität unter PBS ist hier nicht zu beobachten. Möglicherweise wird die Verschiebung hier durch einen zusätzlichen Effekt überdeckt. Denkbar ist in diesem Fall eine Anhäufung positiver Ladung durch eine Injektion von Löchern aus dem p^+ -Poly-Gate. Dieser Injektionsprozess ist in MOSFETs mit p^+ -Poly-Gates gegenüber MOSFETs mit n^+ -Poly-Gates begünstigt [247, 248]. Zusätzlich zu der absoluten Verschiebung von V_{fb} wird in allen Messungen jeweils zwischen der positiven ($\delta V+$) und der negativen ($\delta V-$) Sweep-Richtung eine zunehmende Differenz in der Flachbandspannung mit steigender Stressladungsdichte $|Q_{inj}|$ beobachtet. Dieser Unterschied wird als Hysterese der Flachbandspannung bezeichnet [236, 249] und ist in dieser Arbeit

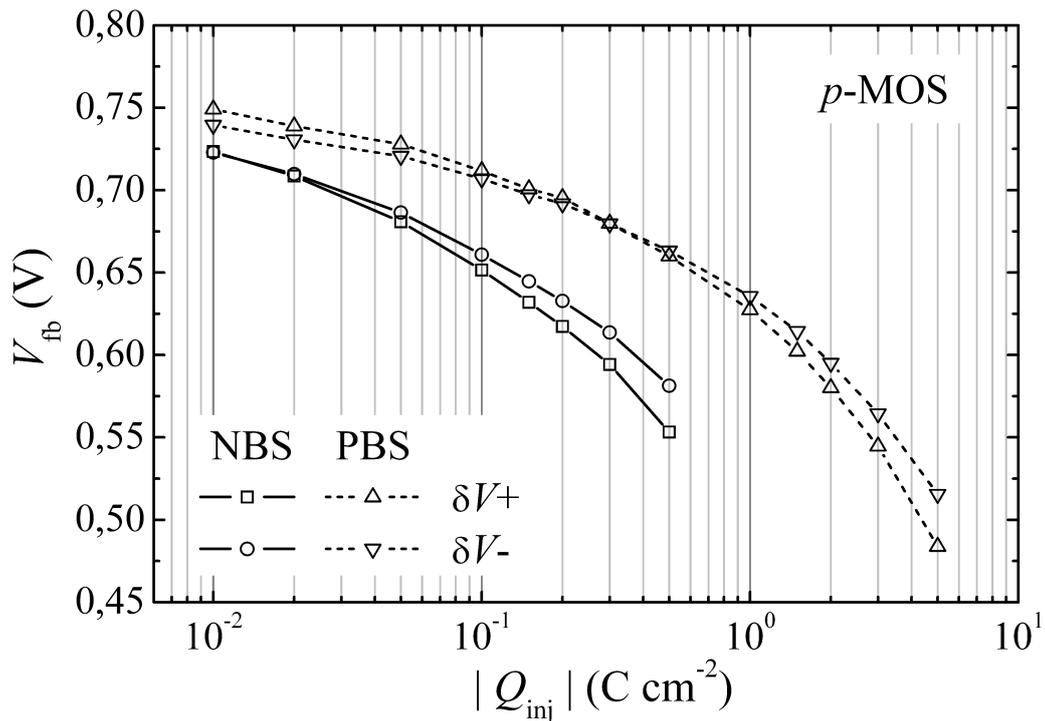


Abbildung 4.13: Verlauf der Flachbandspannung der p -MOS-Kapazität während des FN-Stressexperiments, analog zu Abbildung 4.12. Die Flachbandspannung der ungestressten Kapazität beträgt $V_{fb,0} = 0,75$ V. Unter NBS zeigt sich auch hier eine deutliche Verschiebung von V_{fb} zu negativeren Spannungen. Die absolute Verschiebung ist zudem größer als bei der n -MOS-Kapazität. Auffällig ist die Flachbandspannung unter PBS. Bei positiver Sweep-Richtung und kleinen injizierten Stressladungsdichten ist sie zunächst größer als bei negativer Sweep-Richtung. Für größere Stressladungsdichten kehrt sich die Lage um. Probenparameter: $d_{ox} = 6,4$ nm, $L_G = W_G = 200$ μ m.

definiert als

$$\Delta V_{fb} := V_{fb}(\delta V-) - V_{fb}(\delta V+) \quad . \quad (4.20)$$

Ursache für die Hysterese sind umladbare Störstellen im Oxidvolumen nahe der beiden Grenzflächen des Isolators. Diese grenzflächennahen Störstellen (NIOTs) können Ladungen mit dem Substrat beziehungsweise mit dem Gate-Kontakt austauschen. Dabei hängt der Besetzungszustand von der Lage des Quasi-Fermi-Niveaus und damit von der Gate-Spannung ab. Beim Start einer $C(V)$ -Messung entweder in Akkumulation oder Inversion werden diese Störstellen mit Elektronen gefüllt oder geleert.

Die jeweiligen Situationen sind in Abbildung 4.14 veranschaulicht. Bei der unter (a) gezeigten Situation mit negativer Gate-Spannung, die den Ausgangspunkt für den positiven Sweep darstellt, werden die gate-seitigen Störstellen mit Elektronen gefüllt, die substratseitigen dagegen entleert. In (b) ist die Situation mit positiver Gate-Spannung gezeigt, die den Ausgangspunkt für den negativen Sweep darstellt. Die gate-seitigen Störstellen werden entleert, während sie

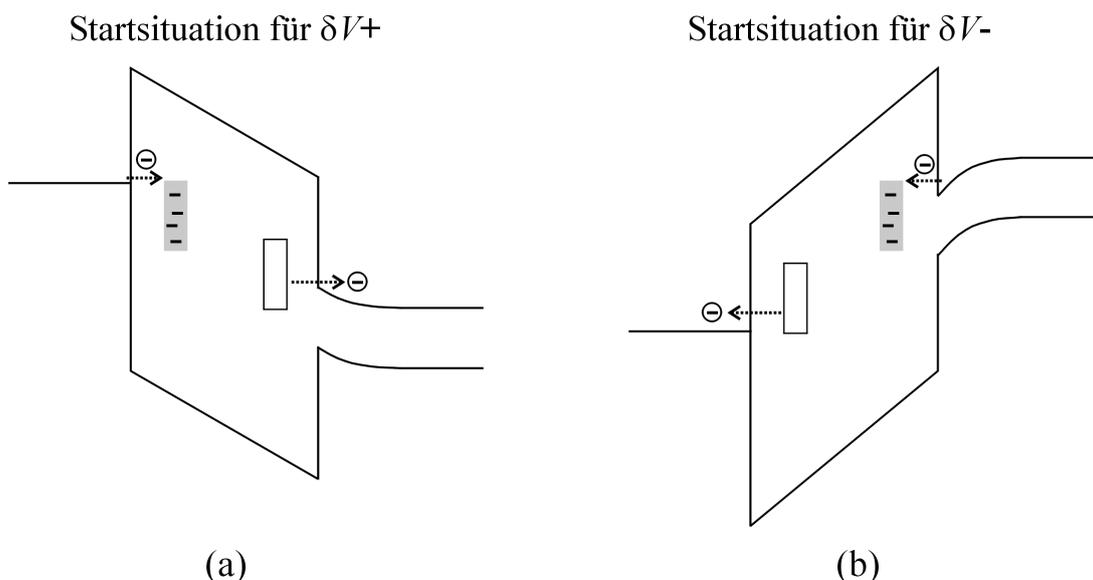


Abbildung 4.14: Darstellung der Ausgangssituationen beim Verfahren der Gate-Spannungen während der $C(V)$ -Messungen. (a) Unter negativer Gate-Spannung werden die grenzflächennahen Störstellen auf der Gate-Seite mit Elektronen geladen und die auf der Substratseite entleert. (b) Unter positiver Gate-Spannung stellt sich der entgegengesetzte Ladungszustand ein.

auf der Substratseite mit Elektronen besetzt werden². Wird die Gate-Spannung schnell genug verfahren, so kann davon ausgegangen werden, dass sich der Besetzungszustand während des Messens nicht ändert. Aufgrund der unterschiedlichen Ladungszustände in der jeweiligen Anfangssituation liegen die gemessenen $C(V)$ -Kurven zueinander verschoben und zeigen eine Hysterese.

Nach Gleichung 4.19 haben die substratseitigen Ladungen einen stärkeren Einfluss auf V_{fb} als die gate-seitigen. Daraus ergibt sich unter der Annahme ähnlich hoher Störstellendichten auf beiden Isolatorseiten, dass die $C(V)$ -Kurve und damit die Flachbandspannung der ($\delta V-$)-Messung relativ zu der der ($\delta V+$)-Messung zu positiveren Spannungen verschoben liegt (siehe Abbildung 4.12). Die Hysterese nach Definition von Gleichung 4.20 ist somit stets positiv.

In Abbildung 4.13 für die p -MOS-Kapazität zeigt sich jedoch unter PBS eine anfängliche negative Hysterese. Bis zu einer Stressladungsdichte von $|Q_{inj}| \simeq 0,2 \text{ C cm}^{-2}$ ist die Flachbandspannung der ($\delta V-$)-Messung kleiner als die der ($\delta V+$)-Messung. Bei höheren injizierten Stressladungen stellt sich wieder eine positive Hysterese ein. Das Auftreten einer negativen Hysterese mit späterem Vorzeichenwechsel wurde in der Literatur bisher nicht beobachtet.

²An dieser Stelle wurde von Störstellen ausgegangen, die Elektronen einfangen beziehungsweise abgeben. Die Betrachtungen bleiben ebenso gültig, falls die Störstellen Löcher einfangen oder abgeben. Wichtig ist nur, dass der Ladungszustand der Störstelle beim Einfang eines Elektrons beziehungsweise bei Abgabe eines Loches um eine Elementarladung negativer wird. Bei Abgabe eines Elektrons beziehungsweise Einfang eines Loches wird der Ladungszustand entsprechend um eine Elementarladung positiver. Die Auswirkung auf die Flachbandspannung bleibt die gleiche.

In Abbildung 4.15 ist die Hysterese ΔV_{fb} für beide Leitungstypen und beide Stresspolaritäten über der injizierten Stressladungsdichte dargestellt. Hier zeigt sich deutlich die negative Hysterese für die p -MOS-Kapazität unter PBS, die allerdings auch unter negativer Stresspolarität auftritt. Jedoch fällt sie unter NBS geringer aus und wechselt schon kurz nach Beginn des Stresses ($|Q_{inj}| \approx 0,02 \text{ C cm}^{-2}$) ihr Vorzeichen.

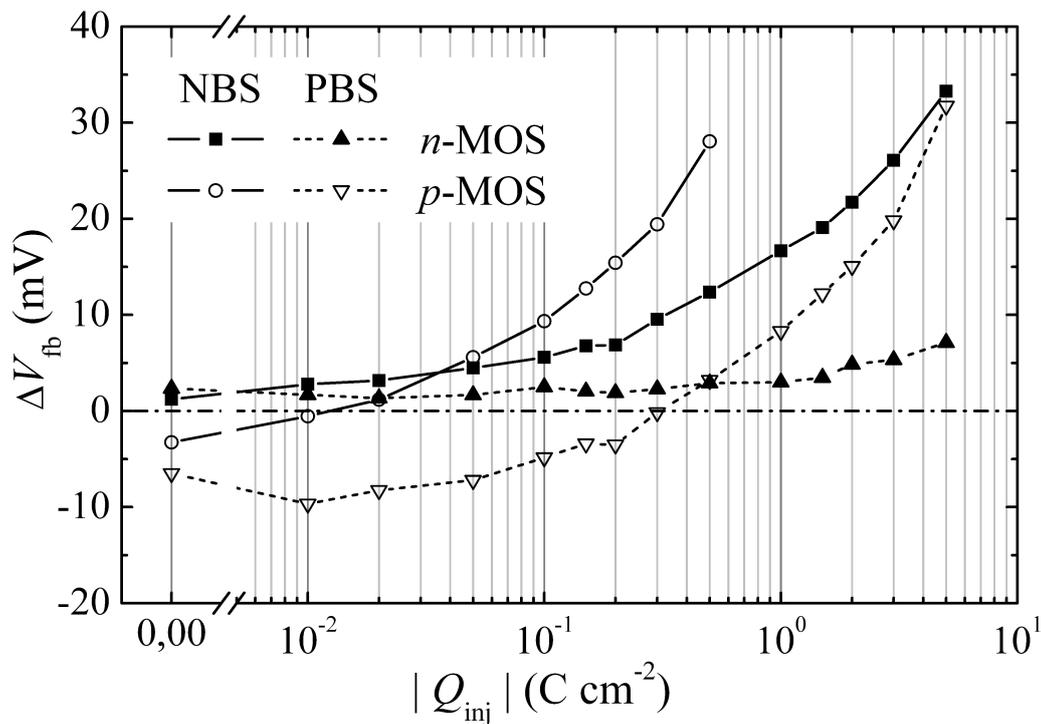


Abbildung 4.15: Hysterese ΔV_{fb} der Flachbandspannung über der injizierten FN-Stressladungsdichte. Unter PBS zeigt sich eine negative Hysterese, die für die p -MOS-Kapazität deutlich ausgeprägter ist. Mit steigender Stressladungsdichte kehrt sich das Vorzeichen um. Probenparameter: $d_{ox} = 6,4 \text{ nm}$, $L_G = W_G = 200 \mu\text{m}$.

Die Beobachtung einer negativen Hysterese bedeutet, dass die gate-seitigen Störstellen einen stärkeren Einfluss auf die Flachbandspannung ausüben, als die substratseitigen. Aufgrund ihrer räumlichen Lage wäre dies nach Gleichung 4.19 nicht der Fall. Eine plausible Erklärung kann daher nur unter der Annahme gegeben werden, dass die Zahl der gate-seitigen Störstellen im Isolator der p -MOS-Kapazität wesentlich höher ist, als die Zahl der substratseitigen. Insbesondere ist dies für das ungestresste beziehungsweise geringfügig gestresste Bauelement der Fall. Im Laufe des Stressexperiments schwächt sich das Ungleichgewicht in der Störstellenkonzentration zwischen Gate- und Substratseite offensichtlich aus und das Vorzeichen der Hysterese wird wieder positiv.

4.3.5 Schlussfolgerung

Die Ursache für die extrem hohe Störstellendichte auf der Gate-Seite muss durch den Herstellungsprozess bedingt sein. Da für beide Leitungstypen der gleiche Oxidationsprozess stattgefunden hat, die negative Hysterese aber nur bei der p -MOS-Kapazität auftritt, liegt die Vermutung nahe, dass dies mit dem p^+ -Poly-Gate zusammenhängt. Hier könnte die Eindiffusion des Dotierstoffs Bor in das Gate-Oxid eine entscheidende Rolle spielen. Bor im Gate-Isolator bewirkt eine erhöhte Effizienz in der Generation von Störstellen [250]. Diese dienen vorwiegend als Einfangzentren für Elektronen, sind verantwortlich für SILC und setzen die Lebensdauer von p -MOSFETs wesentlich herab [251, 252]. Zwar wird die Bordiffusion der allgemeinen Auffassung nach durch eine Nitridierung herabgesetzt [124, 125, 253], jedoch ist sie nicht vollständig auszuschließen. Insbesondere bei höheren Temperaturen in nachfolgenden Prozessschritten kann eine Bordiffusion auftreten.

Die Ergebnisse der in dieser Arbeit untersuchten Bauelemente legen daher nahe, dass eine prozessbedingte Generation von Störstellen auf der Gate-Seite der p -MOS-Bauelemente stattgefunden hat. Es ist nicht auszuschließen, dass im Laufe der Prozessierung ein Prozessschritt mit erhöhter Temperatur zur Diffusion von Bor aus dem p^+ -Poly-Gate ins Siliziumdioxid geführt hat. Es konnte nicht abschließend geklärt werden, auf welche Weise und in welcher Reihenfolge die Poly-Gates hergestellt und aktiviert wurden. Denkbar ist auch, dass die p^+ -Poly-Gates zuerst prozessiert und aktiviert wurden und bei der späteren Aktivierung der n^+ -Poly-Gates eine weitere Temperung erfahren haben.

Bezüglich der schwachen beziehungsweise fehlenden Polarisationsabhängigkeit bei der Generation von Grenzflächenzuständen unter FN-Stress wird vermutet, dass hier der Einfluss der unterschiedlichen Messmethoden eine Rolle spielt. Die in der Literatur häufig angewendeten $C(V)$ -Methoden werden in der Regel bei niedrigeren Messfrequenzen durchgeführt als die hier für die Bestimmung der Grenzflächenzustände verwendeten Frequenzen. Daher beinhaltet die aus $C(V)$ -Messungen bestimmten Zustandsdichten möglicherweise auch Anteile von grenzflächennahen Oxidstörstellen, die somit die Asymmetrie im Generationsverhalten erzeugen.

In dieser Arbeit wurden zu den schnellen Grenzflächenzuständen nur die Störstellen gerechnet, die mit Frequenzen größer als 2 MHz kommunizieren (siehe Abschnitt 4.3.2). Damit wird eine schärfere Trennung zwischen Grenzflächenzuständen und grenzflächennahen Oxidstörstellen ermöglicht, die zu einer detaillierteren Betrachtung der Störstellengeneration führt.

4.4 Messaufbau für Rauschmessungen

Zur eingehenderen Untersuchung der MOSFETs kommen in diesem Abschnitt Rauschmessungen zum Einsatz. Diese werden hauptsächlich im zweiten Teil der Arbeit (Kapitel 5) zur Charakterisierung von NROM-Speicherzellen verwendet. Hier soll jedoch zunächst untersucht werden, inwieweit sich der FN-Stress auf das $1/f$ -Rauschen eines MOS-Transistors auswirkt. Gleichzeitig dienen diese Messungen als Überprüfung der Rauschmessmethode, denn die stressin-

duzierten NIOTs und Grenzflächenzustände sollten mit einem Anstieg der Spektraldichte des Stromrauschens korreliert sein.

Für die Rauschmessungen wurde der Messaufbau aus den Charge-Pumping-Experimenten (Abbildung 4.7) im Wesentlichen beibehalten und nur leicht modifiziert. Die benötigten Drain- und Gate-Spannungen werden vom Parameter-Analyzer bereitgestellt, der in der Lage ist, gleichzeitig den Drain-Strom aufzuzeichnen. Der Pulsgenerator und das Elektrometer werden für Rauschmessungen nicht benötigt. Die Gate-Spannung wird über die SMU 3 angelegt, die Drain-Spannung über die SMU 1. Der Source-Kontakt ist mit der SMU 2 verbunden, die konstant 0 V ausgibt. Der Well-Kontakt (B) ist über die Ground-Unit (GND) des Parameter-Analyzers geerdet. Abbildung 4.16 zeigt den leicht modifizierten Aufbau mit der entsprechenden Konfiguration der Schaltmatrix.

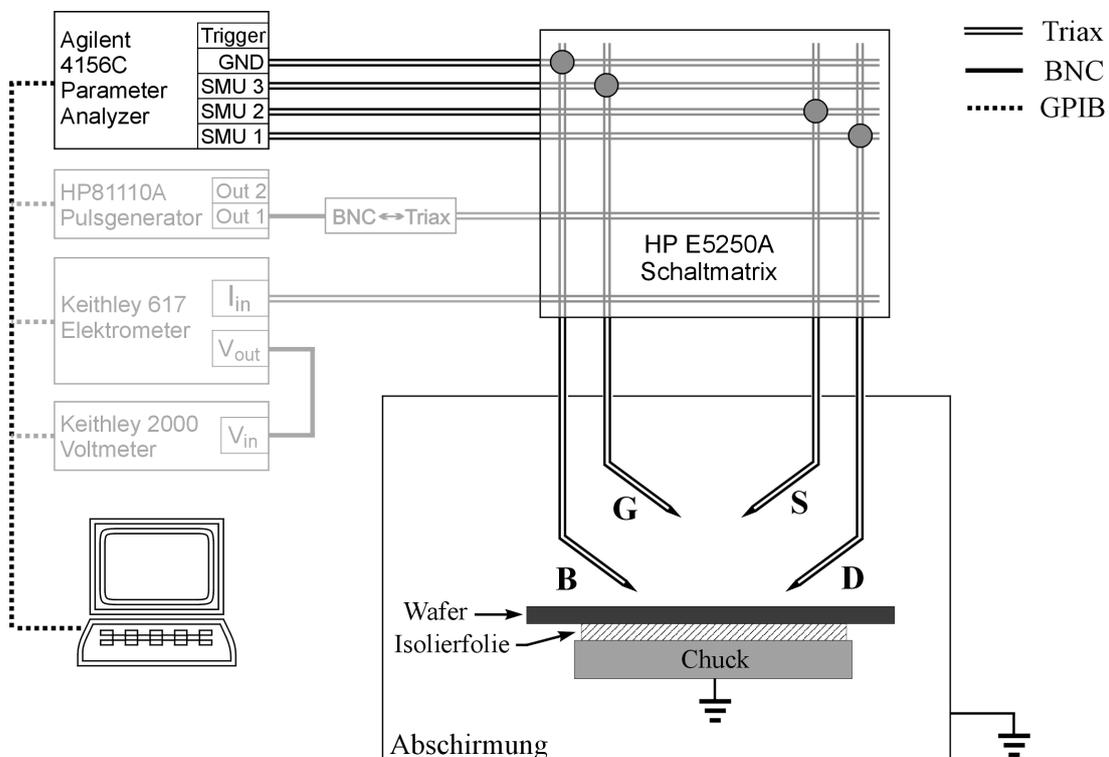


Abbildung 4.16: Schematische Darstellung des Messaufbaus für $1/f$ -Rauschanalysen. Die Rauschmessungen werden mit dem Parameter-Analyzer durchgeführt, der Pulsgenerator, das Elektrometer und das Voltmeter werden nicht benötigt. Die Konfiguration der Schaltmatrix ist durch Punkte gekennzeichnet.

4.5 Messungen des $1/f$ -Rauschens

Für die im Folgenden dargestellten Rauschmessungen wurden die gleichen Transistoren untersucht wie im Abschnitt zuvor. Das Stressexperiment bestand wiederum aus sich abwechselnden Stress- und Messphasen. Um eine Korrelation der Ergebnisse zu den vorangegangenen Charge-Pumping-Untersuchungen zu ermöglichen, wurden dieselben Stresszustände Q_{inj} für die Charakterisierungsmessungen gewählt. Der Messablauf war so gestaltet, dass zunächst eine Transferkennlinie mit $V_D = 10$ mV gemessen wurde, danach das $1/f$ -Rauschen und schließlich weiterer FN-Stress bis zur nächsten geplanten Stressunterbrechung eingepreßt wurde. Dies wurde solange wiederholt, bis der Betrag des kumulierten Stresses 15 C cm^{-2} betrug.

Aus der gemessenen Transferkennlinie wurde jeweils durch Interpolation die Gate-Spannung bestimmt, die einem konstanten Drain-Strom von 700 nA entspricht. Der Transistor befindet sich unter diesen Bedingungen im Unterschwellbereich ($V_G - V_t \simeq -0.15$ V), in dem der Beitrag von Rauschquellen zum Drain-Stromrauschen maximal ist [205].

Zur Bestimmung der Spektraldichte mittels einer FFT wurde der Drain-Strom mit $V_D = 10$ mV bei der zuvor bestimmten Gate-Spannung gemessen. Es wurden bei jeder dieser Messungen 10001 Strommesswerte mit einem zeitlichen Abstand von $\Delta t = 1,28$ ms aufgezeichnet. Dies führt zu einer gesamten Messdauer des Drain-Stromsignals von 12,8 s. Abbildung 4.17 zeigt eine typische Messung. Die Integrationszeit an jedem Messpunkt betrug 1,04 ms, welches in etwa 81% von Δt entspricht. Dadurch wird der Aliasing-Effekt (siehe Abschnitt 3.3.2) weitgehend unterdrückt.

Das Einprägen des FN-Stresses erfolgte analog zu dem in Abschnitt 4.3.3 beschriebenen Stressexperiment mit einer konstanten Stromdichte von $j_{inj} = 0,1 \text{ mA cm}^{-2}$. Es wurden n - und p -MOSFETs mit jeweils positiver (PBS) als auch negativer (NBS) Polarität der Gate-Spannung gestresst.

Die Analyse der Messdaten und die FFT erfolgten mit einem in LabVIEW erstellten Auswertungsprogramm. Die Software stellt die numerischen Algorithmen für die FFT bereit, die hierbei zum Einsatz kamen. Aus der gesamten Zeitspanne des aufgezeichneten Drain-Stromsignals

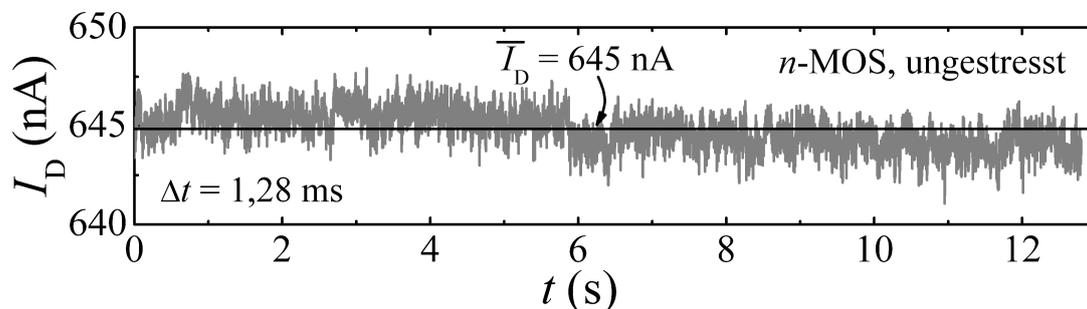


Abbildung 4.17: Typische Messung des Drain-Stroms über der Zeit zur Rauschanalyse. Es werden 10001 Messpunkte mit einer Zeitauflösung von $\Delta t = 1,28$ ms aufgezeichnet. \bar{I}_D ist der Strommittelwert der Messung. Probenparameter: siehe Tabelle 4.1.

von $T = 12,8$ s ergibt sich durch die Fourier Transformation eine Minimalfrequenz von $f_{\min} = 1/T = 78$ mHz. Entsprechend bestimmt sich die Maximalfrequenz nach Gleichung 3.43 zu $f_{\max} = 390$ Hz, welches der Nyquist-Grenzfrequenz f_c entspricht. Hierbei wurde mit der Software stets die einseitige Spektraldichte S_{I_D} nach Gleichung 3.42 berechnet. Die Messungen wurden zwar alle bei einem Drain-Strom von $I_D \simeq 700$ nA durchgeführt, dennoch wurde die Spektraldichte zur besseren Vergleichbarkeit der Daten mithilfe des Strommittelwertes zum Quadrat ($\overline{I_D}^2$) über

$$S'_{I_D}(f) := \frac{S_{I_D}(f)}{\overline{I_D}^2} \quad (4.21)$$

normiert. Die normierte Spektraldichte wird im Folgenden mit S'_{I_D} bezeichnet. Es zeigte sich bei dieser Vorgehensweise eine starke Fluktuation von S'_{I_D} im untersuchten Frequenzbereich, die eine Analyse erschwerte. Um die Spektren zu glätten, wurde die Drain-Strommessung in jedem Stresszustand 50 Mal durchgeführt und transformiert. Aus den 50 Einzelspektren wurde der Mittelwert $\overline{S'_{I_D}}(f)$ gebildet.

In Abbildung 4.18 sind die gemittelten Spektren des Drain-Stromrauschens eines n -Kanal-MOSFETs für verschiedene Stresszustände unter NBS dargestellt. Die Frequenzabhängigkeit kann sehr gut durch die Form $1/f^\gamma$ mit $\gamma \gtrsim 1$ wiedergegeben werden. Mit steigendem Betrag der injizierten Stressladungsdichte Q_{inj} nimmt die Spektraldichte zu.

Im Vergleich dazu ist das Rauschspektrum eines Widerstandes mit $R = 100$ k Ω als Referenz dargestellt (gepunktete Kurve). Der Widerstand wurde direkt anstelle des MOSFETs zwischen die Source- und Drain-Messnadeln des Messaufbaus eingebaut und der Strom bei einer Spannung von 10 mV gemessen. Für Frequenzen kleiner als 4 Hz ergibt sich wie beim MOSFET annähernd ein $1/f$ -Verlauf, der um etwa 2 bis 3 Größenordnungen unter dem des Transistors liegt. Für Frequenzen größer als 4 Hz zeigt sich ein konstanter Verlauf der Spektraldichte. Dies wird durch das thermische Rauschen bei Raumtemperatur ($T = 300$ K) hervorgerufen, dessen Spektraldichte durch Gleichung 3.21 gegeben ist. Die normierte Spektraldichte ergibt in diesem Fall einen Wert von $S'_{I_D} = 1,6 \cdot 10^{-11}$ Hz $^{-1}$, welcher etwa um einen Faktor 5 kleiner ist als in Abbildung 4.18 abgelesen werden kann. Der Grund hierfür sind vermutlich die internen Widerstände der Messgeräte, die ebenfalls ein weißes Grundrauschen besitzen. Auffällig an den gezeigten Spektren sind auch die auftretenden Spitzen in der Spektraldichte besonders bei 50 Hz und ganzzahligen Vielfachen dieser Frequenz. Sie werden ganz offensichtlich durch Einstreuungen der Netzfrequenz über die Messgeräte hervorgerufen.

Der Vergleich mit dem 100 k Ω -Widerstand zeigt, dass die mit diesem einfachen Messansatz und Auswerteverfahren bestimmten Rauschspektren nicht für quantitative Aussagen verwertbar sind. Dennoch hebt sich das Rauschen des MOSFETs um mehrere Größenordnungen vom Grundrauschen ab und es lässt sich ein deutlicher Einfluss des eingepprägten FN-Stresses in den Spektren erkennen. Somit ist dieser Messansatz durchaus für qualitative Bewertungen der Degradation geeignet und kann für die Untersuchung relativer Änderungen sehr wohl verwendet werden.

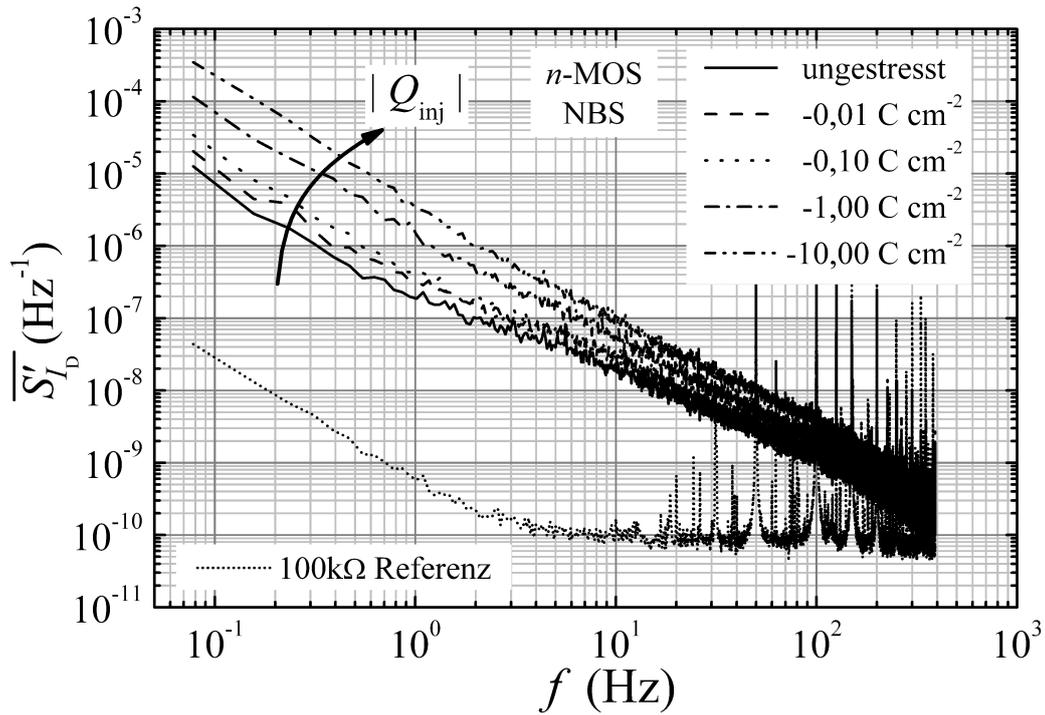


Abbildung 4.18: Spektraldichte eines *n*-MOS-Transistors unter NBS für verschiedene FN-Stresszustände. Die Spektren zeigen annähernd einen $1/f$ -Verlauf. Mit zunehmendem Betrag der injizierten Stressladungsdichte Q_{inj} steigt die Spektraldichte im gesamten gezeigten Frequenzbereich an. Zum Vergleich ist das Spektrum eines 100 k Ω -Widerstandes dargestellt, das unterhalb von ≈ 4 Hz ebenfalls ein $1/f$ -Verhalten und oberhalb davon thermisches (weißes) Rauschen zeigt. Probenparameter: siehe Tabelle 4.1.

4.5.1 Ergebnisse der Rauschmessungen

Die in Abbildung 4.18 gezeigten Spektren wurden für die nachfolgenden Untersuchungen im Frequenzbereich von 0,1 Hz bis 10 Hz durch die Form $S'_{ID} = S'_{ID,1} f^{-\gamma}$ genähert. Der Vorfaktor $S'_{ID,1}$ gibt dabei die Spektraldichte bei 1 Hz an und wird im Folgenden als Maß für das Rauschen verwendet.

Ziel der hier vorgestellten Experimente ist es, eine Korrelation zwischen der gemessenen Spektraldichte und den in Abschnitt 4.3.4 bestimmten Störstellendichten D_{it} und D_{ot} aufzuzeigen. Nach Gleichung 3.39 hängt die Spektraldichte linear von der Störstellendichte nahe der Fermi-Energie ab.

In Abbildung 4.19 ist die Änderung der Spektraldichte $\Delta S'_{ID,1} := S'_{ID,1} - S'_{ID,1}(Q_{inj} = 0)$ über der Änderung der Oxidstörstellendichte $\Delta D_{ot} := D_{ot} - D_{ot}(Q_{inj} = 0)$ dargestellt. Es sind hier nur die Ergebnisse des *p*-MOSFETs unter NBS und PBS, sowie des *n*-MOSFETs unter NBS gezeigt. Der *n*-MOSFET unter PBS lieferte keine verwertbaren Daten. Das Einprägen des FN-Stresses führte schon nach sehr kurzer Stresszeit zu einem extremen Anstieg der Spannung am Gate. Nach 100 s betrug die Stressspannung 8 V und nach 1000 s etwa 9 V. Im weiteren Verlauf

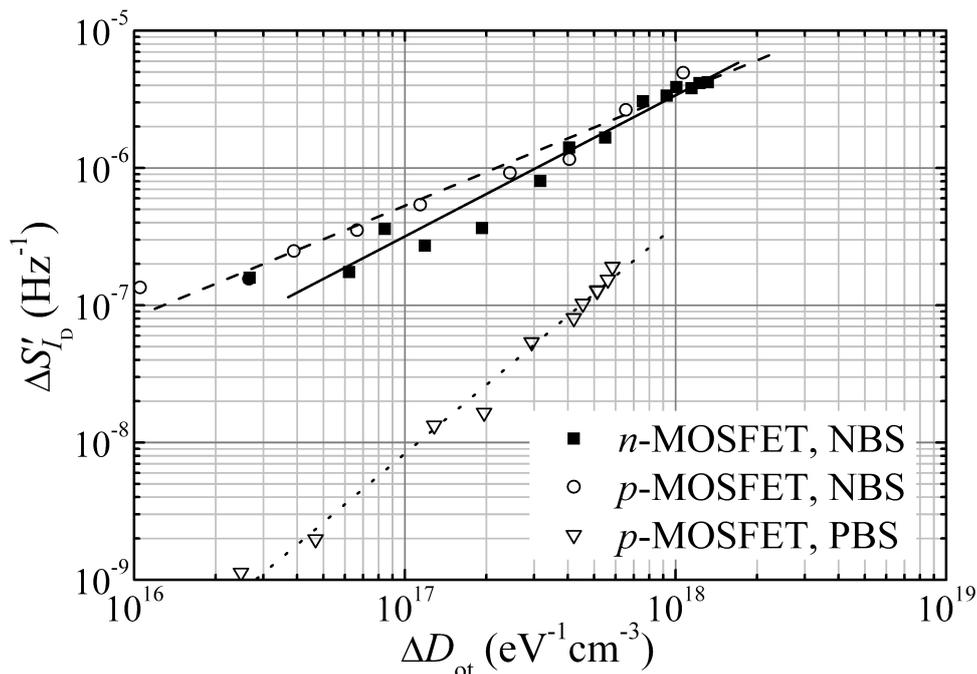


Abbildung 4.19: Änderung der normierten Spektraldichte $\Delta S'_{I_{D,1}}$ über dem Anstieg ΔD_{ot} der Dichte der Oxidstörstellen. Sowohl der n - als auch der p -MOSFET zeigen unter NBS eine gute Korrelation mit der Steigung Eins, welches den linearen Zusammenhang zwischen Spektraldichte und Störstellendichte belegt. Der p -MOSFET unter PBS zeigt einen deutlich geringeren absoluten Anstieg der Spektraldichte bei vergleichbarer Menge an generierten Oxidstörstellen. Zudem ist die Steigung größer als Eins. Probenparameter: siehe Tabelle 4.1.

wurde die Spannungsbegrenzung des Parameter-Analyzers erreicht, die auf 12 V eingestellt war. Während des Stressexperiments für die CP-Untersuchungen lag die Stressspannung annähernd konstant bei 6,1 V (siehe Tabelle 4.5). Die Ursache hierfür konnte nicht gefunden werden. Eine Wiederholung des Experiments an einem zweiten n -MOSFET auf dem Wafer zeigte dasselbe Resultat. Unter NBS trat dieses Phänomen nicht auf, der Verlauf der Stressspannung über der Zeit war hier vergleichbar mit dem aus den CP-Untersuchungen.

Eine grafische Darstellung der Spannungsverläufe während des FN-Stresses befindet sich in Anhang A.3.

Für die übrigen drei in Abbildung 4.19 gezeigten Experimente ergibt sich jedoch eine sehr gute Korrelation zwischen dem Anstieg der Spektraldichte und der Zunahme der Oxidstörstellendichte.

In Abbildung 4.20 sind die gleichen Werte $\Delta S'_{I_{D,1}}$ über der Änderung der Grenzflächenzustandsdichte $\Delta D_{it} := D_{it} - D_{it}(Q_{inj} = 0)$ aufgetragen. Hier zeigt sich ein ähnliches Bild. Für die NBS-Bedingung ergibt sich in der doppeltlogarithmischen Auftragung für beide Transistortypen eine Steigung nahe Eins, welches den linearen Zusammenhang zwischen der Spektraldichte und der Störstellendichte widerspiegelt. Die Steigung für den p -MOSFET unter PBS

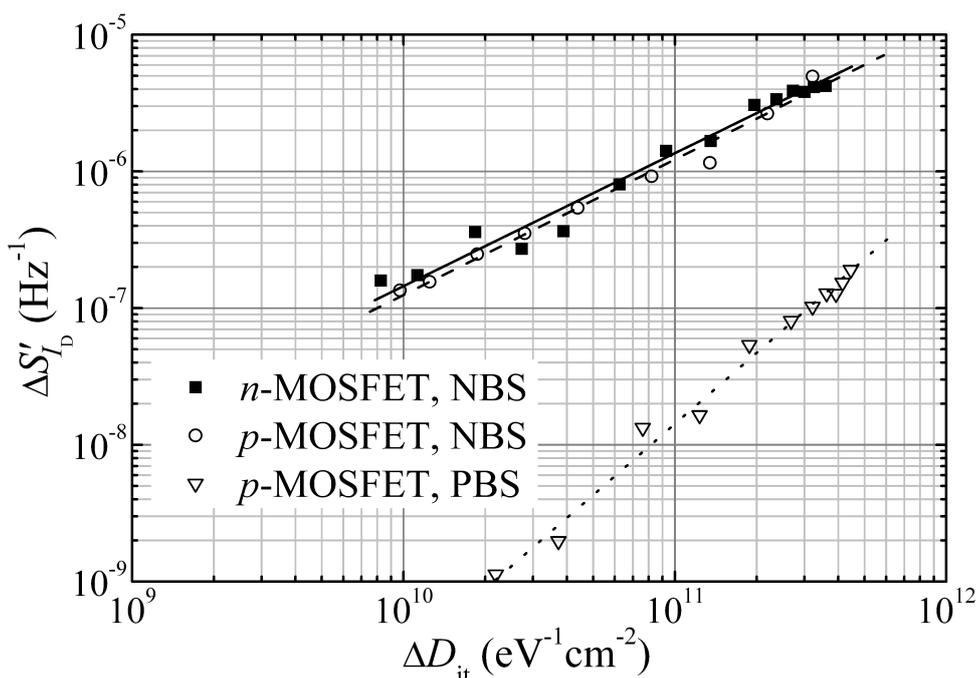


Abbildung 4.20: Änderung der normierten Spektraldichte über dem Anstieg der Grenzflächenzustandsdichte ΔD_{it} . Die Abhängigkeiten sind ähnlich wie in Abbildung 4.19. Die Rauschmessungen korrelieren demnach offensichtlich gleichermaßen mit Störstellen an der Grenzfläche (D_{it}) und mit NIOTs (D_{ot}). Probenparameter: siehe Tabelle 4.1.

scheint dagegen größer als Eins zu sein. Dies widerspricht dem erwarteten linearen Zusammenhang. Eine Ursache hierfür könnte sein, dass der Term $(\frac{1}{n} \pm \alpha\mu)^2$ aus Gleichung 3.39 während des Stressperiments nicht konstant bleibt. Zudem zeigt sich bei den Rauschmessungen kein Unterschied zwischen Grenzflächenzuständen und Oxidstörstellen (NIOTs). Die Spektraldichte korreliert gleichermaßen mit den jeweiligen aus CP-Untersuchungen bestimmten Dichten.

Des Weiteren fällt auf, dass die Spektraldichte unter PBS 1,5 bis 2 Größenordnungen niedriger liegt als unter NBS. Im PBS-Experiment zeigt sich also nur eine verschwindend geringe Zunahme des Rauschens, obwohl Störstellen in einer vergleichbaren Menge wie bei NBS generiert wurden. Dies könnte eine direkte Auswirkung der Stresspolarisation sein und würde bedeuten, dass die unter PBS erzeugten Störstellen zwar mit Charge-Pumping aber weniger mit Rauschanalysen erfasst werden können. Leider fehlt an dieser Stelle für eine genauere Aussage die Messung des n -Kanal-MOSFETs unter PBS. Zudem könnte eine leicht unterschiedliche Lage des Fermi-Niveaus während der Rauschmessungen eine Rolle spielen. Jedoch wurde die Differenz $V_G - V_t$ bei allen Messungen ähnlich eingestellt, so dass die Lagen des Fermi-Niveaus nicht wesentlich voneinander abwichen.

Eine abschließenden Klärung der Ursache konnte im Rahmen dieser Arbeit nicht mehr durchgeführt werden.

5 Charakterisierung von NROM-Speicherzellen

Dieses Kapitel beschreibt die elektrischen Untersuchungen an NROM-Speicherzellen, deren Funktionsweise bereits in Abschnitt 2.5.1 erläutert wurde. Die beiden Bits der NROM-Zelle sind jeweils nahe des Source und Drain lokalisiert und werden durch heiße Elektronen programmiert beziehungsweise durch heiße Löcher gelöscht. Bei diesen Prozessen überwinden die heißen Ladungsträger die Potenzialbarriere des Tunneloxids und werden in der Siliziumnitridschicht von tiefen Störstellen eingefangen. Im Zuge wiederholter Programmier- und Löszyklen (Stressen) degradiert dabei sowohl das Tunneloxid als auch die Grenzfläche zum Substrat, welches zu einer Verschlechterung der Eigenschaften des Speichers, wie zum Beispiel Ladungserhaltung (*engl.*: retention) oder Lebensdauer (*engl.*: endurance) führt.

Für die Bestimmung der Degradation des Bauelements werden $1/f$ -Rauschmessungen angewendet, die bereits in Kapitel 4.5 beschrieben wurden. Hierbei soll der Anstieg der Spektraldichte in Abhängigkeit der Zahl der Programmier- und Löszyklen (*kurz*: P/L-Zyklen) untersucht werden. Die durch das wiederholte Programmieren und Löschen hervorgerufenen Schädigungen im Tunneloxid und an der Grenzfläche zum *Si*-Substrat bewirken einen Anstieg des $1/f$ -Rauschens [254–256]. Daher eignen sich Analysen des $1/f$ -Rauschens sehr gut zum Nachweis von Schädigungen durch heiße Ladungsträger in MOSFETs [257–259]. Da heiße Ladungsträger während des Programmierens und Löschens der Speicherzelle nahe des Drains auftreten, steht bei den folgenden Experimenten die Entwicklung einer Methode zur lokalisierter Detektion der Oxidschädigung im Transistorkanal im Vordergrund.

5.1 Verwendete Probenstrukturen

Für die in diesem Kapitel dargestellten Experimente standen NROM-Teststrukturen der Firma Qimonda auf zwei Wafern mit unterschiedlichen ONO-Schichtstapeln zur Verfügung. Diese wurden in 75 nm-Technologie gefertigt und weisen eine nominelle Gate-Länge von $L_G = 115$ nm und eine nominelle Gate-Weite von $W_G = 75$ nm auf. Auf dem einen Wafer setzt sich der ONO-Stapel aus 4 nm Tunneloxid, 6 nm Siliziumnitrid (Si_3N_4) und 12 nm Zwischenisolator zusammen, der zweite Wafer weist entsprechende Schichtdicken von 3 nm, 6 nm und 9 nm auf. Im Folgenden werden die beiden Wafer entsprechend ihrer ONO-Schichtstapel als 4/6/12-ONO beziehungsweise 3/6/9-ONO bezeichnet. Die Schwellspannung der Transistoren

wurde im linearen Bereich mit $V_D = 20$ mV bei einem konstanten Drain-Strom $I_{D,t}$ ermittelt. Für das 4/6/12-ONO ergibt sich eine Schwellspannung von $V_t = 4,2$ V bei $I_{D,t} = 10$ nA und für das 3/6/9-ONO ein Wert von $V_t = 3,6$ V bei $I_{D,t} = 40$ nA. Der Gate-Kontakt besteht aus n^+ -dotiertem Poly-*Si*. In Tabelle 5.1 sind alle wichtigen Probenparameter zusammengefasst.

	4/6/12-ONO	3/6/9-ONO
Kanalweite W_G	75 nm	75 nm
Kanallänge L_G	115 nm	115 nm
Tunneloxiddicke d_{bottom}	4 nm	3 nm
Siliziumnitridicke $d_{\text{Si}_3\text{N}_4}$	6 nm	6 nm
Zwischenisolatordicke d_{top}	12 nm	9 nm
Schwellspannung V_t	4,2 V	3,6 V
Poly- <i>Si</i> -Gate	n^+	n^+

Tabelle 5.1: Probenparameter der untersuchten NROM-Zellen

Auf den Wafern sind jeweils mehrere NROM-Zellen in einem so genannten Makro integriert. Die Transistoren verfügen über individuelle Gate-, Source- und Drain-Kontakte, sowie über einen gemeinsamen Wannenschluss (Well). Eine Besonderheit besteht in der Herstellung der Gate-Kontakte. Einzelne Gate-Finger mit den hier beschriebenen Strukturbreiten können mit den abbildenden Verfahren der Lithografie nicht hergestellt werden. Daher wurde das Layout entsprechend angepasst und mehrere Gate-Finger parallel mit einem Abstand von 75 nm angeordnet. Diese Gate-Finger werden als Dummy-Gate bezeichnet und sind untereinander elektrisch verbunden. Sie werden in jedem Makro zu einem gemeinsamen Kontakt geführt. Während aller hier durchgeführten Experimente waren die Dummy-Gates über die Ground-Unit des Parameter-Analyzers geerdet.

5.2 Stresseinprägung durch wiederholte P/L-Zyklen

Die Bits einer NROM-Zelle werden programmiert beziehungsweise gelöscht, indem heiße Elektronen und Löcher in die Nitridschicht injiziert werden. Dabei tritt jedoch nicht nur die gewünschte Verschiebung der Schwellspannung durch die Speicherladungen ein, sondern im Laufe wiederholter P/L-Zyklen auch eine Schädigung des Tunneloxids und dessen Grenzfläche zum Substrat durch heiße Ladungsträger. Diese Schädigung äußert sich in einer Anhäufung von festen Ladungen im Oxid und durch einen Anstieg der Grenzflächenzustandsdichte [21, 60, 159, 260]. Jede Anwendung einer Programmier- oder Löschoperation auf die Speicherzelle stellt somit Stress dar und führt zu einer Degradation der Transistoreigenschaften. Die NROM-Zelle

verliert über lange Sicht an Leistungs- und Funktionsfähigkeit. Hierzu zählen vor allem der Ladungsverlust durch das Tunneloxid über störstellenunterstütztes Tunneln [261, 262].

Dieser Abschnitt stellt den detaillierten Ablauf der Programmier- und Löschoptionen vor, der für die vorliegenden Teststrukturen definiert und angewendet wurde. Wie bereits in Abschnitt 2.5.2 beschrieben, ist der MOSFET, aus dem die NROM-Zelle aufgebaut ist, symmetrisch bezüglich Source- und Drain-Kontakt. Im Folgenden wird mit dem Drain stets der Kontakt bezeichnet, der neben dem zu programmierenden Bit liegt. Der Source-Kontakt liegt demnach am entgegengesetzten Ende des Kanals in der Nähe des nicht zu programmierenden Bits.

Jedes Bit wird über eine Serie von Spannungspulsen am Drain-Kontakt programmiert beziehungsweise gelöscht. Dabei wird die Amplitude der Spannungspulse sukzessive erhöht. Sobald die gewünschte Schwellspannungsverschiebung erreicht ist, wird die Serie abgebrochen. Die Spannungen am Source-, Gate- und Well-Kontakt werden während der Pulse auf konstanten Werten gehalten.

Programmieren

Zum Programmieren eines Bits der NROM-Zelle werden Spannungspulse mit einer Länge von 250 ns verwendet. Am Source-Kontakt liegt eine konstante Spannung von 0,3 V und am Gate eine Spannung von 9,0 V an. Der Well-Kontakt sowie die Dummy-Gates sind geerdet. Die Pulsamplitude beträgt am Anfang der Programmieroperation 3,0 V. Nach jedem Puls wird der Drain-Strom im *reverse read mode* bei angelegter Lesespannung von $V_S = 1,3$ V und bei einer definierten Gate-Spannung $V_G = V_{t,p}$ gemessen. Aufgrund der negativen injizierten Ladung verschiebt sich die Transferkennlinie mit jedem Puls schrittweise zu positiven Spannungen.

Abbildung 5.1 zeigt schematisch die Transferkennlinien während einer Programmieroperation. Im unprogrammierten Zustand ergibt sich eine Schwellspannung $V_{t,0}$, die bei einem konstanten Drain-Strom $I_{D,t}$ bestimmt wird. Die am Gate angelegte Spannung $V_{t,p}$ bezeichnet die angestrebte Schwellspannung, die im vollständig programmierten Zustand des Bits erreicht sein soll. Als Entscheidungskriterium dafür dient wiederum der zuvor definierte Schwellstrom $I_{D,t}$. Nach jedem Programmierpuls nimmt der bei $V_G = V_{t,p}$ gemessene Drain-Strom ab. Unterschreitet er den Schwellstrom, ist die Transferkennlinie ausreichend weit zu positiven Spannungen verschoben. Die Serie von Spannungspulsen wird dann abgebrochen und die Programmierung des Bits ist abgeschlossen. In dieser Arbeit wird ein Bit als programmiert definiert, sobald die Schwellspannung im *reverse read mode* den Wert $V_{t,p} = V_{t,0} + 2,0$ V überstiegen hat.

Löschen

Das Löschen eines Bits der NROM-Speicherzelle erfolgt analog zum Programmieren. Jedoch sind die Spannungspulse mit 100 ms hier deutlich länger. Die Spannung am Source-Kontakt wird beim Löschen auf konstant 2,0 V gesetzt. Dadurch wird der Potenzialabfall entlang des

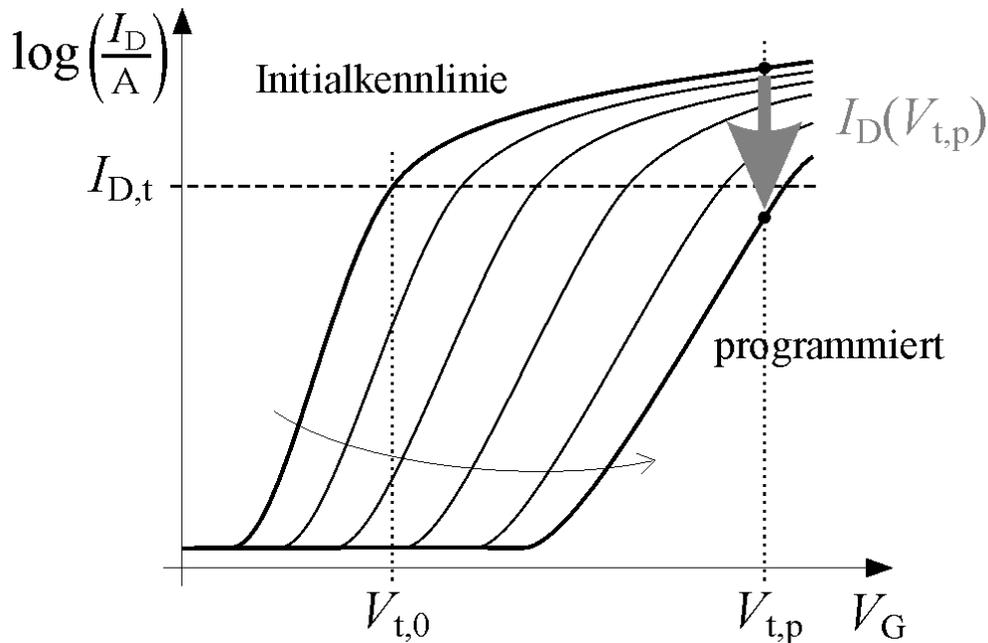


Abbildung 5.1: Schematische Darstellung der Verschiebung der Transferkennlinie einer NROM-Zelle während gepulster Programmieroperationen. Der Drain-Strom $I_D(V_{t,p})$ an der zu programmierenden Schwellspannung $V_{t,p}$ nimmt mit zunehmender Verschiebung der Kennlinie ab. Die Zelle ist programmiert, wenn $I_D(V_{t,p})$ unter einen definierten Schwellstrom $I_{D,t}$ fällt.

Kanals reduziert und eine Generation von heißen Löchern auf der Source-Seite verhindert, die den Zustand des entgegengesetzten Bits beeinflussen würden. Am Gate wird eine konstante Spannung von $-7,0\text{ V}$ angelegt und die Amplitude der Drain-Spannungspulse beträgt am Anfang der Löschoption $3,0\text{ V}$. Durch diesen hohen Potenzialunterschied zwischen Gate und Drain setzt eine ausreichend hohe Bandverbiegung am Drain-Kontakt ein, so dass Band-zu-Band-Tunneln stattfinden kann.

Nach jedem Puls wird wiederum der Drain-Strom im *reverse read mode* bei angelegter Lesespannung von $V_S = 1,3\text{ V}$ gemessen. Die Gate-Spannung beträgt hierbei jedoch $V_G = V_{t,g}$. Bei dieser Spannung liegt der Drain-Strom in der Größenordnung des Leckstroms weit unterhalb des Schwellstroms $I_{D,t}$.

Abbildung 5.2 zeigt schematisch die Rückverschiebung der Transferkennlinie während einer Löschoption zu niedrigeren Spannungen. Der gemessene Drain-Strom bei $V_G = V_{t,g}$ steigt nach jedem Spannungspuls an. Wird der Schwellstrom überschritten, ist die Transferkennlinie weit genug zurückverschoben. Das entsprechende Bit gilt als gelöscht und die Löschoption bricht ab. Bei den in dieser Arbeit durchgeführten Löschoptionen wurden die Bits nie vollständig in den Ausgangszustand zurückgesetzt. Das entsprechende Bit wird als gelöscht definiert, sobald die Schwellspannung im *reverse read mode* unter den Wert $V_{t,g} = V_{t,0} + 0,5\text{ V}$ gefallen ist. Auf diese Weise wird ein Programmierfenster von mindestens $\pm 1,5\text{ V}$ erzielt.

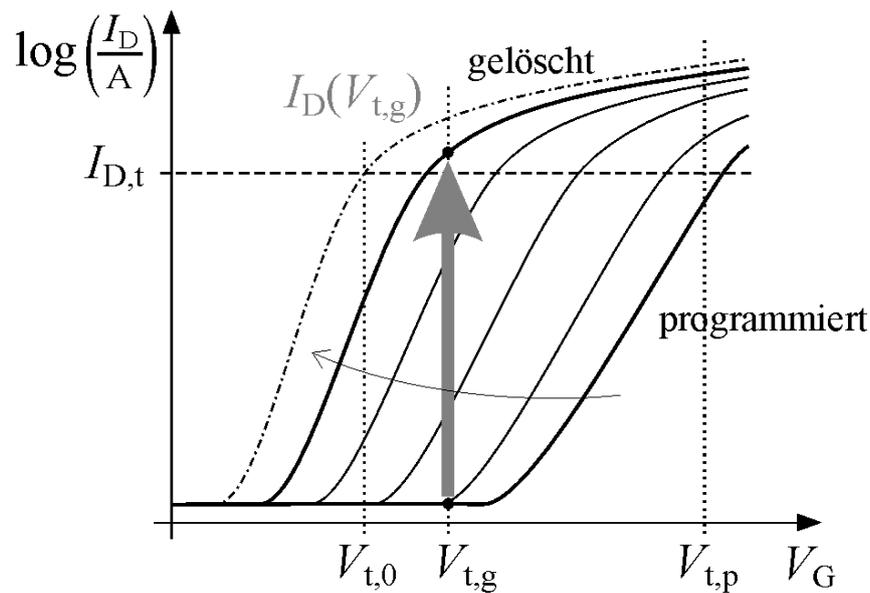


Abbildung 5.2: Schematische Darstellung der Verschiebung der Transferkennlinie einer NROM-Zelle während gepulster Löschoptionen. Der Drain-Strom $I_D(V_{t,g})$ steigt mit zunehmender Verschiebung der Kennlinie an. Die Zelle gilt als gelöscht, sobald $I_D(V_{t,g})$ größer als der definierte Schwellstrom $I_{D,t}$ ist.

Wiederholtes Programmieren und Löschen

Das zuvor beschriebene Programmieren und Löschen wurde mithilfe der Software LabVIEW in einem Computerprogramm realisiert. Dies ermöglichte die automatische Ausführung von aufeinanderfolgenden Programmier- und Löschoptionen, wodurch der Transistor gestresst wird. Ein solcher Zyklus besteht dabei aus dem Programmieren und anschließenden Löschen eines Bits. Um die einzelnen Operationen zu beschleunigen, wurde die Amplitude des Drain-Spannungspulses nach jedem Puls um $+0,1$ V erhöht. Die anfänglichen Amplituden wurden bewusst zu niedrig gewählt, so dass beim ersten Puls noch nicht genügend heiße Ladungsträger zum vollständigen Programmieren beziehungsweise Löschen der Zelle generiert werden. Auf diese Weise tastet sich der Algorithmus von unten an die am Drain-Kontakt notwendigen Programmier- und Löschoptionen heran, wodurch eine Überprogrammierung beziehungsweise ein Überlösen verhindert wird. Dies wird auch als Stepping-Algorithmus bezeichnet [228].

Zusätzlich wird jeweils der Spannungswert der vorletzten Amplitude gespeichert. Wird das Bit nach einer vorangegangenen Löschoption erneut programmiert, startet die Spannungspulsserie am Drain mit der zuvor gespeicherten Amplitude. Hierdurch konnte der Prozess des Herantastens an die entsprechende Programmierspannung etwas verkürzt werden. Analog wurde dasselbe Verfahren auf die Pulsserie beim Löschen angewendet. Auf diese Weise konnten bis zu 280 P/L-Zyklen pro Stunde realisiert werden. Dies ermöglicht 10000 Zyklen in einer überschaubaren Zeit von etwa 1,5 Tagen.

In Anhang B.1 sind die maximalen Amplituden der Spannungspulse, die zum Programmieren beziehungsweise Löschen notwendig waren, in Abhängigkeit der Zyklenzahl dargestellt.

5.3 Untersuchungen des $1/f$ -Rauschens

Für die $1/f$ -Rauschmessungen wurden NROM-Zellen mit unterschiedlich vielen Zyklen gestresst und anschließend das $1/f$ -Rauschen analysiert. Dazu ist der Messaufbau aus Abbildung 4.16 verwendet worden. Die Spannungspulse zum Programmieren beziehungsweise Löschen wurden durch den integrierten Pulsgenerator generiert, welcher in der Lage ist, in Abhängigkeit eines Triggers Spannungspulse einzeln zu erzeugen. Durch diese Funktion konnte das abwechselnde Pulsen und Bestimmen der Schwellspannung während einer Programmier- beziehungsweise Löschoperation mit dem bestehenden Messaufbau realisiert werden. Zusätzlich konnten mit dem Aufbau die Rauschmessungen durchgeführt werden. Die Drain-Strommessungen wurden mit dem Parameter-Analyzer, das Umschalten zwischen Pulsgenerator und Parameter-Analyzer mit der Schaltmatrix durchgeführt.

Für die nachfolgend beschriebenen Rauschmessungen wurde eine Zeitauflösung von $\Delta t = 50$ ms gewählt. Es wurden 10001 Messwerte aufgezeichnet, so dass die gesamte Messdauer 500 s betrug. Daraus ergibt sich nach einer FFT die Spektraldichte des Rauschens im Frequenzbereich von 2 mHz bis 10 Hz.

Abbildung 5.3 zeigt eine typische Messung des Drain-Stroms im Bereich von 0 bis 500 s bei $V_D = 20$ mV. Diese wurde an einer ungestressten Zelle mit 3/6/9-ONO durchgeführt. Der Transistor wurde hierbei im linearen Bereich mit $V_G = V_t + 0,6$ V betrieben, in dem sich ein für die spätere Normierung der Spektraldichte benötigter Strommittelwert von $\bar{I}_D = 204,4$ nA ergab. Über eine FFT wurde die Spektraldichte bestimmt und anschließend normiert.

Für drei verschiedene Stresszustände ist in Abbildung 5.4 die normierte spektrale Dichte S'_{I_D} des

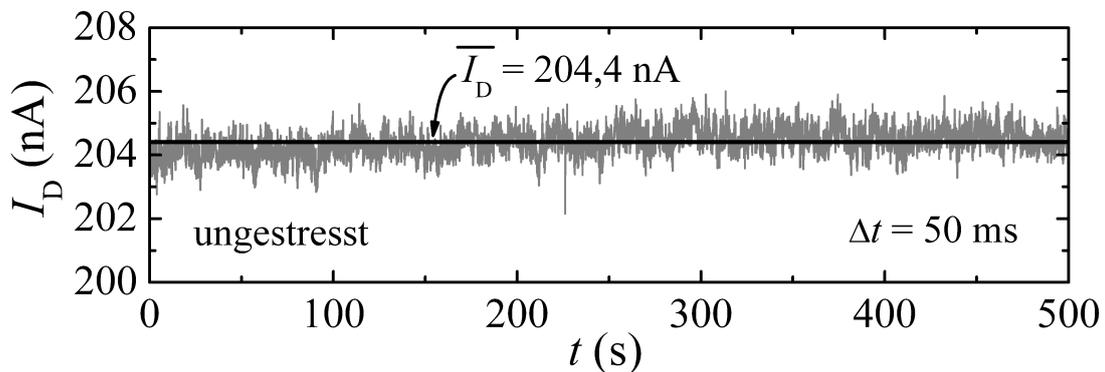


Abbildung 5.3: Typische Drain-Strommessung über der Zeit. Die Messwerte wurden im Bereich von 0 bis 500 s mit einer Zeitauflösung von 50 ms bei einer Drain-Spannung von $V_D = 20$ mV aufgenommen. Der Strommittelwert beträgt $\bar{I}_D = 204,4$ nA.

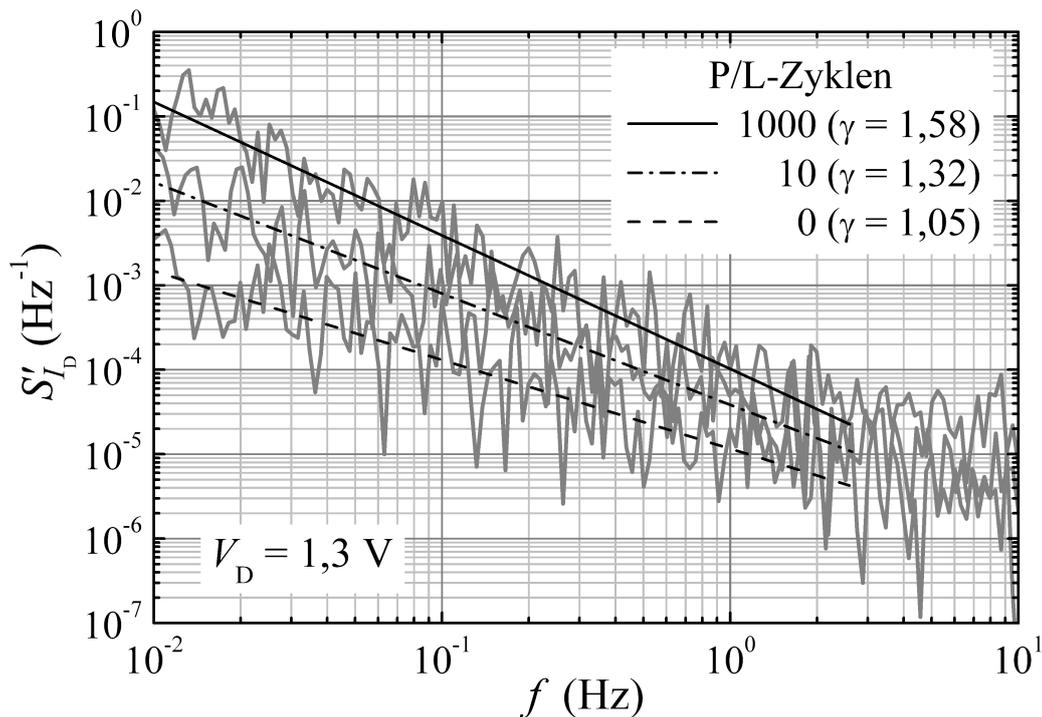


Abbildung 5.4: Rauschspektren in Abhängigkeit der Zyklenzahl im *reverse read mode* bei $V_D = 1,3 \text{ V}$. Mit zunehmender Zyklenzahl steigt die absolute Spektraldichte an. Die Steigung der Spektren steigt ebenfalls von $\gamma \simeq 1,05$ für die ungestresste Speicherzelle über $\gamma \simeq 1,32$ nach 10 Zyklen auf $\gamma \simeq 1,58$ nach 1000 Zyklen an.

Drain-Stromrauschen im Bereich von 0,01 Hz bis 10 Hz dargestellt. Um die Messzeiten bei den jeweiligen Stresszuständen möglichst gering zu halten, wurde jede Drain-Strommessung nur einmal durchgeführt, so dass anschließend nicht über mehrere Spektren gemittelt werden konnte. Daraus resultieren die starken Schwankungen in den dargestellten Spektren. Die Integrationszeit des Parameter-Analyzers konnte bei diesen Messungen maximal auf 20 ms eingestellt werden, welches nur 40% des Zeitintervalls Δt zwischen zwei Messpunkten entspricht. Dadurch macht sich hier der Aliasing-Effekt am oberen Ende des gemessenen Frequenzbereiches für $f \gtrsim 3 \text{ Hz}$ bemerkbar, in dem die Spektraldichte scheinbar erhöht ist. In diesem Fall verlaufen die Spektren dort annähernd horizontal und folgen nicht mehr dem erwarteten $1/f^\gamma$ -Verhalten (vergleiche Abbildung 3.9b). Für die nachfolgende Anpassung der Spektren durch die Form $S'_{ID} = S'_{ID,1} f^{-\gamma}$ wurde daher nur der Bereich von 0,01 Hz bis 2,5 Hz gewählt. Als Ergebnis zeigt sich in Abbildung 5.4 ein deutlicher Anstieg der absoluten Spektraldichte mit zunehmender Zahl der Programmier- und Löschzyklen. Durch die heißen Ladungsträger beim Programmieren und Löschen werden Grenzflächenzustände generiert [159, 260]. Es konnte in der Literatur gezeigt werden, dass eine Zunahme der Grenzflächenzustände auch einen Anstieg des niederfrequenten Rauschens bewirkt [254–256]. Der in dieser Arbeit verwendete Experimentaufbau in Kombination mit den zur Anwendung kommenden Auswerteverfahren

ist demnach prinzipiell geeignet, die Degradation einer NROM-Zelle durch Programmier- und Löschyklen zu detektieren.

Gleichzeitig nimmt bei den in Abbildung 5.4 gezeigten Ergebnissen die Steigung γ der Spektren zu. Bei der ungestressten Speicherzelle beträgt $\gamma \simeq 1,05$, nach 10 Zyklen 1,32 und nach 1000 Zyklen 1,58. Die Ursache hierfür liegt in der extrem kleinen Gate-Fläche der untersuchten MOSFETs ($A_G \simeq 9 \cdot 10^{-11} \text{ cm}^2$) mit einer sehr geringen absoluten Zahl an Inversionsladungsträgern. In diesen Dimensionen können Einzelelektroneneffekte sichtbar sein, die sich in diskreten Sprüngen im Drain-Strom äußern. Durch Tunneln eines einzelnen Ladungsträgers in eine Störstelle nahe der Grenzfläche wird die Ladungsträgerzahl im Kanal um Eins herabgesetzt. Bei der sehr geringen absoluten Zahl von Ladungsträgern im Kanal (~ 100) bewirkt dies eine signifikante Änderung im Drain-Strom. Gleichzeitig kann durch den eingefangenen Ladungsträger ein Streuzentrum für die übrigen Kanalladungsträger entstehen, das eine Änderung der Mobilität und damit wiederum eine Änderung des Drain-Stroms bewirkt [263, 264]. Beide Effekte addieren sich zu einer Drain-Stromänderung ΔI_D , die einige Prozent betragen kann [265]. Dies ist als Random Telegraph Signal (RTS) des Drain-Stroms messbar [266]. Aus der Rauschanalyse des RTS ergibt sich ein Lorentzspektrum, das zu hohen Frequenzen mit $1/f^2$ abfällt (siehe Kapitel 3.3) [267]. Zwischen reinem $1/f$ -Rauschen und RTS können sich durch Überlagerung Spektren mit Exponenten γ zwischen 1 und 2 ergeben [268], welches in den hier gezeigten Ergebnissen der Fall ist. Allgemein werden RTS Fluktuationen hauptsächlich beobachtet, wenn der Transistor im Unterschwellbereich arbeitet, während sich die Drain-Stromänderungen in Inversion eher als $1/f$ -Rauschen manifestieren [269]. Daher wurden die NROM-Zellen während der Rauschmessungen in den folgenden Experimenten in Inversion betrieben.

Der beim Programmieren und Löschen eines Bits wirkende Stress wird, wie bereits erwähnt, durch heiße Ladungsträger hervorgerufen, die überwiegend an den Kanalenden nahe der Source- und Drain-Kontakte entstehen. Daher ist zu erwarten, dass auch die damit verbundene Schädigung der Grenzfläche und des Tunneloxids verstärkt in diesen Bereichen des Kanals auftritt. Aus diesem Grund ist es für Zuverlässigkeitsuntersuchungen besonders interessant, wenn der Grad der Schädigung mittels elektrischer Messungen lokal aufgelöst werden könnte. Eine solche Methode wäre in der Lage, die Länge der geschädigten Region entlang des Transistorkanals zu bestimmen. Die resultierende Frage hierbei wäre, ob der Stress auf die Region des jeweiligen Bits beschränkt ist oder ob er auf den gesamten Kanal ausgeweitet ist. Im Folgenden werden zwei Ansätze verfolgt, eine Lokalisierung der Schädigung mittels $1/f$ -Rauschmessungen zu realisieren.

Der erste Ansatz nutzt den *reverse read mode*, wie er auch zum Auslesen eines Bits der NROM-Zelle verwendet wird. Hierbei finden die Rauschmessungen bei der entsprechenden Lese-Spannung von $V_D = 1,3 \text{ V}$ statt. Der Drain-Strom hängt in diesem Modus empfindlich von der Zahl der gespeicherten Ladungen in dem Bit ab, das sich nahe der Source befindet. Jede Änderung der Ladungszahl würde eine entsprechende Stromfluktuation hervorrufen, die sich im Rauschen des Drain-Stroms bemerkbar macht. Dies trifft nicht nur auf Änderungen der Speicherladung zu, sondern auch auf Kanalladungsträger, die mit Grenzflächenzuständen beziehungsweise grenz-

flächennahen Oxidstörstellen (NIOTs) durch Tunnelprozesse kommunizieren [270, 271]. Im Folgenden werden diese Ereignisse auch als Rauschquellen bezeichnet. Vergleichbare Ladungsträgerfluktuationen nahe des Drains haben dagegen einen sehr geringen Einfluss auf den Drain-Strom und schlagen sich kaum im gemessenen Rauschen nieder [272–274].

Im zweiten Ansatz zur lokalisierten Detektion der Schädigung wird die Programmierfähigkeit des Transistors genutzt. Werden negative Ladungen durch die Programmierung eines Bits in die Nitridschicht injiziert, wird das Oberflächenpotenzial in dem darunter liegenden Kanalbereich durch die elektrostatische Wechselwirkung erhöht. Die lokale Schwellspannung steigt in dieser Region an, so dass sich der Transistor dort im Unterschwellbereich befindet. In der Literatur wurde gezeigt, dass die normierte Spektraldichte des $1/f$ -Rauschens im Unterschwellbereich maximal ist und im Inversionsbereich um mehrere Größenordnungen abfällt [205]. Analog zum ersten Ansatz wird hierdurch die Wirkung von fluktuierenden Ladungen auf das gemessene Rauschen erhöht, während Beiträge aus anderen Kanalbereichen einen verhältnismäßig geringen Einfluss haben. Im Unterschied zum ersten Ansatz kann die Messung des Drain-Stromrauschens und somit die Detektion der Schädigung jedoch bei einer niedrigeren Drain-Spannung als der Lesespannung gemessen werden. Der Transistor arbeitet hierbei im linearen Bereich und der Inversionskanal ist weitgehend homogen.

In den folgenden zwei Abschnitten werden Ergebnisse dieser beiden Ansätze dargestellt und diskutiert.

5.3.1 Rauschen in Abhängigkeit der Drain-Source-Spannungspolarität

Bei diesem Experiment wird die Spektraldichte des $1/f$ -Rauschens in Abhängigkeit der Polarität der Drain-Source-Spannung V_{DS} untersucht. Als Maß für das Rauschen wird wiederum die Spektraldichte $S'_{ID,1}$ bei 1 Hz herangezogen, die sich aus der Näherung der ermittelten Spektren durch die Form $S'_{ID} = S'_{ID,1} f^{-\gamma}$ ergibt. Da die hier durchgeführten Drain-Strommessungen zur Rauschanalyse in Anlehnung an den Lese-Modus der NROM-Zelle durchgeführt werden, werden die beiden Polaritäten der Drain-Spannung als *forward mode* beziehungsweise *reverse mode* bezeichnet. Die Begriffe *forward* und *reverse* beziehen sich dabei stets auf das gestresste Bit des NROM-Speichers.

Für das Experiment wurden zwei NROM-Zellen mit 3/6/9-ONO verwendet. In beiden Speicherzellen wurde jeweils ein Bit mit 1000 P/L-Zyklen gestresst. Der Stress wurde an mehreren Stellen unterbrochen, um Rauschanalysen durchzuführen. Dazu wurde der Drain-Strom bei einem Transistor mit $V_D = 1,3$ V sowohl im *forward mode* als auch im *reverse mode* gemessen. Die Gate-Spannung war auf $V_G = V_t + 1,5$ V eingestellt, um den Transistor in Inversion zu betreiben. Wegen der hohen Drain-Spannung und der sehr kurzen Kanallänge bildet sich trotz der hohen Gate-Spannung ein stark inhomogener Kanal mit hohem elektrischen Feld nahe des Drains aus.

An einem zweiten Transistor wurde dasselbe Experiment durchgeführt, wohingegen die Drain-Strommessung zur Rauschanalyse bei $V_D = 20$ mV durchgeführt wurde. Die Gate-Spannung

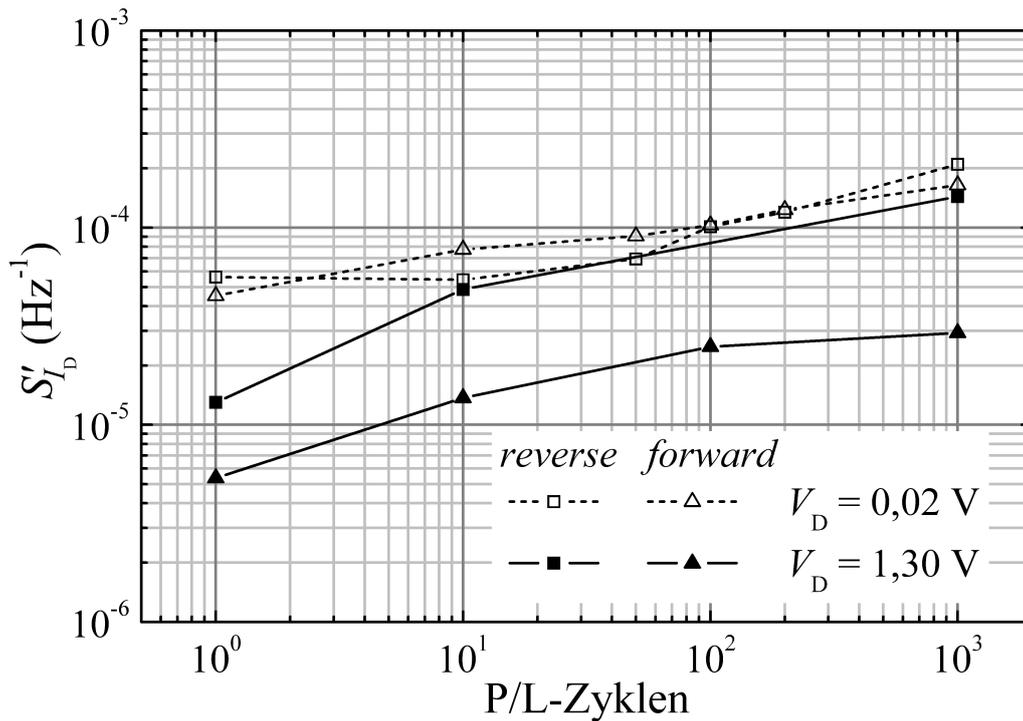


Abbildung 5.5: Normierte Spektraldichte $S'_{I_{D,1}}$ bei 1 Hz in Abhängigkeit der Zyklenzahl und der Drain-Source-Spannungspolarität. Die Messungen wurden sowohl bei kleiner Drain-Spannung von $V_D = 20$ mV als auch bei hoher Drain-Spannung von $V_D = 1,3$ V durchgeführt. Eine Abhängigkeit der Spektraldichte von der Spannungspolarität zeigt sich nur unter hohen Drain-Spannungen, welches durch eine inhomogene Rauschquellendichte entlang des Transistorkanals erklärt werden kann.

wurde hierbei auf $V_G = V_t + 0,6$ V gesetzt, bei der der Inversionskanal weitgehend homogen ist.

Abbildung 5.5 zeigt die normierte Spektraldichte $S'_{I_{D,1}}$ in Abhängigkeit der Zyklenzahl. Alle vier Kurven steigen mit zunehmender Zykelzahl an, welches einen generellen Anstieg der Rauschquellen, also der Dichte von Grenzflächenzuständen beziehungsweise grenzflächennahen Oxidstörstellen bedeutet.

Betrachtet man die Kurven untereinander, so zeigt sich für die kleine Drain-Spannung von $V_D = 20$ mV kein signifikanter Unterschied in der spektralen Dichte zwischen *forward mode* und *reverse mode*. Bei dieser Drain-Spannung befindet sich der Transistor im linearen Bereich und die Inversionsschicht des Kanals ist weitgehend homogen. Auftretende Ladungsträgerfluktuation, die durch Tunneln von Ladungsträgern zwischen dem Kanal und vorhandenen Grenzflächenzuständen beziehungsweise NIOTs hervorgerufen werden, beeinflussen das Rauschen gleichermaßen, unabhängig von ihrer lateralen Position zwischen Source und Drain.

Die beiden Kurven für die hohe Drain-Spannung von $V_D = 1,3$ V zeigen dagegen einen deutlichen Unterschied im absoluten Wert der Spektraldichte, der mit zunehmender Zyklenzahl zudem leicht größer wird. Unter dieser Drain-Spannung ist der Transistorkanal nicht mehr als

homogen zu betrachten. In der Region mit dem hohen elektrischen Feld, die sich nahe dem Kontakt mit der hohen positiven Spannung ausbildet, herrscht an der Grenzfläche eine geringere Stromdichte.

Im Zusammenhang mit RTS Untersuchungen wurde gezeigt, dass die Amplitude von Stromfluktuationen proportional zum Quadrat der Stromdichte ist [275]. Daher beeinflussen Ladungsträgerfluktuationen den Drain-Strom in dieser Region nur geringfügig und ihr Beitrag zum Drain-Stromrauschen ist stark unterdrückt. Im Gegensatz dazu trägt der Kanalbereich auf der anderen Seite, die hierbei als Source fungiert, sehr empfindlich zum Rauschen bei. Dies wurde auch in Simulationen gezeigt [274] und entspricht demselben Effekt, der zum Lesen einer NROM-Zelle ausgenutzt wird [273]. Der zum Rauschen beitragende Kanalbereich wird durch die Hochfeldregion verkürzt und dadurch einige Rauschquellen ausgeblendet, so dass eine geringere Spektraldichte zu erwarten ist. Dieser Zusammenhang wurde auch mit Modellrechnungen sowie Rauschmessungen an Transistoren im Sättigungsbetrieb nachgewiesen [276].

Für das hier gezeigte Experiment liegt die Hochfeldregion unterhalb des gestressten Bits, wenn im *forward mode* gemessen wird. In Abbildung 5.5 liegt genau diese Kurve deutlich niedriger als die im *reverse mode* gemessene. Dies bedeutet, dass im *forward mode* mehr Rauschquellen im Kanal ausgeblendet werden als im *reverse mode*. Andererseits sind die Regionen mit hohem elektrischen Feld in beiden Spannungspolaritäten aufgrund des symmetrischen Aufbaus des Transistors gleich lang. Falls eine homogene Verteilung von Rauschquellen entlang des Kanals vorläge, müssten beide Kurven die gleiche absolute Spektraldichte zeigen. Da diese allerdings einen deutlichen Unterschied aufweisen, der mit zunehmender Zyklenzahl zunimmt, scheint hier eine inhomogene Verteilung von Rauschquellen durch den Stress generiert zu werden.

Aufgrund der Tatsache, dass die Kurve im *forward mode* die geringere Spektraldichte zeigt, müssen vermehrt Rauschquellen nahe des gestressten Bits generiert werden. Diese werden durch die Hochfeldregion im *forward mode* ausgeblendet und tragen nicht zum Rauschen bei. Im restlichen Kanalbereich scheint nur eine geringe Dichte von Rauschquellen vorzuliegen, so dass die gemessene Spektraldichte deutlich unterdrückt ist. Im *reverse mode* dagegen liegen die generierten Rauschquellen auf der entgegengesetzten Seite der Hochfeldregion, auf welcher Ladungsträgerfluktuationen einen empfindlichen Einfluss auf den Drain-Strom und damit auf das Rauschen haben. Der Anteil von Rauschquellen, der bei dieser Messung durch die Hochfeldregion am ungestressten Ende des Kanals ausgeblendet wird, scheint hier verhältnismäßig gering zu sein.

5.3.2 Rauschen in Abhängigkeit des Programmierzustands

Das vorige Experiment zeigte, dass die Quellen des $1/f$ -Rauschens einer NROM-Zelle in der Nähe des gestressten Bits liegen. Zur Messung des Rauschens wurde dazu ein analoges Verfahren wie zum Lesen des Bits verwendet. Der zweite experimentelle Ansatz, der im Folgenden vorgestellt wird, hat ebenfalls die Lokalisierung der Rauschquellen im Transistorkanal zum Ziel. Hierbei wird jedoch ein anderer Messansatz verfolgt.

Die Bits einer NROM-Zelle lassen sich durch die Injektion von Elektronen in die Nitridschicht programmieren. Durch die elektrostatische Wechselwirkung wird das Oberflächenpotenzial für Elektronen unterhalb dieser Region erhöht. Der MOSFET besitzt dort lokal eine höhere Schwellspannung als im übrigen Kanalbereich. Im *reverse read mode* bestimmt diese Region die gesamte Schwellspannung des Transistors, die dann deutlich höher liegt als im gelöschten Zustand. Im *forward read mode* wird diese Zone durch die Pinch-off Region überdeckt und die gesamte Transistorschwellspannung durch den übrigen Kanalbereich bestimmt. Über die Zahl der gespeicherten Ladungsträger in der Nitridschicht kann die lokale Schwellspannung eingestellt werden.

In der Literatur wurde gezeigt, dass die Spektraldichte des Drain-Stromrauschens im Unterschwellbereich maximal ist und in Inversion exponentiell um mehrere Größenordnungen abfällt [205, 211]. Deshalb sollte der Beitrag von Rauschquellen zur gemessenen Spektraldichte in der Region mit erhöhter Schwellspannung maximal sein, während der von Rauschquellen im übrigen Kanalbereich eine untergeordnete Rolle spielt. Dieser Effekt wird hier zur Lokalisierung der Oxidschädigung genutzt. Es sollen die Spektraldichten des Drain-Stromrauschens einer gestressten NROM-Zelle mit verschiedenen stark programmierten Bits bestimmt werden.

Für dieses Experiment wurde eine NROM-Speicherzelle mit 4/6/12-ONO verwendet. Zunächst wurde ein Bit mit 1000 P/L-Zyklen gestresst, um ausreichend Rauschquellen an der darunter liegenden Grenzfläche und im Tunneloxid zu erzeugen. Danach wurde dasselbe Bit mithilfe des in Kapitel 5.2 beschriebenen Programmieralgorithmus in verschiedene Programmierzustände versetzt. Der Grad der Programmierung wurde im *reverse read mode* mit $V_D = 1,3 \text{ V}$ anhand der Schwellspannungsverschiebung $\Delta V_t = V_t - V_{t,0}$ bestimmt. Hierbei ist $V_{t,0} = 4,2 \text{ V}$ die für das 4/6/12-ONO bestimmte Schwellspannung bei $V_D = 20 \text{ mV}$ und $I_{D,t} = 10 \text{ nA}$. Die erzeugten Programmierzustände lagen im Bereich $0 \leq \Delta V_t \leq 1,5 \text{ V}$. Für jeden Programmierzustand wurden analog zum ersten Experiment Drain-Strommessungen im *reverse mode* und im *forward mode* sowie eine anschließende FFT durchgeführt, um die Spektraldichte des Rauschens zu bestimmen. Die Drain-Spannung betrug dabei $V_D = 0,2 \text{ V}$ und die Gate-Spannung wurde auf $V_G = V_{t,0} + 0,9 \text{ V}$ eingestellt.

Zum Vergleich der Ergebnisse wurden die Messungen anschließend an demselben Transistor wiederholt, allerdings wurde hierbei das andere, ungestresste Bit schrittweise programmiert.

Die normierte Spektraldichte $S'_{I_D,1}$ wurde analog zum ersten Experiment für alle oben beschriebenen Messungen bestimmt und ist in Abbildung 5.6 über dem Grad der Programmierung ΔV_t dargestellt. Das Bit 2 ist das zuvor mit 1000 P/L-Zyklen gestresste Bit. Im Folgenden beziehen sich die Bezeichnungen *reverse* und *forward* auf das jeweils programmierte Bit und nicht wie im vorigen Experiment auf das gestresste Bit. Die Messung im *reverse mode* an Bit 2 (gefüllte Dreiecke) zeigt eine deutlich erhöhte Spektraldichte im Gegensatz zu den anderen Messkurven. Bei dieser Messung liegt der durch die Programmierladungen erzeugte Bereich mit erhöhter Schwellspannung in der Region des gestressten Bits 2. Dieser Fall ist in Abbildung 5.7a schematisch dargestellt.

Der Transistorkanal ist in zwei Bereiche A und B eingeteilt. Im Bereich A liegt die lokale

Schwelspannung unverändert bei $V_{t,0}$, im Bereich B dagegen ist sie aufgrund der negativen Ladungen in Bit 2 größer als $V_{t,0}$. Die Rauschquellen im Tunneloxid beziehungsweise an der Grenzfläche zum Substrat, die durch die P/L-Zyklen generiert wurden, sind durch Kreuze (\times) gekennzeichnet und liegen in der Region B, in der die lokale Schwelspannung höher ist als im übrigen Kanalbereich. Ihr Beitrag zur gemessenen gesamten Spektraldichte ist maximal, da sich der Transistor hier im Unterschwellbereich befindet. Im *forward mode* wird zwar eine etwas geringere Spektraldichte gemessen (leere Dreiecke in Abbildung 5.6), jedoch liegt diese für große ΔV_t trotzdem höher als die beiden übrigen Kurven, bei denen Bit 1 schrittweise programmiert wurde (Quadrate).

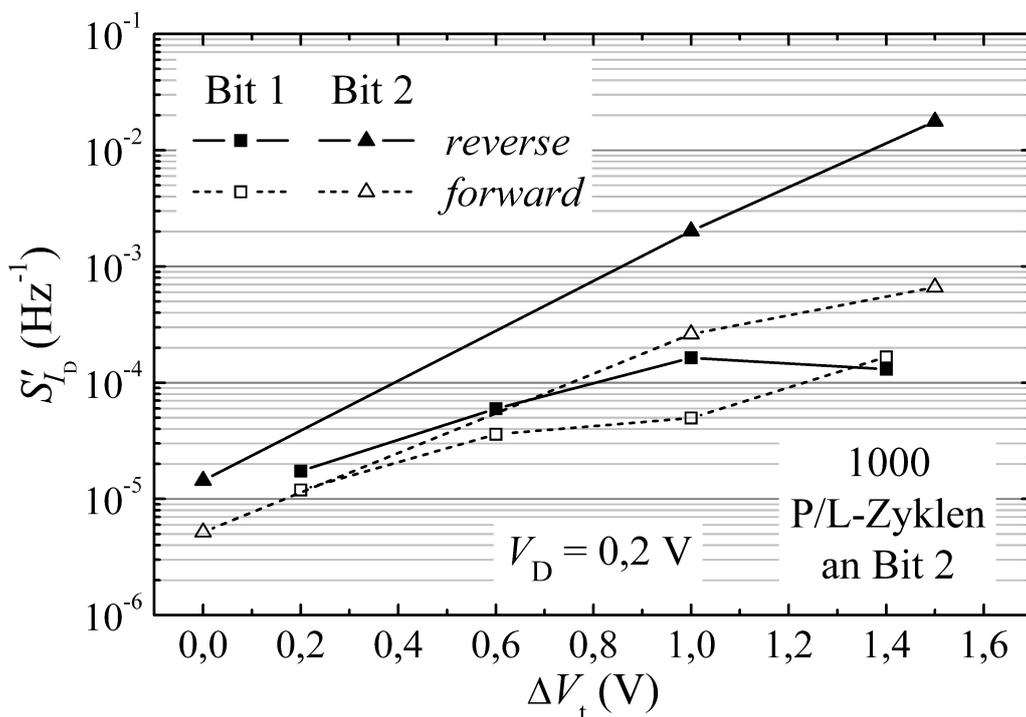


Abbildung 5.6: Normierte Spektraldichte $S'_{I_D,1}$ bei 1 Hz in Abhängigkeit des Programmierzustands ΔV_t eines Bits. Wird das zuvor gestresste Bit 2 programmiert (Dreiecke), liegt die gemessene Spektraldichte deutlich höher als bei der Programmierung des ungestressten Bits 1 (Quadrate). Zusätzlich ist ein Unterschied zwischen *forward mode* und *reverse mode* sichtbar, ähnlich wie in Abbildung 5.5 bei $V_D = 1,3$ V.

Wird statt dem Bit 2 das Bit 1 programmiert, liegt der Bereich mit erhöhter Schwelspannung auf der entgegengesetzten Seite der Rauschquellen. Dieser Fall ist in Abbildung 5.7b dargestellt. Die Rauschquellen liegen im Bereich B mit $V_t = V_{t,0}$ und haben daher nur einen vernachlässigbar kleinen Einfluss auf die gemessene Spektraldichte. Abbildung 5.6 zeigt für diesen Fall eine deutlich verringerte Spektraldichte sowohl im *reverse mode* als auch im *forward mode* (Quadrate). Dieses Ergebnis bestätigt auf gleiche Weise wie das erste Experiment, dass die durch

P/L-Zyklen generierten Rauschquellen in der Nähe des gestressten Bits liegen. In der übrigen Kanalregion ist die Zahl der Rauschquellen offenbar gering.

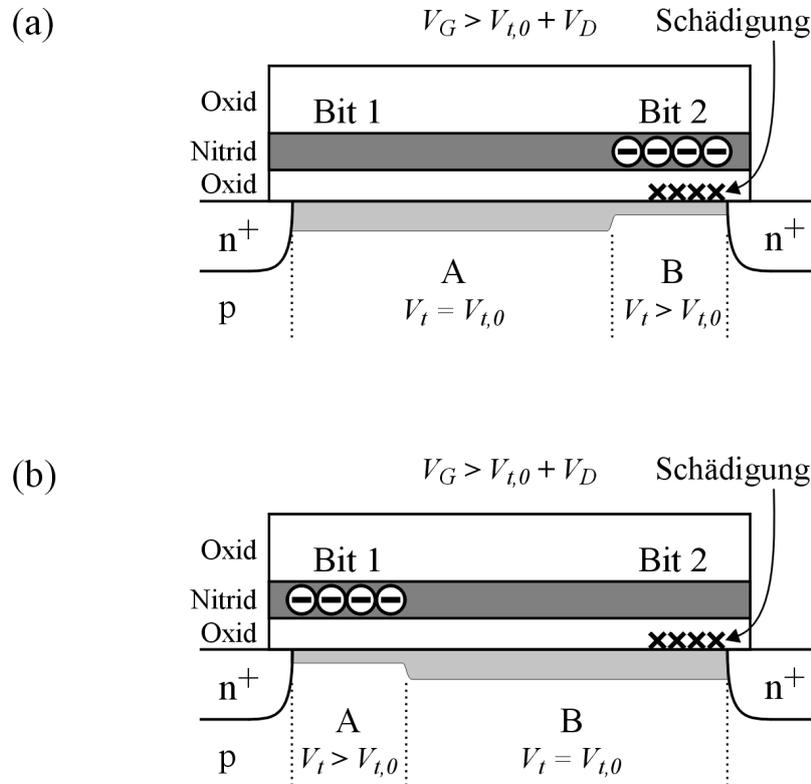


Abbildung 5.7: Darstellung einer gestressten NROM-Zelle mit gleichzeitig programmiertem Bit während der Rauschmessungen. (a) Das programmierte Bit 2 liegt oberhalb der Region mit geschädigtem Tunneloxid. In dieser Region tragen Rauschquellen besonders effektiv zur gemessenen Spektraldichte bei. (b) Bit 1 wurde programmiert, während Bit 2 gelöscht ist. Die Rauschquellen liegen auf der entgegengesetzten Seite und haben nur einen sehr geringen Einfluss auf die gemessene Spektraldichte.

Mit zunehmendem Grad der Programmierung nimmt auch die Stärke des Beitrages von Rauschquellen zur gemessenen Spektraldichte zu. Dies zeigt sich in Abbildung 5.6 dadurch, dass alle Kurven mit zunehmendem ΔV_t ansteigen. Für den Fall, in dem das gestresste Bit 2 programmiert wird, ist dieser Anstieg erwartungsgemäß größer als bei Bit 1.

Des Weiteren fällt auf, dass es einen Unterschied in der gemessenen Spektraldichte zwischen *forward mode* und *reverse mode* gibt. Dieser wurde unter den hier verwendeten Messbedingungen nicht erwartet, denn im Gegensatz zum Lese-Verfahren wurde der Drain-Strom nicht in Sättigung, sondern im linearen Arbeitsbereich des Transistors gemessen. Allerdings ist die verwendete Drain-Spannung von $V_D = 0,2 \text{ V}$ für die hier vorliegenden Kanallängen von $L_G = 115 \text{ nm}$ recht hoch, so dass die Inversionsschicht (Bereich A in Abbildung 5.7a) nicht mehr näherungsweise als homogen angesehen werden kann. Daher scheint in den hier gezeigten Ergebnissen

zusätzlich die Asymmetrie zwischen *forward mode* und *reverse mode*, die im ersten Experiment genutzt wurde, effektiv zu sein. Dies könnte auf der anderen Seite aber auch bedeuten, dass die Zone der Schädigung durch die heißen Ladungsträger beim Programmieren beziehungsweise Löschen weiter in den Kanal hinein reicht, als der Bereich der negativen Programmierladung (Bereich B in Abbildung 5.7a).

5.4 Simulation einer NROM-Zelle

Im vorangegangenen Kapitel wurden experimentelle Ergebnisse von Rauschanalysen an NROM-Speicherzellen vorgestellt. Diese zeigten in einem ersten Experiment eine deutliche Abhängigkeit der gemessenen Spektraldichte von der Drain-Spannungspolarität bei stark inhomogenen Inversionskanälen. In einem zweiten Experiment wurde die Abhängigkeit der Spektraldichte vom jeweiligen Programmierzustand eines Bits einer gestressten Speicherzelle untersucht. Die Ergebnisse und Schlussfolgerungen des ersten Experiments sollen im nachfolgenden Abschnitt durch 3D-Simulationen unterstützt und verifiziert werden.

Ziel der Simulationen ist es, die Auswirkungen eines lokalisierten Elektrons im grenzflächennahen Bereich des Tunneloxids auf den Drain-Strom zu bestimmen. Das Elektron stellt dabei eine besetzte Störstelle im Tunneloxid dar. Über die Simulation wird der Drain-Strom $I_{D,1}$ ermittelt, der in Anwesenheit dieser zusätzlichen Ladung im Oxid durch das Bauelement fließt. Dieser Strom wird anschließend mit dem Drain-Strom $I_{D,0}$ verglichen, der aus einer weiteren Simulation ohne die Elementarladung im Oxid bestimmt wurde. Hieraus wird dann die (negative) relative Änderung des Drain-Stroms

$$-\frac{\Delta I_D}{I_{D,0}} = \frac{I_{D,1} - I_{D,0}}{I_{D,0}} \quad (5.1)$$

berechnet. Diese wird im Folgenden als Drain-Stromempfindlichkeit bezeichnet und ist mit den relativen Amplituden von etwa 1 bis 10 % vergleichbar, die typischerweise in RTS-Untersuchungen ermittelt wurden [275, 277, 278]. Über das Rauschmodell der kombinierten Einflüsse von Beweglichkeitsfluktuationen und der Änderung der Ladungsträgerzahl im Kanal (siehe Abschnitt 3.3.1) kann die relative Stromänderung auch durch

$$\frac{\Delta I_D}{I_{D,0}} = \frac{1}{A_G} \left(\frac{1}{n} \pm \alpha \mu \right) \quad (5.2)$$

ausgedrückt werden [279, 280]. Im Vergleich mit Gleichung 3.39 ergibt sich, dass die normierte Spektraldichte S'_{I_D} des $1/f$ -Rauschens proportional zum Quadrat der relativen Stromänderung ist. Mithilfe dieser einfachen Betrachtung kann die Simulation eines Elektrons im Oxid und der daraus bestimmten relativen Stromänderung qualitativ mit Ladungsträgern verglichen werden, die zwischen Transistorkanal und Oxidstörstellen fluktuieren. Dieser Ansatz wurde bereits in ähnlichen Simulationsstudien gewählt [265, 281].

Die relative Änderung des Drain-Stroms soll in Abhängigkeit der angelegten Drain- und Gate-Spannungen, sowie des lateralen Ortes x_L des Elektrons im Kanal durch Simulation untersucht werden. Als Ausgangspunkt für die Simulationen dient ein zweidimensionales Modell der NROM-Zelle, das durch eine Prozesssimulation generiert und von der Firma Qimonda zur Verfügung gestellt wurde. Dieses Modell weist die gleiche Geometrie wie die im vorangegangenen Abschnitt experimentell untersuchte NROM-Zelle auf. Es enthält zudem die Dotierprofile der Source- und Drain-Gebiete, sowie die des Kanal- und Substratgebietes und spiegelt somit sehr genau die Gegebenheiten der realen Speicherzelle wider. Das Gate-Oxid des Modells ist ein 4/6/12-ONO.

Die zweidimensionale Simulation des Drain-Stroms unter dem Einfluss des Elektrons im Oxid würde zu fehlerhaften Ergebnissen führen. Das Potenzial der zusätzlich eingebrachten Ladung erzeugt in zweidimensionaler Darstellung eine Barriere für alle Kanalladungsträger und reduziert den Stromfluss in Folge dessen übermäßig stark. Das Potenzial entspricht in diesem Fall dem einer Linienladung und folgt anderen Gesetzmäßigkeiten. In der dreidimensionalen Realität jedoch können die Ladungsträger um die punktförmige Potenzialbarriere herum fließen. Daher liefert eine dreidimensionale Simulation realistischere Ergebnisse. Das verfügbare 2D-Modell musste zu diesem Zweck in die dritte Raumdimension erweitert werden.

5.4.1 3D-Simulationen mit ATLAS

Die Simulationen wurden mit dem Bauelementesimulator ATLAS der Firma Silvaco durchgeführt. Dieser ist Bestandteil der Virtuellen Wafer Fab (VWF), in der verschiedene TCAD-Simulatoren und Werkzeuge vereint sind. Der Bauelementesimulator ATLAS löst zum einen die Halbleitergrundgleichungen (Poisson-Gleichung und Kontinuitätsgleichung) für Elektronen und Löcher und zum anderen die vom Modell (hier Drift-Diffusion) abhängigen Transportgleichungen in drei Dimensionen. Dabei kommen die allgemeinen Diskretisierungsalgorithmen sowohl der Finiten-Differenzen- als auch der Finiten-Elemente-Methode zum Einsatz. Eine detaillierte Beschreibung des numerischen Lösungsverfahrens findet sich in [282].

Für die hier durchgeführten Simulationen wurde das zur Verfügung stehende zweidimensionale Modell in die dritte Raumdimension (y -Richtung, vergleiche Abbildung 2.4) erweitert und die Weite dabei auf 75 nm festgelegt. An den Rändern des Simulationsmodells werden automatisch nur die Tangentialkomponenten berechnet, so dass Spiegelsymmetrie vorliegt. Wichtig für die Definition der Struktur ist die Wahl der Vernetzung. Diese muss an den Stellen besonders fein gewählt werden, die für die Simulationsergebnisse den größten Einfluss haben. Bei den hier zu simulierenden Oxidladungen ist dies das Tunneloxid entlang des Kanals und der grenzflächennahe Substratbereich. Da die Punktladungen in Richtung der Transistorweite (y -Richtung) gesehen mittig platziert werden, wurde auch diese Zone feiner vernetzt. Andere Bereiche des Modells wiesen ein gröberes Gitter auf, um die Rechenzeit zu minimieren.

Das zu simulierende Elektron ist im Idealfall punktförmig und stellt somit eine Singularität dar, die nicht direkt im Simulationsmodell platziert werden kann. Dieses Problem wurde durch

die Platzierung eines Ladungsvolumens umgangen, dessen integrale Ladungsmenge Q einer negativen Elementarladung entspricht. Hierzu wurde ein durch benachbarte Gitternetzpunkte definiertes, quaderförmiges Raumelement der Länge $\Delta x = 1,8\text{nm}$, der Weite $\Delta y = 0,9\text{nm}$ und der Höhe $\Delta z = 0,5\text{nm}$ im Tunneloxid ausgewählt und dessen Volumen $V_L = \Delta x \Delta y \Delta z$ berechnet. In diesem Volumen wurde eine konstante Ladungsdichte von

$$\rho_L = -\frac{e}{V_L} \quad (5.3)$$

definiert, wobei e die Elementarladung darstellt. Das Ladungspaket wurde mittig bezüglich der Transistorweite platziert und dehnte sich innerhalb des Tunneloxids im Abstand von 0 bis 0,5 nm zur Substratgrenzfläche aus. Dadurch liegt der Ladungsschwerpunkt in einem Abstand von 0,25 nm von der $Si-SiO_2$ -Grenzfläche entfernt. In aufeinanderfolgenden Simulationen wurde das Ladungspaket an verschiedene Stellen entlang des Transistorkanals zwischen Source und Drain platziert und jedes Mal der Drain-Strom simuliert.

5.4.2 Simulationsergebnisse

Zum Test der Simulation wurde zunächst das Oberflächenpotenzial ψ_s entlang des Transistorkanals simuliert. Dazu wurde der Schwerpunkt des Ladungspakets an drei verschiedene Stellen zwischen Source und Drain bei $x_L/L = 0,116$, $x_L/L = 0,348$ und $x_L/L = 0,832$ platziert. Hierbei bezeichnet $L = 86\text{ nm}$ die effektive Kanallänge, die anhand des zur Verfügung stehenden zweidimensionalen Modells bestimmt werden konnte. Abbildung 5.8 zeigt die Potenzialverläufe sowohl für eine kleine Drain-Spannung von $V_D = 0,02\text{ V}$ als auch für eine große Drain-Spannung von $V_D = 1,30\text{ V}$. Das Potenzial ψ_s ist in der Abbildung so dargestellt, dass es nach unten zunimmt. Aus Sicht der Elektronen bedeutet dies, dass die potenzielle Energie $E_{\text{pot}} = q\psi_s$ nach oben zunimmt. Die Elementarladung erzeugt einen deutlichen lokalen Anstieg der potenziellen Energie um 0,35 eV mit einer Halbwertsbreite von etwa 6 % der effektiven Kanallänge. Dies entspricht ungefähr dem dreifachen Gitternetzabstand Δx im grenzflächennahen Oxidbereich.

Dies stellt für das hier simulierte Ladungspaket etwa 0,4 % der Gate-Fläche dar. Durch die Energieerhöhung entsteht ein Loch in der x - y -Ebene des Inversionskanals, das den Fluss der Elektronen von Source nach Drain beeinträchtigt. Der Drain-Strom sinkt dadurch im Vergleich zum ungestörten Kanal ab, so dass stets $I_{D,1} < I_{D,0}$ gilt. Aus diesem Grund wird im Folgenden stets die negative relative Stromänderung nach Gleichung 5.1 ermittelt. In der Literatur wurden die Auswirkungen von solchen Störpotenzialen auf den Inversionskanal eingehender untersucht [263, 264, 271].

Abbildung 5.9 zeigt die relative Stromänderung entlang der Ortskoordinate x_L in Abhängigkeit der Drain-Spannung zwischen 0,02 und 1,30 V. Die Gate-Spannung war hierbei auf $V_G - V_t = 0,8\text{ V}$ eingestellt. Jeder Punkt einer Kurve stellt dabei eine einzelne Simulation dar, bei der der Drain-Strom $I_{D,1}$ mit einer Elementarladung an der entsprechenden Stelle x_L/L simuliert

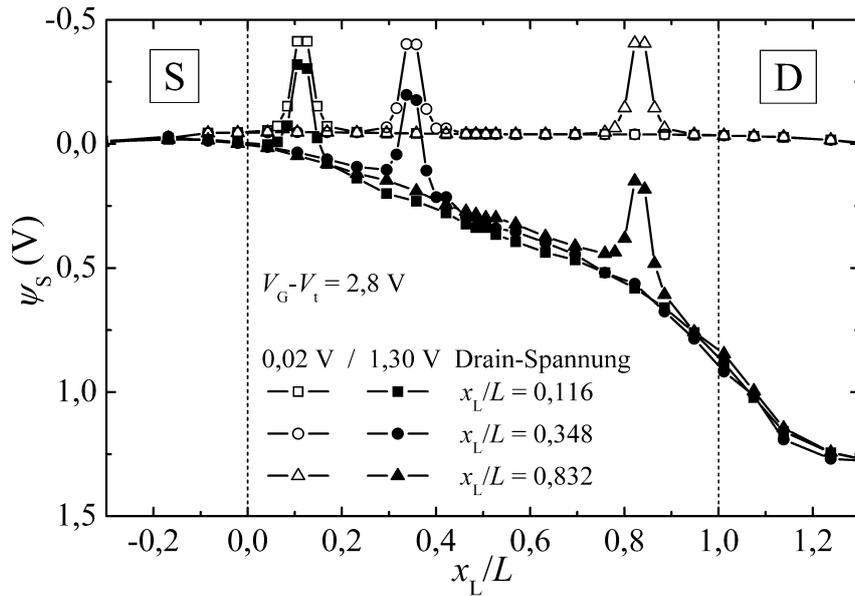


Abbildung 5.8: Simuliertes Oberflächenpotenzial ψ_s einer NROM-Zelle mit einem scharf definierten grenzflächennahen Ladungspaket im Tunneloxid. Das Potenzial steigt in der Auftragung nach unten an, die potenzielle Energie $E_{\text{pot}} = q\psi_s$ nach oben. Die Simulation wurde sowohl bei $V_D = 0,02$ V als auch bei $V_D = 1,30$ V durchgeführt. Der Schwerpunkt des Ladungspaketes wurde nacheinander bei $x_L/L = 0,116$, $x_L/L = 0,348$ und $x_L/L = 0,832$ platziert.

wurde. Aufgetragen ist jeweils die relative Änderung des Drain-Stroms (Empfindlichkeit) im Vergleich zu einem zuvor ohne Elementarladung simulierten $I_{D,0}$ in Abhängigkeit der Ladungsposition zwischen Source ($x_L/L < 0$) und Drain ($x_L/L > 1$).

Für kleine Drain-Spannungen ($V_D \leq 0,2$ V) zeigt sich eine annähernd symmetrische Verteilung der Empfindlichkeit mit jeweils einem Maximum von etwa 10 % nahe Source und Drain. In der Kanalmitte beträgt die Empfindlichkeit dagegen nur etwa 6 %, während sie zu den Kontaktgebieten hin deutlich auf nahezu Null abfällt. Bei Drain-Spannungen oberhalb von 0,2 V entsteht eine deutliche Asymmetrie der Empfindlichkeit mit einer Zunahme des Maximums bei Source auf über 15 %, wohingegen das lokale Maximum am Drain abnimmt und ab $V_D = 0,8$ V dort völlig verschwunden ist. Die Empfindlichkeit am Drain bei $V_D = 1,3$ V ist dabei auch deutlich unter die des homogenen Transistorkanals für $V_D = 0,02$ V gefallen und beträgt für $x_L/L > 0,7$ nur noch maximal 2 %.

Die Simulationen zeigen sehr deutlich die Asymmetrie der Drain-Stromempfindlichkeit auf Oxidladungen bei hohen Drain-Spannungen. Der hier beobachtete Effekt lässt sich mit dem DIBL vergleichen, der bereits in Abschnitt 2.1.3 behandelt wurde. Beim DIBL greift das durch eine hohe Drain-Spannung erzeugte elektrische Feld auf den Source-Kontakt über und reduziert dort die Potenzialbarriere für die Elektronen. In Folge der abgesenkten Barriere fließen mehr Ladungsträger in den Kanal und tragen zu einem erhöhten Drain-Strom bei. Der Stromfluss wird in diesem Zustand nur durch die Höhe der Barriere bestimmt. Bei den hier gezeigten Si-

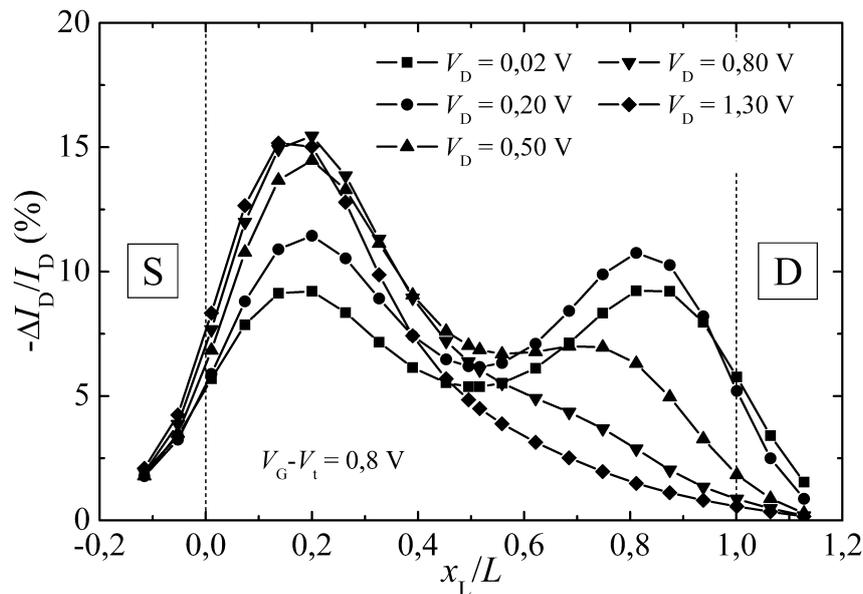


Abbildung 5.9: Simulierte relative Drain-Stromänderung in Abhängigkeit der lateralen Position des Ladungspakets für verschiedene Drain-Spannungen. Mit zunehmender Drain-Spannung nimmt die Asymmetrie der Stromempfindlichkeit zwischen Source- und Drain-Bereich zu. Die Region nahe Source zeigt dabei einen deutlichen Anstieg der Empfindlichkeit, während diese auf der Drain-Seite abfällt.

mulationen ist das Gegenteil der Fall. Das zusätzliche Ladungspaket im Oxid erhöht lokal die potenzielle Energie für die Elektronen und damit die Barriere. Der Drain-Strom, der wiederum durch die Barriere bestimmt wird, sinkt im Gegensatz zum DIBL ab. Folglich sind die Auswirkungen von Oxidladungen auf den Stromfluss nahe Source am Größten.

Diese Empfindlichkeit des Drain-Stroms auf ein im Oxid eingebrachtes Ladungspaket lässt sich mit den oben getroffenen Plausibilitätsannahmen direkt auf fluktuierende Ladungen und somit auf die Rauschanalysen übertragen. Ein Inversionsladungsträger, der mit einer freien Störstelle im Oxid kommuniziert, hat je nach lateraler Position im Kanal unterschiedliche Auswirkungen auf den Drain-Strom. Bei hohen Drain-Spannungen liegt der Einfluss auf den Drain-Strom bei etwa 15 %, wenn die Ladung nahe des Sources fluktuiert. Der Einfluss einer nahe des Drains fluktuierenden Ladung ist dagegen sehr gering (< 2 %). Die Simulation zeigt zudem sehr deutlich, dass die Empfindlichkeit am Drain sogar noch unterhalb dem Wert liegt, der für eine kleine Drain-Spannung und damit für einen annähernd homogenen Inversionskanal bestimmt wurde. Das bedeutet, dass in den experimentellen Rauschanalysen die drain-seitigen Rauschquellen sehr effektiv durch eine hohe Drain-Spannung unterdrückt werden können.

Für Drain-Spannungen von 0,2 V und kleiner sind die Auswirkungen auf den Drain-Strom annähernd gleich groß, unabhängig davon, ob die Ladungsfluktuationen nahe Source oder Drain stattfinden. Die Variation in der Stromempfindlichkeit, die dennoch in der Simulation zu beobachten ist, entsteht vermutlich durch ein inhomogenes Dotierprofil entlang des Kanals. In der Nähe der Source- und Drain-Gebiete ist die Dotierstoffkonzentration des Siliziumsubstrats

etwa fünf Mal höher als in der Mitte des Kanals. Das Oberflächenpotential, welches in Inversion ungefähr $2\psi_F \simeq \frac{2k_B}{q} \ln\left(\frac{N_A}{n_i}\right)$ beträgt, liegt daher in etwa um den Wert $\frac{2k_B}{q} \ln(5) \simeq 83$ mV niedriger. Für die Elektronen bedeutet dies eine um 83 meV höhere potenzielle Energie in der Nähe der Source- und Drain-Gebiete und damit einen erhöhten Widerstand. Ein zusätzliches Ladungspaket in dieser Zone wirkt sich daher empfindlicher auf den Drain-Strom aus als in der Mitte des Kanals.

In einer zweiten Simulationsreihe wurde die Gate-Spannung variiert. Abbildung 5.10 zeigt die Empfindlichkeit des Drain-Stroms sowohl für $V_D = 0,2$ V als auch für $V_D = 1,3$ V in Abhängigkeit der Position x_L des Ladungspakets. Mit zunehmender Gate-Spannung sinkt die Empfindlichkeit insgesamt ab. Für $V_G - V_t = 2,8$ V liegt die maximale relative Stromänderung nur noch zwischen 2 und 5 %. Die Asymmetrie bei hoher Drain-Spannung im Gegensatz zur kleinen Drain-Spannung ist aber auch hier sichtbar. Der Transistor befindet sich unter diesen Spannungsbedingungen in starker Inversion und es herrscht eine hohe Ladungsträgerdichte im Kanal. Diese kann einerseits das Potenzial der Punktladung im Tunneloxid stärker abschirmen, wodurch deren Wirkung auf den Drain-Strom reduziert ist. Andererseits scheint der Inversionskanal trotz der hohen Gate-Spannung inhomogen zu sein, so dass die Empfindlichkeit des Drain-Stroms asymmetrisch bezüglich Source und Drain ist.

Wird die Gate-Spannung bei $V_D = 1,3$ V bis auf $V_G - V_t = 0,3$ V reduziert, ergeben die Simulationen in Abbildung 5.10 einen deutlichen Anstieg der absoluten Empfindlichkeit bis auf über

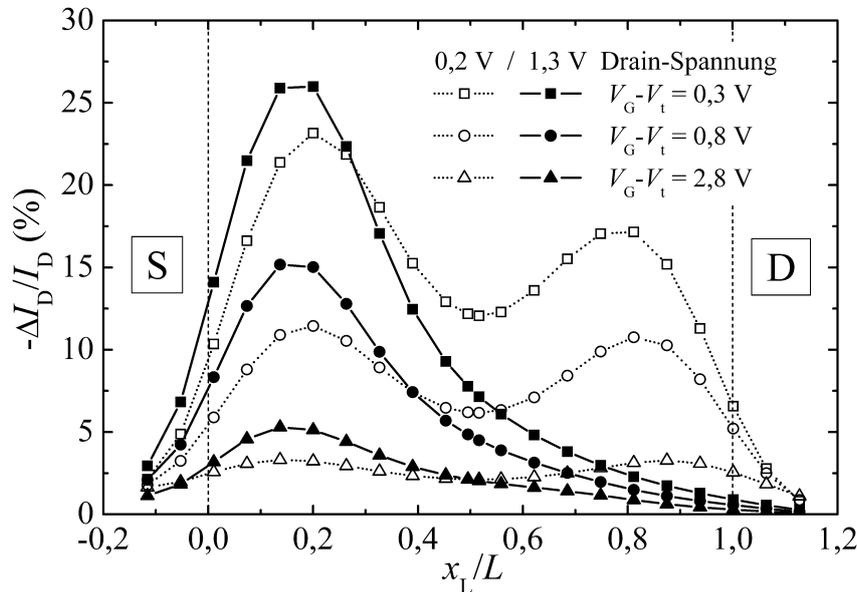


Abbildung 5.10: Simulierte relative Drain-Stromänderung in Abhängigkeit der lateralen Position des Ladungspakets für verschiedene Gate-Spannungen $V_G - V_t$. Die absolute Empfindlichkeit fällt mit zunehmender Gate-Spannung sowohl für kleine ($V_D = 0,2$ V) als auch große Drain-Spannungen ($V_D = 1,3$ V) ab. Für kleine Drain-Spannungen ist mit fallendem $V_G - V_t$ eine Zunahme der Asymmetrie zwischen source- und drain-seitiger Empfindlichkeit zu beobachten.

25 % nahe dem Source. Betrachtet man die Empfindlichkeit unter $V_D = 0,2 \text{ V}$, so zeigt sich mit abnehmender Gate-Spannung ebenfalls eine Zunahme, gleichzeitig bleibt aber die Symmetrie zwischen der Empfindlichkeit nahe Source und der nahe Drain nicht erhalten. Es bildet sich eine Asymmetrie aus, die eine stärkere Empfindlichkeit nahe dem Source zeigt. Die Drain-Spannung ist hier vergleichbar mit $V_G - V_t$ und der Transistor befindet sich damit nicht mehr im linearen Bereich, sondern im Übergangsbereich zur Sättigung. Mit der damit verbundenen inhomogenen Ladungsträgerdichte im Kanal lässt sich die Asymmetrie der Empfindlichkeit analog zu den in Abbildung 5.9 dargestellten Ergebnissen erklären.

5.5 Schlussfolgerung

In Kapitel 5.3 wurden experimentelle Ergebnisse von $1/f$ -Rauschanalysen an NROM-Speicherezellen gezeigt und erläutert. Der Bereich der Schädigung des Tunneloxids durch heiße Ladungsträger konnte mit den Untersuchungen auf die jeweilige Drain-Region des Transistorkanals eingeschränkt werden. Ähnliche Untersuchungen zur Lokalisierung von Störstellen wurden bereits mithilfe von RTS-Analysen durchgeführt [275, 283].

In dieser Arbeit wurden $1/f$ -Rauschanalysen zur Lokalisierung von Oxidschädigungen an Transistoren mit effektiven Kanallängen unterhalb von 100 nm angewendet. Trotz der stark fluktuierenden Rauschspektren konnte eine qualitative Analyse durchgeführt werden. Für quantitative Bestimmungen reichen die Ergebnisse allerdings nicht aus. Hierfür müssten mehrere identische Spektren gemessen und gemittelt werden, ähnlich wie in Kapitel 4.5 beschrieben. Auch der Einsatz eines Spektrumanalysators könnte hier von Vorteil sein, denn damit sind in der Regel Mittelungen mehrerer schnell hintereinander aufgenommener Spektren automatisiert durchführbar. Möglicherweise könnte mit diesem Messansatz eine genauere Bestimmung des geschädigten Kanalbereichs vorgenommen werden. Durch Variation der Drain-Spannung kann die Ausdehnung der Hochfeldregion am Drain verändert werden. Je nach Länge dieser Zone würden dann mehr oder weniger Rauschquellen ausgeblendet werden. Durch Differenzbildung der Spektraldichten könnte anschließend der Anteil von Rauschquellen und damit die Zahl der generierten Störstellen in der jeweiligen Kanalregion bestimmt werden. Dazu müsste der gesamte Versuchsaufbau vorher mit geeigneten Rauschquellen kalibriert werden.

In Kapitel 5.4 wurden 3D-Simulationen einer NROM-Zelle durchgeführt. Für die in Abschnitt 5.3.1 gezeigten Experimente wurde eine Gate-Spannung von $V_G - V_t = 0,6 \text{ V}$ gewählt. Dies entspricht in etwa den Simulationsbedingungen in Abbildung 5.9 bei $V_G - V_t = 0,8 \text{ V}$. Die Simulationsergebnisse bestätigen sehr gut die im experimentellen Teil der Arbeit getroffenen Aussagen bezüglich der Asymmetrie der Rauschempfindlichkeit zwischen Source und Drain. Bei einer sehr kleinen Drain-Spannung von $V_D = 0,02 \text{ V}$ tragen Ladungsfluktuationen auf der Source-Seite ebenso stark zum gesamten Rauschen bei wie Ladungsfluktuationen auf der Drain-Seite. Unter einer hohen Drain-Spannung von $V_D = 1,30 \text{ V}$ wirken sich Fluktuationen auf der Source-Seite wesentlich stärker auf das Drain-Stromrauschen aus, wohingegen sie auf

der Drain-Seite weitgehend unterdrückt sind.

Ähnliche Untersuchungen wurden bereits mit Monte-Carlo-Simulationen an MOSFETs mit Kanallängen und -weiten unterhalb von 30 nm durchgeführt [265, 281]. In diesen Studien wurden ebenfalls relative Drain-Stromänderungen bestimmt, die in einer vergleichbaren Größenordnung liegen, wie bei den hier gezeigten 3D-Simulationen.

In dieser Arbeit wurde darüber hinaus die laterale Position der Ladung im Transistorkanal variiert. Neben der bereits erwarteten Asymmetrie der Drain-Stromempfindlichkeit in Sättigung [277, 283], zeigten die hier durchgeführten Simulationen, dass diese Asymmetrie bereits knapp unterhalb des Sättigungsbetriebs des MOSFETs bei $V_D = 0,2$ V einsetzt (offene Quadrate in Abbildung 5.10). Somit werden die getroffenen Interpretationen der Messergebnisse im experimentellen Teil dieser Arbeit untermauert.

Über eine einfache Plausibilitätsbetrachtung wurde das während der Simulation im Tunneloxid platzierte Ladungspaket und die damit verbundene Drain-Stromänderung mit dem Einfang eines Ladungsträgers in eine Oxidstörstelle verglichen. Die simulierte Stromänderung wurde dann mit der Spektraldichte aus Rauschanalysen in Verbindung gebracht. Hierbei ist zu beachten, dass der Beitrag zur Stromänderung lediglich durch das zusätzliche Potenzial der Punktladung und der damit verbundenen Beweglichkeits- beziehungsweise Leitwertänderung des Kanals hervorgerufen wird. In das vereinheitlichte Modell des Stromrauschens (Gleichung 3.37) gehen neben den Beweglichkeitsfluktuationen $\Delta\mu$ aber auch die Fluktuationen der Ladungsträgerzahl ΔN ein. Letztere werden in den hier durchgeführten Simulationen allerdings nicht berücksichtigt.

6 Zusammenfassung

In dieser Arbeit wurden Gate-Oxide, die in modernen Halbleitertechnologien eingesetzt werden, mithilfe von Charge-Pumping-Messungen und $1/f$ -Rauschanalysen untersucht.

Der erste Teil der Arbeit beschäftigte sich mit n - und p -Kanal-MOSFETs mit nitridiertem Gate-Oxid. Die Bauelemente wurden dabei gezielt Fowler-Nordheim-Stress ausgesetzt, der durch das Einprägen eines konstanten Gate-Stroms erzeugt wurde. Mittels der frequenzabhängigen Charge-Pumping-Methode war es möglich, eine getrennte Bestimmung der grenzflächennahen Oxidstörstellendichte und der Dichte von Grenzflächenzustände vorzunehmen. Erstmals wurde hierbei der Einfluss der Gate-Spannungspolarität beim Stressen auf die Generation von sowohl Grenzflächenzuständen als auch grenzflächennahen Oxidstörstellen getrennt voneinander untersucht.

Zunächst wurden dazu die MOS-Transistoren mit Standard-Charge-Pumping-Methoden analysiert und die für die weiteren Untersuchungen notwendigen Parameter bestimmt. Danach konnten mithilfe theoretischer Betrachtungen die Tiefenbereiche ermittelt werden, aus denen Störstellen zum Charge-Pumping beitragen. Diese Tiefenbereiche hängen im Wesentlichen von der verwendeten Messfrequenz ab. Für die vorliegenden Bauelemente wurde anhand der Messbedingungen und der bestimmten Transistorparameter die Trennung zwischen Grenzflächenzuständen und Oxidstörstellen bei einer Tiefe von etwa 0,8 nm vorgenommen. Alle Störstellen mit einem größeren Abstand von der $Si-SiO_2$ -Grenzfläche wurden demnach den grenzflächennahen Oxidstörstellen zugeordnet.

Die Ergebnisse der Charge-Pumping-Untersuchungen zeigten einen Unterschied im Degradationsverhalten zwischen Grenzflächenzuständen und grenzflächennahen Oxidstörstellen, die erstmals beobachtet wurde. Die in der Literatur vielfach berichtete Polarisationsabhängigkeit der Störstellengeneration von der Stressspannung zeigte sich in dieser Arbeit deutlich für die grenzflächennahen Oxidstörstellen. Die Dichte der Grenzflächenzustände dagegen wies im Falle der p -MOSFETs keine beziehungsweise für die n -MOSFETs nur eine schwache Polarisationsabhängigkeit auf. In der Literatur wurden zur Bestimmung der Störstellendichten oftmals $C(V)$ -Messungen durchgeführt. Diese verwenden in der Regel niedrigere Frequenzen als die hier eingesetzte Charge-Pumping-Methode, wodurch der Tiefenbereich, aus dem Störstellen beitragen, größer ist. Somit liefern die über $C(V)$ -Methoden bestimmten Störstellendichten eine Mischung aus Grenzflächenzuständen und Oxidstörstellen. Die in dieser Arbeit eingesetzte frequenzabhängige Charge-Pumping-Methode ist dagegen in der Lage, eine Trennung zwischen Grenzflächenzuständen und grenzflächennahen Oxidstörstellen vorzunehmen. Dies ermöglichte neue Einblicke in das Generationsverhalten von Störstellen und zeigte erstmals Unterschiede zwischen den beiden räumlich getrennten Störstellenarten auf.

Darüber hinaus konnte in dieser Arbeit über die Messung der Flachbandspannung eine extrem hohe Dichte von gate-seitigen Störstellen am ungestressten p -MOSFET nachgewiesen werden. In den Stressexperimenten führte dies zu einem frühen elektrischen Durchbruch des Gate-Isolators. Die Ursache für die hohe Störstellendichte wurde dem Herstellungsprozess zugeordnet. Eine genauere Analyse in Kooperation mit dem Hersteller konnte im Rahmen dieser Arbeit nicht mehr durchgeführt werden.

Der zweite Teil der Arbeit widmete sich der Untersuchung von NROM-Speicherzellen mithilfe von $1/f$ -Rauschanalysen des Drain-Stroms. Die Speicherzellen wurden während der Experimente durch wiederholte Programmier- und Löschoptionen gezielt gestresst. Dazu wurde ein entsprechender Algorithmus in die Messsoftware implementiert. Im Vordergrund der Untersuchungen stand dabei die Entwicklung einer Methode zur orts aufgelösten Detektion der Schädigung im Transistorkanal, die durch heiße Ladungsträger beim Programmieren und Löschen der Speicherzelle hervorgerufen wird.

In einem ersten Experiment wurde für die Messungen des $1/f$ -Rauschens erstmals ein Verfahren verwendet, das sich an das Leseverfahren von NROM-Zellen anlehnt. Der Drain-Strom wurde sowohl im *forward read mode* als auch im *reverse read mode* gemessen und anschließend mithilfe einer Fast-Fourier-Transformation die Spektraldichte des Rauschens bestimmt. Mögliche Fehlerquellen, die bei der Bestimmung der Spektraldichte auftreten können, wurden durch Wahl der experimentellen Parameter weitgehend unterdrückt oder minimiert. Dennoch zeigte sich, dass mit dem hier verwendeten Messaufbau nur qualitative Aussagen möglich sind. Für quantitative Bestimmungen der Schädigungen durch wiederholte Programmier- und Löschoptionen reichte die Qualität des Messaufbaus allerdings nicht aus.

Die Ergebnisse zeigten, dass in Folge des Stresses eine vermehrte Generation von Störstellen nahe des jeweiligen Drains, also dem Entstehungsort der heißen Ladungsträger stattgefunden hat.

Ein zweites Experiment zur Bestimmung des Schädigungsortes nutzte die Programmierfähigkeit der Speicherzellen aus. Eine Injektion von Ladungen in die Speicherschicht der NROM-Zelle erzeugt einen lokalen Anstieg der Schwellspannung in der entsprechenden Kanalregion. Dadurch wird der Beitrag von Störstellen zum $1/f$ -Rauschen in diesem Bereich erhöht. Dieser Effekt, dass die gemessene Spektraldichte im Unterschwellbereich von MOS-Transistoren am größten ist, wurde bereits in der Literatur gezeigt und hier erstmals zur Untersuchung nichtflüchtiger Speicherzellen mit ultrakurzen Transistorkanälen angewendet.

In Übereinstimmung mit dem ersten Experiment zeigt auch dieser Ansatz, dass eine erhöhte Schädigung im Entstehungsort der heißen Ladungsträger zu verzeichnen ist.

Um die experimentellen Ergebnisse zu verifizieren wurden 3D-Simulationen eines zusätzlichen Ladungspakets im grenzflächennahen Oxidbereich einer NROM-Speicherzelle durchgeführt. Dieses eingebrachte Ladungspaket stellt eine besetzte Oxidstörstelle dar. Über die Simulation wurde der Drain-Strom unter Einfluss dieser zusätzlichen Ladung bestimmt und mit dem Strom verglichen, der ohne Ladungspaket simuliert werden konnte. Aus dem relativen Unterschied der Drain-Ströme wurde eine Empfindlichkeit definiert, deren Amplitude mit der Spektraldich-

te des $1/f$ -Rauschens in Verbindung steht. Nachfolgend wurde die Abhängigkeit der Drain-Stromempfindlichkeit von den angelegten Drain- und Gate-Spannungen untersucht. Dabei wurde insbesondere auch der laterale Ort des Ladungspakets im Kanal variiert und somit die Ortsabhängigkeit des Einflusses von Störstellen auf das Rauschen ermittelt.

Die Annahmen aus den experimentellen Rauschuntersuchungen konnten durch die Simulation bestätigt werden. Unter hohen Drain-Spannungen wird der Beitrag zur Spektraldichte von Rauschquellen nahe des Drains stark unterdrückt. Bei niedrigen Drain-Spannungen ist die Wirkung von Rauschquellen sowohl auf der Source-Seite als auch auf der Drain-Seite vergleichbar. Eine Wiederholung der Simulationsreihe bei gleichzeitig programmierter Speicherzelle, die der Situation im zweiten Experiment entspräche, konnte im Rahmen dieser Arbeit nicht mehr durchgeführt werden. Dies stellt daher den Ausgangspunkt für mögliche Folgesimulationen dar. Durch einen verbesserten Messaufbau, zum Beispiel durch den Einsatz von Signalanalysatoren und Vorverstärkern, könnte eine genauere Analyse des $1/f$ -Rauschens vorgenommen werden. Möglicherweise wären dann auch quantitative Bestimmungen der Schädigung durchführbar.

A CMOS-Charakterisierungsmessungen

A.1 $C(V)$ -Messungen an Kapazitäten mit nitridiertem Gate-Oxid

Die Substratdotierung einer MOS-Kapazität kann über die Messung ihrer hochfrequenten $C(V)$ -Kurve bestimmt werden. Das Verfahren ist genauer in [5] beschrieben. Die Dotierstoffkonzentration N_{Dot} in Abhängigkeit der Tiefe $-z$ im Substrat ist gegeben durch

$$N_{\text{Dot}}(-z) = -2q\epsilon_s \frac{\partial C_{\text{hf}}^{-2}}{\partial V_G} . \quad (\text{A.1})$$

Hierbei ist q die Elementarladung, $\epsilon_s = \epsilon_{\text{Si}}\epsilon_0 = 1,05 \cdot 10^{-8} \text{ F cm}^{-2}$ und C_{hf} die gemessene Hochfrequenzkapazität. Unter der Annahme, dass das Dotierprofil homogen ist, entfällt die Abhängigkeit von der Tiefe z . Die Dotierkonzentration ist dann proportional zur Steigung der $1/C_{\text{hf}}^2$ -Kurve im Verarmungsbereich.

Die gemessenen $C(V)$ -Kurven und die jeweiligen $1/C_{\text{hf}}^2$ -Kurven sind für die n -MOS-Kapazität in Abbildung A.1 und für die p -MOS-Kapazität in Abbildung A.2 dargestellt. Für die n -MOS-Kapazität ergibt sich demnach eine Substratdotierung von $N_A = 3,4 \cdot 10^{17} \text{ cm}^{-3}$ und für die p -MOS-Kapazität von $N_D = 6,1 \cdot 10^{16} \text{ cm}^{-3}$. Aus der flächennormierten Akkumulationskapazität von jeweils etwa $C'_{\text{akk}} = 0,5 \mu\text{F cm}^{-2}$ ergibt sich über

$$d_{\text{ox}} = \frac{\epsilon_{\text{ox}}}{C'_{\text{akk}}} \quad (\text{A.2})$$

eine Oxiddicke von 6,9 nm. $\epsilon_{\text{ox}} = \epsilon_{\text{SiO}_2}\epsilon_0$ ist hierbei die Dielektrizitätskonstante im Siliziumdioxid.

Mithilfe einer zusätzlichen Messung der quasistatischen Kapazitätskurve $C_{\text{qs}}(V_G)$ kann über das Berglund-Integral [238]

$$\psi_s(V_G) = \int_{V_0}^{V_G} \left(1 - \frac{C_{\text{qs}}}{C_{\text{ox}}}\right) dV + \psi_s(V_0) \quad (\text{A.3})$$

das Oberflächenpotenzial ψ_s bei einer vorgegebenen Gate-Spannung V_G berechnet werden. Dabei ist der Startwert V_0 eine genügend weit in Akkumulation liegende Gate-Spannung und $C_{\text{ox}} = C_{\text{qs}}(V_0)$ die dazugehörige quasistatische Kapazität, die der Oxidkapazität entspricht.

In Abbildung A.3 sind die Oberflächenpotenziale sowohl für die n - als auch die p -MOS-Kapazität dargestellt. Für einen Wert des Oberflächenpotenzials von $\psi_s = 0$ lässt sich eine Flachbandspannung von $V_{\text{fb}} = -1,0 \text{ V}$ für den n -MOSFET und $V_{\text{fb}} = 1,0 \text{ V}$ für den p -MOSFET ablesen.

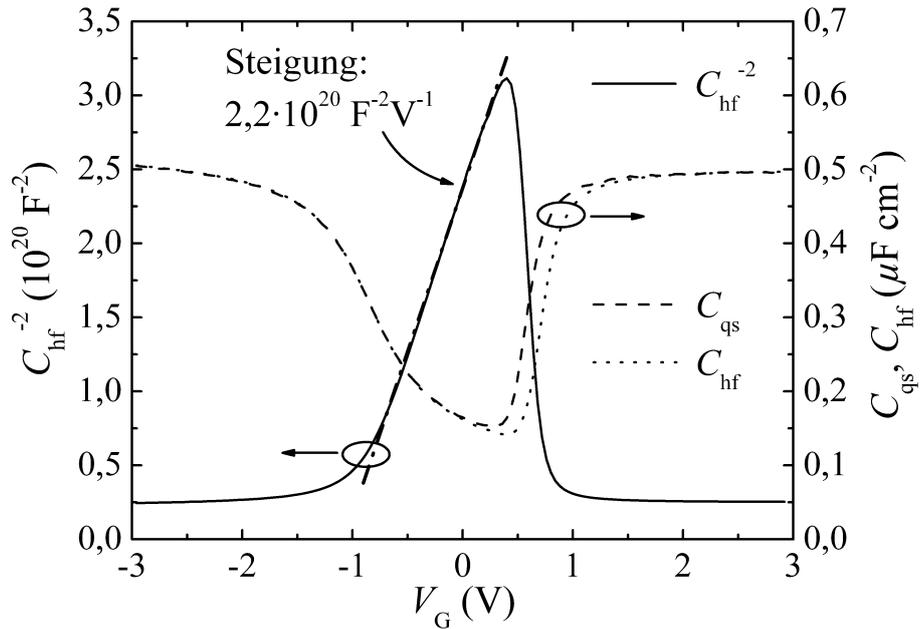


Abbildung A.1: Gemessene quasistatische und hochfrequente $C(V)$ -Kurven für die n -MOS-Kapazität. Die Funktion $1/C_{\text{hf}}^2$ ist aus der hochfrequenten $C(V)$ -Kurve berechnet worden und ebenfalls dargestellt. Aus der Steigung der Kurve im Verarmungsbereich ($-0,8 < V_G < 0,4$) ergibt sich eine Substratdotierung von $N_A = 3,4 \cdot 10^{17} \text{ cm}^{-3}$.

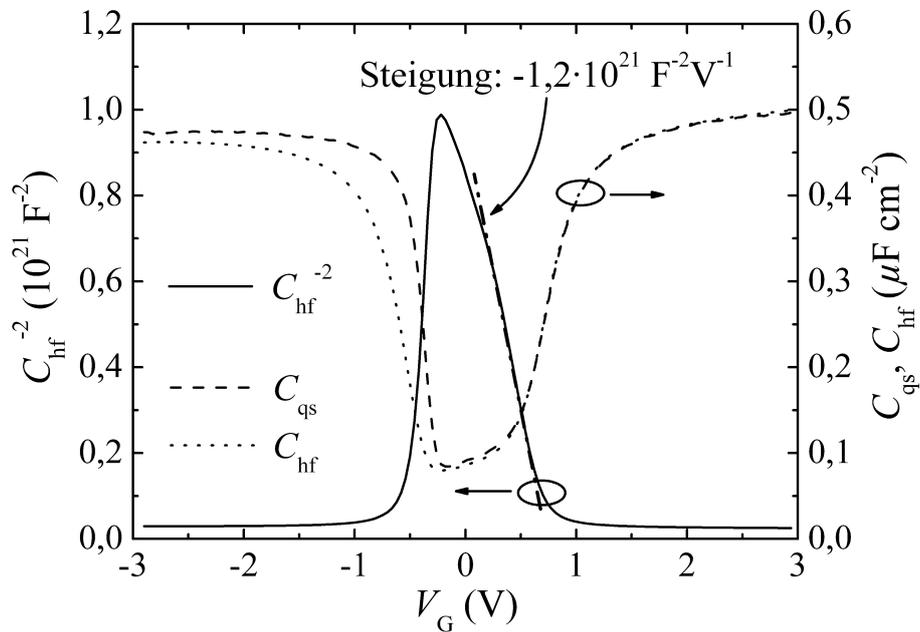


Abbildung A.2: Gemessene quasistatische und hochfrequente $C(V)$ -Kurven für die p -MOS-Kapazität. Die Funktion $1/C_{\text{hf}}^2$ ist aus der hochfrequenten $C(V)$ -Kurve berechnet worden und ebenfalls dargestellt. Aus der Steigung der Kurve im Verarmungsbereich ($0,2 < V_G < 0,6$) ergibt sich eine Substratdotierung von $N_D = 6,1 \cdot 10^{16} \text{ cm}^{-3}$.

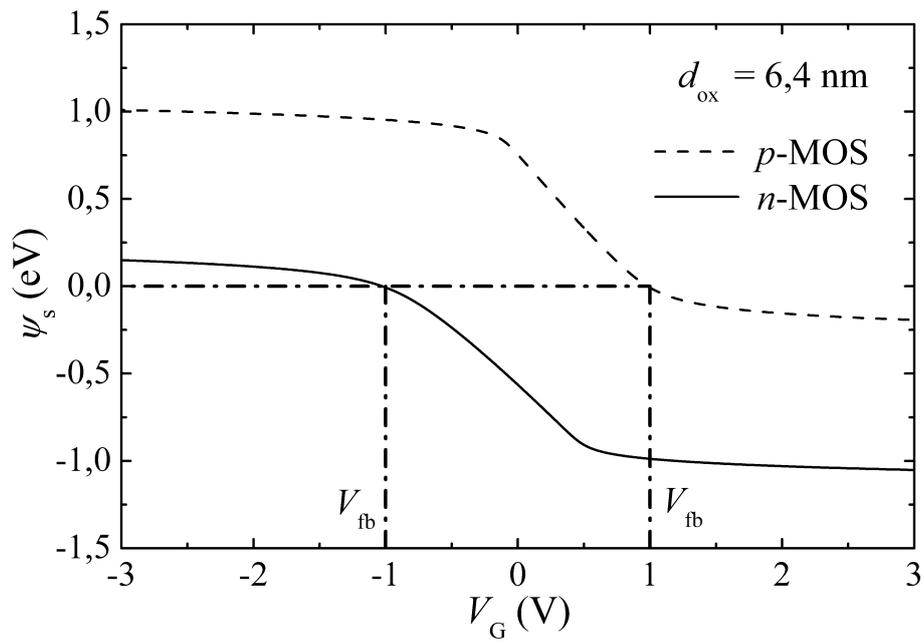


Abbildung A.3: Oberflächenpotenzial in Abhängigkeit der Gate-Spannung, berechnet über das Berglund-Integral aus Gleichung A.3 mithilfe einer quasistatischen Kapazitätsmessung.

A.2 Bestimmung von f_0 mittels frequenzabhängiger CP-Messung

Zur Bestimmung von f_0 wird der maximale CP-Strom bei verschiedenen Frequenzen gemessen. Der Spannungspuls am Gate ist dabei nicht rechteckförmig, wie bei den in dieser Arbeit sonst durchgeführten Messungen, sondern dreieckförmig. Die Anstiegs- und Abfallflanken entsprechen stets 50% der Pulsperiode, während die Amplitude konstant bei 3,0 V gehalten wird. Die Basisspannung beträgt beim n -Kanal-MOSFET $-2,0$ V und beim p -Kanal-MOSFET $-0,4$ V. In Abbildung A.4 ist die während eines Pulses maximale rekombinierte Ladung $Q_{CP,max} = I_{CP,max}/f$ für beide Transistortypen über der Frequenz dargestellt.

Durch eine lineare Anpassung der Messpunkte lässt sich die Frequenz f_0 als Schnittpunkt mit der Frequenzachse bestimmen. Für den n -Kanal-Transistor ergibt sich $f_0 = 14,96$ Hz, für den p -Kanal-Transistor $f_0 = 54,28$ Hz.

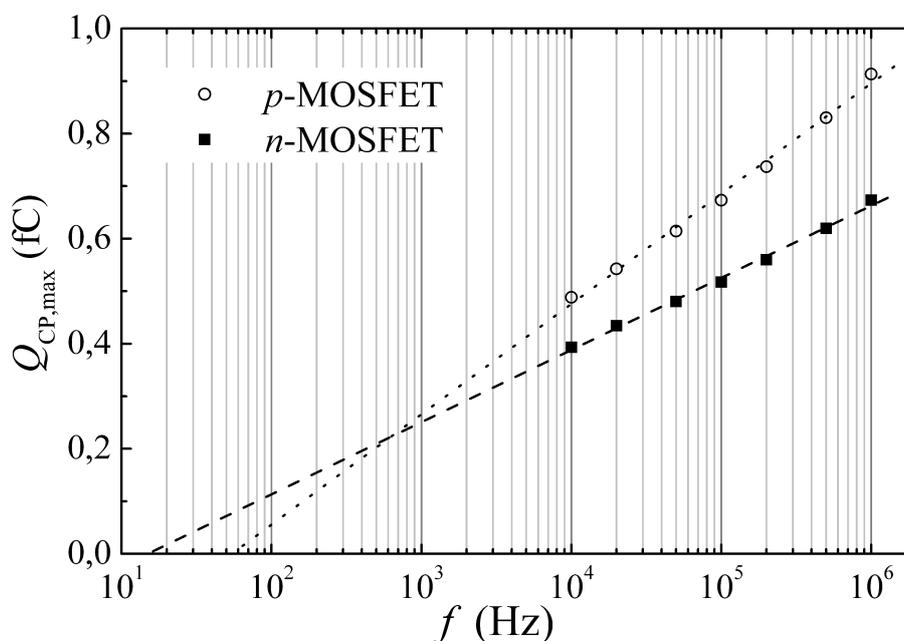


Abbildung A.4: Auftragung von $Q_{CP,max}$ über der Frequenz f . Der Schnittpunkt mit der Frequenzachse bestimmt die Frequenz f_0 , aus der das geometrische Mittel $\sqrt{\sigma_n \sigma_p}$ der Einfangquerschnitte der Störstellen bestimmt werden kann.

A.3 Verlauf der Gate-Spannung beim FN-Stress

In Abbildung A.5 ist die FN-Stressspannung sowohl für den n -MOSFET als auch für den p -MOSFET unter NBS über dem Betrag der injizierten Stressladungsdichte Q_{inj} dargestellt. Für den n -Kanal-Transistor verläuft die Spannung über die gesamte Stressphase sehr konstant. Die Spannung am Gate beträgt hier in etwa $-6,8$ V. Mit zunehmendem Stress zeigen sich vermehrt Einbrüche der Stressspannung. Diese treten bei Wiederaufnahme des Stresses nach einer Charakterisierungsmessung auf und werden durch Umladungseffekte generierter Störstellen hervorgerufen.

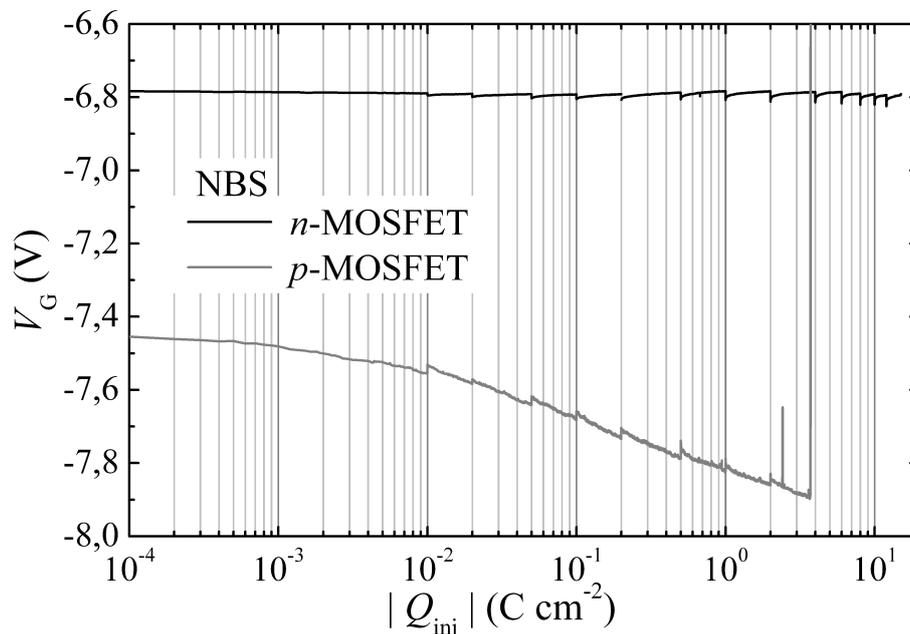


Abbildung A.5: Verlauf der FN-Stressspannung am Gate über der injizierten Stressladungsdichte unter NBS. Am n -MOSFET zeigt sich über die gesamte Stressphase ein sehr konstanter Wert der Gate-Spannung von $-6,8$ V. Am p -MOSFET fällt die Stressspannung von $-7,45$ V kontinuierlich bis auf einen Wert von $-7,9$ V ab, beim dem ein elektrischer Durchbruch des Oxids auftrat.

Die Stressspannung am p -Kanal-Transistor fällt während der Stressphase kontinuierlich von $-7,45$ V auf $-7,9$ V ab. An diesem Punkt fand ein elektrischer Durchbruch des Gate-Isolators statt, woraufhin die Spannung zusammenbrach. Auch hier zeigen sich Umladungseffekte an den Unterbrechungspunkten des FN-Stresses.

In Abbildung A.6 ist analog die FN-Stressspannung unter PBS dargestellt. Beide Transistortypen zeigen einen relativ konstanten Verlauf der Spannung über der gesamten Stressphase. Die Gate-Spannung am n -MOSFET liegt dabei etwa bei $6,1$ V, die am p -MOSFET um $6,9$ V. Umladungseffekte sind unter PBS nicht zu beobachten.

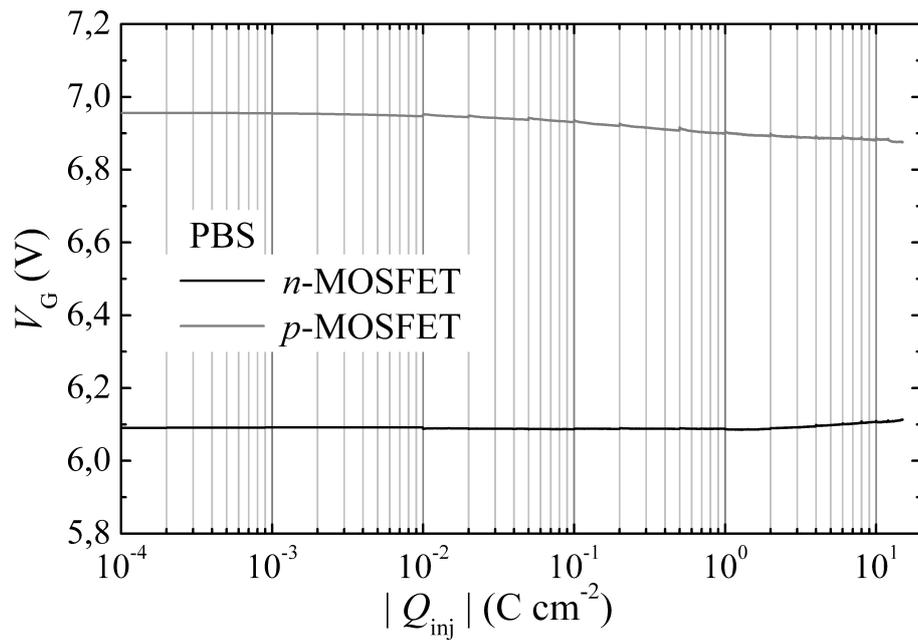


Abbildung A.6: Verlauf der FN-Stressspannung am Gate über der injizierten Stressladungsdichte unter PBS. Beide Transistortypen zeigen über der gesamten Stressphase einen relativ konstanten Verlauf der Spannung von 6,1 V beim n -MOSFET beziehungsweise 6,9 V beim p -MOSFET.

B NROM-Charakterisierungsmessungen

B.1 Spannungsverläufe beim Programmieren und Löschen

Wie bereits in Kapitel 5.2 erwähnt, wurde das Programmieren und Löschen eines Bits einer NROM-Zelle durch eine Serie von Spannungspulsen am Drain realisiert. Die Amplitude der Pulse wurde dabei schrittweise um 0,1 V erhöht. Wenn die notwendige Schwellspannungsverschiebung erreicht war, wurde die Pulsserie abgebrochen und der Spannungswert der zuletzt verwendeten Amplitude gespeichert.

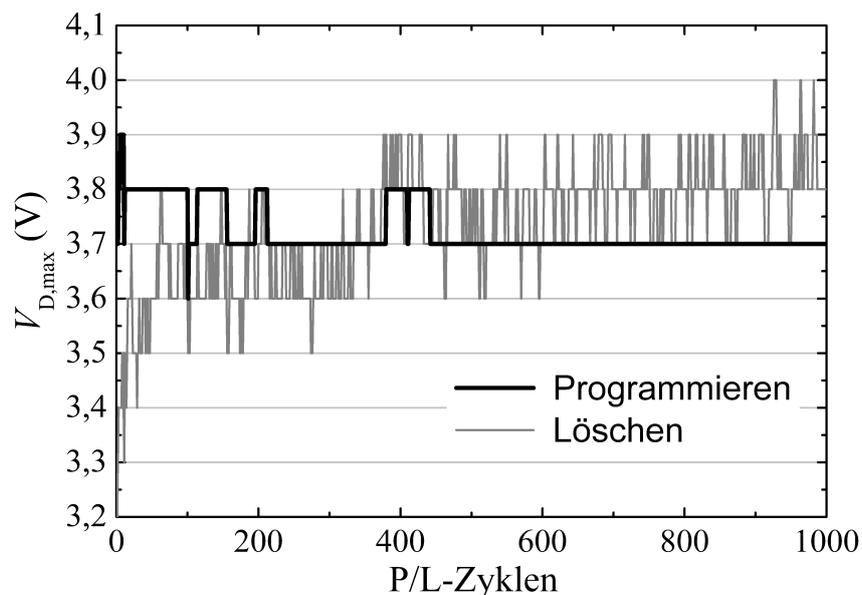


Abbildung B.1: Maximale Amplitude $V_{D,max}$ der Spannungspulse am Drain beim Programmieren und Löschen eines Bits einer NROM-Zelle. Mit steigender Anzahl von Zyklen steigt auch die am Drain notwendige Löschspannung von anfänglich 3,2 V auf etwa 3,8 V. Die zum Programmieren notwendige Amplitude bleibt während des gesamten Stressens nahezu konstant bei 3,7 V.

In Abbildung B.1 sind diese maximalen Amplituden in Abhängigkeit der bereits durchgeführten P/L-Zyklen dargestellt. Die Spannungswerte wurden sowohl beim Programmieren als auch beim Löschen bis zu einer Zyklenzahl von 1000 aufgezeichnet. Die maximale Amplitude, die zum Programmieren eines Bits notwendig war, liegt relativ konstant um 3,7 V. Zum Löschen

wird mit zunehmender Anzahl von Zyklen eine immer größere Spannung benötigt. Anfangs reicht eine Spannung von 3,2 V aus, um die definierte Schwellspannungsverschiebung von +2,0 V zu erreichen. Innerhalb der ersten 50 Zyklen steigt diese Spannung auf 3,6 V an, ab etwa 400 Zyklen müssen mindestens 3,8 V zum Löschen des Bits angelegt werden.

Symbol- und Abkürzungsverzeichnis

Symbole

A_T	Konstante beim direkten Tunneln
A_G	Gate-Fläche
B_T	Konstante beim direkten Tunneln
C'_{akk}	Flächennormierte Akkumulationskapazität
C_{hf}	Hochfrequenzkapazität
C_{ox}	Oxidkapazität
C_{qs}	Kapazität unter quasistatischen Bedingungen
c_n	Einfangrate für Elektronen
c_p	Einfangrate für Löcher
D_{it}	Mittlere Grenzflächenzustandsdichte
D_{ot}	Mittlere Volumendichte von Oxidstörstellen
D_T	Volumendichte von Störstellen
d_{ox}	Dicke des Gate-Oxids
E_{pot}	Potenzielle Energie eines Elektrons
$E_{\text{em,e}}$	Energiegrenze für die Emission von Elektronen
$E_{\text{em,h}}$	Energiegrenze für die Emission von Löchern
E_g	Energiebandlücke im Silizium
E_F	Fermi-Niveau oder Fermi-Energie
$E_{F,\text{akk}}$	Quasi-Fermi-Niveau in Akkumulation
$E_{F,\text{inv}}$	Quasi-Fermi-Niveau in Inversion
E_i	Intrinsisches Energieniveau
E_L	Energie der Leitungsbandkante
\mathcal{E}_{ox}	Elektrische Feldstärke im Oxid
E_T	Energie einer Störstelle
E_V	Energie der Valenzbandkante
\mathcal{E}_{\parallel}	Parallele elektrische Feldkomponente
F_n	Besetzungswahrscheinlichkeit für Elektronen
F_p	Besetzungswahrscheinlichkeit für Löcher
f	Frequenz
f_0	Grenzfrequenz oder Startfrequenz
f_c	Grenzfrequenz oder Nyquist-Frequenz

f_{\max}	Maximale Frequenz
f_{\min}	Minimale Frequenz
g	Gewichtungsfunktion oder Wahrscheinlichkeitsdichte
I	Strom
\bar{I}	Mittelwert des Stroms
I_1, I_2, I_3, I_4	Stromkomponenten beim Charge-Pumping
I_B	Substrat- beziehungsweise Wannennstrom
I_{CP}	Charge-Pumping-Strom
$I_{CP,\max}$	Maximaler Charge-Pumping-Strom
I_D	Drain-Strom
$I_{D,0}$	Simulierter Drain-Strom ohne Punktladung
$I_{D,1}$	Simulierter Drain-Strom mit Punktladung
$I_{D,\text{leak}}$	Drain-Leckstrom
$I_{D,\text{sat}}$	Drain-Sättigungsstrom
$I_{D,t}$	Drain-Strom an der Schwellspannung
I_{in}	Eingangsstrom
I_S	Source-Strom
i	Index oder imaginäre Einheit
j_D	Stromdichte beim direkten Tunneln
j_{FN}	Stromdichte beim Fowler-Nordheim-Tunneln
j_{inj}	Stressstromdichte
K	Konstante beim direkten Tunneln
k_B	Boltzmann-Konstante ($k_B = 1,3804 \cdot 10^{-23} \text{ J K}^{-1}$)
L	Kanallänge
L'	Kanallänge in Sättigung
L_D	Debye-Länge
L_G	Physikalische Gate-Länge
M	Anzahl
m^*	Effektive Elektronenmasse (in SiO_2 : $m^* \simeq 0,5m_e$)
m_e	Masse eines Elektrons ($m_e = 9,1095 \cdot 10^{-31} \text{ kg}$ [2])
N	Teilchenzahl
N_A	Dotierstoffkonzentration im p -Halbleiter
N_D	Dotierstoffkonzentration im n -Halbleiter
N_{Dot}	Dotierstoffkonzentration
n	Index oder Ladungsträgerdichte
n_i	Intrinsische Ladungsträgerkonzentration (in Si bei 300 K: $9,65 \cdot 10^9 \text{ cm}^{-3}$ [2])
$n_{p,0}$	Gleichgewichtsladungsträgerdichte von Elektronen
n_s	Elektronendichte an der Halbleiteroberfläche
n_t	Dichte der mit Löchern besetzten Grenzflächenzustände

$p_{p,0}$	Gleichgewichtsladungsträgerdichte von Löchern
p_s	Löcherdichte an der Halbleiteroberfläche
Q	Ladung
Q_B, Q'_B	Ladungsdichten pro Flächeneinheit im Halbleiter
Q_{CP}	CP-Ladung
Q_{inj}	Injizierte Stressladung
Q_i	Flächenladungsdichte der Inversionsschicht
Q_P	Flächenladungsdichte im Poly-Gate-Kontakt
Q_s	Flächenladungsdichte an der Halbleiteroberfläche
Q_t	Flächenladungsdichte der Grenzflächenzustände
q	Elementarladung ($q = 1,60219 \cdot 10^{-19} \text{ C}$)
$q\chi$	Elektronenaffinität
$q\phi_{Poly}$	Austrittsarbeit im Poly- <i>Si</i>
$q\phi_{Si}$	Austrittsarbeit im Silizium
R	Widerstand
R_x	Widerstand am Ort x
$S'_{I_D,1}$	Normierte Spektraldichte des Drain-Stromrauschens bei 1 Hz
S_X	Spektraldichte der Größe X
S'_X	Auf X^2 normierte Spektraldichte der Größe X
T	Temperatur oder Zeitraum
T_P	Periodendauer
t	Zeit
t_{em}	Dauer der nicht stationären Emission von Ladungsträgern
$t_{em,e}$	Dauer der nicht stationären Emission von Elektronen
$t_{em,h}$	Dauer der nicht stationären Emission von Löchern
t_{Flanke}	Flankendauer eines Pulses
t_f	Dauer der abfallenden Flanke eines Pulses
t_h	Dauer der oberen konstanten Phase eines Pulses
t_{inj}	Stresszeit
t_{int}	Integrationszeit
t_l	Dauer der unteren konstanten Phase eines Pulses
t_r	Dauer der ansteigenden Flanke eines Pulses
V	Spannung
V_0	Startspannung
V_{Ampl}	Amplitude
V_B	Substrat- beziehungsweise Wannenspannung
V_D	Drain-Spannung
V_{DD}	Betriebsspannung in einer CMOS-Schaltung
$V_{D,max}$	Maximale Pulsspannung am Drain
V_{DS}	Spannungsabfall zwischen Drain und Source

$V_{D,sat}$	Drain-Sättigungsspannung
V_{fb}	Flachbandspannung
V_G	Gate-Spannung
$V_{G,max}$	Maximale Pulsspannung am Gate
$V_{G,min}$	Minimale Pulsspannung am Gate
V_{in}	Eingangsspannung
V_L	Raumvolumen eines Ladungspaketes
V_{out}	Ausgangsspannung
V_{ox}	Spannungsabfall am Oxid
V_r	Sperrspannung
V_S	Source-Spannung
V_t	Schwellspannung
$V_{t,0}$	Schwellspannung eines ungestressten Transistors
$V_{t,p}$	Schwellspannung einer programmierten Speicherzelle
$V_{t,g}$	Schwellspannung einer gelöschten Speicherzelle
W	Kanalweite
W_G	Physikalische Gate-Weite
W_R	Ausdehnung der Raumladungszone
$W_{R,max}$	Maximale Ausdehnung der Raumladungszone
x, y, z	Ortskoordinaten
x_L	Laterale Position eines Ladungspaketes
z_e	Eindringtiefe von Elektronen
z_h	Eindringtiefe von Löchern
z_m	Minimale Eindringtiefe
z_{max}	Maximale Eindringtiefe
z_{Tun}	Tunneldistanz
α	Streukoeffizient
α_{geo}	Geometrische Komponente
α_H	Hooge-Parameter
α'_H	Modifizierter Hooge-Parameter
α_{Tast}	Tastverhältnis
β	Kehrwert der thermischen Spannung, $\beta \equiv q/k_B T$ (bei 300 K: $\beta = 38,69 \text{ V}^{-1}$)
ΔE	Energiedifferenz
ΔF	Füllfunktion
ΔI	Fluktuation des Stroms
$\widetilde{\Delta I}$	Fourier-Transformierte von ΔI
$\overline{\Delta I^2}$	Varianz des Stroms
ΔI_D	Drain-Stromdifferenz

ΔN	Fluktuation der Teilchenzahl
$\frac{\Delta N}{N^2}$	Varianz der Teilchenzahl
ΔV_t	Schwellspannungsverschiebung
$\Delta \mu$	Fluktuation der Beweglichkeit
$\Delta \psi_e, \Delta \psi_{ee}, \Delta \psi_h, \Delta \psi_{he}$	Energieintervalle
Δt	Zeitraum
ϵ_0	Elektrische Feldkonstante ($\epsilon_0 = 8,8542 \cdot 10^{-14} \text{ F cm}^{-1}$)
ϵ_{ox}	Dielektrizitätskonstante im Siliziumdioxid ($\epsilon_{\text{ox}} = \epsilon_{\text{SiO}_2} \epsilon_0$)
ϵ_s	Dielektrizitätskonstante im Silizium ($\epsilon_s = \epsilon_{\text{Si}} \epsilon_0$)
ϵ_{Si}	Dielektrizitätszahl für Silizium ($\epsilon_{\text{Si}} = 11,9$ [2])
ϵ_{SiO_2}	Dielektrizitätszahl für Siliziumdioxid ($\epsilon_{\text{SiO}_2} = 3,9$ [2])
γ	Frequenzexponent
λ	Abschwächungskoeffizient
λ_e	Abschwächungskoeffizient für Elektronen
λ_h	Abschwächungskoeffizient für Löcher
μ	Beweglichkeit
μ_l	Beweglichkeit bestimmt durch Gitterstreuungen
μ_n	Elektronenbeweglichkeit
ν_{th}	Mittlere thermische Geschwindigkeit von Ladungsträgern (in <i>Si</i> : $1 \cdot 10^7 \text{ cm s}^{-1}$ [2])
ϕ_B	Energiebarriere (zwischen <i>Si</i> und <i>SiO</i> ₂ : $\phi_B = 3,15 \text{ eV}$ [2])
ϕ_{DIBL}	Potenzialabsenkung beim DIBL
ψ_F	Fermi-Potenzial
ψ_o	Oberflächen-Fermi-Potenzial
ψ_s	Oberflächenpotenzial
ρ_L	Ladungsdichte eines Ladungspaketes
σ_n	Einfangquerschnitt für Elektronen
σ_p	Einfangquerschnitt für Löcher
τ	Zeitkonstante
τ_0	Zeitkonstante an der Grenzfläche
τ_c	Zeitkonstante für den Einfang eines Ladungsträgers
τ_e	Zeitkonstante für die Emission eines Ladungsträgers
τ_{max}	Maximale Zeitkonstante
τ_n, τ_p	Schaltzeiten
τ_t	Zeitkonstante für Elektroneneinfang

Abkürzungen

AHI	Anode Hole Injection
BBT	Band-zu-Band-Tunneln
BNC	Bayonet Nut Connector
CMOS	Complementary Metal-Oxide-Semiconductor
CP	Charge-Pumping
DIBL	Drain-Induced Barrier Lowering
DLTS	Deep Level Transient Spectroscopy
DRAM	Dynamic Random Access Memory
EEPROM	Electrically Erasable Programmable Read Only Memory
EPROM	Electrically Programmable Read Only Memory
ESR	Electron Spin Resonance
FFT	Fast Fourier Transform
FLOTOX	Floating-Gate Tunnel Oxide
FN	Fowler-Nordheim
GPIO	General Purpose Interface Bus
HC	Hot Carrier
HR	Hydrogen Release
II	Impact Ionization
LDD	Lightly Doped Drain
MOS	Metal-Oxid-Semiconductor
MOSFET	Metal-Oxid-Semiconductor Field Effect Transistor
NBTI	Negative Bias Temperature Instability
NIOT	Near-Interface Oxide Trap
NROM	Nitrided Read Only Memory
NVSM	Non-Volatile Semiconductor Memory
ONO	Oxid-Nitrid-Oxid
ROM	Read Only Memory
RTS	Random Telegraph Signal
SILC	Stress-Induced Leakage Current
SIMS	Secondary Ion Mass Spectrometry
SONOS	Silicon-Oxide-Nitrid-Oxide-Semiconductor
TCAD	Technology Computer Aided Design
UV	Ultra Violett
VWF	Virtuelle Wafer Fab
XPS	X-Ray Photoelectron Spectroscopy

Abbildungsverzeichnis

2.1	Aufbau eines MOS-Kondensators	6
2.2	Bandstruktur eines MOS-Kondensators	7
2.3	Bandstruktur in Akkumulation, Verarmung und Inversion	8
2.4	Aufbau eines MOSFETs	10
2.5	Transferkennlinie eines MOSFETs	11
2.6	Ausgangskennlinienfeld eines idealen MOSFETs	12
2.7	Betriebsmodi eines MOSFETs	13
2.8	Parallele elektrische Feldkomponente in einem MOSFET mit LDD	15
2.9	Oberflächenpotenzial in Abhängigkeit der Drain-Spannung (DIBL)	16
2.10	Direktes Tunneln und FN-Tunneln	18
2.11	Klassifizierung von Ladungen im Oxid	20
2.12	Sauerstofffehlstellen und mögliche Konfigurationen	21
2.13	Wasserstoffbrückenbindungen und mögliche Konfigurationen	22
2.14	Chemische Struktur von Grenzflächenzuständen	23
2.15	Zustandsdichten verschiedener Grenzflächendefekte	24
2.16	Schädigungsmechanismen im Siliziumdioxid	26
2.17	Aufbau eines Oxinitrids nahe der <i>Si-SiO₂</i> -Grenzfläche	30
2.18	Schaltbild eines CMOS-Inverters	32
2.19	Aufbau eines Floating-Gate-Speichers und einer NROM-Speicherzelle	33
2.20	Lesen eines Bits einer NROM-Speicherzelle	36
2.21	Transferkennlinien beim Lesen einer NROM-Speicherzelle	37
3.1	Beschaltung eines MOSFETs und Pulsform beim Charge-Pumping	40
3.2	Charge-Pumping-Prozesse an der Grenzfläche	42
3.3	Stromkomponenten und beteiligte Energieintervalle beim Charge-Pumping	44
3.4	Charge-Pumping mit konstanter Amplitude	46
3.5	Charge-Pumping mit variabler Amplitude	47
3.6	Charge-Pumping mit variabler Frequenz	49
3.7	Lorentzspektrum	53
3.8	Rauschspektrum eines Halbleiters	54
3.9	Abtasttheorem und Aliasing-Effekt	58
4.1	Foto des verwendeten Waferprobers	63
4.2	Messaufbau für Charge-Pumping	64

4.3	Pulsparameter beim Charge-Pumping	66
4.4	Charge-Pumping-Kurven für den n - und p -Kanal-MOSFET	67
4.5	Füllfunktion beim Charge-Pumping	71
4.6	Maximale rekombinierte Charge-Pumping-Ladung über der Frequenz	73
4.7	Erweiterter Messaufbau für Charge-Pumping	74
4.8	Stresspolaritäten beim FN-Stressexperiment	75
4.9	Charge-Pumping-Messungen an einem gestressten und ungestressten Transistor	77
4.10	Grenzflächenzustandsdichte unter FN-Stress	78
4.11	Oxidstörstellendichte unter FN-Stress	79
4.12	Verlauf der Flachbandspannung beim n -MOS	82
4.13	Verlauf der Flachbandspannung beim p -MOS	83
4.14	Entstehung einer Hysterese in der Flachbandspannung bei $C(V)$ -Messungen	84
4.15	Hysterese der Flachbandspannung unter FN-Stress	85
4.16	Messaufbau für $1/f$ -Rauschanalysen	87
4.17	Drain-Strommessung über der Zeit für Rauschanalyse	88
4.18	Spektraldichte eines n -MOS-Transistors unter NBS	90
4.19	Korrelation zwischen Spektraldichte und Oxidstörstellen	91
4.20	Korrelation zwischen Spektraldichte und Grenzflächenzuständen	92
5.1	Programmieren einer NROM-Speicherzelle	96
5.2	Löschen einer NROM-Speicherzelle	97
5.3	Drain-Strommessung über der Zeit	98
5.4	Rauschspektren in Abhängigkeit der Zyklenzahl	99
5.5	Spektraldichte in Abhängigkeit der Drain-Source-Spannungspolarität	102
5.6	Spektraldichte in Abhängigkeit des Programmierzustands	105
5.7	Mögliche Zustände einer NROM-Zelle bei Rauschmessungen	106
5.8	Simuliertes Oberflächenpotenzial	110
5.9	Simulierte Drain-Stromänderung in Abhängigkeit von V_D	111
5.10	Simulierte Drain-Stromänderung in Abhängigkeit von V_G	112
A.1	$C(V)$ -Kurven der n -MOS-Kapazitäten	120
A.2	$C(V)$ -Kurven der p -MOS-Kapazitäten	120
A.3	Oberflächenpotenzial als Funktion der Gate-Spannung	121
A.4	Maximale rekombinierte Charge-Pumping-Ladung bei dreieckförmigen Pulsen	122
A.5	FN-Stressspannung unter NBS	123
A.6	FN-Stressspannung unter PBS	124
B.1	Maximale Spannungspulsamplitude beim Programmieren und Löschen	125

Tabellenverzeichnis

4.1	Probenparameter der untersuchten CMOS-Bauelemente	62
4.2	Probenparameter aus Charge-Pumping-Messungen	68
4.3	Probenparameter zur Bestimmung der minimalen Eindringtiefe	72
4.4	$D_{it,0}$ und $D_{ot,0}$ der ungestressten MOSFETs	73
4.5	Gate-Spannungen und elektrische Feldstärken im Oxid	76
5.1	Probenparameter der untersuchten NROM-Speicherzellen	94

Literaturverzeichnis

- [1] BITKOM, *Bundesverband Informationswirtschaft Telekommunikation und neue Medien e. V.*, <http://www.bitkom.org>, Abruf: 22.05.2008.
- [2] S. M. Sze und K. K. Ng, *Physics of Semiconductor Devices*, John Wiley & Sons Inc., Hoboken, New Jersey, Third Edition, 2007.
- [3] H. J. Osten, *Technologie integrierter Bauelemente*, Vorlesung, 2007.
- [4] P. Pavan, R. Bez, P. Olivo und E. Zanoni, „Flash memory cells - an overview”, *Proceedings of the IEEE*, 85(8), S. 1248–1271, August 1997.
- [5] T. Hori, *Gate Dielectrics and MOS ULSIs*, Springer Verlag, Berlin Heidelberg, 1997.
- [6] G. D. Wilk, R. M. Wallace und J. M. Anthony, „High- κ gate dielectrics: Current status and materials properties considerations”, *J. Appl. Phys.*, 89(10), S. 5243–5275, Mai 2001.
- [7] C. Kittel, *Einführung in die Festkörperphysik*, Oldenbourg R. Verlag GmbH, 11. Auflage, 1996.
- [8] L. D. Yau, „A simple theory to predict the threshold voltage of short-channel IGFETs”, *Solid State Electron.*, 17(10), S. 1059–1063, Oktober 1974.
- [9] K. K. Ng und G. W. Taylor, „Effects of hot-carrier trapping in n- and p-channel MOS-FET’s”, *IEEE Trans. Electron Devices*, ED-30(8), S. 871–876, August 1983.
- [10] T. H. Ning, C. M. Osburn und H. N. Yu, „Effect of electron trapping on IGFET characteristics”, *J. Electron. Mater.*, 6(2), S. 65–76, 1977.
- [11] Y. Leblebici und S.-M. Kang, „Modeling of nMOS transistors for simulation of hot-carrier-induced device and circuit degradation”, *IEEE Trans. Electron Devices*, 11(2), S. 235–246, Februar 1992.
- [12] W. Weber, M. Brox, R. Thewes und N. S. Saks, „Hot-hole-induced negative oxide charges in n-MOSFET’s”, *IEEE Trans. Electron Devices*, 42(8), S. 1473–1480, August 1995.
- [13] B. Doyle, M. Bourcerie, J.-C. Marchetaux und A. Boudou, „Interface state creation and charge trapping in the medium-to-high gate voltage range ($V_d/2 \geq V_g \geq V_d$) during hot-carrier stressing of n-MOS transistors”, *IEEE Trans. Electron Devices*, 37(3), S. 744–754, März 1990.

- [14] S. Ogura, P. J. Tsang, W. W. Walker, D. L. Critchlow und J. F. Shepard, „Design and characteristics of the lightly doped drain-source (LDD) insulated gate field-effect transistor”, *IEEE Trans. Electron Devices*, ED-27(8), S. 1359–1367, August 1980.
- [15] R. R. Troutman, „VLSI limitations from drain-induced barrier lowering”, *IEEE Trans. Electron Devices*, ED-26(4), S. 461–469, April 1979.
- [16] R. Degraeve, B. Kaczer und G. Groeseneken, „Degradation and breakdown in thin oxide layers: mechanism, models and reliability prediction”, *Microelectronics Reliability*, 39 (19), S. 1445–1460, Oktober 1999.
- [17] T. H. Ning, „Silicon technology directions in the new millennium”, in *Proceedings of the IRPS*, S. 1–6, April 2000.
- [18] James H. Stathis, „Physical and predictive models of ultrathin oxide reliability in CMOS devices and circuits”, *IEEE Trans. Device Mat. Rel.*, 1(1), S. 43–59, März 2001.
- [19] P. Cappelletti, C. Golla, P. Olivo und E. Zanoni, *Flash Memories*, Kluwer Academic Publishers, Norwell, Massachusetts 02061 USA, 1999.
- [20] K. Naruke, S. Taguchi und M. Wada, „Stress induced leakage current limiting to scale down EEPROM tunnel oxide thickness”, *IEDM Tech. Dig.*, S. 424–427, Dezember 1988.
- [21] A. Modelli, „Reliability of thin dielectric for non-volatile applications”, *Microelectron. Eng.*, 48(1–4), S. 403–410, September 1999.
- [22] M. L. Green, E. P. Gusev, R. Degraeve und E. L. Garfunkel, „Ultrathin (<4 nm) SiO_2 and $Si-O-N$ gate dielectric layers for silicon microelectronics: Understanding the processing, structure, and physical and electrical limits”, *J. Appl. Phys.*, 90(5), S. 2057–2121, September 2001.
- [23] J. H. Stathis, „Reliability limits for the gate insulator in CMOS technology”, *IBM J. Res. Develop.*, 46(2–3), S. 265–286, März–Mai 2002.
- [24] K. Mistry, C. Allen, C. Auth, B. Beattie, D. Bergstrom, M. Bost, M. Brazier, M Buehler, A. Cappellani, R. Chau, C.-H. Choi, G. Ding, K. Fischer, T. Ghani, R. Grover, W. Han, D. Hanken, M. Hattendorf, J. He, J. Hicks, R. Huessner, D. Ingerly, P. Jain, R. James, L. Jong, S. Joshi, C. Kenyon, K. Kuhn, K. Lee, H. Liu, J. Maiz, B. McIntyre, P. Moon, J. Neiryneck, S. Pae, C. Parker, D. Parsons, C. Prasad, L. Pipes, M. Prince, P. Ranade, T. Reynolds, J. Sandford, L. Shifren, J. Sebastian, J. Seiple, D. Simon, S. Sivakumar, P. Smith, C. Thomas, T. Troeger, P. Vandervoorn, S. Williams und K. Zawadzki, „A 45nm logic technology with High-k+Metal gate transistors, strained silicon, 9 Cu interconnect layers, 193nm dry patterning, and 100% Pb -free packaging”, *IEDM Tech. Dig.*, S. 247–250, Dezember 2007.

- [25] J. W. McPherson und R. B. Khamankar, „Molekular model for intrinsic time-dependent dielectric breakdown in SiO_2 dielectrics and the reliability implications for hyper-thin gate oxide”, *Semicond. Sci. Technol.*, 15(5), S. 462–470, Mai 2000.
- [26] M. Depas, B. Vermeire, P. W. Mertens, R. L. van Meirhaeghe und M. M. Heyns, „Determination of tunneling parameters in ultra-thin oxide layer poly- $Si/SiO_2/Si$ structures”, *Solid State Electron.*, 38(8), S. 1465–1471, August 1995.
- [27] R. H. Fowler und L. Nordheim, „Electron emission in intense electric films”, in *Proceedings of the Royal Society of London, Series A*, 119(781), S. 173–181, Mai 1928.
- [28] M. Lenzlinger und E. H. Snow, „Fowler-Nordheim tunneling into thermally grown SiO_2 ”, *J. Appl. Phys.*, 40(1), S. 278–283, Januar 1969.
- [29] M. Lenski, T. Endoh und F. Masuoka, „Analytical modeling of stress-induced leakage currents in 5.1–9.6-nm-thick silicon-dioxide films based on two-step inelastic trap-assisted tunneling”, *J. Appl. Phys.*, 88(9), S. 5238–5245, November 2000.
- [30] D. Ielmini, A. S. Spinelli, M. A. Rigamonti und A. L. Lacaita, „Modeling of SILC based on electron and hole tunneling - part I: Transient effects”, *IEEE Trans. Electron Devices*, 47(6), S. 1258–1265, Juni 2000.
- [31] D. Ielmini, A. S. Spinelli, M. A. Rigamonti und A. L. Lacaita, „Modeling of SILC based on electron and hole tunneling - part II: Steady-state”, *IEEE Trans. Electron Devices*, 47(6), S. 1266–1272, Juni 2000.
- [32] J. Frenkel, „On the theory of electric breakdown of dielectrics and electronic semiconductors”, *Tech. Phys. USSR*, 5, S. 685, 1938.
- [33] J. Frenkel, „On pre-breakdown phenomena in insulators and electronic semiconductors”, *Phys. Rev.*, 54(8), S. 647–648, Oktober 1938.
- [34] H. D. B. Gottlob, T. Echtermeyer, T. Mollenhauer, J. K. Efavi, M. Schmidt, T. Wahlbrink, M. C. Lemme, H. Kurz, M. Czernohorsky, E. Bugiel, H.-J. Osten und A. Fissel, „CMOS integration of epitaxial Gd_2O_3 high- k gate dielectrics”, *Solid State Electron.*, 50(6), S. 979–985, Juni 2006.
- [35] B. E. Deal, „Standardized terminology for oxide charges associated with thermally oxidized silicon”, *IEEE Trans. Electron Devices*, ED-27(3), S. 606–608, März 1980.
- [36] D. M. Fleetwood, „„border traps” in MOS devices”, *IEEE Trans. Nucl. Sci.*, 39(2), S. 269–271, April 1992.
- [37] N. L. Cohen, R. E. Paulsen und M. H. White, „Observation and characterization of near-interface oxide traps with C-V techniques”, *IEEE Trans. Electron Devices*, 42(11), S. 2004–2009, November 1995.

- [38] M. J. Kirton und M. J. Uren, „Noise in solid-state microstructures: A new perspective on individual defects, interface states and low-frequency ($1/f$) noise”, *Advances in Physics*, 38(4), S. 367–468, Januar 1989.
- [39] D. M. Fleetwood, „Fast and slow border traps in MOS devices”, *IEEE Trans. Nucl. Sci.*, 43(3), S. 779–786, Juni 1996.
- [40] B. Djeddar, „What are these border traps: Introduced by radiation and seen by charge pumping technique?”, *IEEE Nuclear Science Symposium Conference Record*, 1, S. 234–239, November 2001.
- [41] T. N. Nicklaw, Z.-Y. Lu, D. M. Fleetwood, R. D. Schimpf und S. T. Pantelides, „The structure, properties, and dynamics of oxygen vacancies in amorphous SiO_2 ”, *IEEE Trans. Nucl. Sci.*, 49(6), S. 2667–2673, Dezember 2002.
- [42] P. E. Blöchl, „First-principles calculation of defects in oxygen-deficient silica exposed to hydrogen”, *Phys. Rev. B*, 62(10), S. 6158–6179, September 2000.
- [43] C. M. Nelson und R. A. Weeks, „Trapped electrons in irradiated quartz and silica: I, optical absorption”, *J. Am. Ceram.*, 43(8), S. 396–399, August 1960.
- [44] R. A. Weeks und C. M. Nelson, „Trapped electrons in irradiated quartz and silica: II, electron spin resonance”, *J. Am. Ceram.*, 43(8), S. 399–404, August 1960.
- [45] P. E. Blöchl und J. H. Stathis, „Hydrogen electrochemistry and stress-induced leakage current in silica”, *Phys. Rev. Lett.*, 83(2), S. 372–375, Juli 1999.
- [46] J. K. Rudra, W. B. Fowler und F. J. Feigl, „Model for the E'_2 center in alpha quartz”, *Phys. Rev. Lett.*, 55(23), S. 2614–2617, Dezember 1985.
- [47] M. F. Beug, *Charakterisierung von EEPROM Tunneloxiden mittels transientser Strom- und Kapazitätsmessungen*, Dissertation, Universität Hannover, 2005.
- [48] W. Füssel, M. Schmidt, H. Angermann, G. Mende und H. Flietner, „Defects at the Si/SiO_2 interface: Their nature and behaviour in technological processes and stress”, *Nuclear Instruments and Methods in Physics Research Section A*, 377(2–3), S. 177–183, August 1996.
- [49] F. J. Himpsel, F. R. McFeely, A. Taleb-Ibrahimi, J. A. Yarmoff und G. Hollinger, „Microscopic structure of the SiO_2/Si interface”, *Phys. Rev. B*, 38(9), S. 6084–6096, September 1988.
- [50] J. H. Stathis und E. Cartier, „Atomic hydrogen reactions with P_b centers at the (100) Si/SiO_2 interface”, *Phys. Rev. Lett.*, 72(17), S. 2745–2748, April 1994.

- [51] E. H. Poindexter und P. J. Caplan, „Electron spin resonance of inherent and process induced defects near the Si/SiO_2 interface of oxidized silicon wafers”, *J. Vac. Sci. Technol. A*, 6(3), S. 1352–1357, März 1988.
- [52] R. Castagné und A. Vapaille, „Description of the SiO_2 - Si interface properties by means of very low frequency MOS capacitance measurements”, *Surf. Sci.*, 28(1), S. 157–193, November 1971.
- [53] J. S. Brugler und P. G. A. Jespers, „Charge pumping in MOS devices”, *IEEE Trans. Electron Devices*, ED-16(3), S. 297–302, März 1969.
- [54] L. D. Thanh und P. Balk, „Elimination and generation of Si - SiO_2 interface traps by low temperature hydrogen annealing”, *J. Electrochem. Soc.*, 135(7), S. 1797–1801, Juli 1988.
- [55] E. Arnold, J. Ladell und G. Abowitz, „Crystallographic symmetry of surface state density in thermally oxidized silicon”, *Appl. Phys. Lett.*, 13(12), S. 413–416, Dezember 1968.
- [56] D. J. DiMaria, D. Arnold und E. Cartier, „Degradation and breakdown of silicon dioxide films on silicon”, *Appl. Phys. Lett.*, 61(19), S. 2329–2331, November 1992.
- [57] E. Harari, „Dielectric breakdown in electrically stressed thin films of thermal SiO_2 ”, *J. Appl. Phys.*, 49(4), S. 2478–2489, April 1978.
- [58] P. Olivo, T. N. Nguyen und B. Ricco, „High-field-induced degradation in ultra-thin SiO_2 films”, *IEEE Trans. Electron Devices*, ED-35(12), S. 2259–2267, Dezember 1988.
- [59] K. F. Schuegraf und C. Hu, „Metal-oxide-semiconductor field-effect-transistor substrate current during Fowler-Nordheim tunneling stress and silicon dioxide reliability”, *J. Appl. Phys.*, 76(6), S. 3695–3700, September 1994.
- [60] D. J. DiMaria und J. W. Stasiak, „Trap creation in silicon dioxide produced by hot electrons”, *J. Appl. Phys.*, 65(6), S. 2342–2356, März 1989.
- [61] D. J. DiMaria, E. Cartier und D. A. Buchanan, „Anode hole injection and trapping in silicon dioxide”, *J. Appl. Phys.*, 80(1), S. 304–317, Juli 1996.
- [62] D. J. DiMaria, E. Cartier und D. Arnold, „Impact ionization, trap creation, degradation, and breakdown in silicon dioxide films on silicon”, *J. Appl. Phys.*, 73(7), S. 3367–3384, April 1993.
- [63] D. J. DiMaria, D. Arnold und E. Cartier, „Impact ionization and positive charge formation in silicon dioxide films on silicon”, *Appl. Phys. Lett.*, 60(17), S. 2118–2120, April 1992.
- [64] I. Yoshii, K. Hama und K. Hashimoto, „Role of hydrogen at poly- Si/SiO_2 interface in trap generation by substrate hot-electron injection”, in *Proceedings of the IRPS*, S. 136–140, 1992.

- [65] E. H. Nicollian und J. R. Brews, *MOS (Metal Oxide Semiconductor) Physics and Technology*, John Wiley & Sons, New York, 1982.
- [66] S. Bernardini, P. Masson, M. Houssa und F. Lalande, „Origin and repartition of the oxide fixed charges generated by electrical stress in memory tunnel oxide”, *Appl. Phys. Lett.*, 84(21), S. 4251–4253, Mai 2004.
- [67] E. H. Poindexter und P. J. Caplan, „Characterizing of *Si/SiO₂* interface defects by electron spin resonance”, *Prog. Surf. Sci.*, 14(3), S. 201–294, 1983.
- [68] R. Gale, F. J. Feigl, C. W. Magee und D. R. Young, „Hydrogen migration under avalanche injection of electrons in *Si* metal-oxide-semiconductor capacitors”, *J. Appl. Phys.*, 54(12), S. 6938–6942, Dezember 1983.
- [69] K. Hess, J. Lee, Z. Chen, J. W. Lyding, Y.-K. Kim, B.-S. Kim, Y.-H. Lee, Y.-W. Kim und K.-P. Suh, „An alternative interpretation of hot electron interface degradation in NMOSFETs: Isotope results irreconcilable with major defect generation by holes?”, *IEEE Trans. Electron Devices*, 46(9), S. 1914–1916, September 1999.
- [70] R. Degraeve, G. Groeseneken, I. De Wolf und H. E. Maes, „Oxide and interface degradation and breakdown under medium and high field injection conditions: A correlation study”, *Microelectron. Eng.*, 28(1–4), S. 313–316, Juni 1995.
- [71] D. J. DiMaria und J. H. Stathis, „Anode hole injection, defect generation, and breakdown in ultrathin silicon dioxide films”, *J. Appl. Phys.*, 89(9), S. 5015–5024, Mai 2001.
- [72] S. Mahapatra, P. B. Kumar und M. A. Alam, „Investigation and modeling of interface and bulk trap generation during negative bias temperature instability of p-MOSFETs”, *IEEE Trans. Electron Devices*, 51(9), S. 1371–1379, September 2004.
- [73] E. Cartier, J. H. Stathis und D. A. Buchanan, „Passivation and depassivation of silicon dangling bonds at the *Si/SiO₂* interface by atomic hydrogen”, *Appl. Phys. Lett.*, 63(11), S. 1510–1512, September 1993.
- [74] J. W. Lyding, K. Hess und I. C. Kizilyalli, „Reduction of hot electron degradation in metal oxide semiconductor transistors by deuterium processing”, *Appl. Phys. Lett.*, 68(18), S. 2526–2528, April 1996.
- [75] J. Wu, E. Rosenbaum, B. MacDonald, E. Li, J. Tao, B. Tracy und P. Fang, „Anode hole injection versus hydrogen release: The mechanism for gate oxide breakdown”, in *Proceedings of the IRPS*, S. 27–32, April 2000.
- [76] K. Hess, I. C. Kizilyalli und J. W. Lyding, „Giant isotope effect in hot electron degradation of metal oxide silicon devices”, *IEEE Trans. Electron Devices*, 45, 1998.

-
- [77] H. Kim und H. Hwang, „High-quality ultrathin gate oxide prepared by oxidation in D_2O ”, *Appl. Phys. Lett.*, 74(5), S. 709–710, Februar 1999.
- [78] Y. Mitani, H. Satake, H. Itoh und A. Toriumi, „Highly reliable gate oxide under Fowler-Nordheim electron injection by deuterium pyrogenic oxidation and deuterated poly-*Si* deposition”, *IEDM Tech. Dig.*, S. 343–346, Dezember 2000.
- [79] Jeff D. Bude D. Esseni und L. Selmi, „On interface and oxide degradation in VLSI MOSFETs-part I: Deuterium effect in CHE-stress regime”, *IEEE Trans. Electron Devices*, 49(2), S. 247–253, Februar 2002.
- [80] Jeff D. Bude D. Esseni und L. Selmi, „On interface and oxide degradation in VLSI MOSFETs-part II: Fowler-Nordheim stress regime”, *IEEE Trans. Electron Devices*, 49(2), S. 254–263, Februar 2002.
- [81] Y. Nissan-Cohen und T. Gorczyca, „The effect of hydrogen on trap generation, positive charge trapping, and time-dependent dielectric breakdown of gate oxides”, *IEEE Electron Device Lett.*, 9(6), S. 287–289, Juni 1988.
- [82] T. Ohmi, K. Nakamura und K. Makihara, „Highly-reliable ultra-thin oxide formation using hydrogen-radical-balanced steam oxidation technology”, in *Proceedings of the IRPS*, S. 161–166, April 1994.
- [83] I. C. Chen, S. Holland, K. K. Young, C. Chang und C. Hu, „Substrate hole current and oxide breakdown”, *Appl. Phys. Lett.*, 49(11), S. 669–671, September 1986.
- [84] H. Uchida und T. Ajioka, „Electron trap center generation due to hole trapping in SiO_2 under Fowler-Nordheim tunneling stress”, *Appl. Phys. Lett.*, 51(6), S. 433–435, August 1987.
- [85] K. R. Mistry, D. B. Krakauer und B. S. Doyle, „Impact of snapback-induced hole injection on gate oxide reliability of N-MOSFET’s”, *IEEE Electron Device Lett.*, 11(10), S. 460–462, Oktober 1990.
- [86] S. Ogawa, N. Shiono und M. Shimaya, „Neutral electron trap generation in SiO_2 by hot holes”, *Appl. Phys. Lett.*, 56(14), S. 1329–1331, April 1990.
- [87] E. Rosenbaum, R. Rofan und C. Hu, „Effect of hot-carrier injection on n- and pMOSFET gate oxide integrity”, *IEEE Electron Device Lett.*, 12(11), S. 599–601, November 1991.
- [88] K. Awazu und H. Kawazoe, „Strained *Si-O-Si* bonds in amorphous SiO_2 materials: A family member of active centers in radio, photo, and chemical responses”, *J. Appl. Phys.*, 94(10), S. 6243–6262, November 2003.
- [89] S. K. Lai, „Interface trap generation in silicon dioxide when electrons are captured by trapped holes”, *J. Appl. Phys.*, 54(5), S. 2540–2546, Mai 1983.
-

- [90] P. M. Lenahan und P. V. Dressendorfer, „Hole traps and trivalent silicon centers in metal/oxide/silicon devices”, *J. Appl. Phys.*, 55(10), S. 3495–3499, Mai 1984.
- [91] F. J. Grunthaler, B. F. Lewis, N. Zamini, J. Maserjian und A. Madhukar, „XPS studies of structure-induced radiation effects at the *Si-SiO₂* interface”, *IEEE Trans. Nucl. Sci.*, NS-27(6), S. 1640–1646, Dezember 1980.
- [92] Y. Nissan-Cohen, J. Shappir und D. Frohman-Bentchkowsky, „Trap generation and occupation dynamics in *SiO₂* under charge injection stress”, *J. Appl. Phys.*, 60(6), S. 2024–2035, September 1986.
- [93] Y. Lu und C.-T. Sah, „Two pathways of positive oxide-charge buildup during electron tunneling into silicon dioxide film”, *J. Appl. Phys.*, 76(8), S. 4724–4727, Oktober 1994.
- [94] J. D. Bude, B. E. Weir und P. J. Silverman, „Explanation of stress-induced damage in thin oxides”, *IEDM Tech. Dig.*, S. 179–182, Dezember 1998.
- [95] P. E. Nicollian, W. R. Hunter und J. C. Hu, „Experimental evidence for voltage driven breakdown models in ultrathin gate oxides”, in *Proceedings of the IRPS*, S. 7–15, April 2000.
- [96] K. O. Jeppson und C. M. Svensson, „Negative bias stress of MOS devices at high electric fields and degradation of MNOS devices”, *J. Appl. Phys.*, 48(5), S. 2004–2014, Mai 1977.
- [97] B. E. Deal, M. Sklar, A. S. Grove und E. H. Snow, „Characteristics of the surface-state charge (Q_{ss}) of thermally oxidized silicon”, *J. Electrochem. Soc.*, 114(3), S. 266–274, März 1967.
- [98] B. E. Deal, „The current understanding of charges in the thermally oxidized silicon structure”, *J. Electrochem. Soc.*, 121(6), S. 198C–205C, Juni 1974.
- [99] C. E. Blat, E. H. Nicollian und E. H. Poindexter, „Mechanism of negative-bias-temperature instability”, *J. Appl. Phys.*, 69(3), S. 1712–1720, Februar 1991.
- [100] M. Makabe, T. Kubota und T. Kitano, „Bias-temperature degradation on pMOSFETs: mechanism and suppression”, in *Proceedings of the IRPS*, S. 205–209, April 2000.
- [101] D. K. Schroder und J. A. Babcock, „Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing”, *J. Appl. Phys.*, 94(1), S. 1–18, Juli 2003.
- [102] C. Choi, C.-S. Kang, C. Y. Kang, S. J. Rhee, M. S. Akbar, S. A. Krishnan, M. Zhang und J. C. Lee, „Positive bias temperature instability effects of *Hf*-based nMOSFETs with various nitrogen and silicon profiles”, *IEEE Electron Device Lett.*, 26(1), S. 32–34, Januar 2005.

-
- [103] F. Crupi, C. Pace, G. Cocorullo, G. Groeseneken, M. Aoulaiche und M. Houssa, „Positive bias temperature instability in nMOSFETs with ultra-thin Hf -silicate gate dielectrics”, *Microelectron. Eng.*, 80, S. 130–133, Juni 2005.
- [104] C.-H. Liu, M. T. Lee, C.-Y. Lin, J. Chen, Y. T. Loh, F.-T. Liou, K. Schroefer, A. A. Katsetos, Z. Yang, N. Rovedo, T. B. Hook, C. Wann und T.-C. Chen, „Mechanism of threshold voltage shift (ΔV_{th}) caused by negative bias temperature instability (NBTI) in deep submicron pMOSFETs”, *Jap. J. Appl. Phys.*, 41, Part 1(4B), S. 2423–2425, April 2002.
- [105] A. T. Krishnan, P. Nicollian S. Chakravarthi, V. Reddy und S. Krishnan, „Negative bias temperature instability mechanism: the role of molecular hydrogen”, *Appl. Phys. Lett.*, 88(15), S. 153518, April 2006.
- [106] H. Küflüoğlu und M. A. Alam, „A generalized reaction-diffusion model with explicit H_2 dynamics for negative-bias temperature-instability (NBTI) degradation”, *IEEE Trans. Electron Devices*, 54(5), S. 1101–1107, Mai 2007.
- [107] M. Ershov, S. Saxena, H. Karbasi, S. Winters, S. Minehane, J. Babcock, R. Lindley, P. Clifton, M. Redford und A. Shibkov, „Dynamic recovery of negative bias temperature instability in p-type metal-oxide-semiconductor field-effect transistors”, *Appl. Phys. Lett.*, 83(8), S. 1647–1649, August 2003.
- [108] S. Mahapatra, D. Saha, D. Varghese und P. B. Kumar, „On the generation and recovery of interface traps in MOSFETs subjected to NBTI, FN, and HCI stress”, *IEEE Trans. Electron Devices*, 53(7), S. 1583–1592, Juli 2006.
- [109] B. Kaczer, V. Arkhipov, R. Degraeve, N. Collaert, G. Groeseneken und M. Goodwin, „Temperature dependence of the negative bias temperature instability in the framework of dispersive transport”, *Appl. Phys. Lett.*, 86(14), S. 143506, April 2005.
- [110] D. S. Ang und S. Wang, „On the non-Arrhenius behavior of negative-bias temperature instability”, *Appl. Phys. Lett.*, 88(9), S. 093506, Februar 2006.
- [111] S. Ogawa, M. Shimaya und N. Shiono, „Interface-trap generation at ultrathin SiO_2 (4–6 nm)- Si interfaces during negative-bias temperature aging”, *J. Appl. Phys.*, 77(3), S. 1137–1148, Februar 1995.
- [112] C. Schlünder, R. Brederlow, P. Wiczorek, C. Dahl, J. Holz, M. Röhner, S. Kessel, V. Herold, K. Goser, W. Weber und R. Thewes, „Trapping mechanisms in negative bias temperature stressed p-MOSFETs”, *Microelectronics Reliability*, 39(6–7), S. 821–826, Juni–Juli 1999.
-

- [113] M. Houssa, V. V. Afanas'ev, A. Stesmans, M. Aoulaiche, G. Groeseneken und M. M. Heyns, „Insights on the physical mechanism behind negative bias temperature instabilities”, *Appl. Phys. Lett.*, 90(4), S. 043505, Januar 2007.
- [114] A. Stesmans, „Dissociation kinetics of hydrogen-passivated P_b defects at the (111) Si/SiO_2 interface”, *Phys. Rev. B*, 61(12), S. 8393–8403, März 2000.
- [115] L. Tsetseris, X. J. Zhou, D. M. Fleetwood, R. D. Schrimpf und S. T. Pantelides, „Physical mechanisms of negative-bias temperature instability”, *Appl. Phys. Lett.*, 86(14), S. 142103, April 2005.
- [116] S.K. Lai, J. Lee und V. K. Dham, „Electrical properties of nitrated-oxide systems for use in gate dielectrics and EEPROM”, *IEDM Tech. Dig.*, S. 190–193, 1983.
- [117] R. Jayaraman und C. G. Sodini, „ $1/f$ noise interpretation of the effect of gate oxide nitridation and reoxidation in dielectric traps”, *IEEE Trans. Electron Devices*, 37(1), S. 305–309, Januar 1990.
- [118] D. J. DiMaria und J. H. Stathis, „Trapping and trap creation studies on nitrated and reoxidized-nitrated silicon dioxide films on silicon”, *J. Appl. Phys.*, 70(3), S. 1500–1509, August 1991.
- [119] Z. Liu, H.-J. Wann, P. K. Ko, C. Hu und Y. C. Cheng, „Improvement of charge trapping characteristics of N_2O -annealed and reoxidized N_2O -annealed thin oxides”, *IEEE Electron Device Lett.*, 13(10), S. 519–521, Oktober 1992.
- [120] T. Ito, T. Nakamura und H. Ishikawa, „Advantages of thermal nitride and nitroxide gate films in VLSI process”, *IEEE Trans. Electron Devices*, ED-29(4), S. 498–502, April 1982.
- [121] J. Ahn, J. Kim, G. Q. Lo und D. L. Kwong, „Suppression of stress-induced leakage current in ultrathin N_2O oxides”, *Appl. Phys. Lett.*, 60(22), S. 2809–2811, Juni 1992.
- [122] X. Guo und T. P. Ma, „Tunneling leakage current in oxynitride: Dependence on Oxygen/Nitrogen content”, *IEEE Electron Device Lett.*, 19(6), S. 207–209, Juni 1998.
- [123] C. Gerardi, M. Melanotte, S. Lombardo, M. Alessandri, B. Crivelli und R. Zonca, „Effects of nitridation by nitric oxide on the leakage current of thin SiO_2 gate oxides”, *J. Appl. Phys.*, 87(1), S. 498–501, Januar 2000.
- [124] M. Herden, A. J. Bauer, M. Beichele und H. Ryssel, „Suppression of boron penetration through thin gate oxides by nitrogen implantation into the gate electrode”, *Solid State Electron.*, 45(8), S. 1251–1256, August 2001.

-
- [125] S. Strobel, A. J. Bauer, M. Beichele und H. Ryssel, „Suppression of boron penetration through thin gate oxides by nitrogen implantation into the gate electrode in PMOS devices”, *Microelectronics Reliability*, 41(7), S. 1085–1088, Juli 2001.
- [126] E. P. Gusev, H.-C. Lu, E. L. Garfunkel, T. Gustafsson und M. L. Green, „Silicon oxynitrides”, *Encyclopedia of Materials: Science and Technology*, S. 8576–8584, 2004.
- [127] F. L. Terry, R. J. Aucoin, M. L. Naiman und S. D. Senturia, „Radiation effects in nitrided oxides”, *IEEE Electron Device Lett.*, EDL-4(6), S. 191–193, Juni 1983.
- [128] T.-Y. Huang, F.-C. Jong, T.-S. Chao, H.-C. Lin, L.-Y. Leu, K. Young, C.-H. Lin und K. Y. Chin, „Improving radiation hardness of EEPROM/Flash cell by N_2O annealing”, *IEEE Electron Device Lett.*, 19(7), S. 256–258, Juli 1998.
- [129] T. Sasaki, K. Kuwazawa, K. Tanaka, J. Kato und D. Kwong, „Engineering of nitrogen profile in an ultrathin gate insulator to improve transistor performance and NBTI”, *IEEE Electron Device Lett.*, 24(3), S. 150–152, März 2003.
- [130] T. Hori, H. Iwasaki und K. Tsuji, „Electrical and physical properties of ultrathin reoxidized nitrided oxides prepared by rapid thermal processing”, *IEEE Electron Device Lett.*, 36(2), S. 340–350, Februar 1989.
- [131] A. Uchiyama, H. Fukuda, T. Hayashi, T. Iwabuchi und S. Ohno, „High performance dual-gate sub-halfmicron CMOSFETs with 6 nm-thick nitrided SiO_2 films in an N_2O ambient”, *IEDM Tech. Dig.*, S. 425–428, Dezember 1990.
- [132] P. T. Lai, J. P. Xu und Y. C. Cheng, „Interface properties of NO -annealed N_2O -grown oxynitride”, *IEEE Trans. Electron Devices*, 46(12), S. 2311–2314, Dezember 1999.
- [133] C. T. Wu, A. Mieczkowski, R. S. Ridley Sr., G. Dolny, T. Grebs, J. Linn und J. Ruzyllo, „Effect of nitridation on the reliability of thick gate oxides”, *Microelectronics Reliability*, 43(1), S. 43–47, Januar 2003.
- [134] H. Wong und V. A. Gritsenko, „Dielectric traps in amorphous silicon oxynitride”, *IEDM Tech. Dig.*, S. 132–139, Juni 2001.
- [135] N. Kimizuka, T. Yamamoto, K. Imai, T. Iizuka, C. T. Liu, R. C. Keller und T. Horiuchi, „NBTI enhancement by nitrogen incorporation into ultrathin gate oxide for 0.10- μm gate CMOS generation”, in *Proceedings of the Symp. on VLSI Tech.*, S. 92–93, Juni 2000.
- [136] Y. J. Lee, Y. C. Tang, M. H. Wu, T. S. Chao, P. T. Ho, D. Lai, W. L. Yang und T. Y. Huang, „NBTI effects of pMOSFETs with different nitrogen dose implantation”, in *Proceedings of the IRPS*, S. 681–682, April 2004.
-

- [137] C. H. Liu, M. T. Lee, C.-Y. Lin, J. Chen, K. Schuefer, J. Brighten, N. Rovedo, T. B. Hook, M. V. Khare, S.-F. Huang, C. Wann, T.-C. Chen und T. H. Ning, „Mechanism and process dependence of negative bias temperature instability (NBTI) for pMOSFETs with ultrathin gate dielectrics”, *IEDM Tech. Dig.*, S. 861–864, Dezember 2001.
- [138] Y. Mitani, M. Nagamine, H. Satake und A. Toriumi, „NBTI mechanism in ultra-thin gate dielectric - nitrogen-originated mechanism in $SiON$ -”, *IEDM Tech. Dig.*, S. 509–512, Dezember 2002.
- [139] J. Ushio, T. Maruizumi und K. Kushida-Abdelghafar, „Interface structures generated by negative-bias temperature instability in Si/SiO_2 and Si/SiO_xN_y interfaces”, *Appl. Phys. Lett.*, 81(10), S. 1818–1820, September 2002.
- [140] V. Huard und M. Denais, „Hole trapping effect on methodology for DC and AC negative bias temperature instability measurements in pMOS transistors”, in *Proceedings of the IRPS*, S. 40–45, April 2004.
- [141] B. Kaczer, V. Arkhipov, M. Jurczak und G. Groeseneken, „Negative bias temperature instability (NBTI) in SiO_2 and $SiON$ gate dielectrics understood through disorder-controlled kinetics”, *Microelectron. Eng.*, 80, S. 122–125, Juni 2005.
- [142] J. B. Yang, T. P. Chen, S. S. Tan und L. Chan, „Analytical reaction-diffusion model and the modeling of nitrogen-enhanced negative bias temperature instability”, *Appl. Phys. Lett.*, 88(17), S. 172109, April 2006.
- [143] N. K. Jha und V. Ramgopal, „A new oxide trap-assisted NBTI degradation model”, *IEEE Electron Device Lett.*, 26(9), S. 687–689, September 2005.
- [144] S. Ogawa, M. Shimaya und N. Shiono, „Impact of negative-bias temperature instability on the lifetime of single-gate CMOS structures with ultrathin (4–6 nm) gate oxides”, *Jap. J. Appl. Phys.*, 35(2B), S. 1484–1490, Februar 1996.
- [145] T. Nigam und E. B. Harris, „Lifetime enhancement under high frequency NBTI measured on ring oscillators”, in *Proceedings of the IRPS*, S. 289–293, März 2006.
- [146] B. Eitan, P. Pavan, I. Bloom, E. Aloni, A. Frommer und D. Finzi, „NROM: A novel localized trapping, 2-bit nonvolatile memory cell”, *IEEE Electron Device Lett.*, 21(11), S. 543–545, November 2000.
- [147] B. De Salvo, C. Garardi, R. van Schaijk, S. A. Lombardo, D. Corso, C. Plantamura, S. Serafino, G. Ammendola, M. van Duuren, P. Goarin, W. Y. Mei, K. van der Jeugd, T. Baron, M. Gély, P. Mur und S. Deleonibus, „Performance and reliability features of advanced nonvolatile memories based on discrete traps (silicon nanocrystals, SONOS)”, *IEEE Trans. Device Mat. Rel.*, 4(3), S. 377–389, September 2004.

- [148] H.-T. Lue, T.-H. Hsu, M.-T. Wu, K.-Y. Hsieh, R. Liu und C.-Y. Lu, „Studies of the reverse read method and second-bit effect of 2-bit/cell nitride-trapping device by quasi-two-dimensional model”, *IEEE Trans. Electron Devices*, 53(1), S. 119–125, Januar 2006.
- [149] B. Eitan und D. Frohman-Bentchkowsky, „Hot-electron injection into the oxide in n-channel MOS devices”, *IEEE Trans. Electron Devices*, ED-28(3), S. 328–340, März 1981.
- [150] A. Furnémont, M. Rosmeulen, K. van der Zanden, J. van Houdt, K. De Meyer und H. Maes, „New operating mode based on electron/hole profile matching in nitride-based nonvolatile memories”, *IEEE Electron Device Lett.*, 28(4), S. 276–278, April 2007.
- [151] G. Ingrosso, L. Selmi und E. Sangiorgi, „Monte carlo simulation of program and erase charge distributions in NROMTM devices”, in *Proc. ESSDERC*, S. 187–190, September 2002.
- [152] A. Shappir, D. Levy, Y. Shacham-Diamand, E. Lusky, I. Bloom und B. Eitan, „Spatial characterization of localized charge trapping and charge redistribution in the NROM device”, *Solid State Electron.*, 48(9), S. 1489–1495, September 2004.
- [153] D. Fuks, A. Kiv, Y. Roizin, M. Gutman, R. Avichail-Bibi und T. Maximova, „The nature of HT V_t shift in NROM memory transistors”, *IEEE Trans. Electron Devices*, 53(2), S. 304–313, Februar 2006.
- [154] L. Larcher, P. Pavan und B. Eitan, „On the physical mechanism of the NROM memory erase”, *IEEE Trans. Electron Devices*, 51(10), S. 1593–1599, Oktober 2004.
- [155] E. G. Stein v. Kamienski, M. Isler, T. Mikolajick, C. Ludwig, N. Schulze, N. Nagel, S. Riedel, J. Willer und K.-H. Küsters, „An overview of Twin-FlashTM technology”, in *Proceedings of the Non-Volatile Memory Technology Symposium*, S. 5–10, November 2005.
- [156] E. Lusky, Y. Shacham-Diamand, I. Bloom und B. Eitan, „Electrons retention model for localized charge in oxide–nitride–oxide (ONO) dielectric”, *IEEE Electron Device Lett.*, 23(9), S. 556–558, September 2002.
- [157] A. Shappir, E. Lusky, G. Cohen, I. Bloom, M. Janai und B. Eitan, „The Two-Bit NROMTM reliability”, *IEEE Trans. Device Mat. Rel.*, 4(3), S. 397–403, September 2004.
- [158] R. Bez, E. Camerlenghi, A. Modelli und A. Visconti, „Introduction to flash memory”, *Proceedings of the IEEE*, 91(4), S. 489–502, April 2003.
- [159] S. J. Wrazien, Y. Wang, B. M. Khan und M. H. White, „Characterizing damage to ONO dielectrics induced during programming SONOS/NROMTM non-volatile semiconductor memory (NVSM) devices”, *Solid State Electron.*, 48(10–11), S. 2035–2039, Oktober–November 2004.

- [160] L. Perniola, S. Bernardini, G. Iannaccone, B. De Salvo, G. Ghibaudo, P. Masson und C. Gerardi, „Electrostatic effect of localised charge in dual bit memory cells with discrete traps”, in *Proc. ESSDERC*, S. 249–252, September 2004.
- [161] D. K. Schroder, *Semiconductor Material and Device Characterization*, John Wiley & Sons, New York, Second Edition, 1998.
- [162] G. Groeseneken, H. E. Maes, N. Beltran und R. F. de Keersmaecker, „A reliable approach to charge-pumping measurements in MOS transistors”, *IEEE Trans. Electron Devices*, ED-31(1), S. 42–53, Januar 1984.
- [163] A. van der Ziel, *Noise in Solid State Devices and Circuits*, John Wiley & Sons, New York, First Edition, 1986.
- [164] M. Schulz und N. M. Johnson, „Transient capacitance measurements of hole emission from interface states in MOS structures”, *Appl. Phys. Lett.*, 31(9), S. 622–625, November 1977.
- [165] M. Schulz und E. Klausmann, „Transient capacitance measurements of interface states on the intentionally contaminated *Si-SiO₂* interface”, *Appl. Phys. A*, 18(2), S. 169–175, Februar 1979.
- [166] T. Endoh T. Chiba, H. Sakuraba, M. Lenski und F. Masuoka, „A quantitative analysis of stress-induced leakage currents and extraction of trap properties in 6.8 nm ultrathin silicon dioxide films”, *J. Appl. Phys.*, 86(4), S. 2095–2099, August 1999.
- [167] M. Baklanov, M. Green und K. Maex, *Dielectric Films for Advanced Microelectronics*, John Wiley & Sons, First Edition, 2007.
- [168] U. Cilingiroglu, „A general model for interface-trap charge-pumping effects in MOS devices”, *Solid State Electron.*, 28(11), S. 1127–1141, November 1985.
- [169] J. G. Simmons und L. S. Wei, „Theory of dynamic charge current and capacitance characteristics in MIS systems containing distributed surface traps”, *Solid State Electron.*, 16(1), S. 53–66, Januar 1973.
- [170] P. Heremans, J. Witters, G. Groeseneken und H. E. Maes, „Analysis of the charge pumping technique and its application for the evaluation of MOSFET degradation”, *IEEE Trans. Electron Devices*, 36(7), S. 1318–1335, Juli 1989.
- [171] A. B. M. Elliot, „The use of charge pumping currents to measure surface state densities in MOS transistors”, *Solid State Electron.*, 19(3), S. 241–247, März 1976.
- [172] J. F. Zhang, H. K. Sii, G. Groeseneken und R. Degraeve, „Degradation of oxides and oxynitrides under hot hole stress”, *IEEE Trans. Electron Devices*, 47(2), S. 378–386, Februar 2000.

- [173] D. Vuillaume, J.-C. Marchetaux, P.-E. Lippens, A. Bravaix und A. Boudou, „A coupled study by floating-gate and charge-pumping techniques of hot-carrier-induced defects in submicrometer LDD n-MOSFET's”, *IEEE Trans. Electron Devices*, 40(4), S. 773–781, April 1993.
- [174] C. H. Ling, S. E. Tan und D. S. Ang, „A study of hot carrier degradation in MOSFET's by gate capacitance and charge pumping current”, *IEEE Trans. Electron Devices*, 42(7), S. 1321–1328, Juli 1995.
- [175] J. Chan, E.-X. Zhao, S. P. Sinha und A. Marathe, „Charge pumping study of hot-carrier induced degradation of sub-100nm partially depleted SOI MOSFETs”, *IEEE Intl. SOI Conference*, S. 43–44, Oktober 2002.
- [176] C.-M. Lai, Y.-K. Fang, W.-K. Yeh, C. T. Lin und T. H. Chou, „The investigation of post-annealing-induced defects behavior on 90-nm in halo nMOSFETs with low-frequency noise and charge-pumping measuring”, *IEEE Electron Device Lett.*, 28(2), S. 142–144, Februar 2007.
- [177] G. V. Groeseneken, I. De Wolf, R. Bellens und H. E. Maes, „Observation of single interface traps in submicron MOSFET's by charge pumping”, *IEEE Trans. Electron Devices*, 43(6), S. 940–945, Juni 1996.
- [178] L. Militaru, P. Masson, V. Célibert und C. Leroux, „Single trap characterization in 50nm MOS transistors by charge pumping measurements”, in *Proceedings of the 31th European Solid State Device Research Conference, Nuremberg, Germany*, S. 459, 2001.
- [179] M. Tsuchiaki, H. Hara, T. Morimoto und H. Iwai, „A new charge pumping method for determining the spatial distribution of hot-carrier-induced fixed charge in p-MOSFET's”, *IEEE Trans. Electron Devices*, 40(10), S. 1768–1779, Oktober 1993.
- [180] C. Chen und T.-P. Ma, „Direct lateral profiling of hot-carrier-induced oxide charge and interface traps in thin gate MOSFET's”, *IEEE Trans. Electron Devices*, 45(2), S. 512–520, Februar 1998.
- [181] Y.-L. Chu, D.-W. Lin und C.-Y. Wu, „A new charge-pumping technique for profiling the interface-states and oxide-trapped charges in MOSFET's”, *IEEE Trans. Electron Devices*, 47(2), S. 348–353, Februar 2000.
- [182] C.-C. Cheng, J. F. Lin, T. Wang, T. H. Hsieh, J. T. Tzeng, Y. C. Jong, R. S. Liou, S. C. Pan und S. L. Hsu, „Physics and characterization of various hot-carrier degradation modes in LD MOS by using a three-region charge-pumping technique”, *IEEE Trans. Device Mat. Rel.*, 6(3), S. 358–363, September 2006.

- [183] W. Chen und T.-P. Ma, „A new technique for measuring lateral distribution of oxide charge and interface traps near MOSFET junctions”, *IEEE Electron Device Lett.*, 12(7), S. 393–395, Juli 1991.
- [184] W. Chen, A. Balasinski und T.-P. Ma, „Lateral profiling of oxide charge and interface traps near MOSFET junctions”, *IEEE Trans. Electron Devices*, 40(1), S. 187–196, Januar 1993.
- [185] A. Melik-Martirosian und T.-P. Ma, „Lateral profiling of interface traps and oxide charge in MOSFET devices: Charge pumping versus DCIV”, *IEEE Trans. Electron Devices*, 48(10), S. 2303–2309, Oktober 2001.
- [186] H.-H. Li, Y.-L. Chu und C.-Y. Wu, „A novel charge-pumping method for extracting the lateral distributions of interface-trap and effective oxide-trapped charge densities in MOSFET devices”, *IEEE Trans. Electron Devices*, 44(5), S. 782–791, Mai 1997.
- [187] M. Rosmeulen, L. Breuil, M. Lorenzini, L. Haspeslagh, J. Van Houdt und K. De Meyer, „Characterization of the spatial charge distribution in local charge-trapping memory devices using the charge-pumping technique”, *Solid State Electron.*, 48(9), S. 1525–1530, September 2004.
- [188] D. Bauza, „Rigorous analysis of two-level charge pumping: Application to the extraction of interface trap concentration versus energy profiles in metal-oxide-semiconductor transistors”, *J. Appl. Phys.*, 94(5), S. 3239–3248, September 2003.
- [189] N. C. Das und V. Nathan, „Hot carrier degradation in MOSFETS: a charge pumping study”, *Semicond. Sci. Technol.*, 8(4), S. 549–554, April 1993.
- [190] R. E. Stahlbush, A. H. Edwards, D. L. Griscom und B. J. Mrstik, „Post-irradiation cracking of H_2 and formation of interface states in irradiated metal-oxide-semiconductor field-effect transistors”, *J. Appl. Phys.*, 73(2), S. 658–667, Januar 1993.
- [191] W. D. Zhang, J. F. Zhang, M. J. Uren, G. Groeseneken, R. Degraeve, M. Lalor und D. Burton, „On the interface states generated under different stress conditions”, *Appl. Phys. Lett.*, 79(19), S. 3092–3094, November 2001.
- [192] R. E. Paulsen, R. R. Siergiej, M. L. French und M. H. White, „Observation of near-interface oxide traps with the charge-pumping technique”, *IEEE Electron Device Lett.*, 13(12), S. 627–629, Dezember 1992.
- [193] P. T. Lai, J. P. Xu, C. K. Poek und Y. C. Cheng, „Analysis on accuracy of charge-pumping measurements with gate sawtooth pulses”, *IEEE Trans. Electron Devices*, 45(4), S. 947–952, April 1998.

- [194] R. E. Paulsen und M. H. White, „Theory and application of charge pumping for the characterization of $Si-SiO_2$ interface and near-interface oxide traps”, *IEEE Trans. Electron Devices*, 41(7), S. 1213–1216, Juli 1994.
- [195] C. E. Weintraub, E. Vogel, J. R. Hauser, N. Yang, V. Misra, J. J. Wortman, J. Ganem und P. Masson, „Study of low-frequency charge pumping on thin stacked dielectrics”, *IEEE Trans. Electron Devices*, 48(12), S. 2754–2762, Dezember 2001.
- [196] D. Bauza und Y. Maneglia, „In-depth exploration of $Si-SiO_2$ interface traps in MOS transistors using the charge pumping technique”, *IEEE Trans. Electron Devices*, 44(12), S. 2262–2266, Dezember 1997.
- [197] Y. P. Tsividis, *Operation and Modeling of the MOS Transistor*, McGraw-Hill, New York, 1987.
- [198] C. Claeys, A. Mercha und E. Simoen, „Low-frequency noise assessment for deep submicrometer CMOS technology nodes”, *J. Electrochem. Soc.*, 151(5), S. G307–G318, März 2004.
- [199] A. Nauen, *Rauschmessungen an mikroskopischen Tunnelsystemen*, Dissertation, Universität Hannover, 2003.
- [200] M. A. Caloyannides, „Microcycle spectral estimates of $1/f$ noise in semiconductors”, *J. Appl. Phys.*, 45(1), S. 307–316, Januar 1974.
- [201] M. B. Das und J. M. Moore, „Measurements and interpretation of low-frequency noise in FET’s”, *IEEE Trans. Electron Devices*, ED-21(4), S. 247–257, April 1974.
- [202] A. van der Ziel, „Some general relationships for flicker noise in MOSFETs”, *Solid State Electron.*, 21(4), S. 623–624, April 1978.
- [203] H. S. Park, A. van der Ziel und S. T. Liu, „Comparison of two $1/f$ noise models in MOSFET’s”, *Solid State Electron.*, 25(3), S. 213–217, März 1982.
- [204] H. Mikoshiba, „ $1/f$ noise in n-channel silicon-gate MOS transistors”, *IEEE Trans. Electron Devices*, ED-29(6), S. 965–970, Juni 1982.
- [205] G. Reimbold, „Modified $1/f$ trapping noise theory and experiments in MOS transistors biased from weak to strong inversion - influence of interface states”, *IEEE Trans. Electron Devices*, ED-31(9), S. 1190–1198, September 1984.
- [206] E. P. Vandamme und L. K. J. Vandamme, „Critical discussion on unified $1/f$ noise models for MOSFETs”, *IEEE Trans. Electron Devices*, 47(11), S. 2146–2152, November 2000.

- [207] A. L. McWorther, *Semiconductor Surface Physics*, Univ. of Pennsylvania Press, Philadelphia, 1956.
- [208] S. Christensson, I. Lundström und C. Svensson, „Low frequency noise in MOS transistors – I theory”, *Solid State Electron.*, 1(9), S. 797–812, September 1968.
- [209] F. Berz, „Theory of low frequency noise in *Si* MOST’s”, *Solid State Electron.*, 13(5), S. 631–647, Mai 1970.
- [210] S. T. Hsu, „Surface state related $1/f$ noise in MOS transistors”, *Solid State Electron.*, 13(11), S. 1451–1459, November 1970.
- [211] Z. Çelik-Butler und T. Y. Hsiang, „Study of $1/f$ noise in N-MOSFET’s: linear region”, *IEEE Trans. Electron Devices*, ED-32(12), S. 2797–2802, Dezember 1985.
- [212] H. E. Maes, S. H. Usmani und G. Groeseneken, „Correlation between $1/f$ noise and interface state density at the fermi level in field effect transistors”, *J. Appl. Phys.*, 57(10), S. 4811–4813, Mai 1985.
- [213] F. M. Klaassen, „Characterization of low $1/f$ noise in MOS transistors”, *IEEE Trans. Electron Devices*, ED-18(10), S. 887–891, Oktober 1971.
- [214] H. S. Fu und C. T. Sah, „Theory and experiments on surface $1/f$ noise”, *IEEE Trans. Electron Devices*, ED-19(2), S. 273–285, Februar 1972.
- [215] F. N. Hooge, „ $1/f$ noise”, *Physica B*, 83(1), S. 14–23, Mai 1976.
- [216] L. K. J. Vandamme, „Model for $1/f$ noise in MOS transistors biased in the linear region”, *Solid State Electron.*, 23(4), S. 317–323, April 1980.
- [217] T. G. M. Kleinpenning und L. K. J. Vandamme, „Model for $1/f$ noise in metal-oxide-semiconductor transistors”, *J. Appl. Phys.*, 52(3), S. 1594–1596, März 1981.
- [218] F. N. Hooge und L. K. J. Vandamme, „Lattice scattering causes $1/f$ noise”, *Phys. Lett. A*, 66(4), S. 315–316, April 1978.
- [219] L. K. J. Vandamme und H. M. M. De Werd, „ $1/f$ noise model for MOSTs biased in nonohmic region”, *Solid State Electron.*, 23(4), S. 325–329, April 1980.
- [220] K. K. Hung, P. K. Ko, C. Hu und Y. C. Cheng, „A unified model for the flicker noise in metal-oxide-semiconductor field-effect transistors”, *IEEE Trans. Electron Devices*, 37(3), S. 654–665, März 1990.
- [221] S. A. Teukolsky und W. H. Press, *Numerical Recipes in C: The Art of Scientific Computing*, Cambridge University Press, Cambridge, England, Second Edition, 1992.

- [222] C. J. Oliver, A. H. James, D. V. McCaughan, J. R. Hill und J. C. White, „Integration-input technique for C.C.D.“, *Electron. Lett.*, 15(16), S. 497–499, August 1979.
- [223] J. Chang, A. A. Abidi und C. R. Viswanathan, „Flicker noise in CMOS transistors from subthreshold to strong inversion at various temperatures“, *IEEE Trans. Electron Devices*, 41(11), S. 1965–1971, November 1994.
- [224] M. N. Ericson, C. L. Britton, Jr., J. M. Rochelle, B. J. Blalock, D. M. Binkley, A. L. Wintenberg und B. D. Williamson, „Flicker noise behavior of MOSFETs fabricated in 0.5 μm fully depleted (FD) silicon-on-sapphire (SOS) CMOS in weak, moderate, and strong inversion“, *IEEE Trans. Nucl. Sci.*, 50(4), S. 963–968, August 2003.
- [225] M. Toita, L. K. J. Vandamme, S. Sugawa, A. Teramoto und T. Ohmi, „Geometry and bias dependence of low-frequency random telegraph signal and $1/f$ noise levels in MOSFETs“, *Fluctuation and Noise Letters*, 5(4), S. L539–L548, Dezember 2005.
- [226] Y. Akue Allogo, M. Marin, M. de Murcia, P. Llinares und D. Cottin, „ $1/f$ noise in 0.18 μm technology n -MOSFETs from subthreshold to saturation“, *Solid State Electron.*, 46(7), S. 977–983, Juli 2002.
- [227] G. Giusi, F. Crupi, C. Pace, C. Ciofi und G. Groeseneken, „Comparative study of drain and gate low-frequency noise in nMOSFETs with hafnium-base gate dielectrics“, *IEEE Trans. Electron Devices*, 53(4), S. 823–828, April 2006.
- [228] S. Prasad, *Persönliche Korrespondenz*, LSI Corporation, Milpitas, Kalifornien.
- [229] P. Habas, G. Groeseneken und G. van den Bosch, „Geometric current component in charge-pumping measurements“, in *Proceedings of the 21st International Conference on Microelectronics*, 2, S. 599–604, September 1997.
- [230] G. van den Bosch, G. Groeseneken und H. E. Maes, „On the geometric component of charge-pumping current in MOSFET’s“, *IEEE Electron Device Lett.*, 14(3), S. 107–109, März 1993.
- [231] N. S. Saks und M. G. Ancona, „Determination of interface trap capture cross sections using three-level charge pumping“, *IEEE Electron Device Lett.*, 11(8), S. 339–341, August 1990.
- [232] X.-J. Yuan, M. Kivi, S. Taylor und P. Hurley, „Effects of Fowler-Nordheim stress on interface trap density and emission cross sections in n-MOSFET’s studied by three-level charge pumping“, *IEEE Electron Device Lett.*, 17(5), S. 239–241, Mai 1996.
- [233] M. Declercq und P. Jespers, „Analysis of interface properties in MOS transistors by means of charge pumping measurements“, *Rev. HF, Acta Tech. Belgica*, 9, S. 244, 1974.

- [234] Y. Maneglia und D. Bauza, „Extraction of slow oxide trap concentration profiles in metal-oxide-semiconductor transistors using the charge pumping method”, *J. Appl. Phys.*, 79(8), S. 4187–4192, April 1996.
- [235] D. Bauza, „Near interface oxide trap capture kinetics in metal-oxide-semiconductor transistors: Modeling and measurements”, *J. Appl. Phys.*, 84(11), S. 6178–6186, Dezember 1998.
- [236] F. P. Heiman und G. Warfield, „The effect of oxide traps on the MOS capacitance”, *IEEE Trans. Electron Devices*, ED-12(4), S. 167–178, April 1965.
- [237] R. A. Wachnik und J. R. Lowney, „A model for the charge-pumping current based on small rectangular voltage pulses”, *Solid State Electron.*, 29(4), S. 447–460, April 1986.
- [238] C. N. Berglund, „Surface states at steam-grown silicon-silicon dioxide interfaces”, *IEEE Trans. Electron Devices*, ED-13(10), S. 701–705, Oktober 1966.
- [239] D. J. DiMaria, „Explanation for the polarity dependence of breakdown in ultrathin silicon dioxide films”, *Appl. Phys. Lett.*, 68(21), S. 3004–3006, Mai 1996.
- [240] R. Degraeve, J. De Blauwe, J. L. Ogier, Ph. Roussel, G. Groeseneken und H. E. Maes, „A new polarity dependence of the reduced trap generation during high-field degradation of nitrated oxides”, *IEDM Tech. Dig.*, S. 327–330, Dezember 1996.
- [241] A. El-Hdiy, G. Salace, M. Jourdain, A. Meinertzhagen und D. Vuillaume, „Stress field polarity effect on defects generation in thin silicon dioxide films”, *Thin Solid Films*, 296, S. 106–109, März 1997.
- [242] T. Matsuoka, S. Taguchi, H. Ohtsuka, K. Taniguchi, C. Hamaguchi, S. Kakimoto und K. Uda, „Hot-carrier-induced degradation of N_2O -oxynitrided gate oxide NMOSFETs”, *IEEE Trans. Electron Devices*, 43(9), S. 1364–1373, September 1996.
- [243] E. Li, S. Prasad, S. Park und J. Walker, „Sub-quarter-micron PMOSFET DC and AC NBTI degradation”, *Meeting of the Electrochemical Society*, April 2003.
- [244] D. J. DiMaria und E. Cartier, „Mechanism for stress-induced leakage current in thin silicon dioxide films”, *J. Appl. Phys.*, 78(6), S. 3883–3894, September 1995.
- [245] V. Huard, F. Monsieur, G. Ribes und S. Bruyere, „Evidence for hydrogen-related defects during NBTI stress in p-MOSFETs”, in *Proceedings of the IRPS*, S. 178–182, April 2003.
- [246] R. E. Stahlbush, E. Cartier und D. A. Buchanan, „Anomalous positive charge formation by atomic hydrogen exposure”, *Microelectron. Eng.*, 28(1–4), S. 15–18, Juni 1995.

- [247] S. Holland, I. C. Chen und C. Hu, „Ultra-thin silicon-dioxide breakdown characteristics of MOS devices with n^+ and p^+ polysilicon gates”, *IEEE Electron Device Lett.*, EDL-8 (12), S. 572–575, Dezember 1987.
- [248] Y. Hiruta, H. Iwai, F. Matsuoka, K. Hama, K. Maeguchi und K. Kanzaki, „Interface state generation under long-term positive-bias temperature stress for a p^+ poly gate MOS structure”, *IEEE Trans. Electron Devices*, 36(9), S. 1732–1739, September 1989.
- [249] M. F. Beug, R. Ferretti und K. R. Hofmann, „Polarity dependent generation of gate-side and substrate-side oxide border traps in nitrided gate oxides”, *Microelectron. Eng.*, 80, S. 444–447, Juni 2005.
- [250] C. Jahan und K. Barla, „Effect of boron penetration on the stress induced leakage current in PMOS structures with p^+ doped polysilicon gate”, *J. Non-Cryst. Solids*, 245(1–3), S. 33–40, April 1999.
- [251] C. Kyono, T. Brozek und V. Ilderem, „Boron penetration effect on gate oxide reliability of 50Å PMOS devices”, *Microelectronics Reliability*, 40(4–5), S. 637–640, April 2000.
- [252] T. Brozek, C. Kyono und V. Ilderem, „Effect of boron on gate oxide degradation and reliability in PMOS devices”, *Solid State Electron.*, 45(8), S. 1293–1297, August 2001.
- [253] R. B. Fair, „Physical models of boron diffusion in ultrathin gate oxides”, *J. Electrochem. Soc.*, 144(2), S. 708–717, Februar 1997.
- [254] M. Stegherr, „Flicker noise in hot electron degraded short channel MOSFETs”, *Solid State Electron.*, 27(12), S. 1055–1056, Dezember 1984.
- [255] P. Vasina, E. Simoen, C. Claeys und J. Sikula, „A low-frequency noise study of hot-carrier stressing effects in submicron *Si* p-MOSFETs”, *Microelectronics Reliability*, 38 (1), S. 23–27, Januar 1998.
- [256] L. Vandamme und X. Li, „Change in DC and $1/f$ noise characteristics of n-submicron MOSFETs due to hot-carrier degradation”, *Microelectronics Reliability*, 38(1), S. 29–35, Januar 1998.
- [257] M.-H. Tsai und T.-P. Ma, „ $1/f$ noise in hot-carrier damaged MOSFET’s: Effects of oxide charge and interface traps”, *IEEE Electron Device Lett.*, 14(5), S. 256–258, Mai 1993.
- [258] J. Kolhatkar, E. Hoekstra, A. Hof, C. Salm, J. Schmitz und H. Wallinga, „Impact of hot-carrier degradation on the low-frequency noise in MOSFETs under steady-state and periodic large-signal excitation”, *IEEE Electron Device Lett.*, 26(10), S. 764–766, Juli 2005.

- [259] S. Dey und M Agostinelli, „Characterization and modeling of low frequency noise degradation due to NMOS hot electron stress”, in *Proceedings of the IRPS*, S. 265–268, April 2005.
- [260] Z. Chen, K. Hess, J. Lee, J. W. Lyding, E. Rosenbaum, I. Kizilyalli, S. Chetlur und R. Huang, „On the mechanism for interface trap generation in MOS transistors due to channel hot carrier stressing”, *IEEE Electron Device Lett.*, 21(1), S. 24–26, Januar 2000.
- [261] W. J. Tsai, N. K. Zous, C. J. Liu, C. C. Liu, C. H. Chen, T. Wang, S. Pan, C.-Y-Lu und S. H. Gu, „Data retention behavior of a SONOS type two-bit storage flash memory cell”, *IEDM Tech. Dig.*, S. 719–722, Dezember 2001.
- [262] W. J. Tsai, S. H. Gu, N. K. Zous, C. C. Yeh, C. C. Liu, C. H. Chen, T. Wang, S. Pan und C.-Y. Lu, „Cause of data retention loss in a nitride-based localized trapping storage flash memory cell”, in *Proceedings of the IRPS*, S. 34–38, April 2002.
- [263] E. Simoen, B. Dierickx, C. L. Claeys und G. J. Declerck, „Explaining the amplitude of RTS noise in submicrometer MOSFET’s”, *IEEE Trans. Electron Devices*, 39(2), S. 422–429, Februar 1992.
- [264] H. H. Mueller und M Schulz, „Conductance modulation of submicrometer metal-oxide-semiconductor field-effect transistors by single-electron trapping”, *J. Appl. Phys.*, 79(8), S. 4178–4186, April 1996.
- [265] A. Asenov, R. Balasubramaniam, A. R. Brown und J. H. Davies, „RTS amplitudes in decananometer MOSFETs: 3-D simulation study”, *IEEE Trans. Electron Devices*, 50(3), S. 839–845, März 2003.
- [266] K. K. Hung, P. K. Ko, C. Hu und Y. C. Cheng, „Random telegraph noise of deep-submicrometer MOSFET’s”, *IEEE Electron Device Lett.*, 11(2), S. 90–92, Februar 1990.
- [267] Z. Shi, J.-P. Miéville und M. Dutoit, „Random telegraph signals in deep submicron n-MOSFET’s”, *IEEE Trans. Electron Devices*, 41(7), S. 1161–1168, Juli 1994.
- [268] A. Karmann und M. Schulz, „Characterization of individual defects in MOSFET’s”, *Appl. Surf. Sci.*, 39(1–4), S. 500–507, Oktober 1989.
- [269] C. H. Park und Y. J. Park, „An explicit rederivation of Reimbold’s $1/f$ trapping noise theory in *Si* nMOSFETs at weak inversion”, *Solid State Electron.*, 43(12), S. 2111–2113, Dezember 1999.
- [270] K. S. Ralls, W. J. Skocpol, L. D. Jackel, R. E. Howard, L. A. Fetter, R. W. Epworth und D. M. Tennant, „Discrete resistance switching in submicrometer silicon inversion layers: Individual interface traps and low-frequency ($1/f$?) noise”, *Phys. Rev. Lett.*, 52(3), S. 228–231, Januar 1984.

- [271] P. Restle, „Individual oxide traps as probes into submicron devices”, *Appl. Phys. Lett.*, 53(19), S. 1862–1864, November 1988.
- [272] Z. H. Fang, S. Cristoloveanu und A. Chovet, „Analysis of hot-carrier-induced aging from $1/f$ noise in short-channel MOSFET's”, *IEEE Electron Device Lett.*, EDL-7(6), S. 371–373, Juni 1986.
- [273] L. Larcher, G. Verzellesi, P. Pavan, E. Lusky, I. Bloom und B. Eitan, „Impact of programming charge distribution on threshold voltage and subthreshold slope of NROM memory cells”, *IEEE Trans. Electron Devices*, 49(11), S. 1939–1946, November 2002.
- [274] F.-C. Hou, G. Bosman und M. E. Law, „Simulation of oxide trapping noise in submicron n-channel MOSFETs”, *IEEE Trans. Electron Devices*, 50(3), S. 846–852, März 2003.
- [275] H. H. Mueller und M. Schulz, „Random telegraph signal: An atomic probe of the local current in field-effect transistors”, *J. Appl. Phys.*, 83(3), S. 1734–1741, Februar 1976.
- [276] C.-H. Chen und M. J. Deen, „Channel noise modeling of deep submicron MOSFETs”, *IEEE Trans. Electron Devices*, 49(8), S. 1484–1487, August 2002.
- [277] E. Simoen, B. Dierickx, B. De Canne, F. Thoma und C. Claeys, „On the gate- and drain-voltage dependence of the RTS amplitude in submicron MOSTs”, *Appl. Phys. A*, 58(4), S. 353–358, April 1994.
- [278] C. Leyris, A. Hoffmann, M. Valenza, J. C. Vildeuil und F. Roy, „Evolution of R.T.S. source activities in saturation range in N-MOSFETs for different oxidation temperatures”, *Proceedings of the 18th International Conference on Noise and Fluctuations*, S. 213–216, September 2005.
- [279] K. K. Hung, P. K. Ko, C. Hu und Y. C. Cheng, „A physics-based MOSFET noise model for circuit simulators”, *IEEE Trans. Electron Devices*, 37(5), S. 1323–1333, Mai 1990.
- [280] P. Fantini, A. Ghetti, A. Marinoni, G. Ghidini, A. Visconti und A. Marmiroli, „Giant random telegraph signals in nanoscale floating-gate devices”, *IEEE Electron Device Lett.*, 28(12), S. 1114–1116, Dezember 2007.
- [281] C. L. Alexander, A. R. Brown, J. R. Watling und A. Asenov, „Impact of single charge trapping in Nano-MOSFETs – electrostatics versus transport effects”, *IEEE Trans. Nanotechnol.*, 4(3), S. 339–344, Mai 2005.
- [282] S. Selberherr, *Analysis and Simulation of Semiconductor Devices*, Springer Verlag, Wien New York, 1984.
- [283] P. Restle und A. Gnudi, „Internal probing of submicron FETs and photoemission using individual oxide traps”, *IBM J. Res. Develop.*, 34(2–3), S. 227–242, März–Mai 1990.

Publikationen

Teile dieser Arbeit wurden in Fachzeitschriften publiziert beziehungsweise auf internationalen Fachkonferenzen vorgestellt:

1. G. Krause, M. F. Beug, R. Ferretti, S. Prasad und K. R. Hofmann, „High-field degradation of poly-*Si* gate *p*-MOS and *n*-MOS devices with nitrided oxides”, *IEEE Trans. Device Mat. Rel.*, 6(3), S. 473–478, September 2006.
2. G. Krause, K. R. Hofmann, M. F. Beug, T. Müller und T. Mikolajick, „ $1/f$ noise analysis of a 75 nm Twin-FlashTM technology non-volatile memory cell”, *Proceedings of the Non-Volatile Memory Technology Symposium*, S. 12–15, IEEE, Piscataway, NJ, USA, 2006.

Danksagung

Mein besonderer Dank gilt Prof. Dr. Karl R. Hofmann für die Möglichkeit, an seinem Institut mitarbeiten und diese Dissertation anfertigen zu dürfen. Seine stete Diskussionsbereitschaft, sein großes Interesse an dem Thema und die Fähigkeit, seine fundierten Kenntnisse über Halbleiterbauelemente vermitteln zu können, trugen wesentlich zum Gelingen dieser Arbeit bei. Prof. Dr. W. Mathis danke ich für die Übernahme des Koreferats und Prof. Dr. H. Garbe für die Übernahme des Prüfungsvorsitzes.

Dr. Bernhard Schwierzi und Dr. Wolfgang Hasse möchte ich für die Unterstützung bei der Bearbeitung von Industrieprojekten danken. Dr. Rüdiger Ferretti gebührt Dank für die umfangreiche Hilfestellung bezüglich der Messtechnik sowie für die zahlreichen Diskussionen zu Experimenten und Ergebnissen.

Ich danke unseren Projektpartnern von Infineon Technologies und der Qimonda AG für die Bereitstellung der wertvollen Wafer, ohne die diese Arbeit nicht möglich gewesen wäre. Insbesondere möchte ich Dr. Alexander Ruf, Dr. Sabine Kolodinski und Prof. Dr. Thomas Mikolajick für die Ermöglichung der interessanten Projekte danken sowie allen Mitarbeitern und Projektpartnern für die freundliche Unterstützung und die lebhaften Projekttreffen. Dr. Sharad Prasad danke ich für die freundliche Bereitstellung von Wafermaterial und die vielen hilfreichen Diskussionen.

Großer Dank gebührt natürlich auch allen Mitarbeitern des Instituts für MBE sowie des LfI für die kollegiale Atmosphäre. Insbesondere danke ich den Angehörigen der Feinmechanikerwerkstatt, Detlef Zech und Ernst Exter, sowie Bernd Koch von der Elektronikwerkstatt. Ohne ihre Ideen und Arbeiten wäre so mancher Experimentaufbau nicht möglich gewesen und die ein oder andere Löt- und Bastelarbeit im Chaos geendet. Bei Dr. Kirsten Weide-Zaage und Farzan Kashanchi möchte ich mich besonders für die Unterstützung bei den Simulationen bedanken.

Nicht zuletzt danke ich meiner lieben Frau Heike, meiner Familie und allen Freunden und Bekannten für die großartige Unterstützung während dieser Zeit.

Lebenslauf

Name:	Gernot Krause
Geburtsdatum:	03.11.1975
Geburtsort:	Bremen
08/1982 – 08/1986	Grundschule in Oyten
08/1986 – 08/1988	Orientierungsstufe in Oyten
08/1988 – 05/1995	Cato Bontjes van Beek-Gymnasium in Achim
05/1995	Abitur
07/1995 – 06/1996	Grundwehrdienst in Schwanewede
10/1996 – 01/2002	Physikstudium an der Universität Hannover
01/2002	Diplom im Fach Physik Diplomarbeit: „Phononenabsorption in mechanisch strukturierten Quantenpunktkontakten“
06/2002 – 08/2007	Wissenschaftlicher Mitarbeiter am Institut für Materialien und Bauelemente der Elektronik, Universität Hannover
ab 09/2007	Mitarbeiter bei AMD Saxony LLC & Co. KG in Dresden