

Kapitel 2

Nanoanalytik mit Rastersondenmethoden

Die Erkenntnis der Bedeutung der Nanotechnologie führte in Deutschland zur Gründung von sechs Kompetenzzentren {Nanonet}, von denen eines für die Nanoanalytik {Nanoanalytik} zuständig ist. Die Zentren fungieren als Dienstleistungsunternehmen: Sie stellen ihr Know-How im Nanometerbereich Interessenten aus der Industrie zur Verfügung. Konkret sollen – in Zusammenarbeit mit dem Auftraggeber/Partner aus der Industrie – die geeigneten Verfahren entwickelt werden, um eine spezielle Fragestellung zu lösen. Diese Art der Förderung orientiert sich folglich an den aktuellen Anforderungen des Marktes.

Die Knotenpunkte des Kompetenzzentrums für Nanoanalytik sind an den Universitäten in Hamburg (Prof. Wiesendanger) {Hamburg}, München (Prof. Gaub) {München} und Münster (Prof. Fuchs) {Münster} lokalisiert. Der Methodenschwerpunkt der genannten Institute liegt auf den Rastersondenverfahren und unterstreicht noch einmal deren Bedeutung für die Nanotechnologie.

An der Universität Hamburg konzentriert man sich auf die Nanoanalytik von Proben aus dem Bereich der Physik (München ↔ Biologie, Münster ↔ Chemie). Die zu analysierenden Proben stammen größtenteils aus der Halbleiter- und Datenspeicherindustrie.

Auf die Halbleiter-Nanoanalytik wird gleich im Anschluss, auf den Bereich Datenspeicherung auf Nanometerskala im dritten Kapitel näher eingegangen.

2.1 Nanoanalytik von integrierten Schaltungen

Die Halbleiter-Industrie (HL-I) bestätigt seit Jahrzehnten die erstaunlich zutreffende Prognose von Gordon Moore aus dem Jahre 1965, der exponentielles Wachstum bzgl. der Komplexität von integrierten Schaltkreisen (engl. Integrated Circuit, IC) vorhergesagt hat. Mit der Komplexität der ICs verbunden ist eine Verkleinerung der das Halbleiter-Bauelement bestimmenden Strukturgröße. Dies passiert mit jeder neuen IC-Generation, die etwa alle drei Jahre auf den Markt kommt, mit einer durchschnittlichen Rate von 30% und wird solange weitergehen, bis eine technologische, physikalische [Mei96] oder wirtschaftliche Grenze erreicht ist. Auf letztere trifft man dann, wenn das Wachstum der HL-I nicht mehr deutlich schneller als bei anderen Industriezweigen vonstatten geht. Dies wird wahrscheinlich erst in 30 Jahren der Fall sein [Hu99]. Eine technologische Grenze wurde immer wieder für die nächste IC-Generation vorhergesagt (z.B. für die 1 μ m-Generation) und doch immer wieder überwunden. Die Entwicklung einer neuen Generation ist mittlerweile jedoch ein so teures Unterfangen geworden, dass kein einzelnes Industrieunternehmen die Kosten dafür mehr aufbringen kann. Dies führt zu großen projektgebundenen Kooperationen der HL-Hersteller (z.B. kostet das Projekt „Semiconductor 300“ von Infineon Technologies und Motorola zur Einführung des 300mm-Wafers 1.45 Mrd. DM, wovon der Bund und das Land Sachsen knapp die Hälfte tragen [VDI99b]) und zur Forderung einer globalen (also weltweiten) Koordination der Vorgehensweise der HL-I. Zu diesem Zweck brachte die Semiconductor Industry Association (SIA) – der HL-I-Verband der USA – 1992 die erste National Technology Roadmap for Semiconductors (NTRS) mit einer Prognose der wahrscheinlichen Entwicklungen in der HL-I in den nächsten 15 Jahren heraus. Bei der neuesten Ausgabe – der International Technology Roadmap for Semiconductors: 1999 (ITRS) – arbeiteten alle bedeutenden nationalen HL-I Vereinigungen aus Europa, Japan, Korea und Taiwan mit {ITRS}. Die bemerkenswerteste Aussage betrifft die Abweichung vom Mooreschen Gesetz bei der vorhergesagten Entwicklung der DRAM-Speicherchips (engl. Dynamic Random Access Memory): weil es ökonomisch günstiger sei, solle der 1Gbit DRAM erst 2003 (also mit einem Jahr „Verspätung“) auf den Markt kommen. Nichtsdestotrotz hält der Zwang zur Miniaturisierung an, so dass – bei heute noch unbekanntem Fertigungstechnologien – eine Gate-Länge von 20nm für das Jahr 2014 prognostiziert wird (siehe Tabelle 2). Ein anderer Trend sind „System-

on-Chips (SoC)“ – also komplette Systeme auf einem Chip –, die zudem einen extrem niedrigen Stromverbrauch haben sollen.

| Jahr | 1999 | 2003 | 2005 | 2011 | 2014 |
|--|------|-------|------|-------|-------|
| DRAM | | | | | |
| halbe Zelle [nm] (=Chip-Generation) | 180 | 120 | 100 | 50 | 35 |
| Kapazität [GB] | 1 | 4 | 8 | 64 | 64 |
| Fläche [mm ²] | 400 | 480 | 526 | 691 | 792 |
| MPU | | | | | |
| Gate-Länge [nm] | 140 | 85-90 | 65 | 30-32 | 20-22 |
| Transistoren/Chip [M] | 23.8 | 95.2 | 190 | 1523 | 4308 |
| Frequenz [GHz] | 1.25 | 2.49 | 3.5 | 10 | 13.5 |
| Verdrahtungsebenen | 6-7 | 8 | 8-9 | 9-10 | 10 |

Tabelle 2: Auszug aus der ITRS1999.

Diese Forderungen an zukünftige IC-Generationen führen zu höchsten Ansprüchen an dringend benötigte Analysegeräte mit einer lateralen Auflösung von besser als 100nm. Nur so können die hergestellten Strukturen überhaupt identifiziert, analysiert und die Performance gesteigert werden. Viele klassische Analyseverfahren liegen bei einer lateralen Auflösung von 0.5µm (Abbèsches Kriterium für die Fernfeld-Lichtmikroskopie) und können heutige Feinstrukturen (140nm Gate-Länge) nicht mehr abbilden. Zudem verlangen die hohen Anforderungen oftmals die Messung neuer – früher nicht berücksichtigter – Parameter. So können parasitäre Effekte – hervorgerufen durch die Strukturverfeinerung und Packungsdichte der Bauelemente – eine dominierende Rolle bei der IC-Entwicklung einnehmen. Hierbei spielen besonders kapazitive Kopplungen zwischen Bauelementen und Leiterbahnen eine große Rolle. Diese Kapazitäten müssen bereits beim CAD-Entwurf (engl. Computer-Aided Design) berücksichtigt werden (dafür gibt es ebenfalls ein großes Industrieprojekt: „Parasitäre Effekte in integrierten Schaltungen“ [VDI99a]). Ein weiterer zentraler Punkt ist das Messen dieser Streukapazitäten mit hoher Orts- und Femto-Farad-Auflösung, damit die CAD-Ergebnisse überprüft und die Simulatoren geeicht werden können. Solche Streukapazitätsmessungen könnten zz. nur mit einem SCM durchgeführt werden.

Auch für die Messung anderer – für den IC-Designer – wichtiger Größen bieten sich Rastersondenmethoden, mit einer lateralen Auflösung von unter 100nm an. Spannungsmessungen bis zu einigen GHz werden durch heterogenes Mischen, spezielle Cantilever (HF-geeignet) und EFM [Böh94, Bri95] ermöglicht. Heiße Stellen (engl. Hot Spots) können mit dem SThM genau lokalisiert werden [Maj99], und die exakte Vermessung des Austrittspunktes einer Laserdiode wurde schon mit dem SNOM realisiert [Bac98, DeA98]. Das SNOM erlaubt zudem die Lokalisierung von solchen Defekten, die unterhalb einer – für die SNOM-Wellenlänge – durchsichtigen Passivierung liegen. Bild 8 zeigt allerdings auch die Schwierigkeiten bei der Interpretation eines SNOM-Bildes; die Erklärung aller SNOM-Strukturen fällt aufgrund der Komplexität ihrer Entstehung schwer. Bei einem Vergleich von defektfreier und defekter Probe ist allerdings die eindeutige Identifizierung von Fehlern möglich.

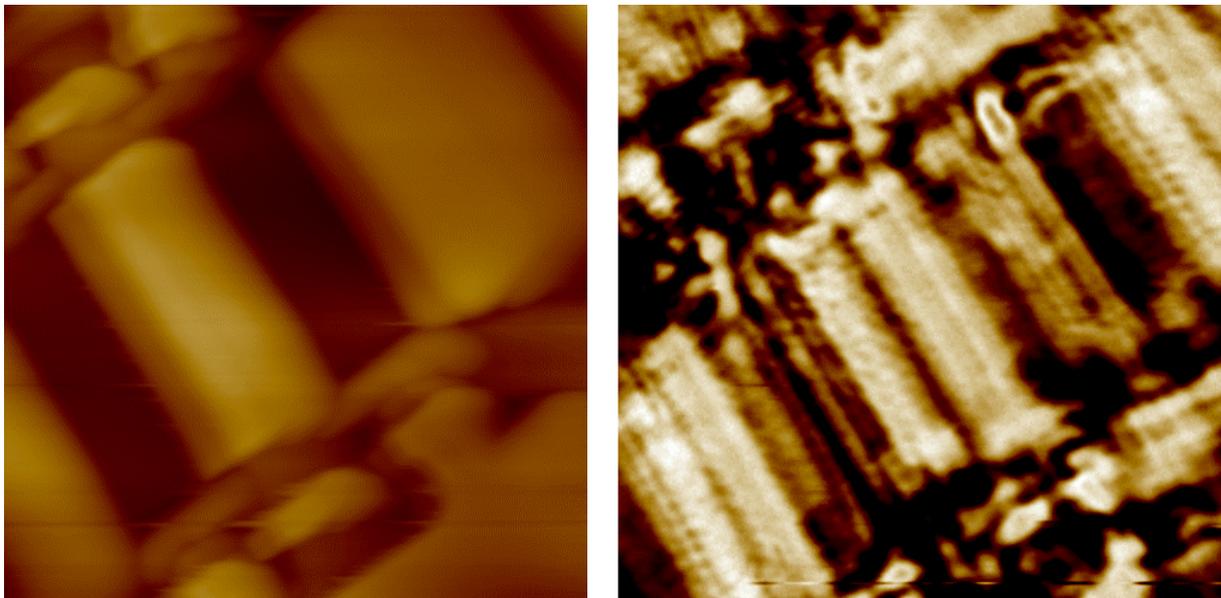


Bild 8: SNOM-Abbildung ($20 \times 20 \mu\text{m}^2$) eines ICs. Links die Topografie und rechts die ortsauflöste Anzahl der reflektierten Photonen.

Als Standardverfahren für topografische Messungen hat sich mittlerweile das SFM bewährt. Im Gegensatz zum Rasterelektronenmikroskop (REM; engl. Scanning Electron Microscope, SEM) kann es gerade die Tiefe einer Struktur sehr genau bestimmen (z-Auflösung besser als 0.1nm). Weiterhin ermöglicht es die Messung der Oberflächenrauigkeit oder die Bestimmung von Korngrößen. Selbst in der automatischen Produktionsüberwachung (engl. online-monitoring) z.B. von magneto-resistiven Schreibleseköpfen werden SFMs eingesetzt.

Als eine der wichtigsten Innovationen für die kommenden Chip-Generationen sieht die ITRS 1999 die Ersetzung des als Isolationsschicht verwendeten Siliziumdioxids

an. Eine einfache Skalierung eines MOSFETs bzgl. der Gatelänge von heute 180nm auf 20nm (im Jahr 2014) würde – bei der Verwendung von Siliziumdioxid – zu einer Isolationsschichtdicke von 0.4nm führen. Ab einer Dicke unterhalb 3nm kommt es allerdings schon zum direkten Tunneln, was einer Nutzung mit der heute üblichen

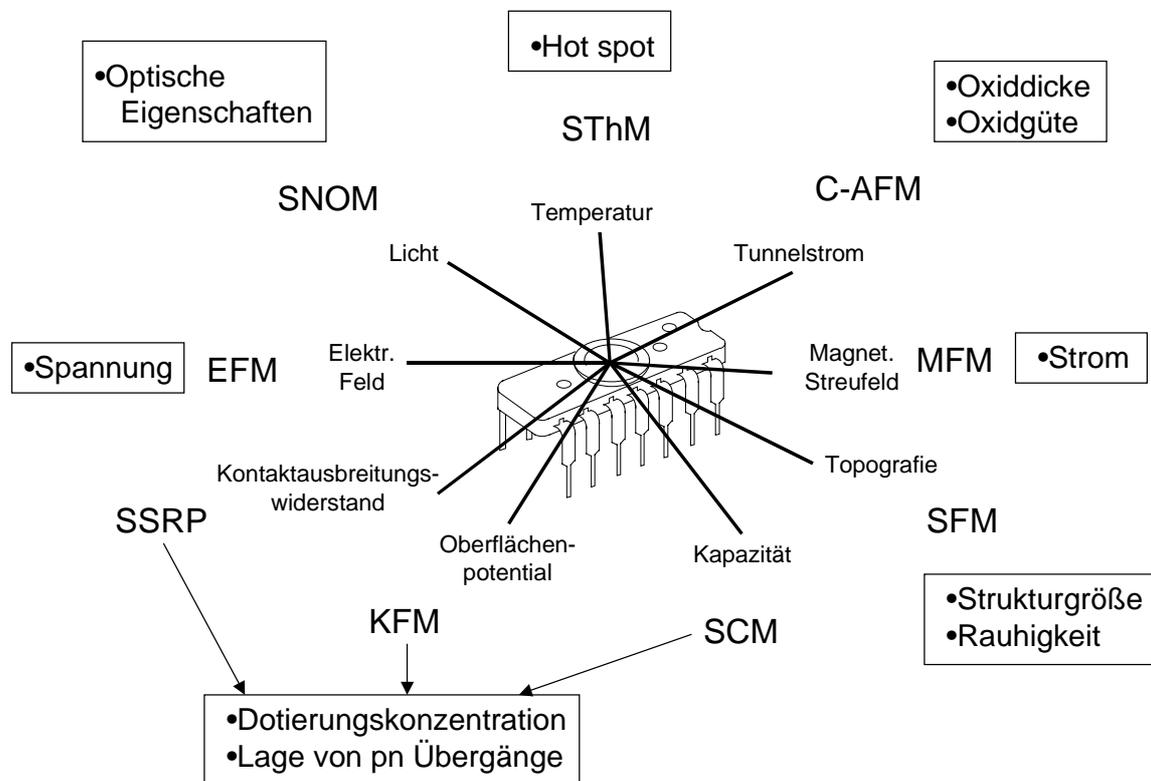


Bild 9: Übersicht IC-Charakterisierung mittels SPM

Funktionalität entgegensteht. Somit steht die Suche von geeigneten neuen Materialien und Bauelement-Designs an oberster Stelle für die Planerfüllung des Mooreschen Gesetzes. Das c-AFM eignet sich dank seiner exzellenten lateralen und Schichtdicken-Auflösung hervorragend zur Charakterisierung der Dicke und Güte von dünnen Siliziumdioxid- oder anderen Isolationsschichten (siehe Bild 10). Für die Messung von Dotierungsgradienten ist es aber weniger geeignet. Aufgrund der fundamentalen Bedeutung der Bestimmung zweidimensionaler Dotierprofile werden große Anstrengungen zur Entwicklung geeigneter Analyseverfahren unternommen. Zz. haben sich drei SPM-Methoden – KFM, SSRM und SCM – herauskristallisiert, die eine laterale Auflösung von unter 50nm versprechen. Auf sie konzentriert sich der nächste Abschnitt dieser Arbeit.

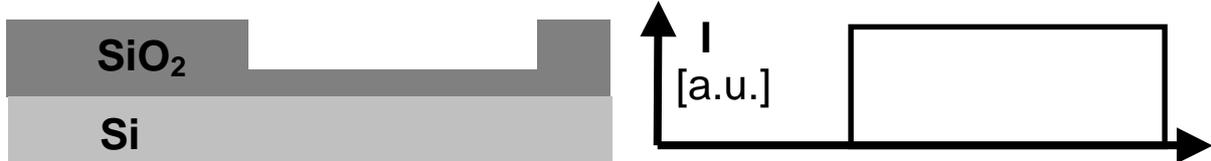
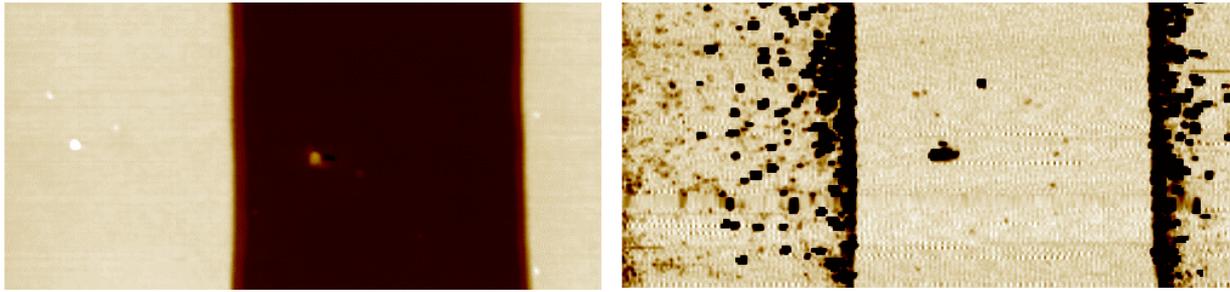


Bild 10: $12 \times 6 \mu\text{m}^2$ große c-AFM Aufnahme (links oben Topografie, rechts oben Strom) einer Siliziumdioxid-Grabenstruktur (links unten schematisch dargestellt). Die theoretisch zu erwartende Rasterlinie im Strombild ist rechts unten skizziert. Experimentell (rechts oben) konnte allerdings im Strombild – im und neben dem SiO_2 -Graben – kein Unterschied festgestellt werden. Dies kann auf einen ionischen Stromanteil zurückgeführt werden, der den Fowler-Nordheim-Tunnelstrom überdeckt. Die auffälligen Strukturen im Strombild, welche an der topografischen Kante liegen, könnten auf den Ätzprozess – der Graben wurde hineingeätzt – zurückzuführen sein. Das Bild soll verdeutlichen, dass das c-AFM eine elektrisch wirksame Oxiddicke und nicht eine strukturelle Dicke misst (die z.B. mittels Ellipsometrie gemessen werden würde). Bei hochqualitativen Oxiden kann aus dem Fowler-Nordheim-Tunnelstrom die lokale absolute Dicke bis auf wenige Ångström genau berechnet werden [Olb99].

Das außergewöhnliche Wachstum des Marktes für mobile elektronische Systeme, wie z.B. Handy/Mobile und Laptop, erfordert die schnelle Entwicklung von leistungsfähigen ICs mit geringem Stromverbrauch. Um die Performance solcher Systeme zu steigern, müssen unerwünschte lokale Stromverbraucher ausfindig gemacht und eliminiert werden. Die Frage, wie viel Strom in welcher Leiterbahn fließt, kann heute von keinem Standardanalyseverfahren beantwortet werden. Solche Strommessungen können allerdings mit einem MFM durchgeführt werden.

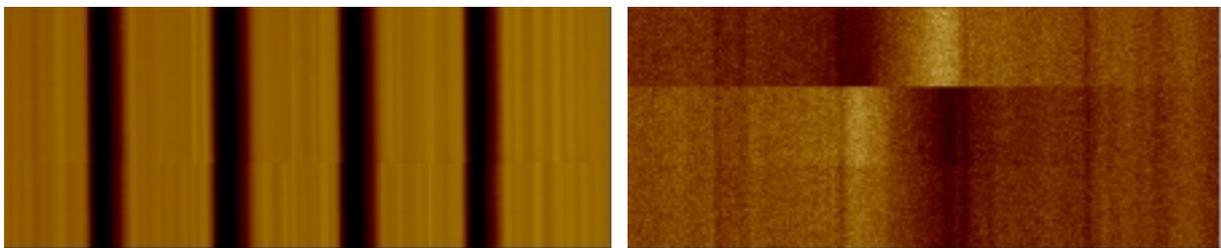


Bild 11: MFM-Response (links Topografie, rechts magnetisches Signal) eines aktiven ICs. Die mittlere – $2 \mu\text{m}$ breite – Leiterbahn trägt einen Strom von -15mA (oberes Drittel) bzw. $+15\text{mA}$ (unterer Bildteil).

Der Vergleich einer Simulation des magnetischen Streufeldes einer Leiterbahn mit dem experimentellen Resultat einer MFM-Messung (Bild 11) ermöglicht eine genaue

Bestimmung des elektrischen Stromes in der Leiterbahn. Dazu muss natürlich die magnetische Sonde des MFM vorher, z.B. an einer Leiterbahn mit bekanntem Strom, charakterisiert und damit geeicht werden (genauer [Blu97]).

Zusammenfassend kann festgehalten werden: der Zwang zur Miniaturisierung in der HL-I führt in den kommenden Jahren zur Nanotechnologie (Top-down-Vorgehensweise). Doch schon heutzutage werden dringend Analysemethoden mit einer lateralen Auflösung von unter 100nm benötigt. Rastersondenmethoden bringen dieses Potential prinzipiell mit sich und erlauben die Messung der benötigten Quantitäten mit hoher lateraler Auflösung. Für den Benutzer in der HL-I geht es jedoch zusätzlich um Zuverlässigkeit bei Messung und Interpretation. Dieses Potential wurde mit dem SFM z.B. im Online-Monitoring in verschiedenen HL-Fabriken demonstriert. Bei allen anderen hier vorgestellten SPM-Methoden muss dies erst noch gezeigt werden. Sowohl die Konversion des gemessenen Signals in die gesuchte physikalische Größe (z.B. bei der MFM-Messung erhält man ein magnetisches Signal, möchte aber die Stromstärke messen), als auch Aussagen zur lateralen Auflösung und Empfindlichkeit der Methode, bedürfen noch eines bedeutenden Fortschrittes. Die ersten Resultate solcher Entwicklungen sind für den – für die HL-I entscheidenden – Bereich der Charakterisierung zweidimensionaler Dotierprofile im folgenden Abschnitt dargelegt.

2.2 Nanoanalytik zweidimensionaler Dotierprofile

Die Eigenschaften eines Halbleiter-Bauelementes werden größtenteils durch sein dreidimensionales Dotierprofil bestimmt. Eine hochaufgelöste Analyse desselben gilt als entscheidender Schritt für die Entwicklung der nächsten Generationen von ICs. Waren die Strukturen früherer Generationen so groß, dass eindimensionale (1D) Dotierprofilmessungen ausreichen, so werden heutzutage zweidimensionale (2D) Messungen notwendig. Einerseits, um TCAD-Systeme (engl. Technology Computer-Aided Design) zu eichen und zu überprüfen [Law96], andererseits, um kleinste fehlerauslösende Strukturen in der Fehleranalyse am realen Bauelement auffinden zu können.

TCAD-Systeme berechnen aus den angenommenen Prozess-Parametern die Dichteverteilung der implantierten Dotieratome, aus ihr die Dichteverteilung der aktivierten Dotieratome (in Abhängigkeit der weiteren Prozessschritte), daraus die Ladungsträger-Dichteverteilung und hieraus wiederum die elektrischen Eigenschaften des Bauelementes. Erst diese können mit den heutigen Analysemethoden gemessen und mit den TCAD-Ergebnissen verglichen werden.

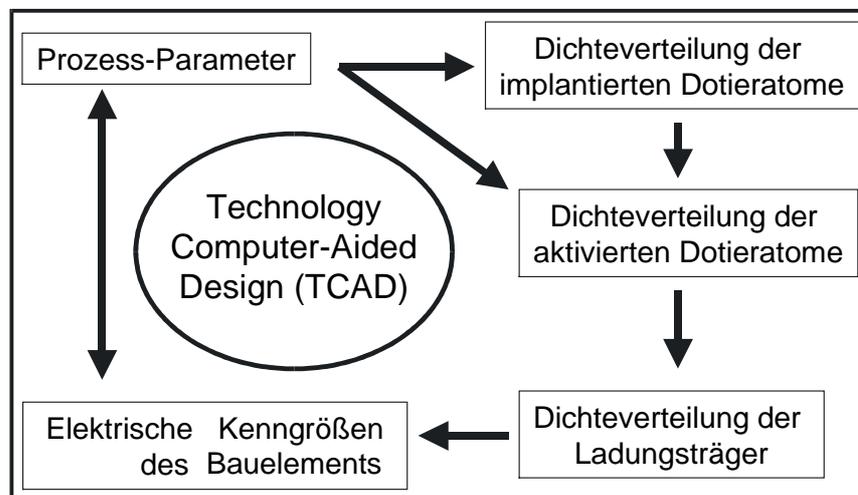


Bild 12: TCAD Schema.

Eine Methode, die die Ladungsträger-Dichteverteilung (also das Ladungsträger-Profil) direkt orts aufgelöst messen könnte, wäre folglich von großem Nutzen für die Entwicklung neuer Bauelemente. Wurden bisher Dotierprofile bzw. Ladungsträgerprofile an speziell hergestellten Teststrukturen (das reale Bauelement wird in bestimmte Richtungen auseinandergezogen, so dass z.B. MOSFETs mit einer Gatelänge von mehreren μm entstehen) mit Sekundärionen-Massenspektroskopie

(engl. Secondary Ion Mass Spectrometry; SIMS) bzw. Ausbreitungswiderstands-Messungen (engl. Spreading Resistance Profiling; SRP) oder C-V-Kurven analysiert, so ist so ein Vorgehen heutzutage durch die Miniaturisierung nicht mehr sinnvoll. Es treten oft Probleme auf, deren Ursache im Dreidimensionalen liegt und nicht durch eindimensionale Teststrukturen modelliert werden kann.

Bei der Fehleranalyse muss die Messung am realen Bauelement durchgeführt werden. Dies macht eine eindimensionale Dotierprofilbestimmung von vornherein unmöglich. Ferner muss die Messmethode schnell und zuverlässig funktionieren. Für den Benutzer sind eine einfache Anwendung und eine einfache Interpretation der Messergebnisse wichtige Eigenschaften. Ein großer Vorteil ist die Verfügbarkeit eines funktionierenden Bauelements, das den Vergleich „gut-schlecht“ erlaubt.

Bei der Dotierprofil-Analyse ist zwischen chemischer und elektrischer Analyse zu unterscheiden. Die chemische misst Dotieratome bzw. deren Konzentration, die elektrische Analyse misst sekundäre Effekte – wie z.B. differentielle Kapazität, Kontaktausbreitungswiderstand und Oberflächenpotential – die auf der Ladungsträgerdichte beruhen. Bild 13 visualisiert den Zusammenhang zwischen der lateralen Auflösung (bei einer Analysetiefe von 2nm)* und der in diesem analysierten Volumen vorhandenen Anzahl an Dotieratomen.

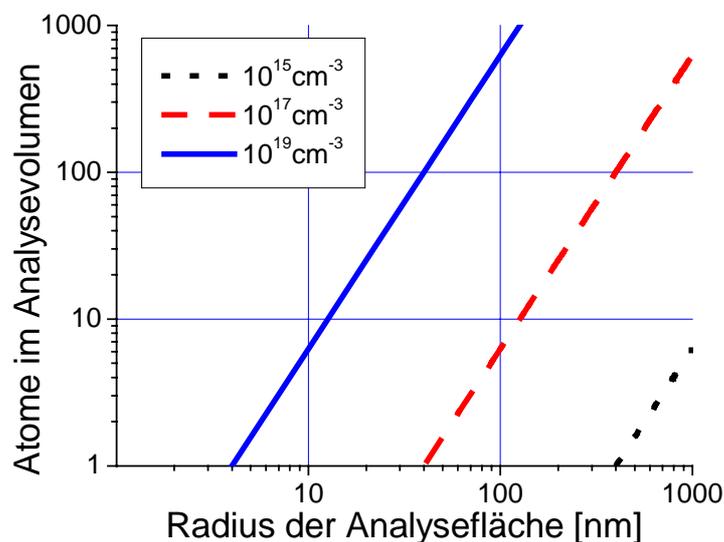


Bild 13: Zusammenhang zwischen lateraler Auflösung (2 Radien) und Anzahl der Dotieratome im analysierten Gebiet bei einer angenommenen Analysetiefe von 2nm.

* Typischer Wert bei einer SIMS-Analyse

Bei einer Dotierkonzentration von 10^{17}cm^{-3} und einer lateralen Auflösung von 100nm (Radius gleich 50nm) ist im Durchschnitt ein Dotieratom im Probenvolumen enthalten! Dadurch muss, um eine ausreichend gute Statistik zu bekommen, ein großer Satz von gleichen Proben analysiert werden. Ansonsten können nur hochdotierte Bereiche vermessen werden. Mit immer kleiner werdenden Bauelementen nimmt die Anzahl der Dotieratome im Kanalbereich so stark ab, dass (unkontrollierbare) Fluktuationen einen Betrieb unmöglich machen (Mikolajick et al. sehen den Einsatz eines MOSFETs für die ULSI Technologie mit einer Gatelänge von 100nm bei einer Siliziumdioxid-Dicke von 3nm schon als kritisch an [Mik97]). Zudem wird durch die statistische Verteilung der Dotieratome die Definition der Lage des metallurgischen p-n-Überganges ($N_A=N_D$) immer sinnloser (entsprechende Simulationen an einem 50nm großen MOSFET wurden z.B. von Asenov durchgeführt [Ase99]).

Dieses Problem besteht bei der Messung der Ladungsträger-Verteilung nicht. Soll diese allerdings in eine Dotierkonzentration konvertiert werden, so setzt die (extrinsische) Debye-Länge, die Längenskala im HL, über der die Ladungsträger bei einer vorgegebenen Dotierung „verschmiert“ sind, eine untere Grenze. Es gilt:

$$L(N) := \sqrt{\frac{k \cdot T \cdot \epsilon_{Si} \cdot \epsilon_o}{e^2 \cdot N}} \quad (14)$$

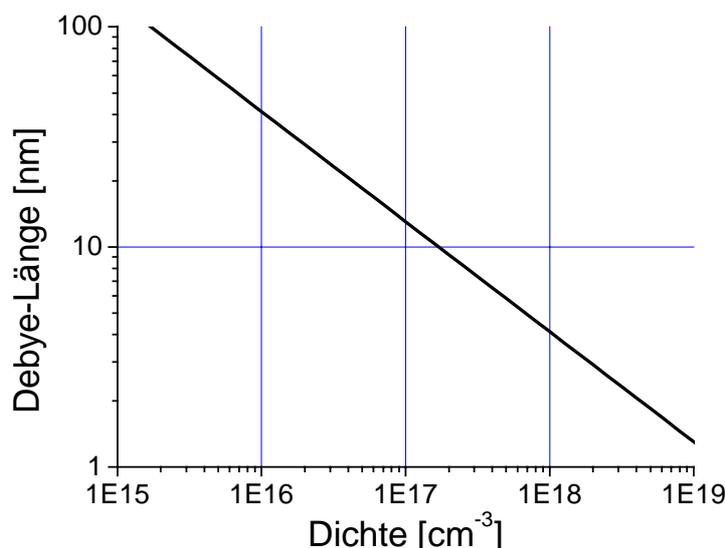


Bild 14: Debye-Länge in Abhängigkeit von der Dotierkonzentration.

Ein Sprung in der Dotierung bei einer Konzentration von 10^{15}cm^{-3} wird über eine Länge von mehr als 100nm ausgeschmiert. Aus verschiedenen Dotierprofilen kann

also ein und dasselbe Ladungsträgerprofil entstehen; dies erschwert die Rückrechnung bzw. macht sie unterhalb der Debye-Länge sogar unmöglich. Nur durch die Annahme eines speziellen Verlaufes (z.B. Gauß-Profil) kann diese Schwelle prinzipiell unterschritten werden.

Für die 1D-Dotierprofilbestimmung haben sich SIMS, SRP und die C-V-Analyse bewährt [Hei96]. Im Folgenden wird speziell auf die C-V-Analyse eingegangen, die die Grundlage des SCMs darstellt. SIMS und SRP werden hingegen nur kurz angerissen. Für eine ausführlichere Behandlung muss auf [Schro90] und die darin enthaltenen Literaturhinweise verwiesen werden.

SIMS führt eine Massenspektroskopie der durch einen Primärionenstrahl herausgeschlagenen Sekundärionen durch und bestimmt das chemische Dotierprofil, d.h. alle – auch die elektrisch nicht aktivierten – Dotieratome. Je nach Energie des Primärionenstrahles werden eine (static SIMS mit einer Ionendichte von 1 nAcm^{-2} und einer Ionenenergie von 1 keV) oder mehrere Monolagen (dynamic SIMS) abgetragen. Aufgrund der element- und umgebungsspezifischen Sputterrate sind meist aufwändige Kalibriermessungen unumgänglich.

SRP misst den Kontaktausbreitungswiderstand zwischen zwei Wolfram-Osmium Spitzen (mit Auflageradien von mehreren μm), die auf die Probe mit großer Kraft aufgesetzt werden, um einen ohmschen Kontakt zu induzieren. Dabei wird davon ausgegangen, dass Silizium bei einem Druck von 12 GPa eine strukturelle Phasenänderung durchläuft [Hu86] und dadurch von einem MOS- bzw. Schottky-Kontakt zu einem ohmschen Kontakt übergeht. Ein Großteil der elektrischen Spannung fällt dann direkt unterhalb der Spitze ab. Der Widerstand R unterhalb der Kontaktfläche mit dem Radius r (der sogenannte Ausbreitungswiderstand) dominiert folglich den gesamten gemessenen Widerstand. Theoretisch lässt er sich durch die Integration über ellipsoidische Equipotentialflächen herleiten [Hol79]:

$$R = \frac{\rho}{4 \cdot r} \quad (15)$$

Bei genauerer Betrachtung muss noch ein Barrieren-Widerstand und ein Korrekturfaktor berücksichtigt werden. Der gemessene Widerstand R ist dann proportional zum spezifischen Widerstand ρ . Dieser ist über

$$\rho = \frac{1}{e \cdot (n \cdot \mu_n + p \cdot \mu_p)} \quad (16)$$

mit der Ladungsträgerdichte n bzw. p verknüpft. SRP wird meist an schräg-geschliffenen Proben eingesetzt, um das Dotierprofil in die Länge zu ziehen. Es ergeben sich Probleme bzgl. des Einflusses der Probenbehandlung (Schrägschliff), der Spitzen (bisher existiert kein ausreichendes physikalisches Kontakt-Modell [Cla96a, Gaw96]) und der Konversion in ein Dotierprofil [Van96, Cla96]. Brauchbare Resultate sind nur mit Kalibrierproben [Cla98a] und aufwändigen Computersimulationen zu erzielen. Eine vergleichende SRP-Studie von 21 internationalen Instituten [Cla98] offenbarte große Unterschiede in den Resultaten und einen großen Bedarf an der Weiterentwicklung dieser Methode.

C-V-Messungen werden im Produktionsprozess eingesetzt, um Dotierprofile überprüfen zu können. Bei genügend großer Gatefläche des aufprozessierten MOS-Kondensators können Randeffekte vernachlässigt [Schro90] und eine 1D-Betrachtungsweise durchgeführt werden. Ein Vergleich von Soll- und Ist- C-V-Kurve erlaubt eine schnelle Überprüfung der Güte der Silizium-Siliziumdioxid Grenzfläche oder des Tiefen-Dotierprofiles. Die Bestimmung eines unbekanntes (Tiefen-) Dotierprofiles aus einer C-V-Kurve ist – in erster Näherung (depletion approximation model) – begrenzt auf einen Bereich von etwa der dreifachen Debye-Länge unterhalb der Probenoberfläche bis zum Bereich des elektrischen Durchbruchs (siehe Bild 15). In diesem Streifen kann das Dotierprofil nur mit einer Genauigkeit der Debye-Länge vertikal (in die Tiefe) aufgelöst werden (siehe Bild 14). Rückschlüsse auf oberflächennahe Dotierungen und Dotierprofile unterhalb der Debye-Länge sind nur mit aufwändigeren Rechnungen [Lin83] oder inverser Modellierung möglich [DeM72]. D.h. es wird von einem bestimmten parametrisierbaren Verlauf des Dotierprofiles ausgegangen (z.B. Gauß-förmig), die dazugehörige C-V-Kurve ausgerechnet und durch Variation der Parameter die simulierte C-V-Kurve an die experimentelle angeglichen. Diese Simulation kann je nach Problemstellung – bei maximal 6 Parametern – zwischen Minuten und Stunden dauern (die Rechenzeit steigt exponentiell mit der Anzahl der Parameter) [Kuz96]. Ein zusätzlicher Einfluss von Ladungen (im Oxid oder an der Grenzfläche) kann eine genaue Bestimmung des Dotierprofiles unmöglich machen. Ist hingegen die Dotierung bekannt, kann z.B. die Dichte der Fehlstellen an der Grenzfläche bestimmt werden. Sie erlaubt Aussagen über die Güte des Prozesses (optimal ist eine Dichte von kleiner als 10^{11}cm^{-2}).

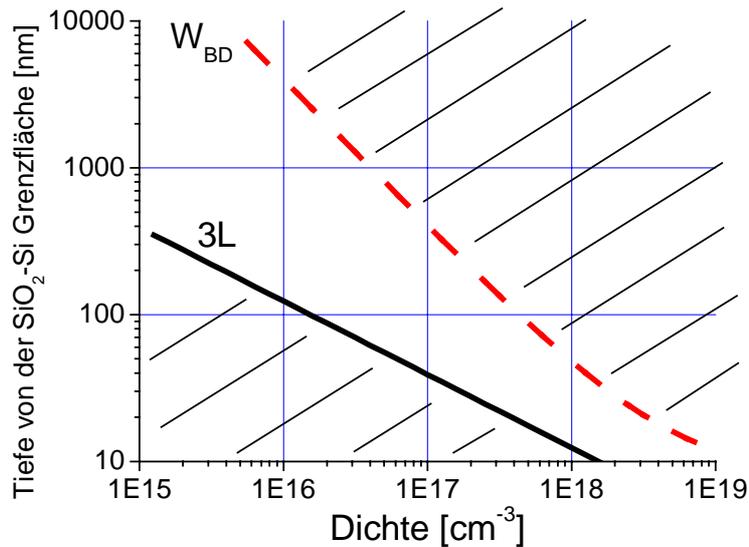


Bild 15: Zugänglicher Bereich bzgl. der direkten Konversion von C-V-Kurven in ein Dotierprofil ($3L$ ist die dreifache Debye-Länge und W_{BD} die Tiefe der Raumladungszone beim Durchbruch).

1D-Messungen des Dotierprofils mittels C-V-Analyse fordern akkurate experimentelle Messungen (Einhalten der Kleinsignal-Bedingung, sorgfältige Probenpräparation), eine genaue Kenntnis der Größe der Gatefläche (die zudem noch ausreichend groß sein muss, d.h. mehrere $10\mu\text{m}$ Durchmesser) und aufwändige Computersimulationen (bei komplizierten Dotierprofilen).

Beruhend auf den genannten 1D Methoden wurden 2D-Analyseverfahren entwickelt [Sub92, Die96, Van97].

Für die chemische 2D-Dotierprofilbestimmung eignen sich „Imaging-SIMS“ [Bry85], lateral-SIMS [vonC94, vonC98], 2D-SIMS [Dow92, Coo96] und STM [Joh93].

Der Ionenstrahl eines SIMS kann auf mehrere 10nm fokussiert und damit auch die laterale Auflösung gesteigert werden (Imaging-SIMS). Dies bedingt allerdings auch ein sehr kleines Analysevolumen (siehe Bild 13) und damit einhergehend ein sehr schlechtes Signal-Rausch-Verhältnis. Für lateral- und 2D-SIMS werden spezielle Proben benötigt, so dass beide Verfahren nicht in der Fehleranalyse eingesetzt werden können.

STM kann zwar einzelne Dotieratome in z.B. GaAs oder InAs abbilden, ist aber für Silizium aufgrund der hohen Oberflächenzustandsdichte, die alle unterliegenden Dotieratome maskiert, für die Bestimmung der Dotieratomdichte (durch Auszählen der Dotieratome innerhalb der obersten zwei Monolagen) ungeeignet.

Für die elektrische 2D-Dotierprofilbestimmung haben sich neben der Elektronen-Holografie [Rau99] – die aufgrund aufwändiger Probenpräparation und mangelnder Sensitivität für die Fehleranalyse z. B. weniger geeignet erscheint – und dem direkten Abbilden mit einem REM (leidet ebenfalls an mangelnder Sensitivität [Ven96]) viele Rastersondenmethoden empfohlen: STM, KFM (EFM), SSRM und SCM.

Das STM eignet sich für die Bestimmung der Lage des elektrischen p-n-Überganges [Hos88, Cha98, Hil98] bei spezieller Probenpräparation (H-terminierte Oberfläche oder dünnes Oxid), die ein Festhalten (engl. Pinning) des Fermi-Niveaus verhindert. Dabei muss auch eine Verschiebung des p-n-Überganges mit der angelegten Tunnelspannung – also eine Beeinflussung der Probe durch die Sonde – berücksichtigt werden [Cha92]. Auch ein MOS-Übergang wurde nach Brechen der Probe im UHV mittels STM untersucht [Kom96]. Durch die hohe Oberflächen-sensitivität reagiert das STM extrem empfindlich auf Kontaminationen, so dass reproduzierbare Messungen nur im Ultra-Hoch-Vakuum (UHV) möglich sind.

Auf die Technik des KFM wurde schon im Kapitel 1.2 eingegangen (bzgl. IC-Charakterisierung per KFM siehe [Hoc96]). Durch eine Bestimmung des Oberflächenpotentials wäre ein Rückschluss auf die Dotierung möglich (eine Änderung der Dotierung um eine Größenordnung entspräche einer Änderung des Oberflächenpotentials um etwa 60mV), wenn die Oberfläche defekt- bzw. kontaminationsfrei wäre. Das native Siliziumdioxid, das sich auf jeder Silizium-Probe unter Umgebungsbedingungen befindet, besitzt jedoch meist neben einer hohen Defektdichte eine negative Oberflächenladung. Dies ist durch die Wechselwirkung der OH-Gruppen mit dem dünnen Wasserfilm bedingt und verhindert eine quantitative Bestimmung der Dotierung. Weiterhin ist die laterale Auflösung durch den relativ großen mittleren Abstand von Sonde und Probe bestimmt, was zu einer Mittelung über einen Radius von über 50nm führt [Ste97]. Anhand von Bild 16 kann die Auflösung einer KFM- mit einer SCM-Abbildung einer Teststruktur verglichen werden. Das Dotierprofil ist eingezeichnet.

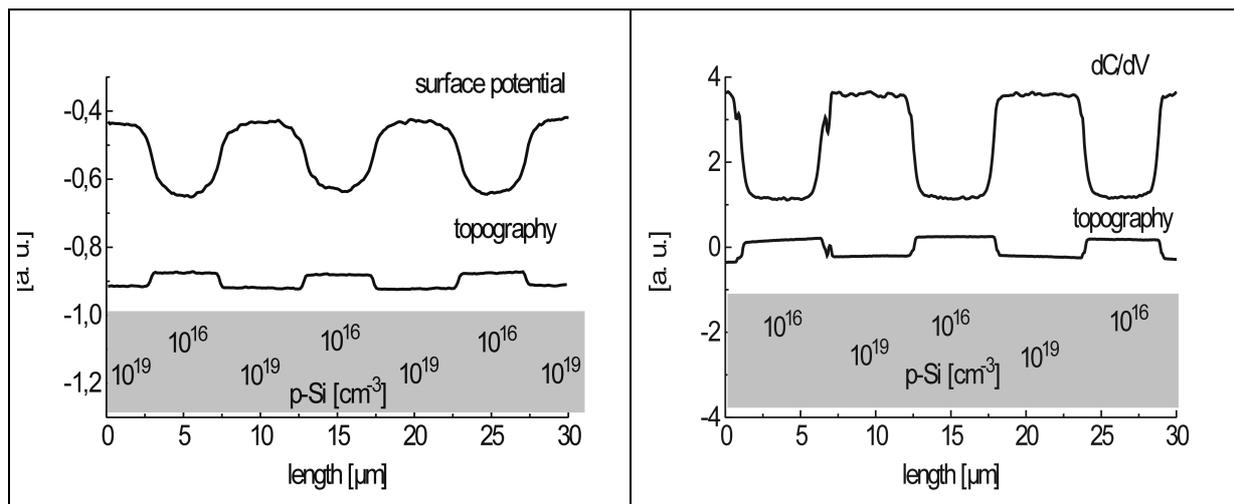


Bild 16: KFM-Abbild einer Teststruktur (Dotierprofil ist eingezeichnet) im Vergleich zum SCM-Bild.

Die p-n-Übergänge sind in der KFM-Aufnahme deutlich ausgeschmierter als im SCM-Bild. Auch wird in der KFM-Messung – im Gegensatz zur SCM-Abbildung – kein konstantes Potential über den homogen dotierten Bereichen erreicht. Damit zeigt sich eindeutig die schlechtere laterale Auflösung des KFM (in diesem Messmodus) gegenüber dem SCM. Bei falscher oder gar keiner Probenpräparation konnte mitunter kein Potentialunterschied zwischen verschiedenen dotierten Bereichen aufgelöst werden. Dies lag dann vor, wenn die Probenoberfläche zu rau (Querschnitts-Präparation), zu verschmutzt oder aufgeladen war. Quantitative Messungen lassen sich nur unter UHV-Bedingungen realisieren.

Das SSRM (auch als nano-SRP, SRM [Sha94] oder CCM [May94] bezeichnet) basiert auf dem SRP. Die eine Elektrode wurde durch einen festen ohmschen Rückkontakt, die andere durch eine leitfähige SFM-Sonde ersetzt (siehe Bild 17). Wie beim SRP muss der nötige Druck ausgeübt werden, um das dünne native Oxid durchzudrücken und einen ohmschen Kontakt zum Silizium aufbauen zu können. Dies stellt extreme Anforderungen an die Sonde. Bei einem Radius der Auflagefläche von 5nm müsste eine Kraft von mindestens 1μN aufgewendet werden, um einen Druck von 12GPa aufzubauen. Für p- und n-Dotierungen im Bereich 10¹⁵cm⁻³ bis 10²⁰cm⁻³ treten typischerweise – bei Kräften von einigen 10μN – Widerstände von 100Ω bis zu 1GΩ und bei einer Spannung von 50mV Ströme von 500μA bis zu 50pA und damit Stromdichten bis zu 1·10¹²A/m² in der Spitze auf. Selbst Sonden, die mit dotiertem CVD-Diamant beschichtet sind, können diese hohen Ansprüche nur selten erfüllen [Tre99]. Meistens existiert statt des gewünschten ohmschen Kontaktes ein Schottky-Kontakt, so dass die Resultate nicht

einfach in ein Dotierprofil zu konvertieren sind und rein qualitativen Charakter besitzen.

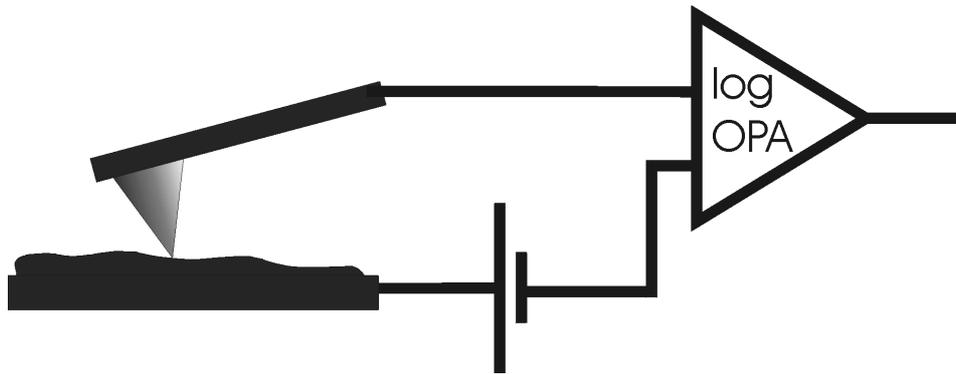


Bild 17: Experimenteller Aufbau des SSRM.

Die Messung einer Teststruktur (Bild 18) zeigt eine gleich gute laterale Auflösung für SSRM und SCM. Simulationen, die mit einem 3D-Device-Simulator [ISE] (auf ihn wird im Kapitel 2.2.2 noch ausführlich eingegangen) durchgeführt wurden, zeigten, dass mittels SSRM theoretisch – bei der Annahme eines idealen ohmschen Kontaktes – bei niedrigen Dotierungen eine höhere laterale Auflösung (bzgl. der direkten Konversion!) als mittels SCM erreichbar wäre (bzgl. verbesserten Konversionsverfahren können sie als gleichwertig angesehen werden). Denn während beim SCM die laterale Auflösung von der Größe der Verarmungsschicht bestimmt wird (näheres im nächsten Kapitel), fallen 70% (90%) des Widerstandes in einem Radius von $2r$ ($6r$) um den Spitzenmittelpunkt ab (dotierungsunabhängig). Hierbei wurde allerdings nicht eine Änderung des Stromflusses durch die induzierte Phasenänderung im Silizium berücksichtigt, da dieser Einfluss noch nicht hinreichend verstanden ist und mit dem Device-Simulator nicht modelliert werden kann (für eine Einführung in die Problematik der Physik des Kontaktes siehe [Cla96]). Heute wird von einem metallischen Bereich (plastisch verformtes Silizium) direkt unter der Spitze ausgegangen, der von einem Schottky-Bereich (elastisch verformtes Silizium) umgeben ist. Die Größe des effektiven Kontaktradius hängt damit nichttrivial von der Größe des Kontaktradius der Spitze, vom Material der Spitze, vom ausgeübten Druck, von der angelegten Spannung, von der Dotierung der Probe und den Eigenschaften der Probenoberfläche (Grenzflächenzustandsdichte) ab. Damit sind theoretische Aussagen zur lateralen Auflösung ebenfalls nichttrivial.

Bei einem reinen Schottky-Kontakt kann von einer vergleichbaren bzw. schlechteren lateralen Auflösung von SSRM gegenüber SCM ausgegangen werden.

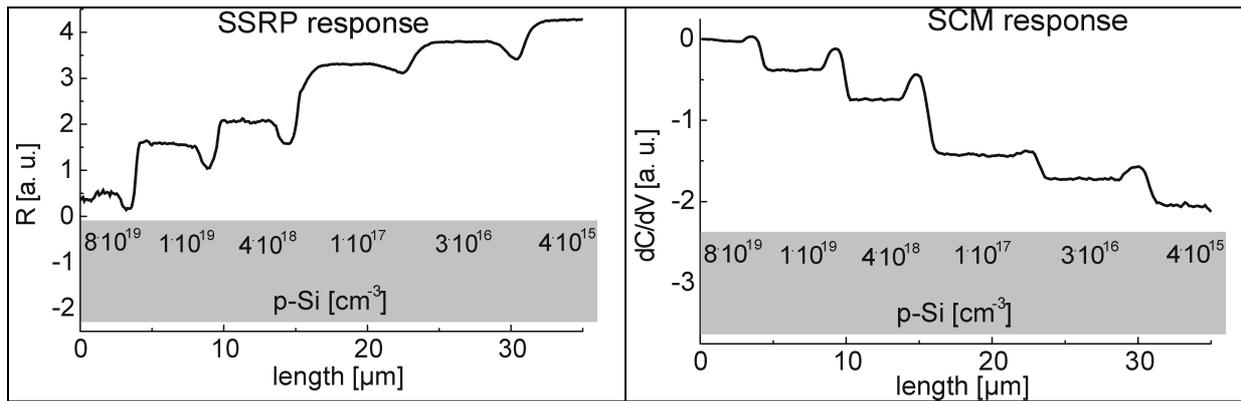


Bild 18: Das SSRM-Abbild im Vergleich zur SCM-Messung zeigt eine gleich gute hohe laterale Auflösung. Zwischen den 5 μm breiten homogen dotierten Bereichen, deren Dotierung angegeben ist, liegen noch 0.5 μm breite höher dotierte Gebiete. Sie wirken sich als Peaks in der SSRP- bzw. SCM-Messung aus.

Mittels Device-Simulator wurde auch die Frage der Auflösung eines p-n-Überganges untersucht. Bild 19 zeigt die Ergebnisse einer 2D-Simulation der Response eines SCMs und eines SSRMs mit gleicher Sondenbreite von einem p-n-Übergang (beide

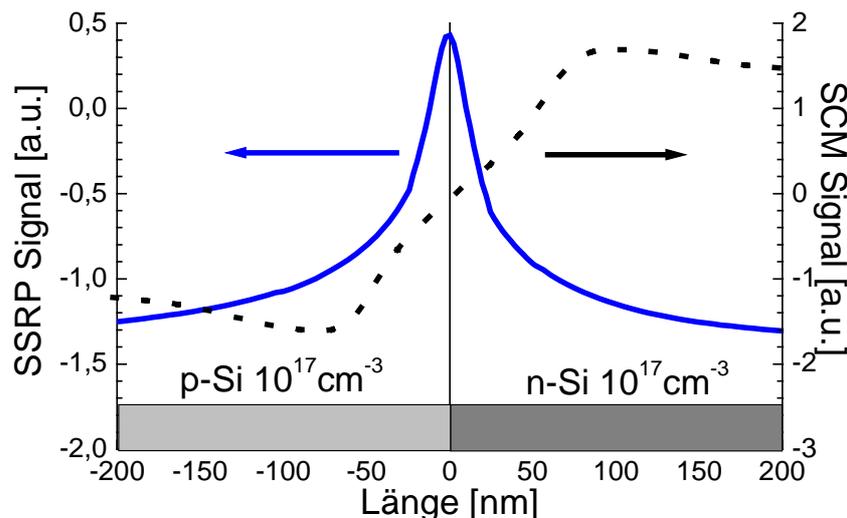


Bild 19: Simulation einer SSRM- und einer SCM-Messung eines symmetrischen p-n-Überganges.

Bereiche mit einer Dotierung von in etwa 10^{17}cm^{-3}). Da beide Methoden die Ladungsträgerdichte messen, ermitteln sie auch die elektrische Position des Überganges und nicht die metallurgische (beide liegen allerdings bei dem hier gezeigten symmetrischen p-n-Übergang bei Null). Beim SSRM wird diese Position durch einen Peak (großer Widerstand), beim SCM durch einen Nulldurchgang (keine Kapazitätsänderung) angezeigt. Im Experiment kommen Ungenauigkeiten bei der Bestimmung der Position beim SSRM durch eine systembedingte Asymmetrie bzgl.

p- und n-Dotierung (SSRM-Kontakt zeigt eine Kombination von ohmschem und Schottky-Verhalten) und beim SCM durch eine mögliche Verschiebung der Position aufgrund einer angelegten Spannung (Näheres im folgenden Kapitel) hinzu.

Zusammenfassend zeigt sich, dass das SSRM theoretisch – im vereinfachten Modell mit rein ohmschem Kontakt – ein ähnlich gutes, wenn nicht sogar besseres Auflösungsvermögen (bzgl. der direkten Konversion!) als das SCM haben sollte. Praktisch traten allerdings eine Vielzahl an Problemen auf:

- Haltbarkeit der Spitze: Trotz großer Bemühungen konnten noch keine geeigneten Spitzen entwickelt werden. Selbst eine gute Spitze hält die hohe Beanspruchung meist nur fünf/zehn Messungen/Bilder aus [DeW98, Tre99].
- Kontaktierung aller Proben-Bereiche mit einem ohmschen Rückkontakt: Da die meiste Spannung direkt unterhalb des Kontaktes abfallen muss, sollte im übrigen Strompfad kein größerer Widerstand mehr auftreten, d.h. keine niedrigere Dotierung und kein p-n-Übergang. Sind die Proben zu dünn oder zu dick, so werden ebenfalls die Messergebnisse verfälscht. Damit ergeben sich Probleme bei der Analyse von Dotierprofilstrukturen, die kleine Abmessungen in allen drei Dimensionen besitzen.
- Kalibrierung der Spitze: Ein Verschleiß der Spitze während der Messung führt zu einer Änderung der Kontaktfläche und kann auch zu einer Änderung der Kontaktart – ohmscher Kontakt wird zum Schottky-Kontakt – führen.
- SSRM-Modellierung: Obwohl es bereits viele Simulationspakete zu SRP gibt, die teilweise für SSRM eingesetzt werden könnten, herrscht Unklarheit über die Güte der Pakete, da meist ein unzureichendes Modell zugrunde gelegt wird [Cla98, DeW98].

Bei der Verfügbarkeit von entsprechenden Spitzen, Kalibrierproben und Konvertier-Programmen könnte das SSRM für grundlegende Untersuchungen neuer Devices eingesetzt werden. Insgesamt ist allerdings noch ein großer Entwicklungsbedarf – gerade bei der physikalischen Modellierung – vorhanden.

Eine weitere 2D-Dotierprofil-Analysemethode beruht auf dem selektiven Ätzen mit anschließender Abbildung der Topografie mittels SFM. Dotierungsätzen wird schon seit den Anfangsjahren der Halbleiter-Analytik in Kombination mit REM oder TEM [Spi96] eingesetzt (eine kurze Übersicht mit den gebräuchlichsten Ätzlösungen liefert [Bec88]). Meist ging es allerdings um eine qualitative Analyse. Erst in jüngster Zeit

wurde die Möglichkeit der quantitativen Auswertung in Kombination mit einer SFM-Topografie-Messung untersucht [Bar96, Alv96]. Hierbei muss die Faltung der Spitze mit der durch Ätzen entstandenen Topografie berücksichtigt werden [Wild96]. Doch auch wenn von Spitzenartefakten abgesehen wird und zwei identische Proben nach einer identischen Probenpräparation untersucht werden, gibt es signifikante – bisher nicht kontrollierbare – topografische Unterschiede [Ukr98]. Eine Reduzierung könnte durch elektrochemische Ätzverfahren erzielt werden [Tre98]. Für eine quantitative Analyse ist allerdings noch viel Entwicklungsarbeit zu leisten.

Zusammenfassend lässt sich festhalten:

1. Die HL-I benötigt dringend Methoden zur Analyse von 2D-Dotierprofilen.
2. Alle bisher entwickelten 2D-Dotierprofil-Methoden benötigen noch einen erheblichen Entwicklungsaufwand. Sie erfüllen zz. noch nicht die Anforderungen der HL-I.
3. Die chemischen 2D-Dotierprofilmethoden (SIMS-Techniken) leiden an aufwändiger Probenpräparation und/oder mangelnder Sensitivität, zudem verliert die chemische Information mit abnehmender Strukturgröße ihre Bedeutung.
4. Bei den elektrischen 2D-Dotierprofilmethoden leiden die Elektronen-Holografie und die direkte REM-Abbildung an aufwändiger Probenpräparation und/oder mangelnder Sensitivität.
5. SPM-Methoden können eine hohe Auflösung unter Umgebungsbedingungen erzielen, es fehlen aber konkrete Aussagen bzgl. des Einflusses der Sonde und der Umgebungs- und Präparationsbedingungen auf das Messergebnis. Dies liegt an dem – für alle SPM-Methoden – noch nicht hinreichend ausgereiften physikalischen Modells der Wechselwirkung.

Daraus folgt:

Für den Einsatz in der (Fehler-) Analyse von HL-Bauelementen bieten sich SPM-Methoden in ausgezeichneter Weise an. Sie bedürfen allerdings eines grundlegenden Verständnisses der Nano-Physik. Aus den obigen Ausführungen geht klar hervor, dass mit den Methoden zur Analyse von

Nanometerstrukturen auch ein Verständnis dafür gefordert wird, dass die zu analysierende Struktur mit dem Analyseinstrument ein Nano-System bildet, das als Ganzes begriffen werden muss.

Dies wird beispielhaft im nächsten Kapitel anhand der Rasterkapazitätsmikroskopie vorgeführt.

2.2.1 SCM (Scanning Capacitance Microscopy)

Die detektierte Wechselwirkung in der Rasterkapazitätsmikroskopie ist die differentielle Kapazität ($C=dQ/dV$) einer Nano-MOS-Struktur. Bevor auf die Besonderheiten des SCMs, infolge der Größe bzw. Geometrie der Gate-Fläche und der Instrumentierung eingegangen wird, soll ein Resümee der klassischen Bestimmung der differentiellen Kapazität von MOS-Strukturen gezogen und damit zugleich auch die Anforderungen an die Vorgehensweise bei der Untersuchung der Physik des SCM herausgearbeitet werden.

Bei der klassischen Messung von MOS-Strukturen [Nic82] wird die differentielle Kapazität meist in Form des Verschiebestromes mittels eines Impedanzwandlers (R) und eines phasenempfindlichen Verstärkers (Lock-In Verstärker) gemessen (siehe Bild 20). Dadurch erhält man die Response einer angelegten Wechselspannung (V_{prob}) mit einer typischen Amplitude von 10mV (zur Einhaltung der Kleinsignal-Bedingung) in Abhängigkeit von einer Biasspannung (V_{bias}) – die sogenannte C-V-Kurve.

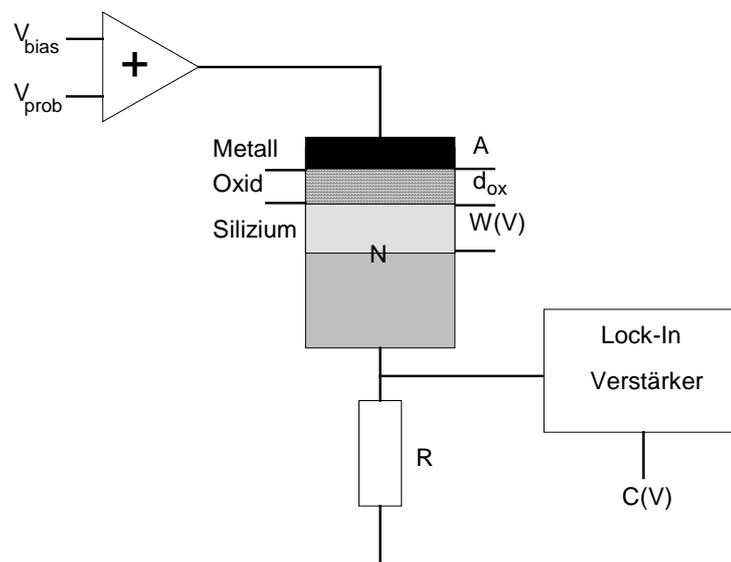


Bild 20: Klassische Bestimmung der C-V-Kurve einer MOS-Struktur.

Je nach Frequenz (f_{prob}) der Wechselspannung und Rampengeschwindigkeit der Gleichspannung können drei Grundtypen von C-V-Kurven beobachtet werden (siehe Bild 21). Alle drei stimmen im Bereich der Akkumulation (negative Biasspannung im Falle einer p-Dotierung) und Verarmung der Majoritätsladungsträger überein und unterscheiden sich nur im Bereich der Inversion. Nur hier können Minoritätsladungsträger (Elektronen bei p-Si) ihren Einfluss geltend machen. Bei der Niederfrequenz-

kurve (engl. Low Frequency, LF) beträgt die Frequenz maximal einige Hertz, so dass die Minoritätsladungsträger im Bereich der Inversion genug Zeit haben, eine Inversionsschicht im Takte der Wechselfspannung auf- und abzubauen. Diese Zeit haben sie bei $f_{\text{prob}}=1\text{MHz}$ nicht mehr. Hier können sie nur der Gleichspannung folgen. Dementsprechend langsam baut sich eine Inversionsschicht an der Grenzfläche ($\text{SiO}_2\text{-Si}$) auf, was zu einer konstanten Kapazität im Bereich der Inversion bei der Hochfrequenzkurve (engl. High Frequency, HF) führt. Bei der LF- und der HF-Kurve wird von Gleichgewichtsbedingungen ausgegangen. Die Rampengeschwindigkeit muss deshalb kleiner als 1V/s sein. Bei höheren Werten können die Minoritätsladungsträger auch der Gleichspannung nicht mehr folgen. Es werden Kurven der tiefen Verarmung (engl. Deep Depletion, DD) gemessen.

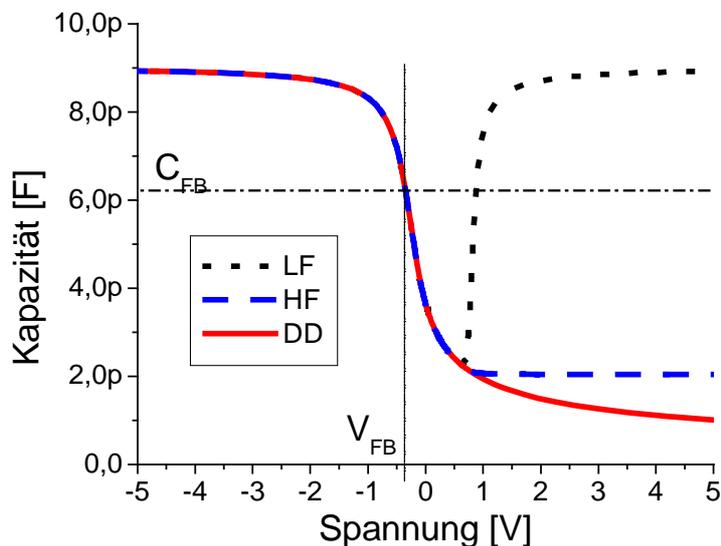


Bild 21: C-V-Kurven einer 1D-MOS-Struktur mit einem Gate-Flächenradius von $R=50\mu\text{m}$, einer Oxiddicke von $d_{\text{ox}}=30\text{nm}$ und einer homogenen p-Dotierung von 10^{16}cm^{-3} .

Die Abhängigkeit der C-V-Kurven von der Dotierung bildet die Basis der Dotierungsmessungen mit dem SCM. Bild 22 veranschaulicht die Unterschiede zwischen p- bzw. n- und niedriger (10^{16}cm^{-3}) bzw. hoher (10^{18}cm^{-3}) Dotierung. Im linken Bild sind C-V-Kurven dargestellt, im rechten dC/dV -V-Kurven. Die C-V-Kurven einer betragsmäßig gleichen p- bzw. n-Dotierung unterscheiden sich nur aufgrund der Biasspannung. Sie können durch eine Spiegelung an der V-Achse (für $V_{\text{bias}}=0$) ineinander überführt werden. Dies gilt allerdings nur, wenn die Austrittsarbeit der Gate-Elektrode (Sonde) auf dem Potential des intrinsischen Siliziums liegt (\equiv absolut

Null). Ansonsten muss eine zusätzliche Spannung V_{offset} angelegt werden (dazu später mehr). Bei allen Simulationen wurde von absolut Null ausgegangen.

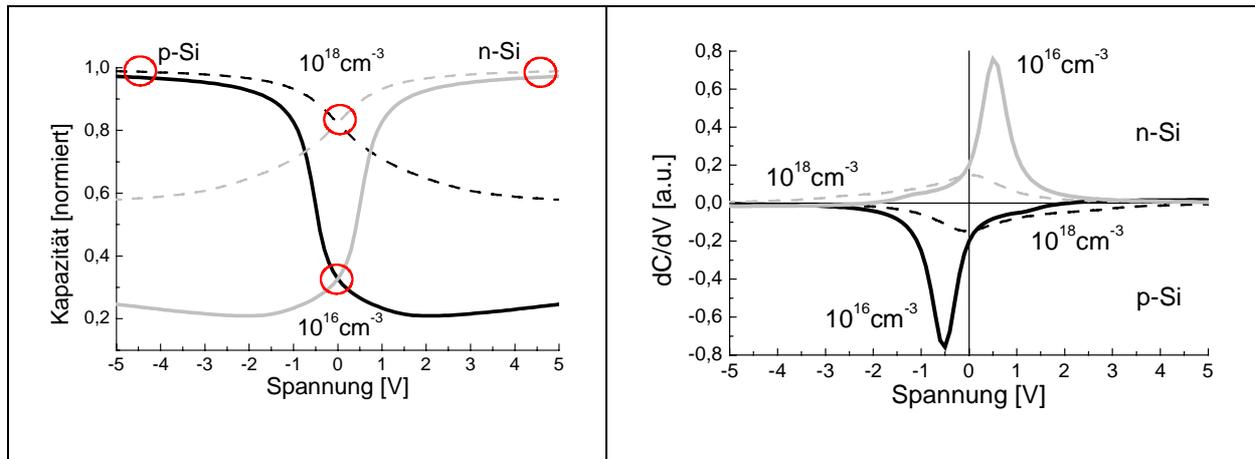


Bild 22: C-V- (links) und dC/dV-V-Kurven (rechts) einer 1D-MOS-Struktur. Dargestellt für p- und n-Si mit zwei verschiedenen Dotierungen (10^{16}cm^{-3} und 10^{18}cm^{-3}).

Als Faustformel gilt:

p-Si liefert negative, n-Si positive, eine geringe Dotierung große und eine hohe Dotierung kleine dC/dV-Werte.

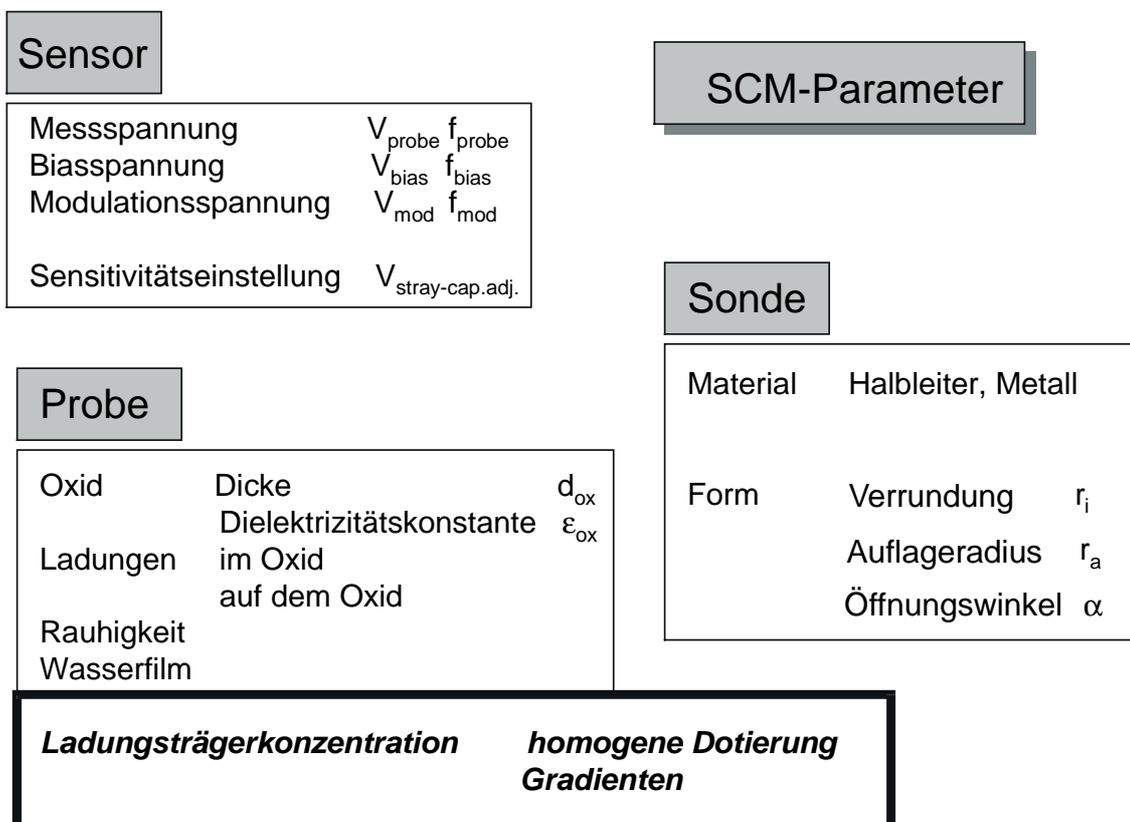


Bild 23: Übersicht der wichtigsten SCM-Parameter.

Dies ist nur als grobe Richtlinie aufzufassen, da die C-V- bzw. dC/dV-V-Kurven beim SCM von einer Vielzahl von Parametern abhängen (Instrumenten-, Sonden- und Proben-Parametern, siehe Bild 23).

Die idealen C-V-Kurven (Bild 21) können zudem durch andere Ladungen – als die bisher ausschließlich berücksichtigten Ladungsträger im HL – stark verändert werden. In der klassischen C-V-MOS-Analyse wird von vier weiteren möglichen Ladungstypen ausgegangen: feste Ladungen an der SiO₂-Si-Grenzfläche (engl. fixed oxide charge, Q_f), umladbare Fehlstellen an der SiO₂-Si-Grenzfläche (engl. interface trap charge, Q_{it}), feste Ladungen im Oxid (engl. bulk oxide trapped charge, Q_{ot}) und bewegliche Ladungen im Oxid (engl. mobile ionic charge, Q_m) [Nic82].

Während Q_f und Q_{ot} für eine Verschiebung der Flachbandspannung der C-V-Kurve sorgen, bewirkt Q_m (meistens Natrium- und Kalium-Ionen) eine Hysterese im Bereich der Inversion [Meg90]. Technologisch besonders wichtig ist die Bestimmung von Q_{it}, die Aussagen über die Interfacegüte erlaubt und die für eine „Verschmierung“ der C-V-Kurve sorgt. Sind Dotierung, Fläche und Oxiddicke bekannt, so kann aus einem Vergleich von Messung und Theorie die Dichte dieser umladbaren Fehlstellen in der Bandlücke des Halbleiters (wird meist angegeben in cm⁻²eV⁻¹) bestimmt werden (näheres siehe [Nic82]). Auch aufwändigere Messtechniken (Messung der HF- und LF-C-V-Kurve) machen eine Bestimmung der Fehlstellendichte experimentell direkt möglich. Steht hingegen nur die HF-C-V-Kurve, und keine weiteren Informationen hinsichtlich Dotierprofil und Dichte der Oxidladungen zur Verfügung, können diese Informationen nur indirekt bestimmt werden. D.h. es muss von einem Dotierprofil und einer Dichte der Oxidladungen ausgegangen und daraus die C-V-Kurve bestimmt werden. Bei dieser inversen Vorgehensweise (engl. inverse modeling technique) werden die Annahmen solange variiert, bis simulierte und experimentelle C-V-Kurven übereinstimmen [Kuz96].

Die Auflösung des Ladungsträgerprofils mit der C-V-Technik ist bestimmt durch [Schro90]:

$$p(W) = \frac{\epsilon_{Si} \cdot \epsilon_o}{e} \cdot \frac{C}{W^2} \cdot \frac{\Delta V}{\Delta C} \quad (17)$$

Sie ist folglich proportional zur Kapazität (C) und zur Sensitivität des Messgerätes (ΔV/ΔC). Die Kapazität ist proportional zur Fläche und umgekehrt proportional zur Oxiddicke. Aus diesem Grunde sind große Flächen und dünne Oxidschichten zu

wählen. Unterhalb von 2-3nm Oxiddicke kann es aber zu starken Veränderungen in der C-V-Kurve kommen [Cho99].

Eine Erhöhung der Messspannung ΔV steigert die Auflösung hinsichtlich der Ladungsträgerdichte, vergrößert aber auch den Bereich, über den gemittelt wird, und senkt somit die Tiefenauflösung. Ferner sinkt die Auflösung umgekehrt quadratisch zur Tiefe (W). Eine hohe Auflösung ist also besonders nahe an der Oberfläche zu erzielen.

Ein großes – oft vernachlässigtes – Problem können Randeffekte darstellen. Umso kleiner die Fläche (A) der Gate-Elektrode wird, desto größer wird der Quotient von effektiver Fläche und Gatefläche (A_{eff}/A). Der Flächenradius sollte deshalb nicht kleiner als

$$r_{\text{min}} = 3.7 \cdot 10^{-2} \cdot \frac{N}{10^{16}}^{-0.35} \quad (18)$$

gewählt werden [Schro90]. Er hängt damit von der Dotierung ab. Bei 10^{15}cm^{-3} ergibt sich ein r_{min} von $830 \mu\text{m}$. Diese Randbereiche können zudem, durch eine Aufladung der Oberfläche der Probe einen großen Einfluss auf die C-V-Kurve ausüben [Nic82].

Zentrale Forderungen für 1D-C-V-Bestimmung des Dotierprofils sind daher:

- die Verwendung von Messgeräten mit großer Sensitivität ($\Delta V/\Delta C$). Dabei sollte ΔV nicht zu groß gewählt werden, da es sonst zu Nichtlinearitäten kommt und die Ortsauflösung (bzw. Tiefenauflösung) wie auch die Ladungsträgerdichte-Auflösung sinkt. Aufgenommen werden sollten DD-C-V-Kurven, um den Einfluss der Minoritätsladungsträger zu minimieren.
- die Verwendung von Proben mit geringer Oxidladungsdichte, da diese einen großen Einfluss auf die C-V-Kurve hat. Ferner sollte die Oxiddicke um die 3nm und eine möglichst große Fläche gewählt werden.
- die Verwendung von Simulationsprogrammen zur Konversion der gemessenen Kurven in ein Dotierprofil mittels inverser Methoden.

Damit ist das Potential klassischer C-V-Messung aufgezeigt. Das Verlangen nach immer höherer lateraler Auflösung führt uns direkt zum Rasterkapazitätsmikroskop. Das wachsende Interesse an dieser SPM-Spielart schlägt sich in der steigenden Anzahl von Publikationen nieder (Bild 24). Besonders in den USA wurde die Weiter-

entwicklung dieser Methode in den letzten Jahren massiv – von industrieller wie von staatlicher Seite – unterstützt und mündete schließlich 1997 in die Entwicklung eines kommerziellen Gerätes (D3000 von D.I.), das auch in der vorliegenden Arbeit eingesetzt wurde. Daneben wurde mit einem SCM von S.I.S. gearbeitet, das mittlerweile ebenfalls kommerziell erhältlich ist.

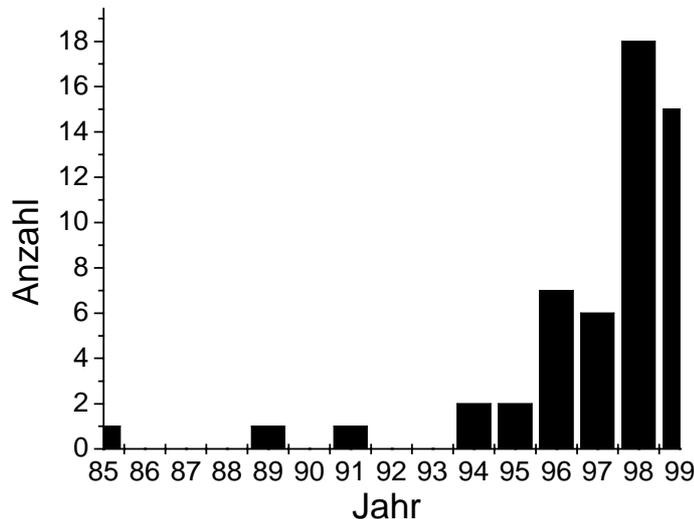


Bild 24: Anzahl der SCM-Veröffentlichungen pro Jahr (begründet auf einer Recherche in der STN Datenbank des FIZ Karlsruhe).

Trotz der großen Zahl an Veröffentlichungen gibt es kaum fundierte Aussagen über die Physik, die hinter der Messmethode steht. Dies kann auf der einen Seite auf die Neuheit, auf der anderen Seite auf die scheinbare Unkompliziertheit der Methode zurückgeführt werden. Ohne große Kenntnisse lassen sich schnell ansehnliche Messergebnisse erzielen. Dabei kann relativ sorglos bzgl. Sensoreinstellung, Sondenauswahl (Hauptsache leitend), Probenpräparation und Probenkontaktierung vorgegangen werden. Selbst wenn die Proben nicht angeschlossen sind (z.B. bei eingebetteten Schliffen), können Dotierstrukturen aufgelöst werden. Dadurch liefert das SCM schnell Resultate, die nicht weiter hinterfragt werden.

Sollen allerdings quantitative und reproduzierbare Messergebnisse erzeugt werden, muss das System systematisch analysiert werden. Dies war eine der zentralen Aufgaben der vorliegenden Arbeit.

Der genauen Analyse sollen aber einige allgemeine Bemerkungen zum SCM vorangestellt werden:

Der typische SCM-Messmodus liefert ein Signal, das proportional (der Proportionalitätskonstanten entspricht die Sensitivität) zur Ableitung der differentiellen Kapazität ist:

$$\text{SCM - Signal} = K_{\text{sens}} \cdot \frac{\Delta}{\Delta V_{\text{mod}}} \left(\frac{\Delta Q}{\Delta V_{\text{prob}}} \right) \Big|_{V_{\text{bias}}} \quad (19)$$

Bei einer konstanten Gleichspannung (Biasspannung V_{bias}) wird die Oberfläche zeilenweise abgerastert. Für jeden Punkt der Topografie wird zusätzlich ein kapazitives (SCM-) Signal aufgenommen. Die Kapazität des Sonde-Probe-Kontaktes wird von einem höchst empfindlichen Sensor gemessen, welcher an einem leitfähigen Cantilever angeschlossen ist. Wird die Kapazität ($C = \Delta Q / \Delta V_{\text{prob}}$) direkt gemessen, so kann das kapazitive Signal aufgrund einer großen Rauschamplitude (großer Einfluss der Streukapazitäten) nicht aufgelöst werden. Erst die Lock-In-Technik ($\Delta C / \Delta V_{\text{mod}}$) erlaubt eine zufriedenstellende Auflösung des SCM-Signals.

Um die Komplexität des Messsystems besser verstehen und die einzelnen Effekte besser unterscheiden zu können, wurden umfangreiche Simulationen mit einem Device-Simulator durchgeführt (TCAD-Programmpaket von {ISE}).

Wichtig für das grundsätzliche Verständnis der SCM-Messung bleibt, dass in das SCM-Signal nicht die Dotierung, sondern die Ladungsträgerdichte eingeht. In der Literatur – wie auch in dieser Arbeit – werden beide Begriffe mitunter synonym benutzt. Deshalb sei an dieser Stelle noch einmal hervorgehoben:

Das SCM misst die Ladungsträgerdichte-Verteilung!

Für eine systematische SCM-Analyse haben sich die folgenden Punkte als wesentlich herausgestellt:

- Eine hohe Auflösung geht einher mit einer hohen Sensitivität des Messgerätes. Der Einfluss von diesem auf das Messergebnis muss deshalb intensiv untersucht und genau verstanden werden. Instrumentierung
- Der Probenpräparation kommt besondere Bedeutung zu. Probe
- Der Einfluss der 3D-Geometrie der Sonde kann nur per Computersimulationen adäquat erfasst werden. Sonde
- Die maximal mögliche Auflösung der Dotierung muss untersucht werden. Dotierungs-Auflösung

- Die mögliche laterale Auflösung muss untersucht werden. laterale Auflösung
- Es muss eine Methode zur Eichung des Gerätes entwickelt werden. Eichung
- Die Messmethode muss an realistischen Aufgabenstellungen aus dem Bereich der HL-Analyse getestet werden. Beispiele aus der Fehleranalyse
- Der Einfluss der Umgebung muss intensiv untersucht und verstanden werden. SCS
- Auch der Einfluss der Sonde auf inhomogene Ladungsträgerprofile kann nur mittels Computersimulationen adäquat erfasst werden. Quantitative SCM

Ziel der nun folgenden Abschnitte (2.2.1.n) ist das Verstehen des Nanosystems, das die Sonde-Probe-Wechselwirkung des SCM dominiert.

2.2.1.1 Instrumentierung

Mit kleiner werdender Fläche nimmt auch die Kapazität und somit das Signal ab. Für Gateflächen mit einem Radius von einigen 10nm beträgt die Kapazitätsänderung nur noch einige Atto-Farad. Dies muss der für SCM verwendete Kapazitätssensor auflösen können (bzgl. der Einschätzung der Sensitivität von verschiedenen Kapazitätssensoren siehe Anhang A1).

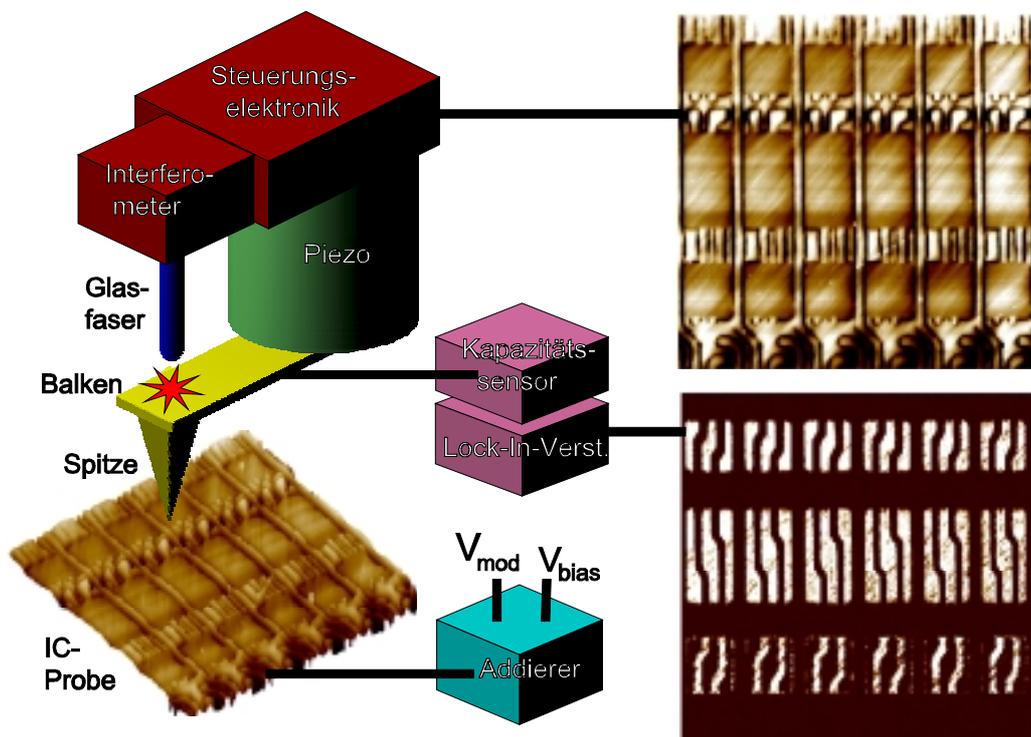


Bild 25: Schematischer Aufbau eines SCM.

Bild 25 verdeutlicht, dass das SCM der Kombination eines SFMs mit einem hochempfindlichen Kapazitätssensor entspricht. Dieser ist an eine leitfähige Sonde angeschlossen und liefert eine Spannung, die proportional zur differentiellen Kapazität ($C(V) = dQ(V)/dV_{prob}$) ist. Das Ausgangssignal hängt extrem empfindlich von Streukapazitäten (C_{stray}) ab, die um mehrere Größenordnungen über der eigentlich zu messenden Kapazität (C_{prob}) liegen. Sie sind zudem starken Schwankungen unterworfen (Bewegungen im Raum, Messungen in der Nähe einer Probenkante usw.). Aus diesem Grund wird das Ausgangssignal mit einem phasenempfindlichen Verstärker (Lock-in-amplifier, LIA) weiterverarbeitet. Dieser liefert – durch eine

Modulation der Biasspannung mit der Spannung V_{mod} – ein Signal, das proportional zur Ableitung der differentiellen Kapazität ($dC(V)/dV_{\text{mod}}$) ist und das eigentliche SCM-Signal darstellt. Der verwendete Kapazitätssensor basiert auf dem Prinzip des Resonanzkreises, der über eine Varaktordiode mit der Spannung V_{adj} noch zusätzlich verstellt werden kann. Damit ergeben sich vier Spannungen: V_{prob} , V_{bias} , V_{mod} und V_{adj} , die alle einen Einfluss auf das Messsignal haben. Welchen Einfluss sie haben, soll im Folgenden geklärt werden, zuerst soll allerdings untersucht werden, über welchen Bereich die Spannungen abfallen.

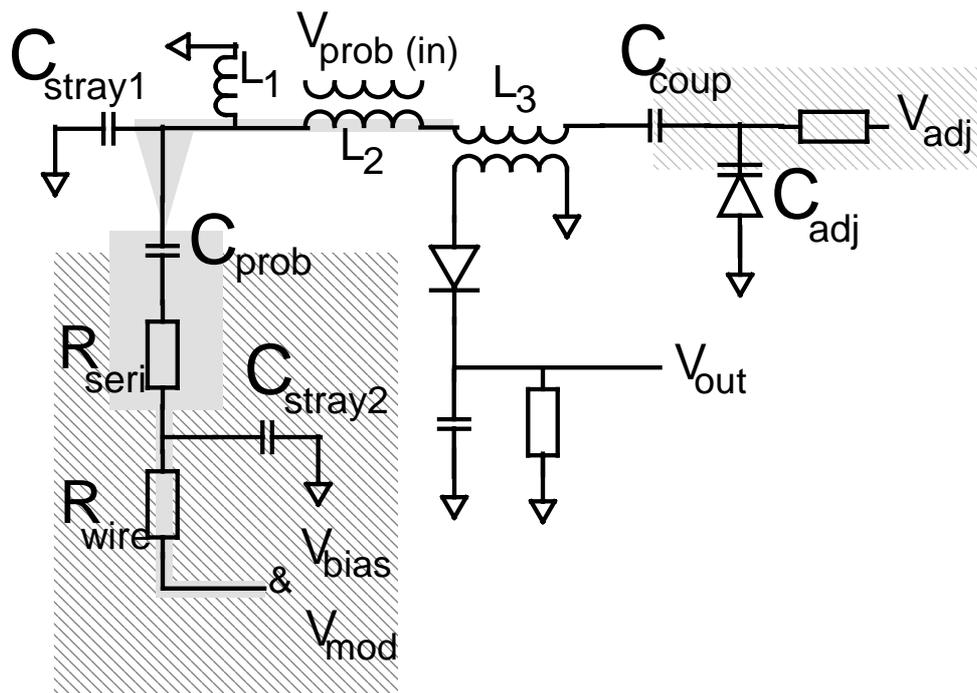


Bild 26: Schematischer Schaltplan vom Kapazitätssensor mit folgenden Bezeichnungen: Spannungen beim SCM V_{prob} , V_{bias} , V_{mod} und V_{adj} , Eigenschaften der Zuleitungen zum Probenrückkontakt R_{wire} und C_{stray2} , der Probe R_{seri} , des Sonde-Probe-Kontakte C_{prob} , der Sonde und Zuleitung zum Sensor C_{stray1} und des Sensors selbst L_{1-3} , C_{coup} , C_{adj} .

Die vom Rückkontakt der Probe eingekoppelten Spannungen V_{bias} und V_{mod} fallen beide an C_{prob} ab. Die Kapazität der Varaktordiode C_{adj} (typ. 18-6pF) kann mit V_{adj} (typ. 0-10V) auf nichtlineare Weise verändert und somit die Resonanzfrequenz des Schwingkreises verstimmt werden. V_{adj} wird vom Koppelkondensator C_{coup} abgeblockt. Der Bereich des Schwingkreises zwischen C_{coup} und C_{prob} , und somit auch die Sonde, liegt über die Ferritperle L_1 virtuell – d.h. für alle Gleichspannungen – auf Massenniveau.

Die hochfrequente Messspannung $V_{\text{prob(in)}}$ (typ. 0.9-1GHz, 5-10V) wird induktiv in den Resonanzkreis eingekoppelt (L_2) und liegt faktisch im gesamten Gerät und Raum an (1GHz ist Mobilfunk-Frequenz, d.h. der Sensor „strahlt“). V_{prob} wird folglich von allen

ingezeichneten Kapazitäten beeinflusst. Das resultierende Spannungssignal (V_{res}) des Resonanzkreises wird mit L_3 ausgekoppelt und anschließend spitzengleichgerichtet, so dass bei V_{out} die Änderung von V_{res} abgegriffen und verstärkt werden kann.

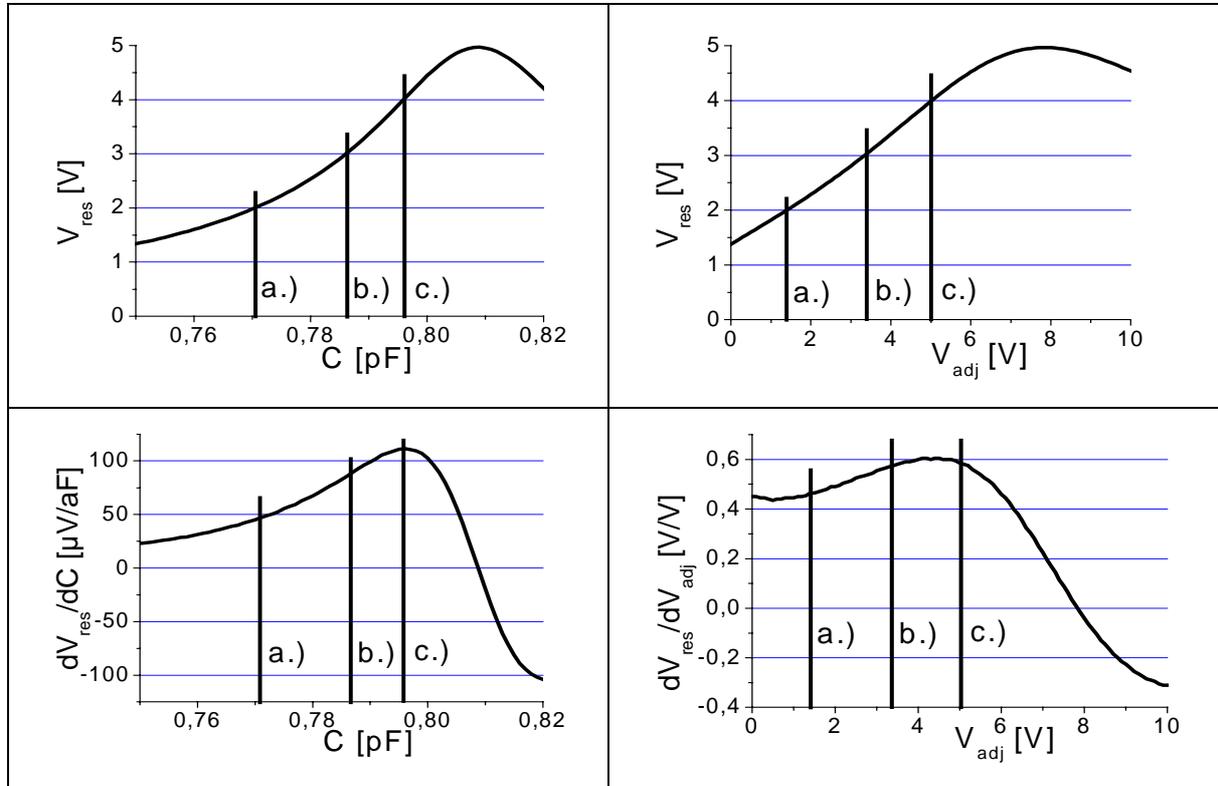


Bild 27: Simulierte Resonanzkurven (oben) und deren Ableitungen (unten) des verwendeten Kapazitätssensors. Links in Abhängigkeit von der Kapazität und rechts von V_{adj} .

Alle eingezeichneten Kapazitäten addieren sich zu einer Gesamtkapazität $C(C_{\text{stray}1,2}, C_{\text{prob}}, C_{\text{coup}}, C_{\text{adj}})$. Somit kann der Kapazitätssensor als LCR-Schwingkreis modelliert werden (typ. $L=40\text{nH}$, $C=0.7\text{pF}$, $R=5\Omega$). Für die Sensitivität des Schwingkreis-Kapazitätssensors ergibt sich in erster Näherung (siehe Anhang A1):

$$\frac{\Delta V}{\Delta C} = 0.22 \cdot \frac{V_p}{C} \cdot Q \quad (20)$$

Die Sensitivität ist somit proportional zur Resonanzpeak-Amplitude (V_p) und Güte (Q) und umgekehrt proportional zur (gesamten) Kapazität des Kreises. In den meisten Veröffentlichungen wird von $V_p=5\text{V}$, $Q=45$ und $C=1\text{pF}$ ausgegangen, so dass sich eine Sensitivität von $50\mu\text{V/aF}$ ergibt. Da das Rauschen typischerweise bei $25\mu\text{V}$ liegt, wäre eine maximale Auflösung von 0.5aF erreichbar. Diese angenommenen Werte können sich allerdings von den tatsächlichen, wegen des Einflusses von Streukapazitäten (z.B. durch wechselnde Probengeometrien, siehe auch Bild 28) und eines falsch eingestellten Sensors stark unterscheiden. Für eine Eichung des

Kapazitätssensors müsste die Ableitung $dV_{\text{res}}(C)/dC$ bei jeder Messung bestimmt werden. Gemessen werden kann jedoch nur $V_{\text{out}}(V_{\text{adj}})$. Der Unterschied – bedingt durch die Nichtlinearität von $C(V_{\text{adj}})$ – ist in Bild 27 zu erkennen. Nur durch Anfitzen der gemessenen Kurve $V_{\text{out}}(V_{\text{adj}})$ mit einer theoretischen Kurve könnte – bei bekanntem R , L , C_{coup} , $C_{\text{adj}}(V_{\text{adj}})$ – die Streukapazität $C_{\text{stray}1}$ und darüber die Kurve $V_{\text{res}}(C)$ annähernd bestimmt werden, aus der sich die Sensitivität errechnen ließe. Dies müsste bei jeder Messung durchgeführt werden, wenn sich die Sonde unmittelbar über der Probenoberfläche befindet, da sich die Resonanzcharakteristik aufgrund der starken Änderung der Streukapazität auf dem letzten 1mm Fahrweg beträchtlich ändern kann. Somit kann die Sensitivität noch nicht genau bestimmt werden und liegt bei einigen $10\mu\text{V}/\text{aF}$ je nach Probengeometrie und V_{adj} .

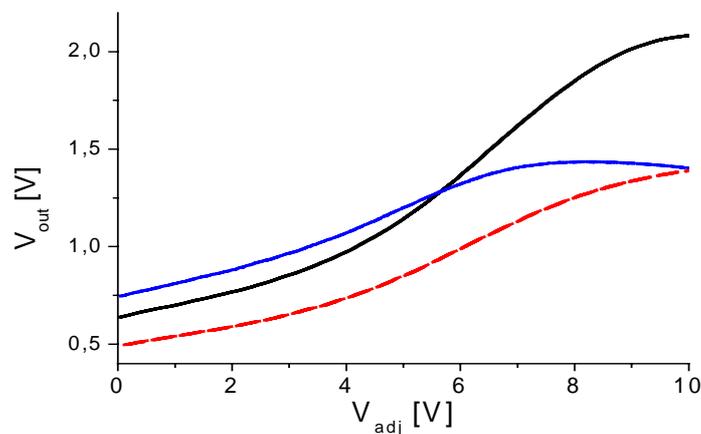


Bild 28: Experimentelle Resonanzkurven bei verschiedenen Streukapazitäten.

Bei ungünstiger Justierung des Kapazitätssensors und großer Messkapazität (C_{prob}) zeigen die aufgenommenen C-V- (bzw. dC/dV -V-) Kurven beträchtliche Abweichungen von den theoretisch zu erwartenden Kurven. Simuliert wurde der Einfluss des Sensors bzw. der Einstellung von V_{adj} auf das Messsignal. Die Messung wurde mit einem gewöhnlichen Kobalt-Chrom-beschichteten Silizium-Cantilever (CoCr-Si) durchgeführt. Durch den längeren Gebrauch war die Spitze stark abgeplattet, so dass sich ein großer Auflageradius und damit auch eine große Kapazitätsänderung $C_{\text{prob}}(V)$ ergab. Dies zeigte sich in der Beeinflussung des Kapazitätssensors durch die Messkapazität und damit in dem resultierenden Messsignal. Dies überrascht nicht, da in der Nähe des Resonanzpeaks gemessen wurde (siehe im folgenden Bild oben links).

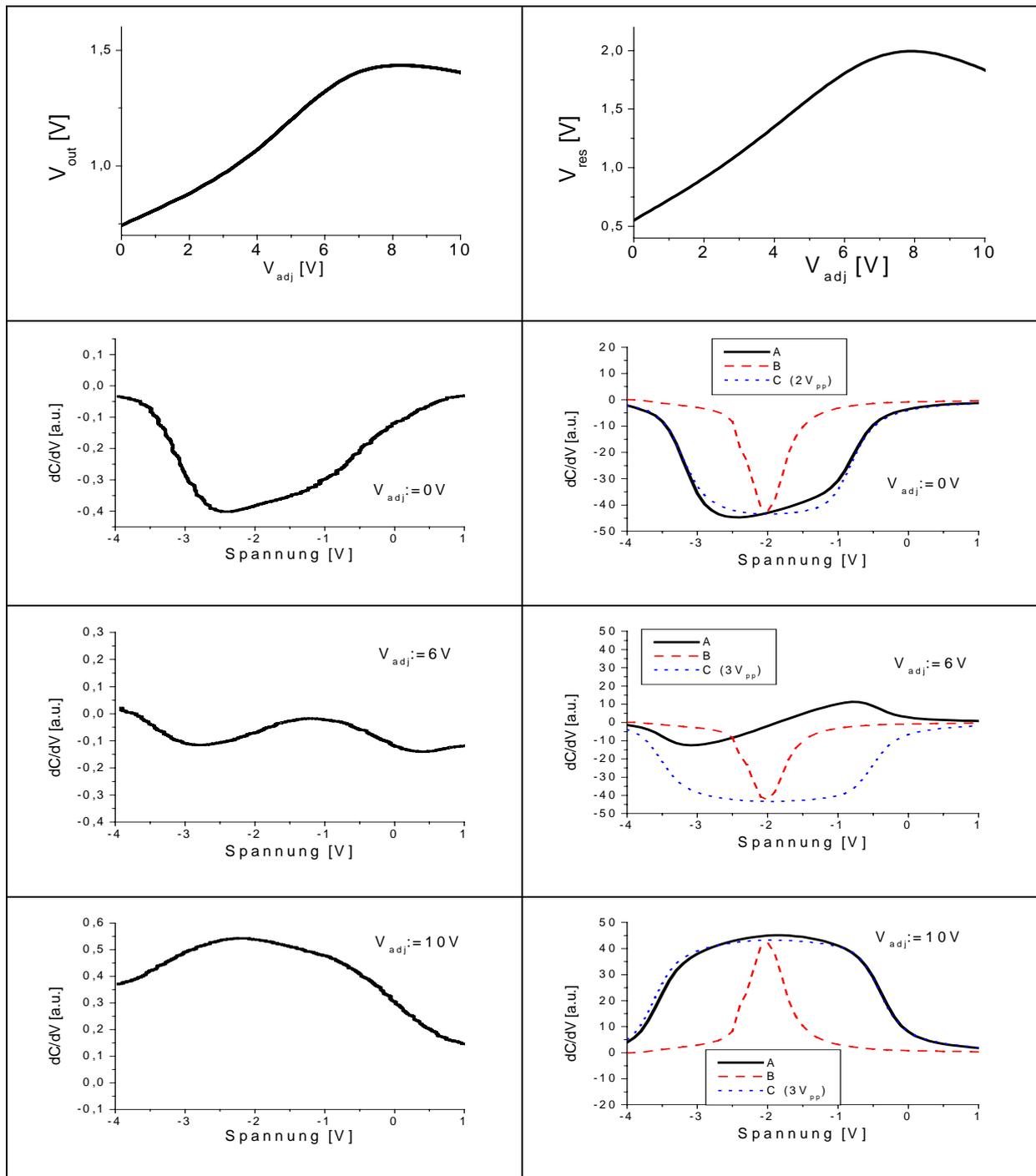


Bild 29: Gemessene (links) und simulierte (rechts) Resonanzkurve des Kapazitätssensors (oben). In Abhängigkeit von V_{adj} wurden dC/dV - V -Kurven gemessen (links, mit $V_{mod}=0.1V$) bzw. simuliert (rechts).

Allerdings ist es durchaus erstaunlich, wie früh die Beeinflussung stattfindet (schon bei $0V_{adj}$). Zusätzlich zum simulierten Signal (rechts, A) wurde auch die zu erwartende dC/dV - V -Kurve (B) angegeben. Ferner wurde über diese mit einer Amplitude gemittelt (C), die der Resonanzpeak-Kurve bei der entsprechenden Spannung V_{adj} entnommen wurde.

Damit zeigt sich, dass die zu messende Kapazitätsänderung sehr klein gehalten werden sollte, damit keine Beeinflussung der Sensitivität des Sensors durch die Messkapazität stattfinden kann. Ferner sollte der Kapazitätssensor so eingestellt werden, dass kein Peak zu sehen ist und die dC/dV -V-Kurven bei einer Änderung von V_{adj} keine Gestaltänderung erleiden. Zudem zeigt sich, dass der Sensor – bei einer akzeptablen Sensitivität ($100\mu V_{pp}/aF$) – mit einer Amplitude von einigen Volt arbeitet, was dazu führt, dass die C-V-Kurven über den entsprechenden Bereich gemittelt werden. Bild 30 zeigt den Zusammenhang zwischen Messamplitude (V_{res} , mit $V_p=5V$) und der Sensitivität. Auch wenn die Sensitivität auf ein Zehntel abgefallen ist, beträgt die Messamplitude immerhin noch 1V! Mit einer Verkleinerung von V_{prob} geht auch immer eine Reduzierung der Sensitivität einher (siehe Formel 20).

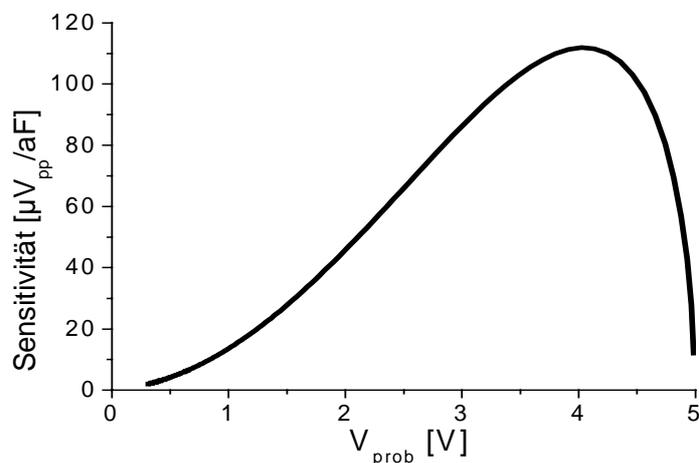


Bild 30: Zusammenhang zwischen Messspannung und Sensitivität des verwendeten Kapazitätssensors.

Da eine hohe Ortsauflösung und damit eine kleine Kapazität gewünscht ist, muss mit einer Sensitivität von mindestens $10\mu V/aF$ gearbeitet werden. Dies führt zur besagten Amplitude von mind. $2V_{pp}$. Um die theoretischen C-V-Kurven mit den experimentellen Kurven vergleichen zu können, müssen sie folglich über $\pm 1V$ gemittelt werden. Die Höhe der Modulationsspannung V_{mod} bewirkt wiederum eine Mittelung (diesmal der dC/dV -V-Kurve), so dass auch sie möglichst klein ($<1V$) gehalten werden sollte.

Den Einfluss der verschiedenen Spannungen (V_{prob} , V_{bias} , V_{mod}) auf das SCM-Signal ($\sim dC/dV-N$) veranschaulichen die Bilder 31 bis 33. Bild 31 enthüllt die enormen Auswirkungen der Biasspannung V_{bias} . Wie schon erwähnt, wurde als Bezugspunkt

für die Biasspannung das intrinsische Silizium gewählt (\equiv absoluter Null-Wert), d.h. es wurde angenommen, dass bei $V_{\text{bias}}=0\text{V}$ kein Kontaktpotential zwischen der Sonde und einer nicht-dotierten Probe besteht. Im realen Fall muss der bei $V_{\text{bias}}=0\text{V}$ existierende Potentialunterschied mit einer zusätzlichen Biasspannung (V_{offset}) ausgeglichen werden (auf eine experimentelle Bestimmung von V_{offset} wird noch eingegangen). Nur bei dieser Wahl der Biasspannung ist das Betragssignal von dC/dV für p- und n-Dotierung identisch. Bei n-Silizium ist dC/dV positiv, für p-Silizium negativ.

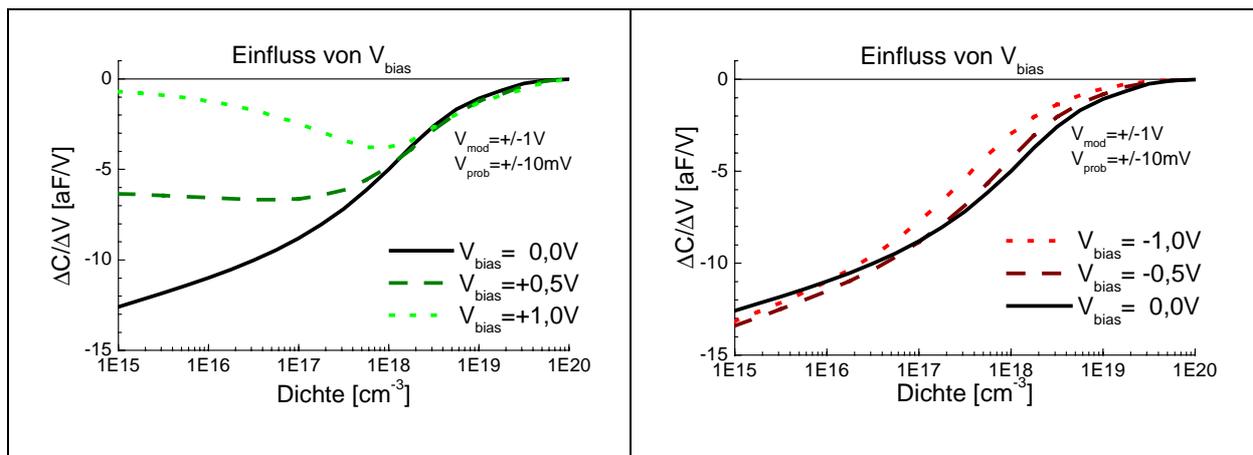


Bild 31: dC/dV -N Kurven bei Variation von V_{bias} .

Bei der Simulation der dC/dV -N Kurven in Bild 31 bis 33 wurde eine Sonde mit $r_a=47\text{nm}$, $r_i=25\text{nm}$ und $\alpha=20^\circ$ verwendet (Näheres zu diesen Parametern im nächsten Abschnitt). Im Bild 31 betrug die Messspannung $V_{\text{prob}}=10\text{mV}$, genauer: Der Simulator erfüllt immer die Klein-Signal-Bedingung. Dies entspricht dem experimentellen Fall, bei dem mit einer kleinen Amplitude von typischerweise 10mV gemessen wird. Die Modulations-Spannung liegt bei $V_{\text{mod}}=\pm 1\text{V}$ ($\Delta C=C(-1\text{V})-C(+1\text{V})$). Schon bei kleinen Abweichungen der Biasspannung vom absoluten Null-Wert zeigen sich große Unterschiede in den dC/dV -N Kurven. Bei positiver Spannung (bei n-Silizium ist es eine negative Spannung) steigt die Kurve nicht mehr monoton an. D.h. hier können einem Signalwert sogar zwei Dotierungswerte zugeordnet werden. Bild 32 veranschaulicht den Einfluss der Amplitude der Modulations-Spannung. Je größer die Amplitude, desto geringer die Abhängigkeit von der Biasspannung. Bei einer Modulations-Spannung von $\pm 0.1\text{V}$ wird eine nicht monoton steigende Funktion erzielt. Dies lässt sich anhand von C-V-Kurven bei unterschiedlicher Dotierung verstehen (siehe Bild 32 rechts). Da bei $V_{\text{bias}}=0\text{V}$ gemessen wurde, ist die Steigung der C-V-Kurven bei niedrigen Dotierungen ($<10^{17}\text{cm}^{-3}$) hier flacher als die der Kurven bei einer größeren Dotierung. Einer Vergrößerung der Modulations-Spannung

entspricht einer Mittelung über einen größeren Bereich, und damit wird auch die prinzipiell größere Steigung der C-V-Kurven bei niedriger Dotierung besser berücksichtigt.

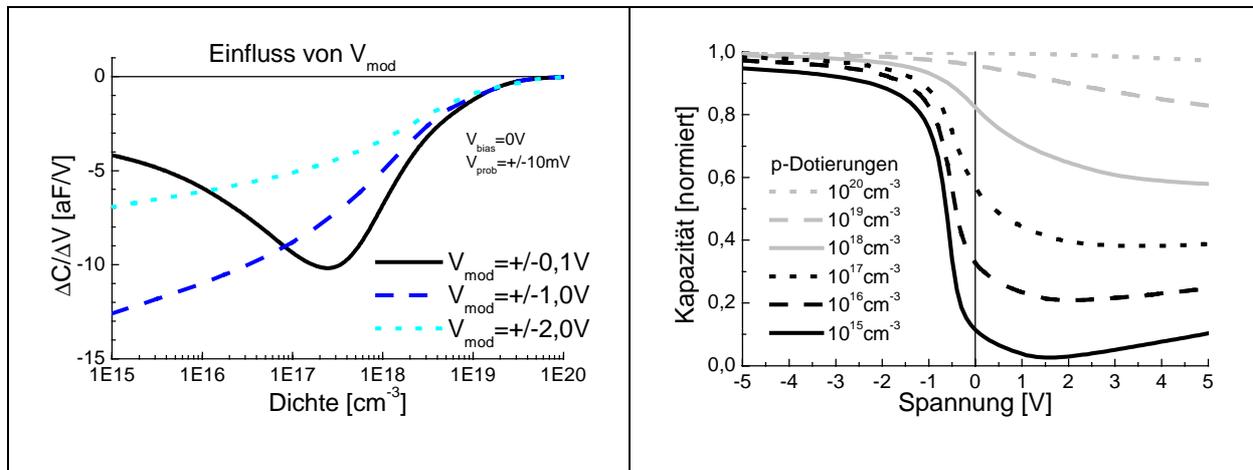


Bild 32: dC/dV -N-Kurven bei Variation von V_{mod} (links) und C-V-Kurven bei Variation von N (rechts).

Ein ähnliches Verhalten kann bei einer Vergrößerung der Messspannung V_{prob} beobachtet werden. Die Simulation einer größeren Messspannung erfolgte über die Simulation von C-V-Kurven bei Einhaltung der Kleinsignal-Bedingung (entspricht $V_{prob}=10\text{mV}$) mit anschließendem Mitteln über einen Spannungsbereich, der der gewünschten Messspannung entspricht. Experimentell konnte die Richtigkeit dieses Ansatzes bestätigt werden, wenn die Amplitude der Messspannung (also der Spannungsbereich über den gemittelt wurde) unterhalb 2V blieb. Oberhalb dieser Größe kam es zu Hysterese-Effekten in den experimentellen C-V-Kurven [Bor96, Seite 33]. Bild 33 verdeutlicht den Einfluss von V_{prob} . Bei einer Biasspannung von 0V tritt eine nicht monoton steigende Funktion nur bei Amplituden kleiner 0.5V auf (bei einer Modulations-Spannung von nur $\pm 0.1\text{V}$). Selbst bei einer Biasspannung von +1V wird bei einer Messspannung von $\pm 2\text{V}$ noch monoton steigendes Verhalten beobachtet.

Bei $V_{prob}=\pm 1\text{V}$ und $V_{mod}=\pm 1\text{V}$ kann die Biasspannung über einen Bereich von $\pm 1\text{V}$ variieren, ohne dass sich ihr monoton steigendes Verhalten ändert, sowohl für p- als auch für n-dotiertes Silizium (beim IMEC wurde hingegen auch in diesem Bereich ein nicht monoton steigendes Verhalten beobachtet [Ste98, Ste00], was meines Erachtens aber auf eine andere Ursache – schlecht präparierte Probe – zurückzuführen ist).

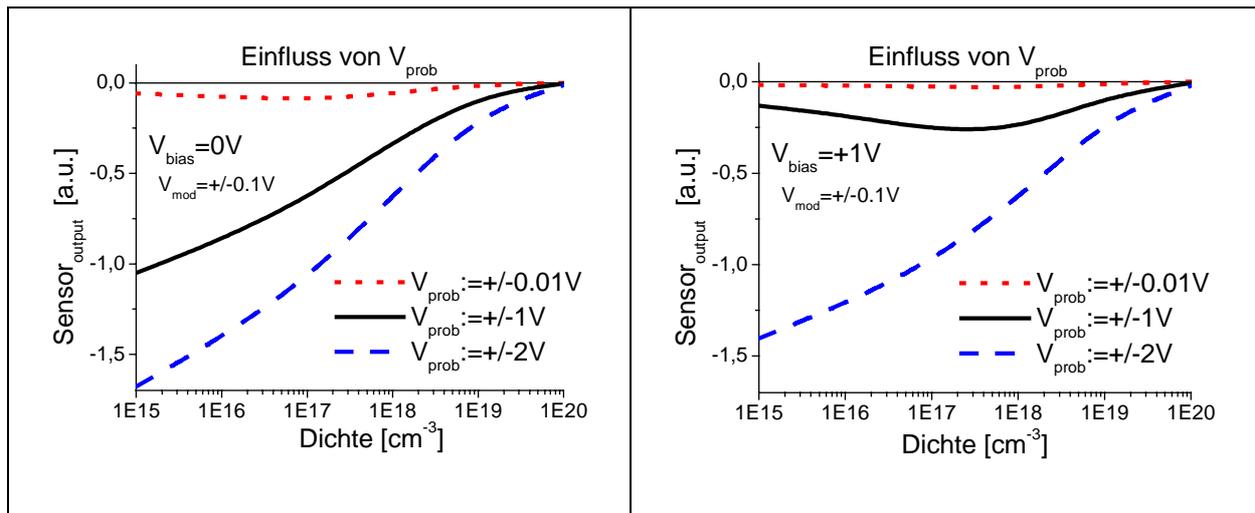


Bild 33: dC/dV -N Kurven bei Variation von V_{prob} bei einer Biasspannung von 0V (links) und +1V (rechts), mit Berücksichtigung der Abhängigkeit der Sensitivität des Sensors von V_{prob} .

Steigt der spezifische Widerstand der Probe R_{seri} (siehe oben Bild 26) über $5\Omega\text{cm}$ an, so bricht die Messspannung V_{prob} langsam zusammen und beeinflusst die zu messende Größe [Pal82]. Dies passiert bei homogenen Dotierungen bei Werten unterhalb 1 bzw. $4 \cdot 10^{15}\text{cm}^{-3}$ (bei n- bzw. p-Si). Ein entsprechendes Signal ist in Bild 34 dargestellt.

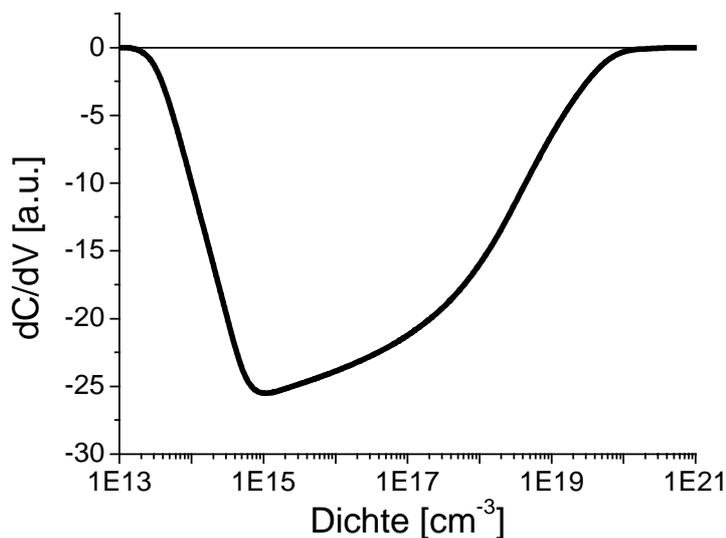


Bild 34: dC/dV -N Kurve bei Berücksichtigung der Eigenschaften des Kapazitätssensors.

Der Bereich der dC/dV -N Kurve zwischen 10^{15} und 10^{21}cm^{-3} kann mit einem 3D-Device-Simulator berechnet werden (bei Berücksichtigung der mittellenden Eigenschaften des Sensors durch die große Messamplitude). Der Bereich unterhalb 10^{15}cm^{-3} ist hingegen eine Auswirkung der Sensor-Eigenschaften und muss als

Instrumentenfunktion in die entsprechende Konversionskurve zusätzlich eingebracht werden.

Ebenso können schlechte Rückkontakte oder lange Signalleitungen zu einer Verfälschung der C-V-Kurve führen. Dies wiederum muss bei einer Messung vermieden werden. Alle zu messenden Dotier-Bereiche sollten mit einem möglichst kleinen Wechselstrom-Widerstand ($<10\text{k}\Omega$) kontaktiert sein.

Befinden sich p-n- oder Schottky-Übergänge in der Zuleitung (z.B. der Rückkontakt der Probe), so haben sie hingegen in der Regel keinen Einfluss auf das SCM-Signal, da die große Kapazität solcher Übergänge (im pF-Bereich) in Serie mit der eigentlichen Messkapazität liegt (im aF-Bereich).

Zusammenfassend zeigte sich, dass:

- eine Eichung des Kapazitätssensors für jede Probengeometrie vorgenommen werden müsste; sie ist zwar aufwändig, aber prinzipiell automatisierbar.
- der Kapazitätssensor das Messsignal stark beeinflussen kann. Im besten Fall resultiert die Beeinflussung in einer Verbreiterung der Messkurve (Mittelung über V_{prob}).
- der lateralen Auflösung aufgrund der Sensitivität des Sensors Grenzen gesetzt sind. Eine Erhöhung der Sensitivität geht nur zu Lasten einer größeren Amplitude, was zu einer größeren Mittelung (und somit zu einer schlechteren lateralen Auflösung) führt.
- bei sehr kleinen Mess- und Modulationsamplituden das SCM-Signal stark von der Biasspannung abhängt, deren absoluter Null-Punkt schwer exakt gemessen werden kann.
- die Sensitivität des Sensors unterhalb einer Dotierung von 10^{15}cm^{-3} abnimmt. Dies muss in einer Instrumentenfunktion berücksichtigt werden.

Verbesserungen sind hier nur durch ein anderes Sensordesign (siehe Anhang A1) und eine automatische Eichprozedur zu erzielen.

2.2.1.2 Sonden

Der größte physikalische Unterschied zwischen SCM und konventioneller C-V-Messung, liegt in der dreidimensionalen Geometrie von Sonde und Probe. Die Theorien konventioneller C-V-Messungen beruhen alle auf einer 1D-Geometrie-Annahme. Mögliche Randeffekte müssen bei der Messung vermieden werden. Beim SCM dominieren genau diese Effekte die Messung. Eine genaue Untersuchung der Auswirkungen stellt folglich einen wichtigen Meilenstein im Verständnis der Rasterkapazitätsmikroskopie dar.

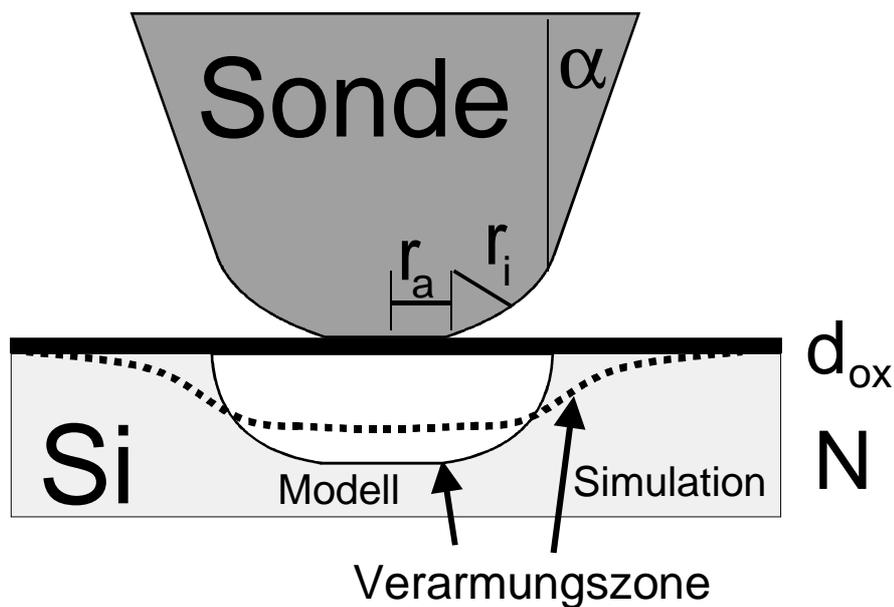


Bild 35: Skizze des verwendeten Sondenmodells.

Wie aus Bild 35 ersichtlich, wurde von einer Sonde ausgegangen, die im Kontaktbereich abgeplattet ist und einen Auflageradius r_a , einen Verrundungsradius r_i und einen Öffnungswinkel α besitzt. Obwohl die meisten Veröffentlichungen eine Kugelgeometrie als Sondenmodell verwenden, zeigen alle Experimente eine starke Abplattung der real eingesetzten metallisierten Silizium-Cantilever, nach einigen Rasterlinien bzw. schon beim ersten Kontakt von Sonde und Probe. Um den Einfluss der Randeffekte verstehen zu können, wurde von einer einfachen Gestalt der Raumladungszone (engl. Space Charge Region, SCR) ausgegangen. Diese kann dann aus einer Kombination von Platten-, Zylinder- und Kugelkondensator erfasst werden [Ise99] und wird im Folgenden als 3D-Theorie bezeichnet. Für eine Modellierung der Streukapazität der Sonde kann ebenfalls auf einfache Geometrien (z.B. Kegel über Ebene [Hud98]) zurückgegriffen werden. Problematisch an dieser

Herangehensweise ist die richtige Erfassung der Tiefe der SCR. Rechnungen mit einer SCR aus einem 1D-Modell konnten prinzipiell alle 3D-Effekte qualitativ richtig produzieren, scheiterten jedoch völlig am quantitativ richtigen Einfluss von r_a , r_i , α und V_{bias} . Dieser Einfluss kann nur mit einem 3D-Device-Simulator modelliert werden. In Bild 35 ist eine typische Simulations-SCR angedeutet. Sie verläuft flacher und ist insgesamt breiter als in der 3D-Theorie. Bild 36 zeigt drei C-V-Kurven, die aus MOS-Strukturen mit einem Radius der Auflagefläche von 47nm, einer Oxiddicke von 10nm und einer homogenen p-Dotierung von 10^{16}cm^{-3} berechnet wurden. Bei der 1D-Kurve (Bild 36, gepunktet) wurden alle Randeffekte vernachlässigt. Die Kreiskurve (Bild 36,

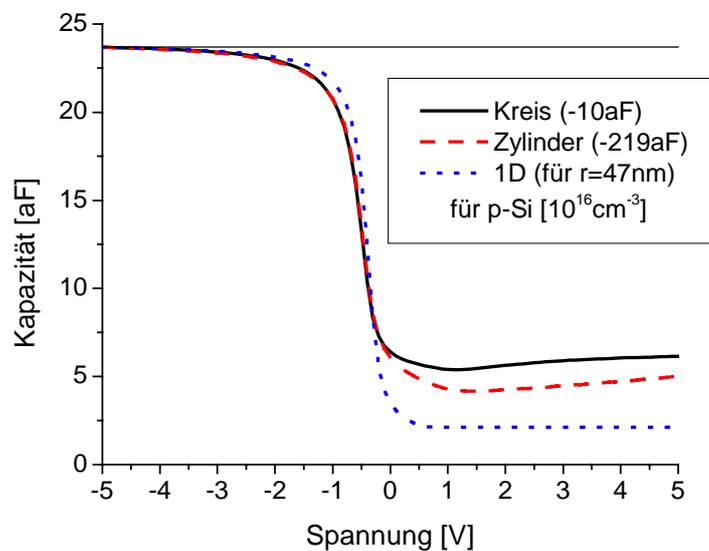


Bild 36: Simulierte C-V-Kurven.

durchgezogen) ergab sich bei einer Kreisscheibe als Gate-Elektrode, und bei der Zylinderkurve (Bild 36, gestrichelt) wurde mit einem 10 μm hohen Zylinder als Gate-Elektrode simuliert. Bei den letztgenannten Kurven wurde ein bestimmter Betrag (Streukapazität) abgezogen, so dass alle Kurven dieselbe Oxidkapazität aufwiesen. Schon die Kreisscheibe führte zu einer Streukapazität von 10aF ($\cong 55\%$ der eigentlichen Messkapazität), der Zylinder zu einem Wert von 219aF ($\cong 1200\%$). Experimentell wurde eine gesamte Streukapazität (Sonde, Cantilever, Zuleitungen) von 1-10pF beobachtet. Nur mit einem Draht als Sonde wird eine Streukapazität unterhalb 0.5pF erreicht [Sak96]. Im Bereich der Inversion konnten die 3D-Kurven nicht die niedrigen Werte der 1D-Kurve erreichen. Der Kapazitätsunterschied (Anreicherung-Inversion) und damit auch die Tiefe der SCR ist im 3D-Fall folglich geringer.

Beim SCM interessiert allerdings nicht so sehr das C-V- als vielmehr das dC/dV -N-Signal, also die Abhängigkeit des SCM-Signals von der Dotierung bzw. Ladungsträgerdichte. Dabei werden alle SCM-Messungen – wenn nicht anders erwähnt – mit $V_{\text{bias}}=0\text{V}$ durchgeführt. Dies ist in Bild 37 für die oben dargelegte 3D-Theorie, für die klassische 1D-Theorie und eine 3D-Simulation dargestellt. Die dotierungsabhängigen Werte wurden aus C(V,N)-Kurven ausgerechnet ($\Delta C=C(-1\text{V},N)-C(+1\text{V},N)$), wobei als Bezugspunkt der Biasspannung das Kontaktpotential des intrinsischen Siliziums gewählt wurde (absolut Null).

Während sich alle Kurven im Bereich hoher Dotierungen annähern, weichen sie im Bereich niedriger Dotierung stark voneinander ab. Die 3D-Theorie-Kurve stimmt mit der 3D-Simulations-Kurve im Bereich hoher bis mittlerer Dotierung annähernd überein. Der Verringerung des Kapazitätsunterschiedes aufgrund der Verbreiterung und Verflachung der SCR in der realistischen Simulation kann sie allerdings unterhalb 10^{18}cm^{-3} nicht mehr folgen. Dies ist im Bild 36 bei 10^{16}cm^{-3} anhand einer C-V-Kurve bzw. im Bild 35 oben schematisch dargestellt.

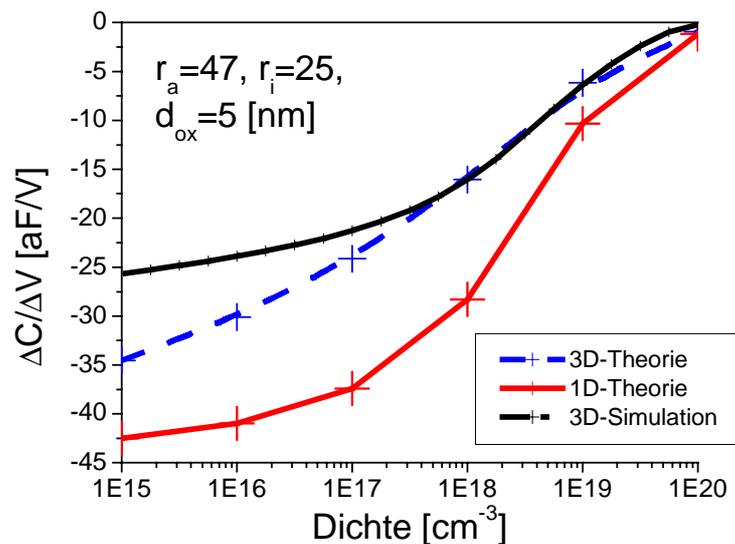


Bild 37: Simulierte bzw. theoretische dC/dV -N-Kurven.

Eine Einschätzung der Auswirkung der zwei dominierenden Sondenparameter r_a und r_i (der Einfluss von α ist verschwindend gering [Ise99]) ermöglicht Bild 38 (mit $d_{\text{ox}}=10\text{nm}$, $V_{\text{bias}}=0\text{V}$, $V_{\text{prob}}=10\text{mV}$ und $V_{\text{mod}}=1\text{V}$).

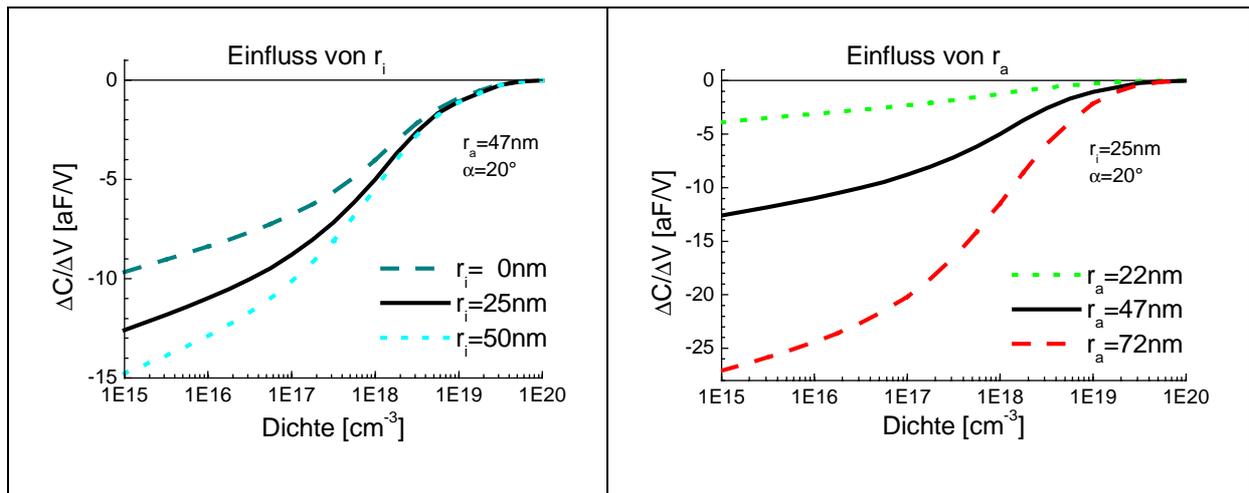


Bild 38: dC/dV -N Kurven bei Variation von r_i (links) und r_a (rechts).

Den größten Einfluss hat demnach die direkte Auflagefläche (r_a). Aus diesem Grund können Simulationen bzw. analytische Berechnungen mit einem Kugel-Ebene-Modell die experimentellen Ergebnisse (abgeflachte Spitze) nicht einmal annähernd richtig erfassen. Solche Kugel-Ebene-Modell-Berechnungen wurden z.B. am National Institute of Standards and Technology (NIST) [Mar98] und in Stanford [Yu99] durchgeführt und bilden dort die Grundlage für eine große Datenbank, mit der SCM-Messwerte in ein Dotierprofil umgerechnet werden. Obwohl der physikalische Ansatz hinter dieser Datenbank sehr fraglich ist, zeigten Vergleiche zwischen SCM- und SIMS-Messungen eine gute Übereinstimmung. Dies lässt sich – meiner Einschätzung nach – durch den dort verwendeten Fit-Mechanismus erklären (die SCM-Daten werden an die SIMS-Daten angepasst, wobei die der SCM-Datenbank zugrunde liegenden physikalischen Parameter wie z.B. der Sondenradius allerdings keine reale Bedeutung besitzen). Auf das Anfitzen von SCM-Daten mit physikalischen Modellen wird noch näher eingegangen.

Ebenso wie die Sondengestalt kann auch das Material der Sonde einen Einfluss auf das SCM-Signal besitzen, und zwar wenn es sich um eine dotierte Silizium-Spitze handelt. Kommerzielle Si-Sonden (von Nanosensors, siehe {Cantilever}) sind n-dotiert mit einer Konzentration von 10^{18}cm^{-3} . Bild 39 zeigt die Auswirkungen einer halbleitenden Sonde (n-Si, 10^{18}cm^{-3}) und halbleitender Proben (links p-Si, 10^{16}cm^{-3} und rechts n-Si, 10^{16}cm^{-3}) auf eine C-V-Kurve. Während es bei entgegengesetzter Dotierung von Sonde und Probe keinen großen Unterschied zu einer C-V-Kurve mit einer metallisierten Sonde gibt (Bild 39, links), da sich die im Vergleich zur Probe kleinere SCR in der Sonde bei negativer Spannung kaum auswirkt, ist bei

gleichnamiger Dotierung ein deutlicher Unterschied zu erkennen. Bei negativer Spannung besteht eine SCR in der Probe (n-Si) und eine Akkumulation in der Sonde (ebenfalls n-Si, aber „entgegengesetzte“ Spannung). Bei positiver Spannung ist es genau umgekehrt. Eine gleichnamige Dotierung von Sonde und Probe bewirkt also eine C-V-Kurve, die weniger stark von der Spannung abhängt, als wenn es sich um eine metallisierte Sonde handelt. Die Biasspannungs-Unabhängigkeit kann z.B. bei der Lokalisierung von p-n-Übergängen ausgenutzt werden.

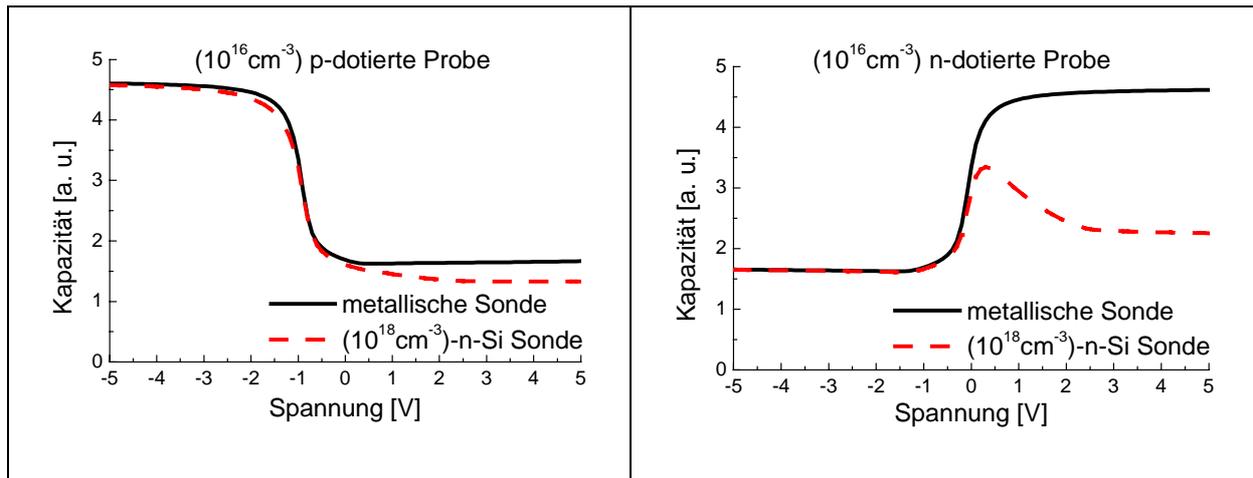


Bild 39: Auswirkung einer halbleitenden n-Si-Sonde auf die C-V-Kurve von einer p-Si- (links) und einer n-Si- (rechts) Probe.

Die Bell-Labs verwenden genau aus diesem Grunde Si-Sonden [OMa99]. Ein weiterer Vorteil besteht in dem meist kleineren Spitzenradius im Vergleich zu dem Radius einer metallisierten Sonde. Problematisch ist m.E. allerdings die Konvertierung der mit Si-Sonden gemessenen SCM-Daten. Zudem sind Silizium-Spitzen recht spröde und können bei kleinen Spannungen oxidieren und somit die SCM-Messung beeinflussen. Ferner ist das auf den Spitzen vorhandene native Oxid von schlechter Qualität und besitzt eine hohe Fehlstellendichte. Dadurch können Ladungen während einer Messung in der Spitze gespeichert werden und die Messung verfälschen. Nur mit gereinigten Silizium-Spitzen (HF-Ätze), die anschließend definiert oxidiert werden, sind m.E. reproduzierbare Messungen überhaupt erst möglich. Außerdem ist die Argumentation: kleiner Spitzenradius gleich hohe laterale Auflösung – gerade bei der Messung von p-n-Übergängen – nur bedingt richtig, da u.a. auch die Höhe der Dotierung die laterale Auflösung bestimmt (mehr dazu im weiteren Verlauf dieses Abschnittes).

In dieser Arbeit wurden nur metallisierte Silizium-Cantilever verwendet. Als Metallisierung wurde meist Kobalt-Chrom (CoCr) gewählt. Daneben wurden aber

auch Sonden mit Wolfram-Carbid (W_2C) mit einem spezifischen Widerstand von kleiner $100\mu\Omega\text{cm}$ [She98] benutzt. Die W_2C -Silizium-Sonden hielten länger, konnten aber anscheinend nicht mit der gleichen Güte wie die CoCr-Silizium-Cantilever hergestellt werden. Mit dotiertem Diamant beschichtete Silizium-Cantilever zeigten ähnliche Effekte in den C-V-Kurven, wie sie in Bild 39 zu sehen sind. Ansonsten konnten mit diesen Spitzen über einen sehr langen Zeitraum stabile Messungen durchgeführt werden.

Neben der Untersuchung der Eigenschaften von konventionellen Sonden, wurden auch die Möglichkeiten von Koaxial-Sonden untersucht. Diese bestehen aus einem Innenleiter (IL) und einem – durch eine dünne Isolationsschicht getrennten – Außenleiter (AL). Koaxial-Sonden lassen sich, in Bezug auf das SCM, für zwei verschiedene Messmodi nutzen, die als aktiver Modus und passiver Modus bezeichnet werden.

Beim aktiven Modus wird das Potential des AL – aktiv – auf dem gleichen Potential wie der IL gehalten. Mit dieser aktiven Abschirmung ist folglich eine Minimierung der Streukapazitäten verbunden. Allerdings müsste das Potential mit der Frequenz der Messspannung mitgeführt werden, was bei nahe 1GHz schwierig zu realisieren ist. Aus diesem Grunde würde sich eine aktive Abschirmung nur mit einem anderen Kapazitätssensordesign realisieren lassen, welches im Anhang vorgestellt wird. Diese Kombination könnte die Möglichkeit eröffnen, die lokale Kapazität auch in Abhängigkeit von der Messfrequenz zu bestimmen.

Beim passiven Modus wird der AL auf einem konstanten Potential gehalten. Je nach Spannung und Dotierung kann somit die SCR fokussiert und damit die laterale Auflösung gesteigert werden. Bild 40 verdeutlicht diesen Effekt für eine p^- -Dotierung (10^{16}cm^{-3}). Indem eine negative Spannung an dem Außenleiter anliegt, wird die – durch die positive Spannung des IL erzeugte – SCR eingeschränkt und damit fokussiert. Die Breite dieser SCR kann fast mit der Breite der Auflagefläche der Sonde (r_a) gleichgesetzt werden (selbst bei dieser niedrigen Dotierung).

Wie unter 2.2.1.5 noch eingehend erklärt wird, hängt die laterale Auflösung stark von der Dotierung ab.

Mit einer Koaxial-Sonde könnte diese Dotierungsabhängigkeit der Auflösung überwunden werden. Damit wäre auch eine einfachere Konversion der SCM-Messdaten in ein 2D-Dotierprofil möglich.

Bei einem p-n-Übergang müsste natürlich jede Rasterlinie mit zwei verschiedenen Spannungen am AL durchgeführt werden. Simulationen dieses Prinzips ergaben eine deutlich höhere Auflösung (siehe [Kra00]).

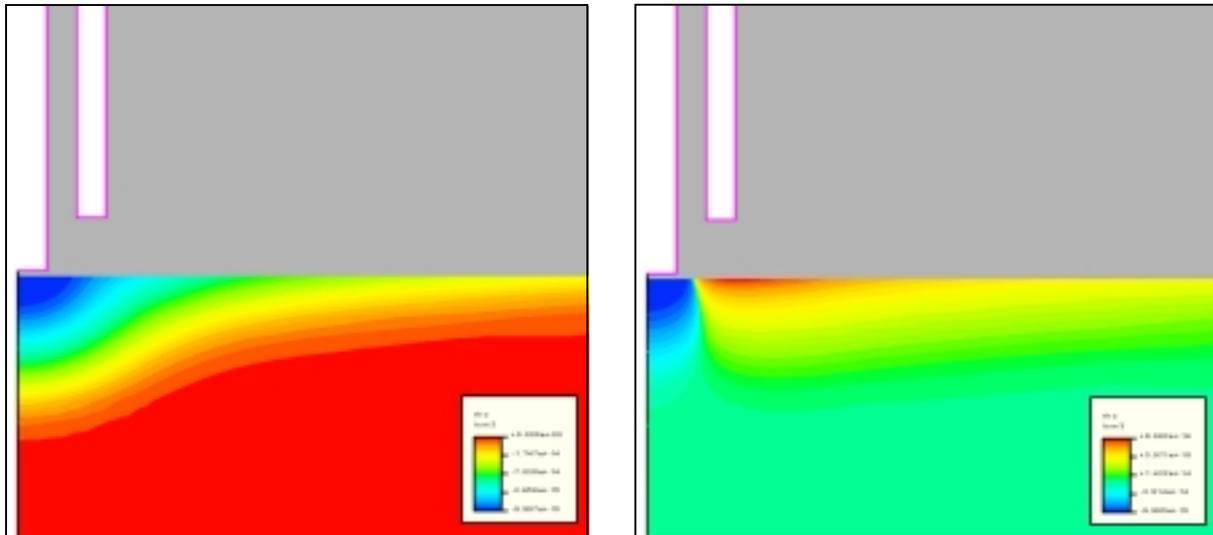


Bild 40: Schnitt durch die SCR einer Koaxial-Sonde mit unterschiedlichen AL-Spannungen (links $V_{AL}=0V$, rechts $V_{AL}=-5V$). Der IL liegt bei beiden Simulationen bei +5V.

In naher Zukunft sollen erste Experimente mit Koaxial-Sonden und einem selbstentwickelten hochsensitiven Kapazitätssensor durchgeführt werden.

2.2.1.3 Probenpräparation

Eine wichtige Voraussetzung für quantitative SCM-Messungen stellt die Proben-Präparation dar (siehe Bild 41). Da in die C-V-Kurve auch die Qualität der Oberfläche bzw. des Oxids eingeht, diese aber für die Dotierungsbestimmung nicht von Bedeutung sein sollte, muss dieser Einfluss minimiert werden.

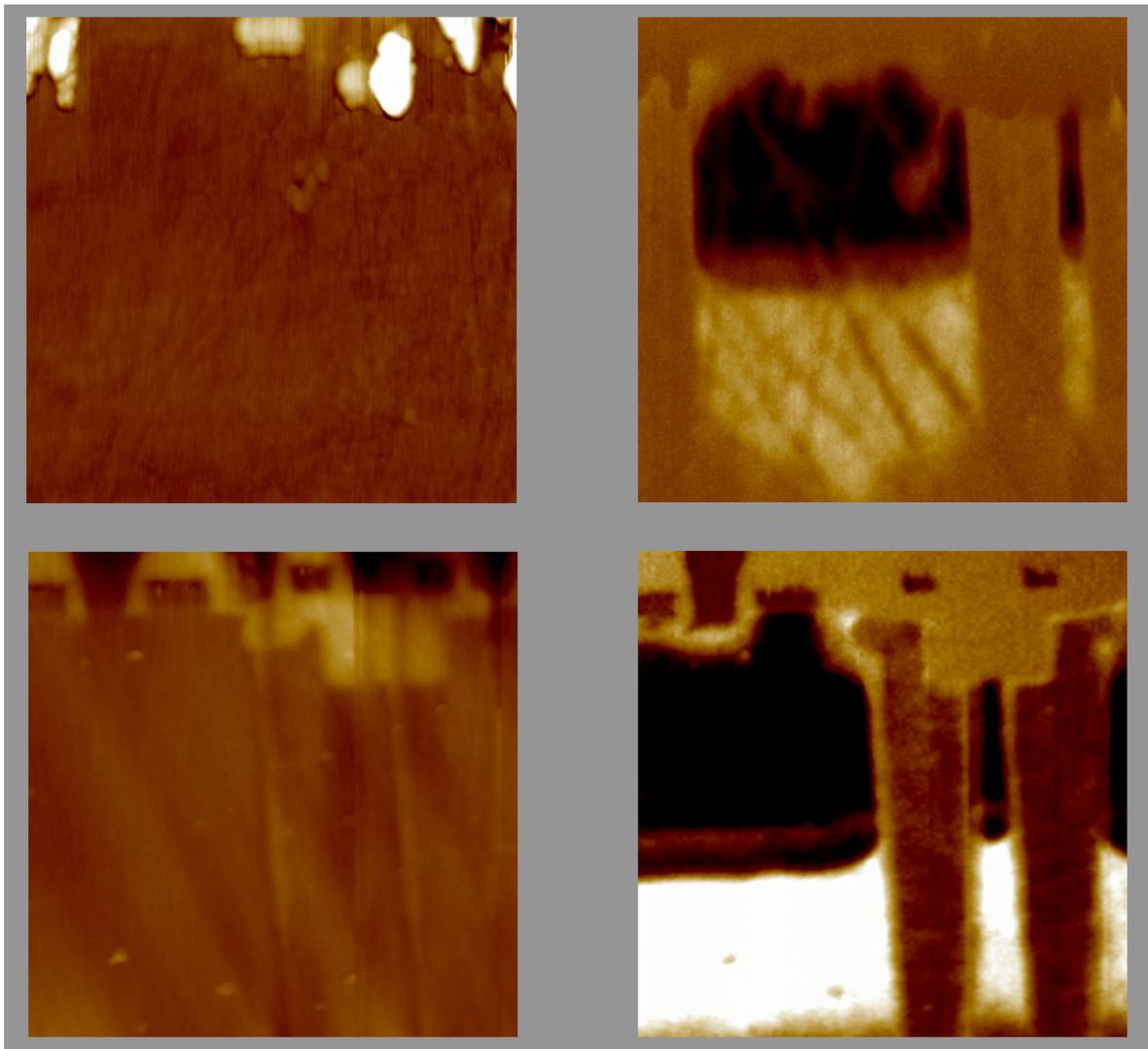


Bild 41: $2 \times 2 \mu\text{m}^2$ großes SCM-Bild einer Proben-Oberfläche (links Topografie, rechts SCM-Signal) nach einer konventionellen (oben) und der hier vorgestellten Präparations-Prozedur (unten).

Bei Proben, die schon mehrere Wochen unter Umgebungsbedingungen lagerten, wurden extreme Hysterese-Effekte beobachtet [Bor96]. Zudem war eine stabile Abbildung fast unmöglich, das SCM-Signal veränderte sich auch auf homogen dotieren Proben während der Messung bzw. wurde schnell schwächer und tendierte gegen Null. Dieselben Effekte konnten auch mit einem c-AFM beobachtet werden.

Nach einer Reinigungs-Prozedur verschwanden die Effekte jedoch sowohl bei der SCM- als auch bei der c-AFM-Messung, so dass sie – im Rahmen dieser Arbeit – nicht näher untersucht wurden.

Auch die in der SFM-Aufnahme deutlich sichtbaren Schmutzpartikel beeinflussen das SCM-Signal und sind mit konventionellen Reinigungsmethoden (Ultraschall-Bad mit Isopropanol oder Aceton [Bec88]) mitunter nur schwer zu entfernen. Gerade der Siliziumstaub, der beim Brechen von Proben auftritt, haftet extrem hartnäckig an der Oberfläche. Eine komplette RCA-Reinigung (siehe z.B. [Rug84]) kann dagegen bei den Proben aus der Fehleranalyse (FA) nicht angewendet werden, da es sich dabei um komplette ICs (mit Metallisierung) handelt.

Nur mit dem SnowJet [She94] {SnowJet} konnten zufriedenstellende Reinigungs-Ergebnisse erzielt werden. Bei diesem Gerät wird CO₂-Gas unter großem Druck durch eine Düse auf die Oberfläche geführt. Das CO₂ kristallisiert zu Eiskristallen. Der Reinigungseffekt wird zum einen durch Impulsübertrag dieser Eiskristalle auf große Schmutzpartikel, zum anderen über das Lösen kleinerer Schmutzpartikel durch eine Übertragung der Verdampfungsenergie des CO₂-Eises erzielt. Um eine dauerhafte Vereisung der Probenoberfläche zu verhindern, muss die Probe z.B. mit einem Heißluftfön, während der gesamten CO₂-Bestrahlung, erwärmt werden. Ein weiteres Problem ergab sich aufgrund der, durch die SnowJet-Bestrahlung auftretende starke Aufladung der Siliziumdioxid-Oberfläche [Hil95]. Mit dem KFM wurden Oberflächenpotentiale von bis zu 20V gemessen. Durch diese Aufladung der Oberfläche können auch die SCM-Signale von der Probe indirekt beeinflusst werden (darauf wird im Kapitel 2.2.2 noch näher eingegangen), so dass die Oberfläche nach der SnowJet-Bestrahlung z.B. per Ethanol-Bad entladen werden muss.

Bei der sogenannten Querschnitts-Präparation (engl. Cross-section) wird die Probe bis zur entsprechenden Ebene quergeschliffen (Zielschliff). Die besten Resultate wurden mit eingebetteten Schliffen erzielt. Dazu wird die Probe in ein spezielles Harz (Araldit XY103 mit Härter HY956 {Ciba-Geigy}) eingegossen und anschließend mit Schleifpapier erst einmal grob vorgeschliffen. Der Feinschliff erfolgt dann mit Diamantpaste mit immer feiner werdender Körnung (typ. 6µm, 3µm, 1µm, 0.25µm) [Bec88]. Auch freistehende Schliffe auf einer Quarzplatte sind möglich und sogar schneller durchzuführen, sie resultieren aber in einer deutlich größeren Rauigkeit.

Insgesamt gelingt nicht jeder Schliff gleich gut. Notfalls muss der Feinschliff mehrmals wiederholt werden (nach jedem Schliff ist eine SFM-Kontrollmessung notwendig). In der Regel konnte aber stets eine mittlere Rauigkeit von $1\text{-}5\text{nm}_{\text{RMS}}$ (SFM-Messung auf einer Fläche von $10\times 10\mu\text{m}^2$) erreicht werden. Diese in der Fehleranalyse verbreitete Standardprozedur reichte allerdings für hochauflösende SCM-Messungen nicht aus (siehe Bild 41 oben). Problematisch wirkt sich nach dieser Probenpräparations-Prozedur besonders die so erhaltene Oxidqualität aus, da die Ladungsdichte des Oxides bzw. der Grenzfläche Siliziumdioxid-Silizium proportional zur Rauigkeit dieser Grenzfläche ist [Yam99]. Sie lässt sich durch die Erhöhung der effektiven Fläche verstehen [Lai99]. Die Grenzfläche sollte daher so glatt wie möglich sein.

Aus diesem Grunde wurde die nach obiger Prozedur vorbehandelte Probe noch mit einer kolloidalen (mittlere Teilchengröße 60nm) Siliziumdioxid-Suspension (Mastermed auf Mastercloth {Buehler}) nachbehandelt. Nach einer Polierzeit von nur einer Minute konnte die mittlere Rauigkeit um eine Größenordnung gesenkt werden. Längere Polituren führten nur zu einer größeren Verrundung der Probenkante, aber nicht zu einer weiteren Reduzierung der Rauigkeit. Zur Politur wurde nur Wasser und keine anderen Ätzlösungen zugegeben, so dass der pH-Wert bei 9.5 lag. Bei diesem pH-Wert wird eine hohe Abtragsrate und eine geringe Oxidationsrate erzielt.

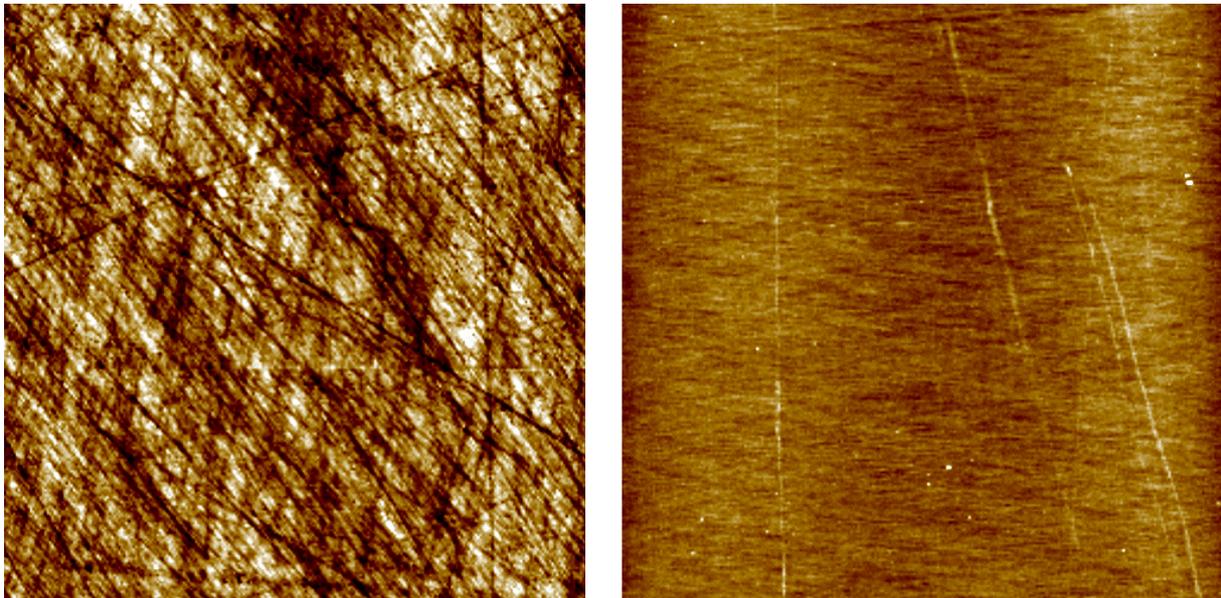


Bild 42: $10\times 10\mu\text{m}^2$ SFM-Bild vor (links) und nach (rechts) der Politur. In beiden Bildern wurde dieselbe Höhen-Skala (3nm) verwendet.

Die so erhaltene Siliziumdioxid-Schichtdicke sollte damit direkt nach der Politur nur eine Monolage betragen. Bild 42 veranschaulicht die Auswirkungen der be-

schriebenen Politur. Durch diese Präparation konnten Rauigkeiten bis zu 1\AA_{RMS} (SFM-Messung auf einer Fläche von $10\times 10\mu\text{m}^2$) erzielt werden. Die Siliziumprobe hat dann eine spiegelähnliche Oberfläche und zeigt keine Schleifrippen mehr. Die sichtbaren Kratzer entstanden durch das Entfernen der Polier-Kolloide mit einem Wattestäbchen, denn die Siliziumdioxid-Kolloide haften extrem hartnäckig an der Probenoberfläche. Erst durch ein Ultraschall-Bad der Probe in einer Seifenlauge (Tickopur R33) mit direkt anschließender SnowJet-Bestrahlung konnten sie gründlich – ohne Oberflächen-Beschädigung – entfernt werden.

Besonderer Wert wurde auf die Entwicklung eines Verfahrens zur Herstellung einer dünnen (3nm) hochqualitativen Siliziumdioxid-Schicht gelegt, denn die Qualität des Oxids spiegelt sich in der C-V-Messung direkt wieder (dafür werden 1D-C-V-Messungen in der klassischen HL-FA auch hauptsächlich eingesetzt). Aufgrund der Schwierigkeiten bei der Herstellung einer geeigneten Isolationsschicht (Passivierung) der HL-Oberfläche war der erste Transistor (1947) auch ein Punktkontakt-Transistor und kein MOSFET, obwohl Shockley diesen schon 1939 entworfen hatte [Que85]. Mittlerweile ist aber die großindustrielle Herstellung solcher idealer Isolationsschichten (für Schichtdicken $>5\text{nm}$) kein Problem mehr. Allerdings wird auch ein entsprechender Aufwand getrieben (so wird z.B. das benötigte Wasser im Oxidationsofen mit einer kontrollierten Knallgas-Reaktion aus spektroskopisch reinen Gasen hergestellt, um die Alkalimetall-Ionenkonzentration in der Oxidschicht zu minimieren). Die beste Oxidqualität wird mittels thermischer Oxidation erzielt. Diese kann allerdings für die mittels SCM zu untersuchenden Proben mit Metallisierung nicht verwendet werden, da Temperaturen über 500°C (typ. sogar mehr als 1000°C) auftreten. Zudem ist die thermisch erzeugte Oxiddicke abhängig von der unterliegenden Dotierkonzentration. Auch dies sollte natürlich bei einer SCM-Probe vermieden werden, da die Oxiddicke auch in die Kapazitätsmessung mit eingeht. Die meisten PVD- (engl. Physical Vapour Deposition) und CVD-Verfahren (engl. Chemical Vapour Deposition) erzeugen hingegen nicht die gewünschte Qualität. Sie sind bzgl. der Schichtdicke (3nm) schwierig zu kontrollieren, es treten teilweise zu große Temperaturen auf und/oder die Technik ist viel zu aufwändig. D.h. sie kann in der FA standardmäßig nicht eingesetzt werden (eine Übersicht über gängige Verfahren zur Herstellung von SiO_2 -Schichten liefern z.B. [Hil96], [Sze88], [Wid96]). Auch die nasschemische Oxidation liefert Oxide mit geringer Dichte (bzgl.

der Abhängigkeit von Ladungsspeicherung und Oxiddichte siehe [Mrs99]) und schlechter Qualität.

Als optimal für die Herstellung von dünnen hochqualitativen Oxidschichten für die SCM-Probenpräparation stellte sich die UV-Ozon-Oxidation heraus. Diese kann relativ einfach bewerkstelligt werden (dazu wurde ein Oxidationsofen aufgebaut), darüber hinaus ist sie zeitlich unkritisch, d.h. die Oxiddicke strebt nach einer Stunde gegen die gewünschten 3nm und bleibt dann stabil, bzw. längere Oxidationszeiten verbessern nur die Qualität des Oxides. Die UV-Ozon-Oxidation wird in der HL-Industrie schon länger eingesetzt, um Silizium-Oberflächen von Kohlenwasserstoff-Kontaminationen zu befreien [Tab84]. Diese können verschiedenen Ursprungs sein (Umgebungsbedingungen, nasschemische Verfahren) und besitzen einen großen Einfluss auf die Qualität der Grenzfläche.

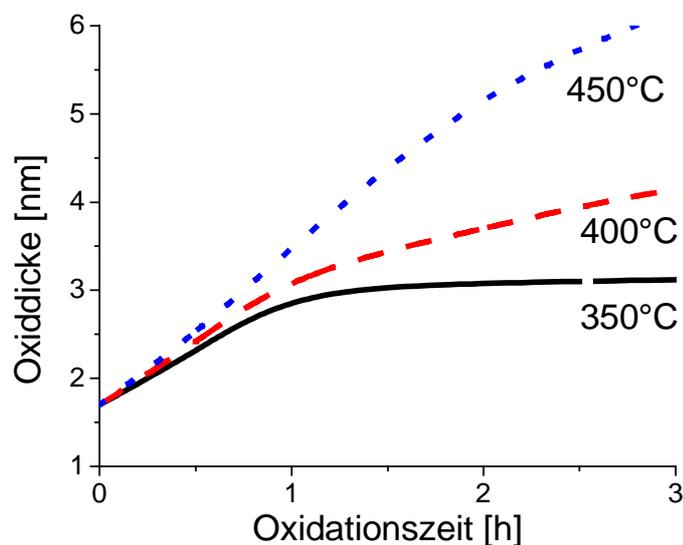


Bild 43: Zeitabhängige Oxiddicke bei der UV-Ozon-Oxidation [Boy93].

Um solche Kontaminationen zu entfernen, wird die Probe in einer Sauerstoff-Atmosphäre mit einer Hg-Niederdruck-Gasentladungslampe bestrahlt. Diese Lampe besitzt einen Quarzglas Kolben, der für UV-C-Strahlung transparent ist. So kann Strahlung mit der Wellenlänge $\lambda=184.9\text{nm}$ und 253.7nm die Lampe verlassen. Die kürzere Wellenlänge wird für die Erzeugung von Ozon und atomarem Sauerstoff verantwortlich gemacht, während die zweite Wellenlänge von vielen Kohlenwasserstoffen absorbiert wird, so dass diese aufbrechen und vom atomaren Sauerstoff zu flüchtigen Bestandteilen (CO , CO_2 , H_2O) oxidiert werden können. Nach einer UV-Ozon-Reinigung einer Silizium-Oberfläche beträgt die dadurch erzeugte

Siliziumdioxid-Schichtdicke 1.2-1.4nm [Pie92] (die ersten 0.5nm bilden sich nach nur 5s [Niw92]). Sie ist damit zu dünn, da diese Oxide bei SCM-Messungen – bei denen Spannungen von 2-3V auftreten – durchbrechen können. Erst eine Erhöhung der Proben-Temperatur auf 350°C während der Oxidation brachte schließlich nach 1h Oxidationszeit die gewünschte Schichtdicke von 3nm (siehe Bild 43 nach [Boy93, Boy97]). Damit wurde das Ziel: Herausarbeitung einer einfach anzuwendenden Methode zur Erzeugung einer 3nm dicken hochqualitativen Oxidschicht, die dotierungsunabhängig in akzeptabler Zeit aufgebracht werden kann, erreicht.

Eingebettete Schiffe dürfen allerdings nur bis maximal 100°C erhitzt werden, da sonst die Probenoberfläche durch das Harz stark kontaminiert wird. Deshalb muss es vor dem Oxidationsprozess entfernt werden. Allerdings verschmutzt die in der HL-FA übliche Prozedur (80°C heiße Ätzlösung) die Probenoberfläche irreversibel, so dass eine andere Methode entwickelt werden musste: Eine 3-5 Minuten dauernde Infrarot-Bestrahlung (mit einer 200W IR-Lampe) macht das Harz so spröde, dass eine kleine mechanische Spannung ausreicht, das Harz zu deformieren und aufzubrechen. So kann die IC-Probe unbeschadet aus dem Harz entfernt werden.

Wird solch eine Probe länger als 1h oxidiert, ändert sich bei 350°C Probentemperatur die Schichtdicke nicht wesentlich (siehe Bild 43) und die Oxidqualität nimmt zu. Soll der Prozess abgebrochen werden, so sollte zuerst die Sauerstoff-Zufuhr unterbrochen, die Kammer mit normaler Atmosphäre (oder feuchtem Sauerstoff) geflutet und erst danach die Ozonlampe abgeschaltet werden. Dadurch wird die Oberfläche mit Silanolgruppen (Si-OH) und einem dünnen Wasserfilm terminiert und eine schnelle Verschmutzung durch Kohlenwasserstoffe vermieden. Der dünne Wasserfilm bewirkt zudem durch eine teilweise Dissoziation ($\text{Si-OH} \rightleftharpoons \text{SiO}^- + \text{H}^+$) eine negative Oberflächenladung, die die Ausbildung von Wasserstoffbrückenbindungen unterdrückt [Pie92]. Die so entstandene hydrophile Oberfläche ist im Gegensatz zu einer hydrophoben Oberfläche sehr stabil gegenüber Kontaminierung aus der Atmosphäre. Erst nach Tagen konnte ein deutlicher Unterschied in den SCM-Messungen festgestellt werden.

Die erarbeitete Probenpräparation lässt sich wie folgt zusammenfassen:

Für Lift-Off-Proben:

(nach der konventionellen HF-Ätze zur Entfernung der Leitungsbahn-Ebenen)

| |
|---|
| 1. Gründliche Reinigung mit dem SnowJet |
| 2. Ethanol-Ultraschall-Bad (1 Minute) |
| 3. UV-Ozon-Oxidation (3h bei 350°C) |

Für Cross-Section-Proben:

(nach dem konventionell durchgeführten eingebetteten Zielschliff)

| |
|--|
| 1. Polieren mit Mastermed |
| 2. Ultraschall-Bad in Tickopur (5 Minuten) |
| 3. Gründliche Reinigung mit dem SnowJet |
| 4. Nach Kontrolle der Topografie (bzw. der erreichten Rauigkeit): Entfernen des Harzes |
| 5. Gründliche Reinigung mit dem SnowJet |
| 6. Ethanol-Ultraschall-Bad (1 Minute) |
| /. UV-Ozon-Oxidation (3h bei 350°C) |

Sollen mehrere Wochen **alte Proben** noch einmal vermessen werden, so sollte dieselbe Reinigung bzw. Präparation wie im Falle der Lift-Off-Proben wiederholt werden. Während die Oxiddicke bei der Wiederholung des 3. Präparationsschrittes kaum geändert wird, wird das Wasser aus dem Oxid entfernt (zum Zusammenhang von Wasser und Ladungsspeicherung im Oxid siehe [Fei81]) und die Oberfläche von Kohlenwasserstoff-Kontaminationen befreit.

Direkt nach der Präparation befindet sich auf der hydrophilen Oberfläche ein Wasserfilm, der einige Monolagen betragen kann, je nach der relativen Luftfeuchtigkeit (relative humidity, r.h.) der Atmosphäre. Dieser Wasserfilm – und damit die Umgebungsbedingung – beeinflusst eine SPM-Messung je nach Methode unterschiedlich stark. Beim STM kann er die Abbildung einer Probe entscheidend verändern (siehe z.B. [Pat97]). Die enorme Leitfähigkeit dieses Wasserfilms (bis zu 10^4 mal größer als die von Volumenwasser [Hei98]) ermöglicht sogar STM-Messungen auf Isolatoren [Guc94]. Beim SFM im Contact-Mode zeigt sich der

Einfluss eines Wasserfilmes besonders deutlich in der Kraft-Distanz-Kurve (siehe Bild 3). Auswirkungen auf Reibungskräfte sind ebenso bekannt [Bin95]. Auch die Beschichtung der Sonde zur Reduzierung der Adhäsionskräfte wurde schon untersucht [Ume98]. Beim SThM und c-AFM wird sogar die laterale Auflösung entscheidend durch diesen Wasserfilm bestimmt [Maj99, Pet99]. Für viele Rastersondenmethoden, die unter Umgebungsbedingungen arbeiten, wurde dieser Einfluss aber überhaupt noch nicht in Betracht gezogen. So gibt es bisher auch für die Rasterkapazitätsmikroskopie keine aussagekräftigen Studien, obwohl eine Beeinflussung der Messung durch den Wasserfilm schon früh erkannt wurde [Bor97]. Die Dicke eines solchen Wasserfilms lässt sich über den sogenannten „disjoining-pressure“ berechnen [Isr92, Maj99]. Es gilt:

$$P(d) := \frac{k_B \cdot T}{v} \ln(r.h.) \quad \text{mit} \quad P(d) := \sum_i P_i(d). \quad (21)$$

Dabei kennzeichnet k_B die Boltzmann-Konstante, T die Temperatur, v das Volumen eines Wassermoleküls ($3 \cdot 10^{-29} \text{m}^3$), $r.h.$ die relative Luftfeuchtigkeit und $P_i(d)$ die verschiedenen Wechselwirkungen, die zu einer Vergrößerung der Wasserfilmdicke führen. Im Einzelnen repräsentieren diese Terme die Summe der Van-der-Waals-Kräfte (P_1), die Bindungsenergie der Wasserstoffbrückenbindungen zwischen

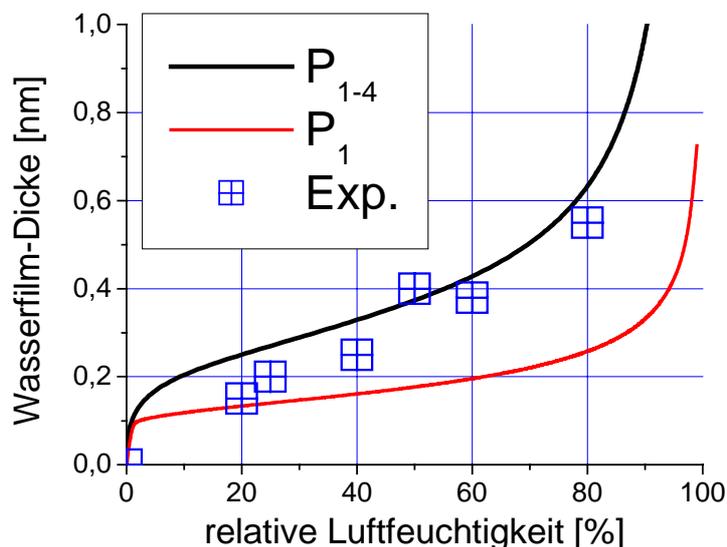


Bild 44: Wasserfilmdicke auf Siliziumdioxid in Abhängigkeit von der $r.h.$

Wasser und Siliziumdioxid (P_2), die elektrostatischen Kräfte zwischen einer Ionenlösung und einer geladenen Oberfläche (P_3) und die Energie, die durch eine Verspannung der Eis-ähnlichen ersten und zweiten Monolage mit der Unterlage

zustande kommt (P_4) (Einzelheiten siehe [Maj99]). Mit diesen Formeln kann die Abhängigkeit der Wasserfilmdicke von der relativen Luftfeuchtigkeit dargestellt werden (Bild 44). Dabei wurden in der Kurve P_{1-4} alle vier oben genannten Terme berücksichtigt, während in P_1 nur der Van-der-Waals-Anteil einfluss. Zudem sind noch einige experimentelle Resultate eingezeichnet, die mit unterschiedlichen Methoden – Ellipsometrie [Hei98] und Scanning Polarization Force Microscopy [Hu95] – gewonnen wurden. Danach kann bei einer r.h. von 50% von einer Wasserfilmdicke von 0.4nm ausgegangen werden. Bei 90% sind es sogar 1nm. Simulationen, die mit dieser Schichtdicke durchgeführt wurden, zeigten keine Auswirkung auf die C-V-Kurve.

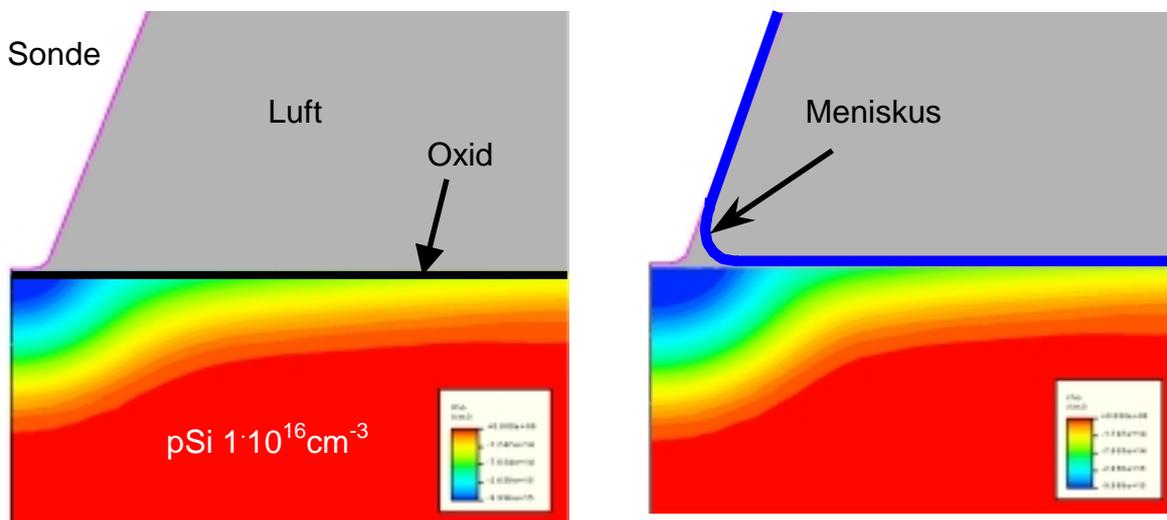


Bild 45: SCR ($V_{bias}=0V$) einer Sonde ($r_a=35nm$, $r_i=40nm$, $\alpha=20^\circ$), ohne und mit einem 3,4nm dicken Wasserfilm.

Dabei wurde allerdings nicht berücksichtigt, dass sich der Wasserfilm durch Kapillarkräfte an der Sonde hochzieht und einen Meniskus bildet, der mehrere nm dick sein kann. Problematisch ist auf der einen Seite die Bestimmung des Profils dieses Meniskus; zwar gibt die Kelvin-Gleichung [Isr92] für die Kugel-Ebene-Geometrie einen Anhaltspunkt (so hat der Meniskus einer Kugel mit einem Radius von 50-100nm einen maximalen Durchmesser von 35-50nm [Maj99] bei r.h.=50% und einem Kontaktwinkel von 0°), aber die eigentliche Sondengeometrie wird hierbei nicht berücksichtigt (siehe auch [Schen98]). Auf der anderen Seite sind grundlegende Rechnungen zu aufwändig [Gao97], so dass ein anderer Weg gewählt wurde: Die Form des Meniskus wurde definiert als die Äquipotentialfläche des elektrischen Feldes, das sich zwischen zwei Elektroden (Sonde und Probe) aufbaut. Diese Fläche kann natürlich mit dem bisher eingesetzten Device-Simulator (Lösung

der Laplace-Gleichung) einfach gewonnen werden. Als Dielektrizitätskonstante der so definierten zusätzlichen Isolatorschicht wurde $\epsilon=81$ gewählt. Dieser Wert gilt normalerweise auch bei 1GHz, aber wahrscheinlich nicht für eine Schichtdicke von 1-2 Monolagen (vermutlich gilt für die erste $\epsilon=6$ und für die zweite Monolage $\epsilon=30$). Andererseits ist der Meniskus mehrere nm dick, so dass obige Festlegung gerechtfertigt erscheint. Auf diese Art und Weise können die Auswirkungen des Wasser-Meniskus in die C-V-Simulation mit einbezogen werden, obwohl hier sicherlich nicht die gesamten physikalischen Eigenschaften (z.B. verändert er sich mit der angelegten Biassspannung) erfasst worden sind.

Solch ein Meniskus und die Auswirkungen auf die SCR sind in Bild 45 anhand eines 3,4nm dicken Wasserfilmes (99% r.h.) dargestellt. Bild 46 verdeutlicht die Auswirkungen eines 1nm bzw. 3,4nm dicken Wasserfilmes (90% bzw. 99% r.h.) auf die C-V-Kurve. Die C-V-Kurve zeigt einen großen Unterschied, der dadurch zustande kommt, dass die Verrundung der Sonde durch den Wasserfilm ausgeschmiert wird. Dies führt zu einer Verbreiterung der effektiven Auflagefläche und damit zu einer Erhöhung der Kapazität.

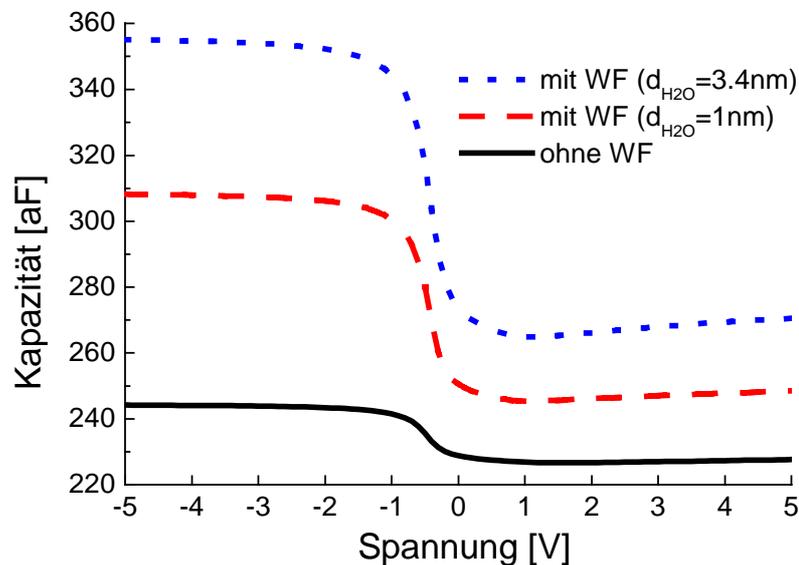


Bild 46: C-V-Kurven einer Sonde ($r_a=35\text{nm}$, $r_i=40\text{nm}$, $\alpha=20^\circ$) ohne und mit einem 1nm bzw. 3.4nm dicken Wasserfilm (WF).

In Bild 45 bzw. Bild 46 wurden extreme Fälle (99% bzw. 90% r.h.) dargestellt, um den Effekt zu verdeutlichen; normalerweise ist mit einer r.h. von 50% zu rechnen (in Reinräumen liegt sie noch darunter). Damit folgt eine Schichtdicke von 0.4nm und die in Bild 45 bzw. 46 dargestellten Effekte reduzieren sich erheblich, so dass der

Einfluss des Wasserfilms auf eine SCM-Messung – bzgl. des hier dargestellten Effektes der Vergrößerung der Auflagefläche – als nicht wesentlich bezeichnet werden kann. (Er kann allerdings auf eine mehr indirekte Art und Weise einen großen Einfluss auf die SCM-Messung haben, wie im Abschnitt 2.2.2 gezeigt werden wird.) Interessanter ist die Frage, ob der Wasserfilm nicht für eine neue verschleißfreie SCM-Messmethode eingesetzt werden kann. Bild 47 (links) zeigt das elektrostatische Potential (aus dem die Gestalt des Wassermeniskus entnommen wurde) einer Sonde, die um 50nm angehoben („geliftet“) wurde und über einen Wassermeniskus mit der Oberfläche verbunden ist.

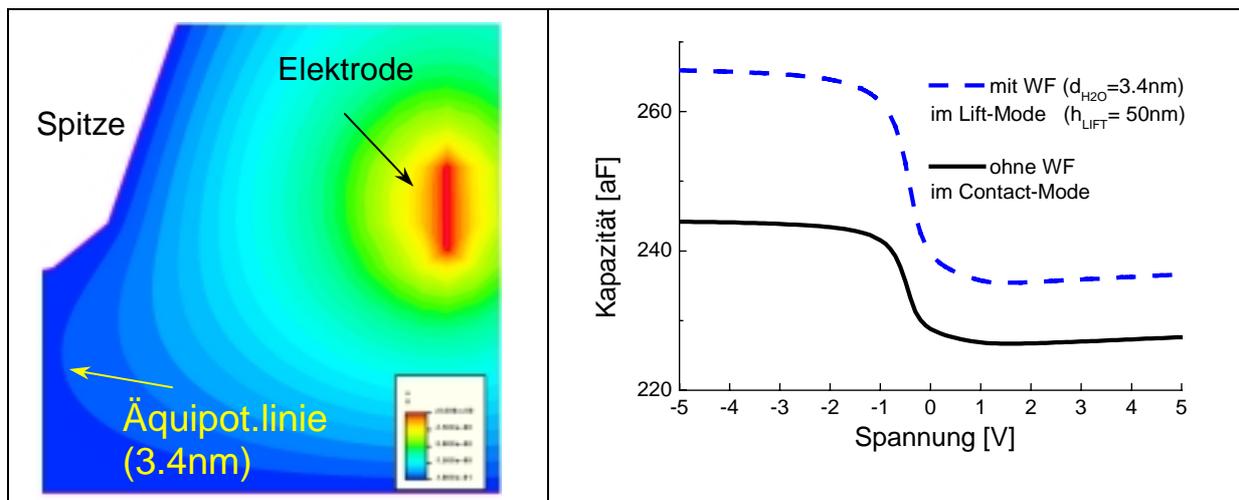


Bild 47: $100 \times 100 \text{ nm}^2$ Querschnittsbild (links) des elektrostatischen Potentials einer Sonde im Lift-Mode mit einer Lifthöhe von 50 nm. Die entsprechenden C-V-Kurven (rechts) sind einmal im Contact-Mode (ohne Wasserfilm) und einmal im Lift-Mode (mit einem 3,4nm dicken Wasserfilm) simuliert.

Das Signal (die C-V-Kennlinien aus Bild 47 rechts) der gelifteten Sonde ist durch den Wassermeniskus größer, als wenn sie sich ohne Wasserfilm im Kontakt mit der Probe befindet. Dies könnte genutzt werden, um die Sonde in einem kleinen Abstand zur Oberfläche zu führen (im Lift-Modus), wenn vorher die Topografie mit dem Tapping-Mode bestimmt wurde. Das SCM-Signal sollte dann nicht kleiner als im Contact-Mode sein, aber die Sonde würde sich bedeutend weniger schnell abnutzen, wodurch ein großes Problem der üblichen SCM-Messungen gelöst wäre. Auch die SCR ist durch den Fokussierungseffekt des Wassermeniskus nicht sonderlich größer, also die laterale Auflösung nicht viel schlechter als im Contact-Mode. Für Experimente muss natürlich ein extrem harter Cantilever ($k > 40 \text{ N/m}$) eingesetzt werden, da sonst die Sonde bis zum Abrisspunkt auf der Oberfläche – aufgrund der großen Adhäsionskräfte des Wassermeniskus ($F = 10\text{-}100 \text{ nN}$ je nach Spitzenform und Abstand von der Oberfläche [Gao97]) – verweilt und sich somit keine

Verschleißminimierung erzielen lässt. Leider zeigten die Experimente, dass eine solche Messung sehr instabil ist. Diese Methode wurde daher auch nicht weiter entwickelt.

Eine weitere Möglichkeit, um den Wasserfilm für verschleißarme SCM-Messungen auszunutzen, besteht in der spannungsinduzierten Ausbildung eines Wasser-meniskus während des Tapping-Modes, was eine bedeutend stabilere Messung erlaubt [Gar98]. Allerdings müssten dazu hohe Biasspannungen (12V) angelegt werden, die wiederum eine SCM-Messung verhindern.

2.2.1.4 Dotierungs-Auflösung

Um Aussagen zur Dotierungs-Auflösung des SCM zu treffen, müssen Simulationsergebnisse (Kapitel 2.2.1.2) mit den Eigenschaften des Sensors (Kapitel 2.2.1.1) kombiniert werden. Mit $r_a=47\text{nm}$, $r_i=25\text{nm}$, $\alpha=20^\circ$ und $d_{\text{ox}}=5\text{nm}$ folgt bei einer Sensitivität des Kapazitätssensors von $20\mu\text{V/aF}$ und einer Rauschamplitude von $50\mu\text{V}_{\text{pp}}$ eine mögliche Auflösung hinsichtlich der Bestimmung der Dotierung, die in Bild 48 dargestellt ist. Das linke Bild zeigt das mit Rauschen belegte Messsignal, das sich aus der Simulation und den Sensor-Kenngrößen ergibt. Das rechte Bild gibt die maximal mögliche Auflösung der Dotierung wieder.

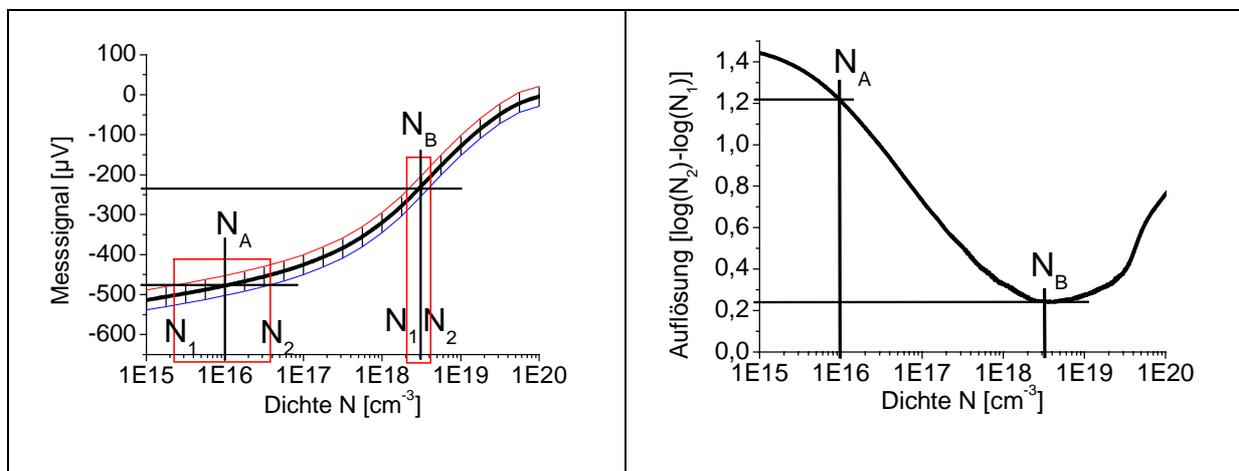


Bild 48: Mit SCM erzielbare Dotierungs-Auflösung.

So kann z.B. eine Dotierung von 10^{16}cm^{-3} (entspricht einem Signal von $480\mu\text{V}$) nur mit einer Genauigkeit von einer Größenordnung ($2 \cdot 10^{15}$ bis $3 \cdot 10^{16}\text{cm}^{-3}$, entspricht einem Rauschen von $50\mu\text{V}$) bestimmt werden. Mit einer Vergrößerung der Sensitivität des Sensors bzw. Verringerung des Rauschens könnte die Dotierung natürlich auch genauer bestimmt werden.

Bild 49 gibt die Dotierungs-Auflösung von zwei verschiedenen Sonden $r_a=47\text{nm}$ (links) und $r_a=22\text{nm}$ (rechts) mit jeweils zwei verschiedenen Sensitivitäten des Sensors $20\mu\text{V/aF}$ (obere Kurve) und $100\mu\text{V/aF}$ (untere Kurve) wieder.

Die beste Auflösung (Minimum) wird zwischen 10^{18} und 10^{19}cm^{-3} erzielt. Insgesamt ergibt sich die Dotierungs-Auflösung aus den Kennwerten der Sonde (Auflagefläche dominiert) und des Sensors (Sensitivität und Rauschen).

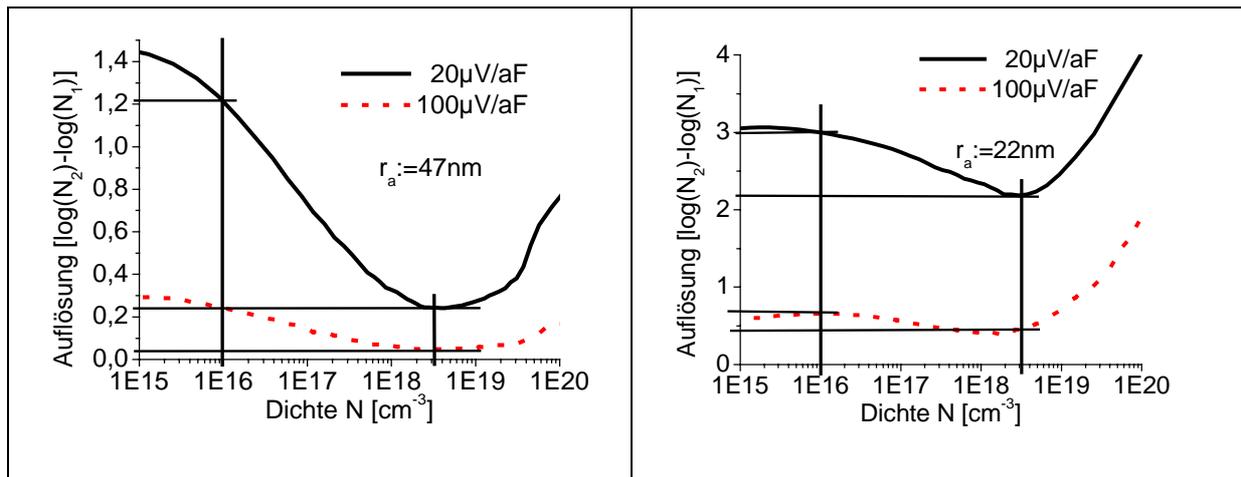


Bild 49: Dotierungs-Auflösung mit unterschiedlichen Sonden und Sensitivitäten.

Bisher wurde davon ausgegangen, dass die Modulationsamplitude V_{mod} konstant gehalten und ein Signal dC/dV_{mod} , also die Kapazitätsänderung ΔC gemessen wurde. Eine andere Möglichkeit ist, das Signal dC/dV_{mod} , also die Kapazitätsänderung ΔC , mit Hilfe eines Regelkreises konstant zu halten und die Modulationsamplitude dem entsprechend nachzuregeln. Dieser Modus wird als $\Delta C = \text{konst.}$ -Modus oder auch als ΔC -Feedback-Modus bezeichnet und wurde an der University of Utah von C.C. Williams (einer der geistigen Väter des Einsatzes des SCMs in der Halbleiteranalytik) entwickelt [Hua95]. Seiner Argumentation nach, um die laterale Auflösung des Systems zu verbessern. Williams argumentiert, dass die laterale Auflösung des SCM durch die Ausdehnung der SCR bestimmt wird und dass diese wiederum durch die Größe der Modulationsspannung dominiert wird. Eine konstante Modulationsamplitude bewirke bei einer niedrigen Dotierung eine viel größere Änderung der SCR als bei einer höheren Dotierung. Um eine gleich große Änderung der SCR und damit auch eine gleich hohe (dotierungsunabhängige) laterale Auflösung zu erzielen, sollte die Modulationsamplitude im $\Delta C = \text{konst.}$ -Modus nachgeregelt werden. M.E. nach greift diese Argumentation nur zum Teil, da die Größe der SCR erst einmal durch die Biasspannung V_{bias} (im Idealfall zwischen -0.5 und 0.5 V, je nach Dotierung) bestimmt wird. Die Änderung dieser SCR wird wiederum von der Messamplitude V_{prob} definiert, die – wie im vorhergehenden Kapitel dargelegt – zwischen $1V_{\text{pp}}$ und $5V_{\text{pp}}$ betragen kann. Erst eine zusätzliche Änderung wird durch die Modulationsspannung V_{mod} bewirkt. Diese hat typischerweise eine Amplitude von kleiner $1V_{\text{pp}}$ und fällt somit nicht sonderlich ins Gewicht. Während auf die laterale Auflösung noch

näher eingegangen wird, soll im Folgenden die Dotierungs-Auflösung des $\Delta C = \text{const.}$ -Modus untersucht werden.

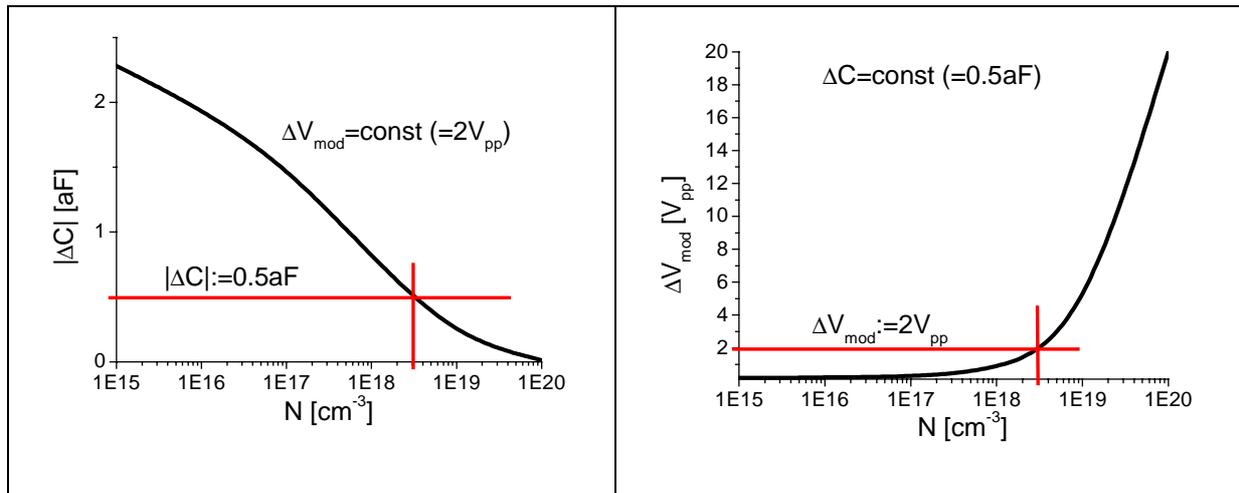


Bild 50: Vergleich der Response des $\Delta V = \text{const.}$ - (links, nur der Betrag ist dargestellt) und des $\Delta C = \text{const.}$ -Modus (rechts).

Typische dotierungsabhängige SCM-Response-Kurven sind in Bild 50 wiedergegeben. Links ist der Betrag der bekannten $\Delta V = \text{const.}$ ΔC - N -Kurve (bisher als dC/dV -Kurve mit absoluten Werten wiedergegeben, d.h. negative Werte im Falle von p-Si) und rechts eine $\Delta C = \text{const.}$ ΔV - N -Kurve dargestellt. Bei letzterer wird eine Modulationsamplitude von $2V_{\text{pp}}$ erst bei einer Dotierung von über 10^{18}cm^{-3} benötigt, um einen Kapazitätsunterschied von 0.5aF zu erzielen. Bei 10^{20}cm^{-3} sind es bereits $20V_{\text{pp}}$. Damit steigt die Kurve fast exponentiell an, was – auf den ersten Blick – auf eine hervorragende Dotierungs-Auflösung im Bereich über 10^{18}cm^{-3} schließen lässt. Dies wurde näher untersucht und zeigt (siehe Bild 51), dass die Dotierungs-Auflösung im $\Delta C = \text{const.}$ -Modus oberhalb 10^{18}cm^{-3} vergleichbar und oberhalb 10^{19}cm^{-3} besser als im $\Delta V = \text{const.}$ -Modus ist. Durch eine Änderung der Einstellung des Regelkreises (ΔC_{offset}) kann der Bereich der höheren Auflösung im $\Delta C = \text{const.}$ -Modus auch in Richtung niedriger Dotierung geschoben werden. Dann können allerdings keine höheren Dotierungen aufgelöst werden, da die maximale Modulationsamplitude auf $20V_{\text{pp}}$ begrenzt ist. Zudem kann die Dotierungs-Auflösung im $\Delta V = \text{const.}$ -Modus natürlich auch durch eine Erhöhung von V_{mod} optimiert werden.

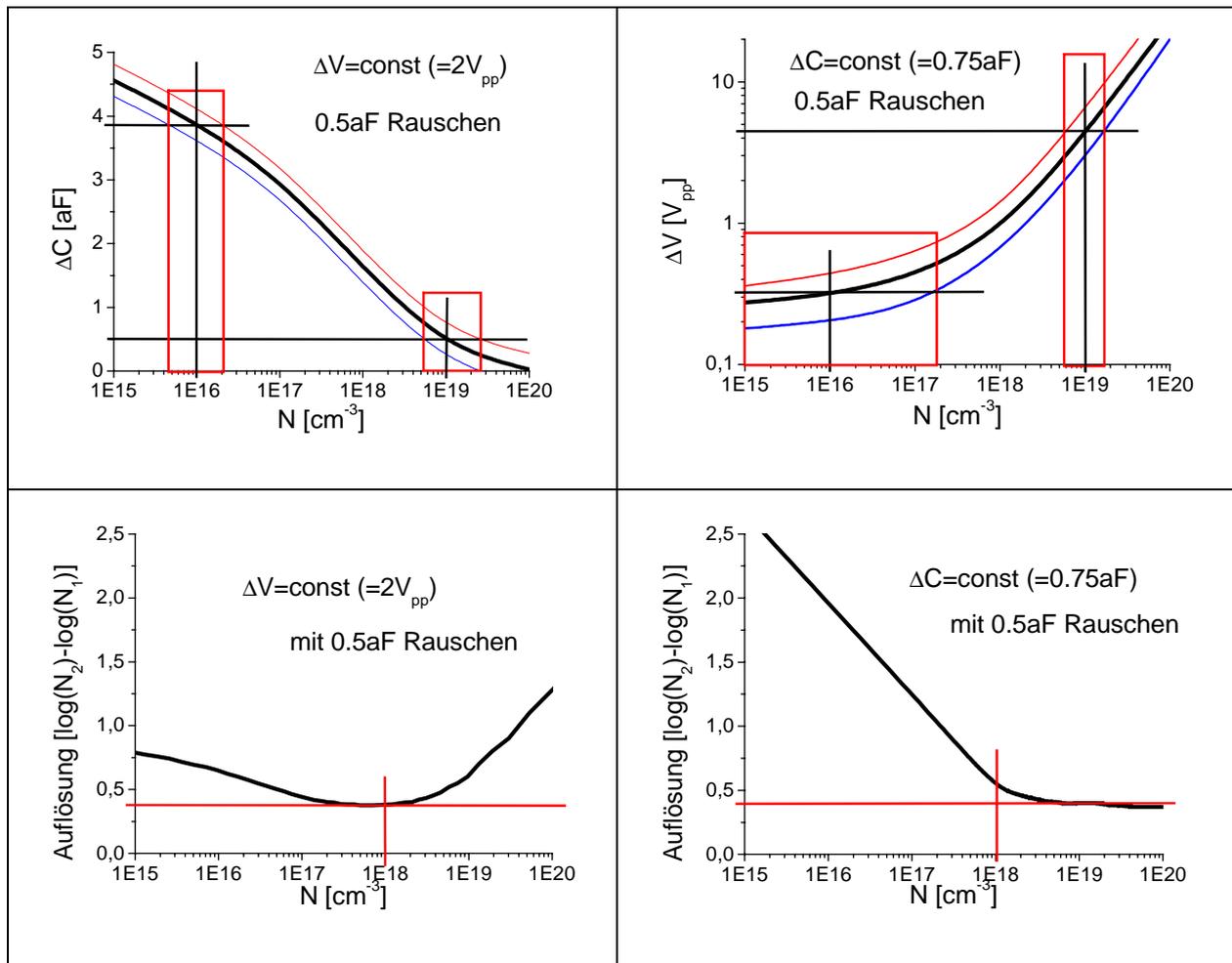


Bild 51: Vergleich der Dotierungs-Auflösung des $\Delta V = \text{const}$ - (links) und des $\Delta C = \text{const}$ - Modus (rechts).

Insgesamt bietet der $\Delta C = \text{const}$ -Modus bei der Dotierungs-Auflösung, wie auch bei der lateralen Auflösung (siehe [Kra00]) und der direkten Konversion, keine wesentlichen Vorteile. Er bietet nur eine begrenzte Verbesserung gegenüber dem $\Delta V = \text{const}$ -Modus, besitzt jedoch eine Reihe von Nachteilen: Es treten bis zu $20V_{pp}$ große Amplituden auf, die zu großen Strömen und damit zu Oxid-Schädigungen führen können. Die Auflösung in Richtung kleinerer Dotierungen nimmt rapide ab und im $\Delta C = \text{const}$ -Modus kann nur der Betrag gemessen werden. Das heißt, wir verlieren die Phaseninformation und damit die Möglichkeit zwischen n- und p-Si überhaupt unterscheiden zu können. Auch experimentell konnte keine höhere Auflösung beobachtet werden. Aus diesem Grunde wurde fast ausschließlich der $\Delta V = \text{const}$ -Modus eingesetzt. Im folgenden wird immer dieser Modus zugrunde gelegt, wenn nicht explizit auf den $\Delta C = \text{const}$ -Modus hingewiesen wird.

2.2.1.5 Laterale Auflösung

Obwohl die instrumentenbedingte laterale Auflösung eine der wichtigsten Kenngrößen der 2D-Dotierprofil-Bestimmung darstellt, gibt es – trotz der großen Anzahl an SCM-Veröffentlichungen – keine konkreten Untersuchungen und damit auch keine genaueren Aussagen über die mögliche laterale Auflösung eines Rasterkapazitätsmikroskops. Dies soll in diesem Abschnitt nachgeholt werden.

Im idealen 1D-Fall bewirkt die Biasspannung eine (vertikale) SCR mit einer scharfen Grenze, die durch $W(V)$ gegeben ist. Indem V_{bias} von dem Bereich der Akkumulation (negative Spannung bei p-Si) in den Bereich der Verarmung (positive Spannung) gefahren wird, wird die SCR vergrößert und die Kapazität verkleinert. Mit einer Änderung der Biasspannung wird die Dotierung in einer bestimmten Tiefe ausgemessen. Diese Tiefe hängt zudem von dem darüber liegenden Dotierprofil ab. Bei einer homogenen Ladungsträgerdichte von 10^{16}cm^{-3} beträgt die Ausdehnung der SCR bei $V_{\text{bias}}=0\text{V}$ $W(V)=120\text{nm}$ (siehe Bild 52). Dies kann als laterale Auflösung bzw. im 1D-Fall als vertikale Auflösung angesehen werden.

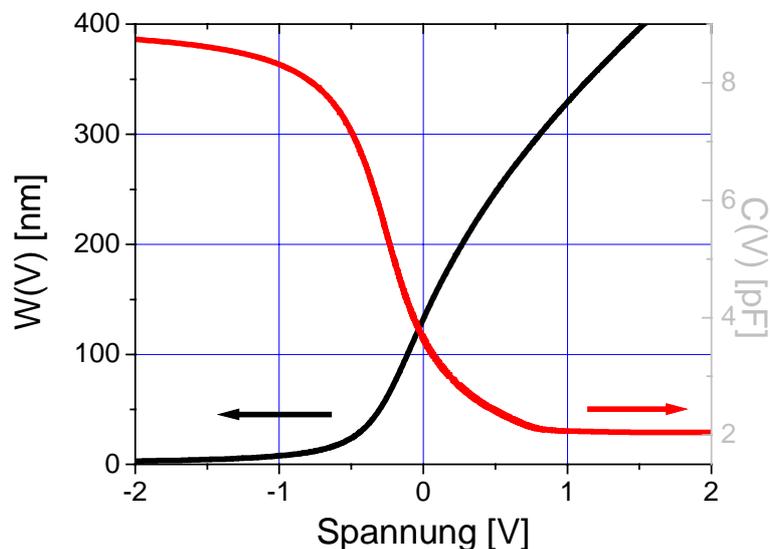


Bild 52: Kapazität $C(V)$ und Tiefe der SCR $W(V)$ im idealen 1D-Modell.

$W(V)$ ist in Abhängigkeit von der Ladungsträgerdichte in Bild 53 dargestellt. Im Flachband-Fall ist die Ausdehnung der SCR identisch mit der Debye-Länge (d_{fb} in Bild 53). Stimmt V_{bias} mit der Flachbandspannung überein, so beträgt die Tiefe der SCR bei einer homogenen Dotierung von 10^{17}cm^{-3} etwas über 10nm. Bei $V_{\text{bias}}=0\text{V}$ bzw. +1V erstreckt sie sich immerhin schon 40nm bzw. 100nm in die Tiefe (Bild 53).

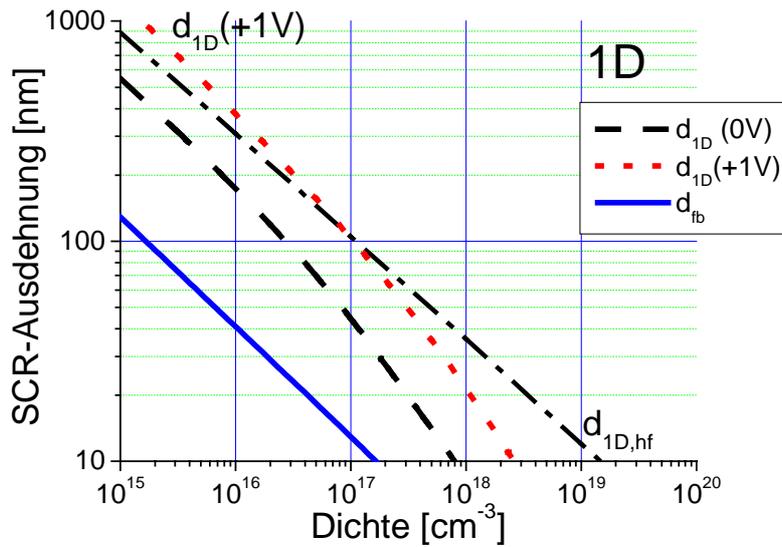


Bild 53: Dotierungsabhängige Tiefe der SCR bei $V_{bias}=0, +1V$ und im Flachband-Fall.

Im realen Fall ist die Grenze der SCR nicht scharf, sondern – auch bei sehr kleinen Messspannungen – über die Debye-Länge verschmiert. Größere Messspannungen V_{prob} und Modulationsspannungen V_{mod} führen zu einer weiteren Mittelung um $W(V)$. Beim SCM betragen beide zusammen typischerweise mehr als $\pm 1V$. Dadurch ergibt sich eine Ausdehnung der SCR bei $V_{bias}=0V$ – am Beispiel von $10^{17}cm^{-3}$ – von über 100nm ($d_{1D}(+1V)$ in Bild 53), über die gemittelt wird.

Der Device-Simulator arbeitet – wie schon mehrfach erwähnt – unter Berücksichtigung der Kleinsignal-Bedingung. Die im Folgenden vorgestellten Ergebnisse beruhen auf solchen Simulationen. Das SCM-Signal wurde bestimmt über $\Delta C=C(-1V)-C(+1V)$. Es wird also mit einer großen Modulationsamplitude und mit einer sehr kleinen Messamplitude gerechnet. Ähnliche Resultate lagen auch dann vor, wenn mit einer großen Messamplitude und einer kleinen Modulationsamplitude kalkuliert wurde.

Bei einer scharfen Sonde (3D) reicht die SCR nicht so tief wie bei einer großen Gate-Elektrode (1D). Dafür müssen aber Randeffekte berücksichtigt werden, d.h. die laterale Ausdehnung der SCR ist größer als die Sondenbreite. Bild 54 veranschaulicht das Ergebnis von Simulationen der dotierungsabhängigen Ausdehnung der SCR einer Sonde mit einem Auflageradius von $r_a=25nm$.

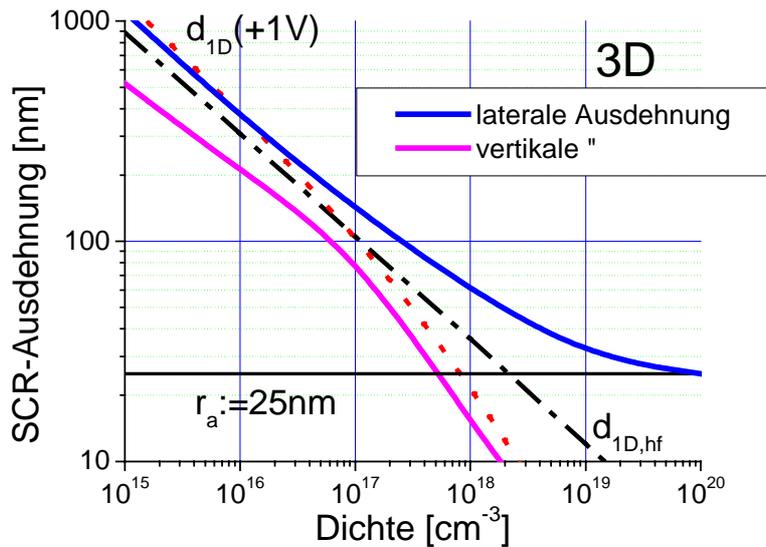


Bild 54: Dotierungsabhängige Ausdehnung der SCR bei einer Sonde mit $r_a=25\text{nm}$.

Die laterale Ausdehnung der SCR bei $V_{\text{bias}}=1\text{V}$ liegt bei 10^{16}cm^{-3} bei über 300nm , während die Tiefe der SCR bei 200nm liegt. Für hohe Dotierungen nähert sich die laterale Ausdehnung der SCR im 3D-Fall der Spitzenbreite und die Tiefe der 1D-Simulation (bei $+1\text{V}$) an. Für niedrige Dotierungen liegt die laterale Ausdehnung allerdings weit über der Spitzenbreite. D.h. in diesem Fall wird die laterale Auflösung von der Dotierung und nicht von der Sondengeometrie bestimmt. Um diese These weiter zu untersuchen, wurden Simulationen an drei Teststrukturen durchgeführt.

Die erste Teststruktur besteht aus einer p^- Dotierung *auf* einer anderen Dotierung (einmal p^+ und ein andermal n^- , Bild 55 oben), wobei die Dicke der aufliegenden Dotierung variiert und die Auswirkung auf eine SCM-Messung untersucht wurde (Bild 56).

Die zweite Struktur besteht aus einer zylinderförmigen p^- -Dotierung *in* einer anderen Dotierung (einmal p^+ und ein andermal n^- , Bild 55 unten), wobei der Durchmesser der innenliegenden Dotierung variiert und wiederum die Auswirkung auf SCM-Messungen untersucht wurde (Bild 57). Der Spitzenradius (r_a) der zylinderförmigen Sonde lag bei der ersten Struktur bei 47nm und bei der zweiten bei 25nm .

Bei der dritten Struktur wurde von einer homogenen p^- -Dotierung (10^{16}cm^{-3}) und einer zylinderförmigen Sonde ausgegangen. Weiterhin wurde in einem bestimmten Abstand um die Sonde eine weitere Elektrode auf der Oberfläche definiert, deren Potential geändert werden konnte (siehe [Kra00]).

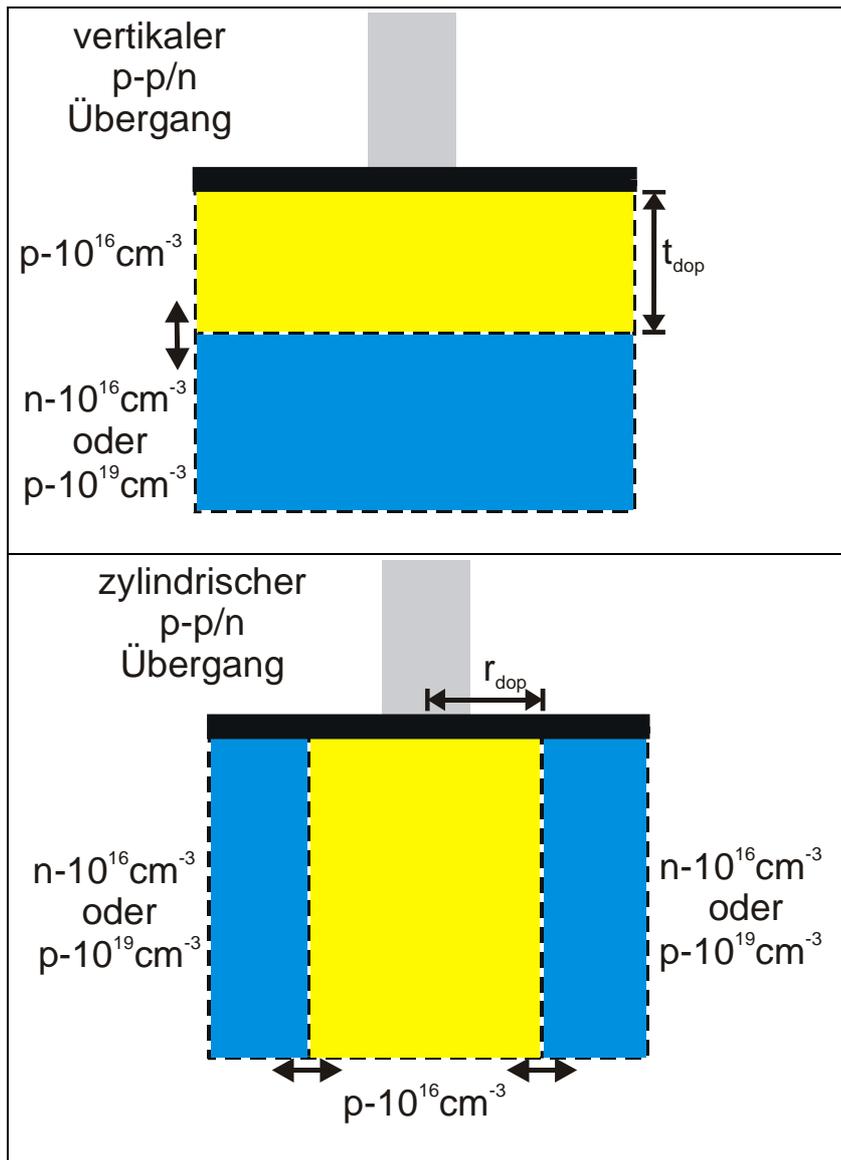


Bild 55: Teststrukturen für die Untersuchung der möglichen Auflösung der SCM .

Bei der ersten Teststruktur zeigten sich die Auswirkungen der unterliegenden p^+ -Schicht auf die darüber liegende p^- -Dotierung erst ab einer p^- -Schichtdicke von 75nm (vgl. Bild 56, Abfall des SCM-Signals um 10%). Obwohl die Ausdehnung der SCR bei 10^{16} cm^{-3} mindestens 200nm betragen sollte, sind die Auswirkungen auf das SCM-Signal vergleichsweise gering, da die oberflächennahe Dotierung stärker die Kapazitätsmessung beeinflusst.

Ganz anders sieht es im Fall $p\text{-Si auf } n\text{-Si}$ aus: Hier wird das Signal deutlich ab einer Tiefe von 200nm beeinflusst. Es klappt von $p\text{-Si-}$ auf $n\text{-Si-}$ Verhalten um. Dabei muss aber auch beachtet werden, dass ein symmetrischer $p\text{-n-}$ Übergang mit einer Dotierung von 10^{16} cm^{-3} auch eine intrinsische SCR der Breite von 405nm besitzt. D.h.: Liegt der abrupte $p\text{-n-}$ Übergang in 200nm Tiefe, so beträgt die Ladungsträgerdichte an dieser Stelle 10^{10} cm^{-3} , an der Oberfläche gerade eben

10^{16}cm^{-3} . Dies spiegelt sich natürlich in der C-V-Kurve und somit in dem SCM-Signal wieder (Bild 56).

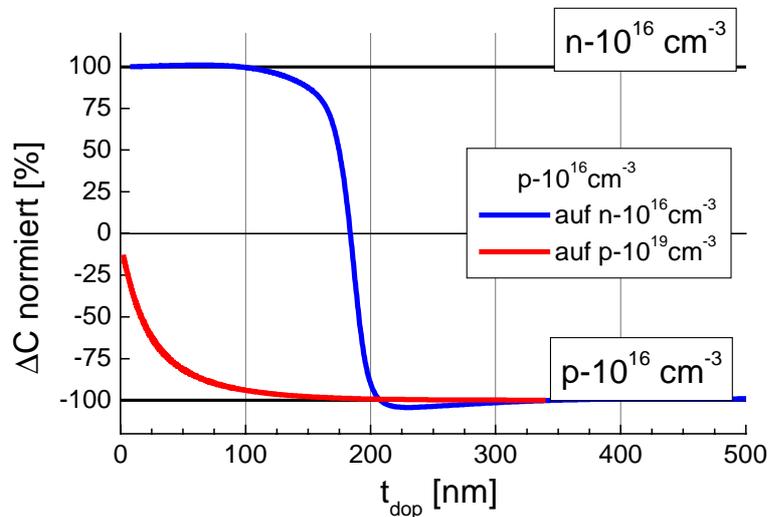


Bild 56: Vertikale Auflösung des SCM bei einer Dotierung von 10^{16}cm^{-3} .

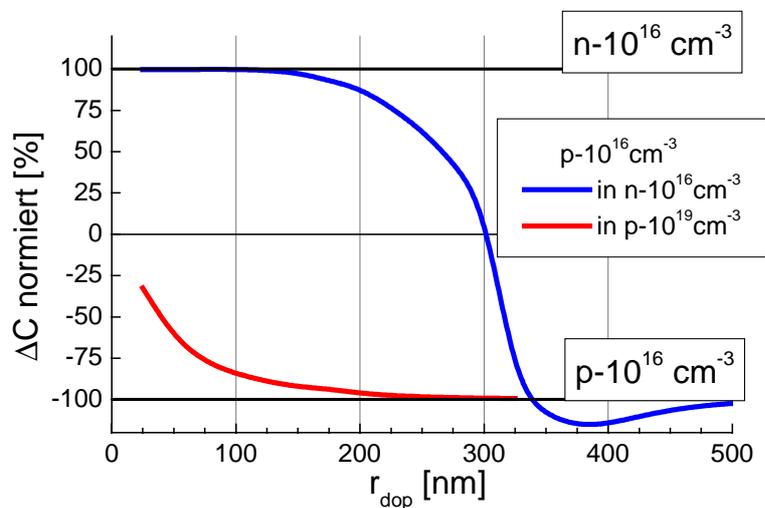


Bild 57: Laterale Auflösung des SCM bei einer Dotierung von 10^{16}cm^{-3} .

Eine Untersuchung der zweiten Teststruktur (p-Si 10^{16}cm^{-3} in einer anderen Dotierung, einmal p^+ , ein andermal n^-), führte zu den Kurven, die in Bild 57 zu sehen sind. Sie sind vergleichbar mit denen aus Bild 56. Nur setzt hier die Beeinflussung schon bei einem größeren Abstand ein.

Bei der dritten Teststruktur hing der Einfluss der äußeren Elektrode auf das SCM-Signal nicht nur von der Entfernung von der Sonde, sondern auch von der an der Elektrode anliegenden Spannung ab. In einer Entfernung von 300nm konnte eine Änderung des Signals erst ab einer Elektrodenspannung von über 1V beobachtet werden. Erst ab dieser Spannung kann sich eine Inversionsschicht bilden, die mit der spitzeninduzierten SCR wechselwirken kann. Dann wird wieder die Oxidkapazität bzw. eine LF-C-V-Kurve gemessen [Kra00].

Daraus ergibt sich folgender Zusammenhang zwischen der Ausdehnung der SCR und der lateralen bzw. vertikalen Auflösung des SCM: Die Kurven aus Abbildung 54 können als Anhaltspunkte für die mit dem SCM erreichbare laterale und vertikale Auflösung gelten. Allerdings unterscheiden sich die erreichbaren Auflösungen bei Messungen von Proben mit einer gleichnamigen (z.B. p^- und p^+) und von Proben mit einer ungleichnamigen (z.B. p^- und n^-) Dotierung (Bild 58).

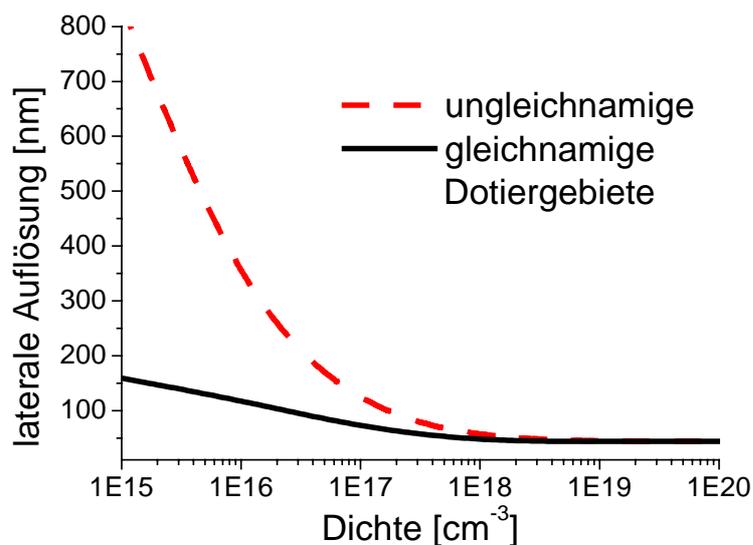


Bild 58: Laterale SCM-Auflösung bei gleichnamiger und ungleichnamiger Probendotierung.

Eine gleichnamige Dotierung, die in einem weiteren Abstand von der Sonde liegt, wirkt sich viel schwächer aus, da die Auflösung umgekehrt proportional zum Quadrat der Ausdehnung der SCR ist (siehe auch oben Formel 17). Eine höhere Dotierung in einer bestimmten Tiefe bewirkt nur ein schnelleres Abfangen der SCR, aber keine große Kapazitätsänderung. Den größten Einfluss hat die Dotierung in der Nähe der Sonde. In diesem Fall besitzt das SCM also eine Auflösung, die weit unterhalb der Kurven in Abbildung 54 liegt.

Liegt hingegen ein p-n-Übergang oder eine permanente Inversionsschicht in Reichweite der SCR, so kommt es zu einer großen Signaländerung, wenn die SCR den Übergang erreicht hat. In diesem Fall wird nämlich die SCR stark beeinflusst. Sie kann sogar ganz verschwinden, da die Spannung, die bei p-Si eine Verarmung bewirkt, bei n-Si eine Akkumulation hervorruft.

Die Kurven aus Bild 58 können somit als Grenzen der unbeeinflussten lateralen Ladungsträgerdichten-Auflösung im normalen SCM-Modus ($V_{\text{prob}}=2V_{\text{pp}}$, $V_{\text{bias}}=0\text{V}$) interpretiert werden.

Die laterale Auflösung kann durch das Verschieben der Biasspannung in den Bereich der Akkumulation (kleinere SCR) verbessert werden. Bild 59 zeigt zwei dC/dV-V-Kurven. Die Erste (Nr. 1) stellt eine dC/dV-V-Kurve mit $V_{\text{prob}}=\pm 10\text{mV}$ und $V_{\text{mod}}=\pm 10\text{mV}$ dar, während die Zweite (Nr. 2) die dC/dV-V-Kurve mit $V_{\text{prob}}=\pm 1\text{V}$ und $V_{\text{mod}}=\pm 1\text{V}$ widerspiegelt. Mit Kurve 1 würde folglich die SCR in einer Tiefe von 120nm (=vertikale Auflösung) abgetastet werden. Dabei wäre das SCM-Signal (dC/dV) um fast die Hälfte – gegenüber seinem Maximal-Wert – abgefallen. Mit Kurve 2 wird bei $0V_{\text{bias}}$ eine laterale Ausdehnung der SCR von 450nm abgetastet. Bei einer Messung mit $V_{\text{bias}}=-2\text{V}$ wird hingegen nur noch über eine Tiefe von 120nm gemittelt. Das SCM-Signal ist dann allerdings um mehr als die Hälfte – gegenüber seinem Maximal-Wert – reduziert.

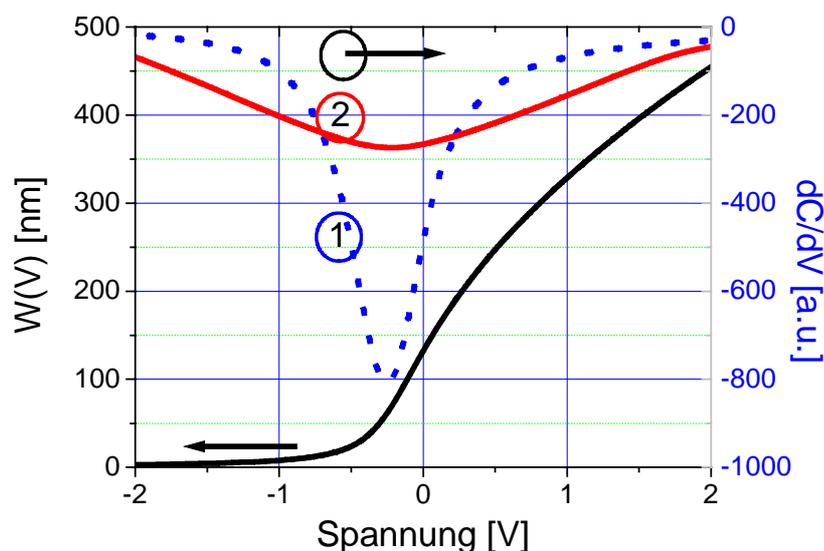


Bild 59: dC/dV-V-Kurven und eine Kurve der Tiefe der SCR (im 1D-Fall) bei einer homogenen Dotierung von 10^{16}cm^{-3} .

Nicht ganz unproblematisch ist dieses Vorgehen im Fall von ungleichnamigen Dotierungen. Hier muss mit zwei Spannungen (eine Rasterlinie mit -2V und dieselbe Linie anschließend mit $+2\text{V}$) gearbeitet werden. Noch besser ist die Aufnahme von kompletten C-V- bzw. dC/dV -V-Kurven. Denn nur mit der Rasterkapazitätsspektroskopie (Scanning Capacitance Spectroscopy, SCS) ist es möglich, die gesamte kapazitive Information aus einer Probe zu gewinnen und zugleich auch eine minimale laterale Auflösung zu erreichen (darauf wird in 2.2.2 näher eingegangen).

Abschließend sei noch einmal deutlich gemacht, dass *die in diesem Kapitel herausgearbeitete vertikale bzw. laterale SCM-Auflösung (Bild 54 und 58) sich auf das durch umliegende Dotiergebiete unbeeinflusste SCM-Signal bezieht*. D.h., wenn wir Proben mit Dotierstrukturen abbilden, deren Größe oberhalb der Werte von Bild 54 liegen, so können die so erhaltenen Messsignale – mit einer Eichstruktur – direkt in ein quantitatives 2D-Dotierprofil konvertiert werden.

An sich können mit dem SCM Dotierstrukturen mit einer Ausdehnung von unter 20nm aufgelöst werden. Für die Konversion dieser SCM-Daten in ein Dotierprofil, muss dann aber ein erheblicher (Simulations- und Rechner-) Aufwand betrieben werden (siehe 2.2.3).

2.2.1.6 SCM-Eichung

Bei der Eichung eines Systems sind zuerst die Parameter herauszuarbeiten, die die entscheidende Rolle spielen (siehe oben Bild 23).

Z.B. spielen bei der Probe (siehe Abschnitt 2.2.1.3) die Rauigkeit, die Oxiddicke, die Oxidgüte, die relative Luftfeuchtigkeit und die Dielektrizitätskonstante des Oxides eine bedeutende Rolle. Durch die entwickelte Probenpräparation konnte der Einfluss dieser Parameter reduziert (Rauigkeit, Oxidgüte), fixiert (Oxiddicke d_{ox} , Dielektrizitätskonstante ϵ_{ox}) bzw. geklärt (relative Luftfeuchtigkeit) werden. Bei der Sonde (siehe Abschnitt 2.2.1.2) konnte gezeigt werden, dass der Auflageradius r_a die dominierende Rolle spielt. Und bei der Instrumentierung (Abschnitt 2.2.1.1) wurde gezeigt, dass die Messspannung V_{prob} , die Biasspannung V_{bias} , und die Sensitivität des Kapazitätssensors α_{sens} die dominierenden Parameter darstellen.

Für eine unbekannte SCM-Konfiguration (ungeeichter Sensor, ungeeichte Sonde und unbekannte Probe) sind folgende Parameter zu bestimmen: r_a , V_{prob} , V_{bias} und α_{sens} . Erst wenn diese Parameter bekannt sind, können die gemessenen Werte in ein Ladungsträgerdichte- bzw. Dotierprofil konvertiert werden. Bei dieser Konversion sind zwei Verfahren zu unterscheiden: einerseits die *direkte Konversion* und andererseits die *korrigierte Konversion*.

Bei der direkten Konversion wird das System mit einer bekannten Dotierstruktur geeicht. Die verwendete Eichstruktur besteht aus sieben – je $5\mu\text{m}$ breiten – homogen dotierten Bereichen und reicht von 10^{15} bis 10^{20}cm^{-3} [Ise99]. Ist die grobe Geometrie von Eichstruktur und unbekannter Probe identisch (z.B. wenn beide Proben Querschnitts-Präparationen sind), so kann von gleicher Sensitivität des Kapazitätssensors ausgegangen werden. Mit der berechtigten Annahme, dass sich r_a , V_{prob} und V_{bias} ebenfalls nicht ändern, kann die Konversionskurve der Eichprobe ($dC/dV-N$) angefitet und für die Konversion der Messdaten hergenommen werden. Damit folgt eine Beschränkung auf Proben mit flachen Dotiergradienten. D.h. die Größe der Dotiergebiete der Probe muss oberhalb der im vorherigen Abschnitt definierten lateralen Auflösung liegen. Bei Proben mit großen Gradienten, wie z.B. p-n-Übergängen, kann die direkte Konversion nicht eingesetzt werden (siehe [Kra00]). Hier muss auf die korrigierte Konversion zurückgegriffen werden. Diese benötigt allerdings zusätzliche Informationen, also die konkreten Werte von r_a , V_{prob} , V_{bias} und

α_{sens} . Doch zuerst sollen die Anforderungen für eine Eichprobe im Falle der direkten Konversion herausgearbeitet werden.

Das Resultat einer SCM-Messung der Eichprobe besteht in einer Zuordnung von 7 SCM-Messsignalen zu 7 Dotierungen der Eichprobe. Anschließend müssen diese 7 Wertepaare angefitet werden. Hierbei stellt sich die Frage, ob es möglich ist, eine Fit-Funktion zu entwickeln, die weitere Aussagen über die physikalisch relevanten Parameter erlaubt. Dabei würde sich als Ausgangspunkt das im Abschnitt 2.2.1.2 vorgestellte 3D-Modell (3D-Theorie) anbieten. Es basiert, wie schon beschrieben, auf der Erfassung der SCR durch eine Kombination von Platten-, Zylinder- und Kugelkondensator und geht bei der Tiefe der SCR von dem 1D-Modell aus. Daraus lässt sich eine Formel für das SCM-Signal entwickeln [Ise99], mit der auch die Messdaten angefitet werden könnten. Leider waren die Resultate nicht überzeugend. Die erhaltenen Fit-Parameter hatten nichts mit den eigentlich verwendeten Parametern zu tun. Zudem reagierte die Fit-Funktion extrem sensitiv auf die Startwerte. Also gab es viele lokale Minima im „Potentialgebirge“ der fünf Parameter der verwendeten Fit-Funktion, und/oder das 3D-Modell war nicht ausreichend. Mit weiterentwickelten Formeln konnten experimentelle und Simulationsdaten noch besser angefitet werden (Bild 60), allerdings verlor man jeglichen Realitätsbezug für die Parameter.

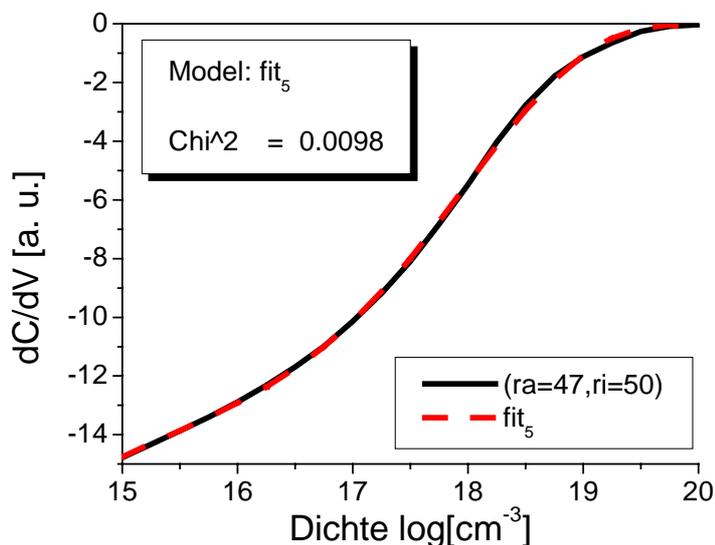


Bild 60: Fit einer dC/dV-N-Kurve mit einer Fit-Funktion mit 5 Parametern.

Insgesamt gelang es nicht, eine Fit-Funktion zu entwickeln, die die Werte ausreichend genau anfitete und zusätzlich noch physikalisch aussagekräftige Parameter lieferte. Aus diesem Grunde wird auf die gewöhnliche Spline-Fit-Methode

zurückgegriffen. Hierbei geht es dann nur noch um eine korrekte Annäherung der Fit-Funktion an die experimentellen Werte ohne zusätzliche Informationen. Und damit stellt sich die Frage, wie die Höhe der homogenen Dotierungen auf der Eichprobe verteilt sein sollten, damit die Annäherung an diese nichtlineare Funktion vertretbar wäre. Zudem sollte der Abfall des Sensorsignals (unterhalb 10^{15}cm^{-3}) mit erfasst werden, der bei den Simulationen nicht berücksichtigt wird. Dies ist in Bild 61 dargestellt. Neun Dotierungsstufen reichen folglich für eine gute Anpassung völlig aus.

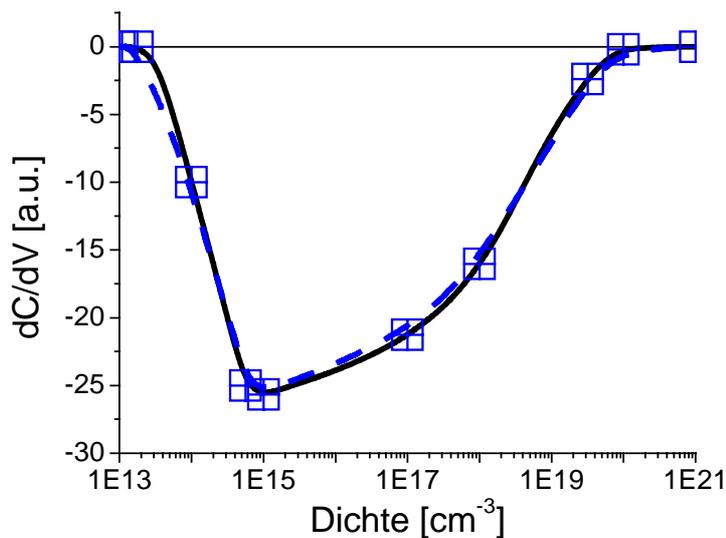


Bild 61: Güte des Spline-Fits von SCM-Daten bei Vorgabe von gut verteilten Dotierungswerten.

Die uns zur Verfügung stehende Eichprobe besitzt, wie erwähnt, sieben Stufen von 10^{15} bis 10^{20}cm^{-3} , aber leider nicht unterhalb 10^{15}cm^{-3} . Dieser Bereich müsste durch eine andere Probe abgedeckt werden (hierbei reicht allerdings auch eine einmalige Eichung, da die Bereiche unterhalb 10^{15}cm^{-3} schlecht aufgelöst werden, für die Technologen von geringem Interesse sind und zudem es nur um einen unkritischen Abfall auf Null geht). Insgesamt eignet sich die benutzte Eichprobe mit Hilfe des Spline-Fits für die direkte Konversion, kann aber keine physikalisch relevanten Parameter liefern.

Da die verwendete Eichprobe nur p-Dotierungen besitzt, stellt sich die Frage nach der Bestimmung einer Biasspannung von (absolut) 0V, denn nur dann verhalten sich p- und n-Dotierung symmetrisch zueinander (beide gleiche Betragswerte). Prinzipiell würde für solch eine Messung (Eichung des Sondenmaterials) ein niedrig dotierter

symmetrischer p-n Übergang benötigt werden. Dann gibt es zwei Möglichkeiten den Nullpunkt der Biasspannung zu finden.

Bei der ersten Methode wird über den p-n-Übergang gerastert und die Biasspannung so justiert, dass der Betrag des SCM-Signals von der p- und n-Dotierung gleich groß ist.

Bei der zweiten Methode werden dC/dV-V-Kurven im p- und n-Bereich der Probe aufgenommen, die Spannungswerte der Extrema der Kurven bestimmt und der Nullpunkt der Biasspannung zwischen diese Werte gesetzt. Der Grund liegt in unseren Simulationen, die zeigten, dass die Extrema der dC/dV-V Kurven für niedrig dotierte Proben mit der Flachbandspannung gleichzusetzen sind. Der absolute Nullpunkt befindet sich natürlich genau dazwischen.

Bei einer definierten Oberfläche von Probe und Sonde muss der Nullpunkt nur einmal, für einen bestimmten Sondentypus bzw. für eine bestimmte Beschichtung, messtechnisch bestimmt werden. Bei den CoCr-Si-Cantilevern lag dieser Nullpunkt bei $V_{\text{bias}}=0\pm 0.2\text{V}$ – nach einer guten Probenpräparation. Bei ungenügender Probenpräparation können leicht Flachbandspannungen von $V_{\text{bias}}=\pm 5\text{V}$ gemessen werden. In diesem Fall kann die SCM-Messung von vornherein nicht zu aussagekräftigen Ergebnissen führen.

Die Messspannung V_{prob} lässt sich nicht direkt messen, da jede Messspitze die Streukapazität des Kapazitätssensors und damit seine Sensitivität als auch seine Messspannung ändern würde. Dies lässt nur eine indirekte Bestimmung zu. Durch einen Vergleich der gemessenen mit den simulierten C-V-Kurven kann auf die Messamplitude zurückgeschlossen werden, wenn die Dotierung bekannt ist. Solche Vergleiche ergaben Werte zwischen 1 und 3V, je nach Einstellung von $V_{\text{stray cap adj}}$ und decken sich damit hervorragend mit der „Sensor-Theorie“ aus Abschnitt 2.2.1.1.

Die Sensitivität des Sensors könnte ebenfalls durch einen Vergleich von simulierter und gemessener C-V-Kurve bestimmt werden. Dazu muss aber der genaue Auflageradius der Sonde bekannt sein, da er entscheidend in die Simulation mit eingeht (wenn die anderen oben genannten Parameter bekannt sind), wie bereits im Abschnitt 2.2.1.2 gezeigt wurde.

Die Bestimmung der Gestalt der Sondenspitze kann mittels hochauflösendem Rasterelektronenmikroskop (REM) oder Transmissionselektronenmikroskop (TEM) durchgeführt werden. Die Güte einer solchen Charakterisierung verdeutlicht Bild 62. Die TEM Aufnahme ermöglicht eine Nanometer-genaue Bestimmung des Spitzenradius. Bei dieser – unbenutzten – CoCr-Si-Spitze liegt er bei 35nm.

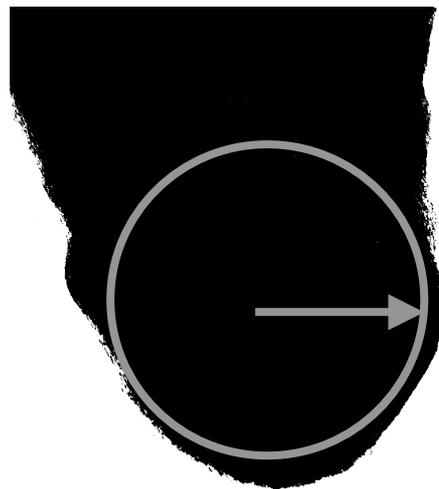


Bild 62: TEM-Bild einer mit CoCr-beschichteten Si-Spitze mit einem Spitzenradius von 35nm.

Beide Methoden sind aber nicht überall und besonders nicht jederzeit verfügbar, so dass sich eine – auf diese Praktiken stützende – Charakterisierung der Sonde als zu aufwändig erweist. Schneller sind Methoden, die sich auf die Entfaltung (Deconvolution) eines SFM-Bildes von einer bekannten Probe mit einer unbekanntem Sonde stützen. Die Methode der Entfaltung von Sonde und Probe stellt für sich schon ein eigenes Thema dar (eine fundierte Einführung und Übersicht wird in [Vil97] gegeben).

Eine durchgeführte Entfaltung mit einer Eichprobe, die aus scharfen pyramidenförmigen Siliziumspitzen besteht, funktionierte zwar tadellos, erbrachte aber nicht die gewünschte Auflösung bzgl. des vordersten Sonden-Bereiches, da der Spitzenradius der Eichprobe nicht genügend klein war (siehe Bild 63). Der Standard bewährte sich aber für eine schnelle Beurteilung der groben Gestalt der Sonde. Mit ihm konnten auch Anlagerungen von Kontaminationen und ein „Abblättern“ der Metallisierung beobachtet werden.

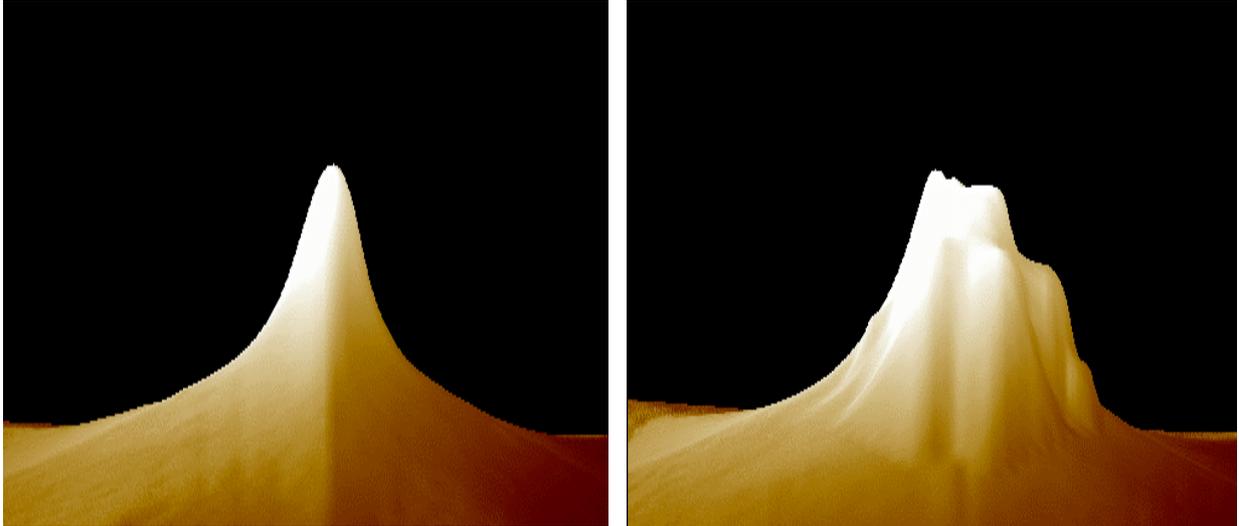


Bild 63: 1 μ m breite 3D-SFM-Abbildung einer Eichprobe. Links wurde diese mit einem unbenutzten Silizium-Cantilever mit einem Spitzenradius von 2nm und rechts mit einem abgenutzten CoCr-Si-Cantilever abgebildet.

Für eine höhere Auflösung der vordersten Sondenspitze eignet sich die Vermessung von Gold-Kolloiden mit einem Durchmesser von 20nm, die mit einem selbst-organisierenden Prozess auf eine Siliziumoxid-Oberfläche aufgetragen wurden (Bild 64, Näheres siehe [Bar00]).

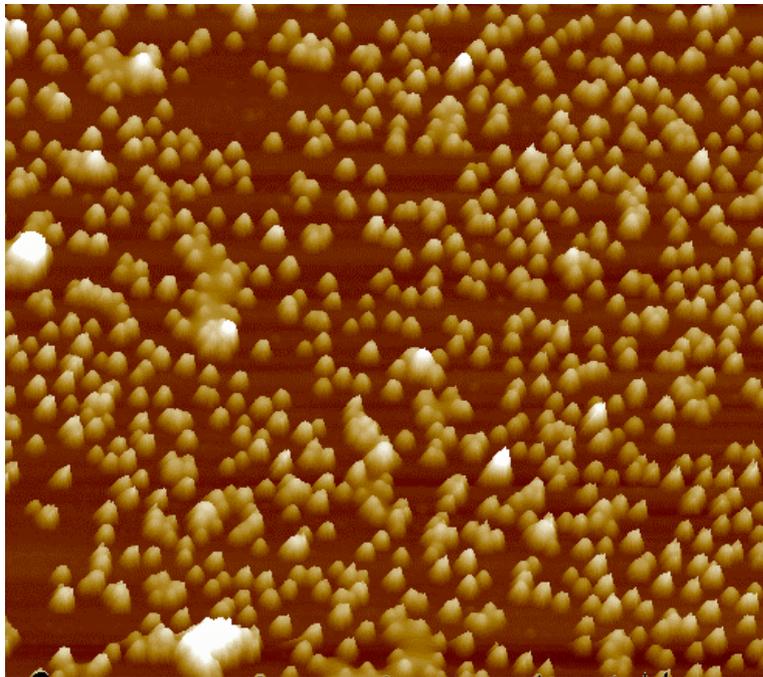


Bild 64: 3D-SFM-Abbildung (2 \times 2 μ m²) der Kolloid-Probe.

Mit der hier entwickelten Technik war es möglich, die Sonde schnell, einfach und mit ausreichender Auflösung zu charakterisieren.

2.2.1.7 Beispiele aus der Fehleranalyse

Ein Ziel dieser Arbeit war die Weiterentwicklung des SCMs zu einem Standardverfahren für den Einsatz in der industriellen Fehleranalyse. Dazu gehören neben der Entwicklung der nötigen Präparations- und Messvorschriften auch Untersuchungen der Verträglichkeit mit anderen Analysemethoden. Es geht also um die Frage: Welche Analyseverfahren dürfen vor einer SCM-Messung eingesetzt werden?

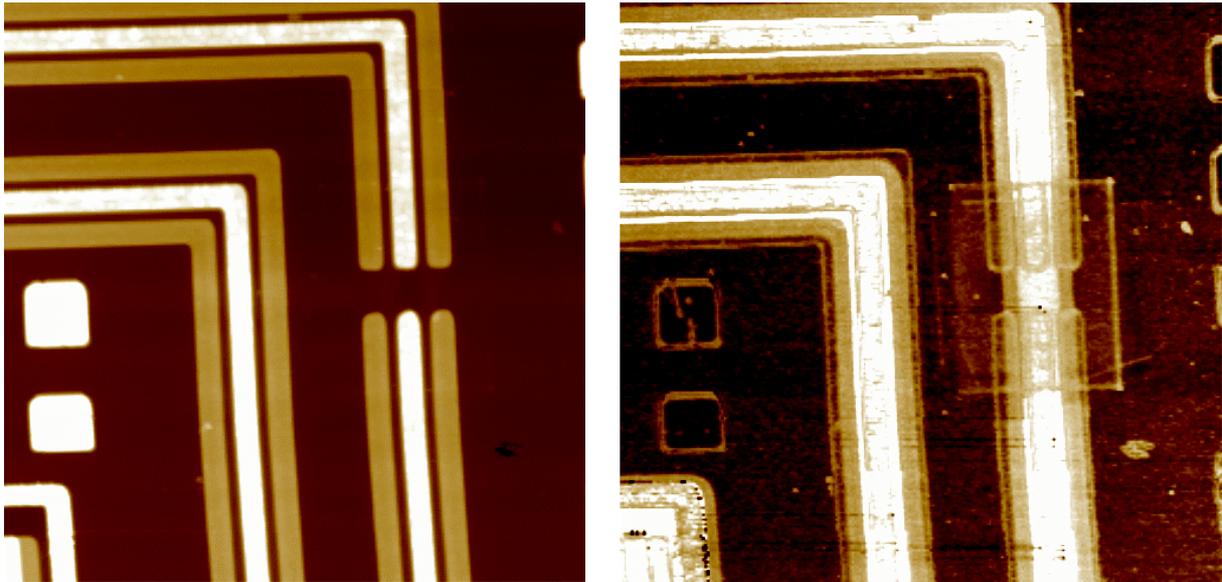


Bild 65: SCM-Bild ($20 \times 20 \mu\text{m}^2$) eines ICs nach einer REM Untersuchung. Im SCM-Signal (rechts) ist deutlich eine hellere rechteckige Fläche zu erkennen, die von der vorher durchgeführten REM Messung stammt.

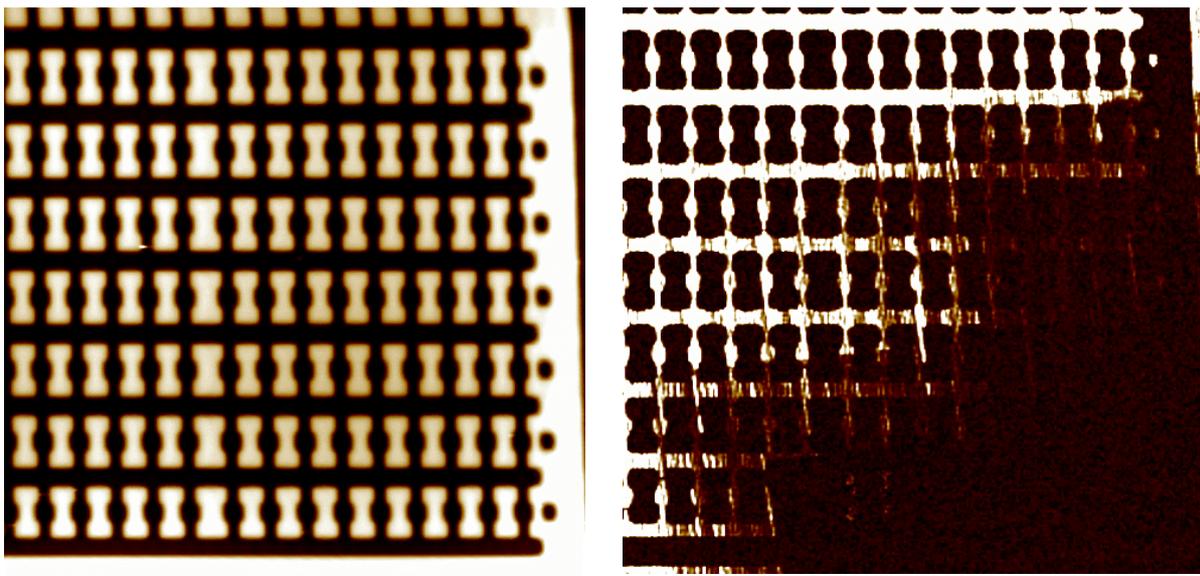


Bild 66: SCM-Bild ($50 \times 50 \mu\text{m}^2$) eines ICs nach einer FIB Untersuchung.

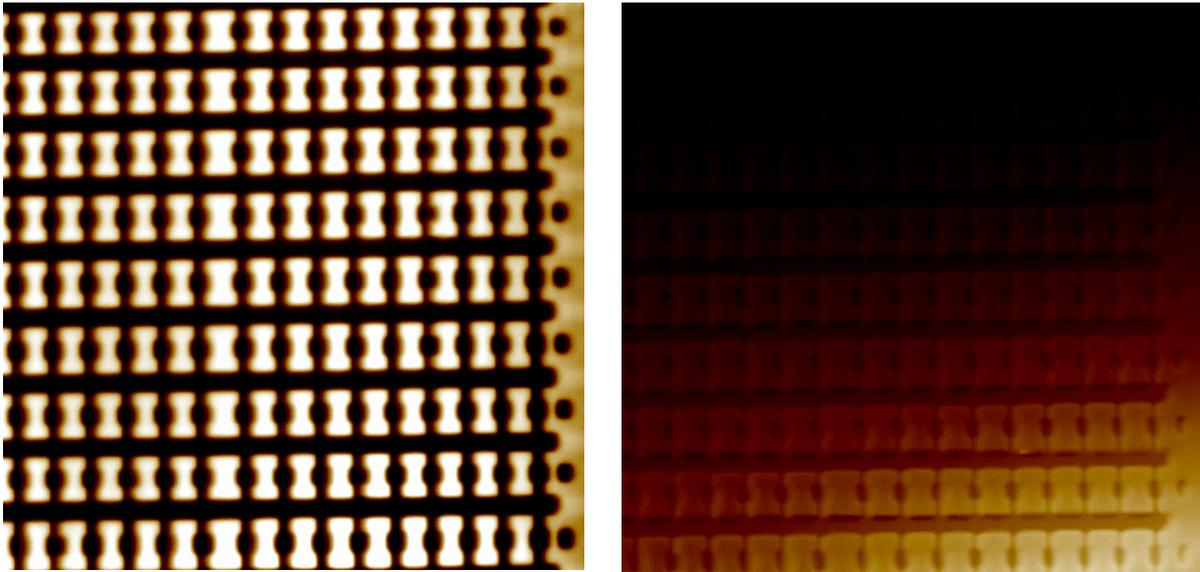


Bild 67: KFM-Bild ($50 \times 50 \mu\text{m}^2$) eines ICs nach einer FIB-Untersuchung.

Zz. gelten als die Arbeitspferde der HL-FA das REM und das FIB (Focussed Ion Beam). Mit beiden Methoden wurden IC-Strukturen untersucht und anschließend mit dem SCM weiter charakterisiert. Die Auswirkungen beider Analyseverfahren auf das SCM-Signal konnten klar nachgewiesen werden. Das REM arbeitet mit Elektronen, das FIB mit Ga-Ionen, die beide die Oberfläche stark aufladen können. Bei der FIB-Messung führt dies sogar zu solch starken Störungen, dass eine weitere SCM-Analyse unmöglich wird. Eine FIB-Untersuchung muss folglich vor einer SCM-Messung unterbleiben. Entsprechend muss auch eine FIB-Präparation von Proben für SCM-Untersuchungen unterbleiben.

Dieselbe – durch das FIB – geschädigte Stelle des ICs wurde nach der SCM-Messung noch mit dem KFM analysiert. Auch das KFM-Bild zeigt eine großflächige Aufladung der Probe (Bild 67), die zu einem Oberflächenpotential von bis zu 7V führte. Die einzelnen FIB-Spuren lassen sich aber nur mit dem SCM auflösen (Bild 66).

Im Folgenden wird eine Auswahl der Ergebnisse von SCM-IC-Charakterisierungen präsentiert, die für die HL-Fehleranalyse durchgeführt wurden. Dabei handelt es sich um Lift-off- (Bild 68-70) oder Querschnittspräparationen (Bild 71-72), wie sie unter 2.2.1.2 beschrieben wurden.

Bild 68 zeigt die SCM-Aufnahme eines elektrostatischen Durchschlags (engl. Electrostatic Discharge, ESD). ESD-Schäden stellen die Hauptursache für das Versagen eines ICs dar.

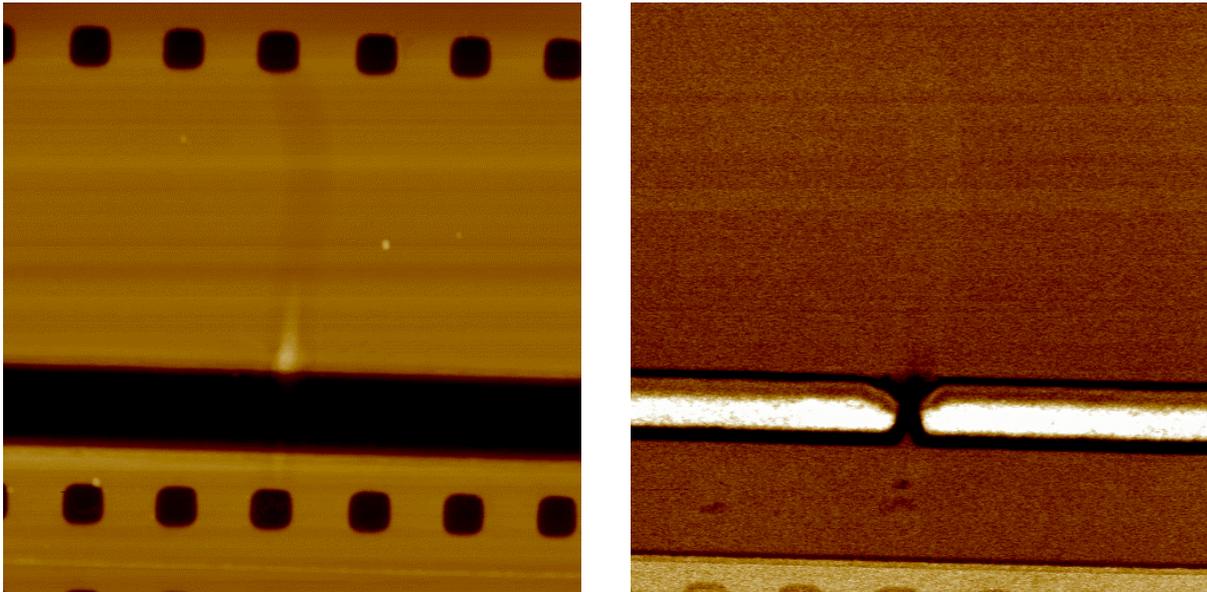


Bild 68: SCM-Bild ($10 \times 10 \mu\text{m}^2$) eines elektrostatischen Durchbruchs.

In der Topografie (Bild 68, links) können die Auswirkungen des ESD zwischen zwei Kontaktlöchern gerade noch aufgelöst werden (mit dem REM war dies nicht mehr möglich). Im SCM-Signal (Bild 68, rechts) konnte deutlich der durch den ESD entstandene Kanal über die n-Wanne (heller Steifen) hinweg aufgelöst werden. Nach der Theorie bewirkt der beim ESD fließende hohe Strom eine lokale Aufheizung der Probe, die zu einem Wandern der Dotieratome und damit zu einer Ausbildung eines leitfähigen Kanals führt. Diese Theorie konnte mit dem SCM-Bild experimentell eindeutig belegt werden.

Ein weiteres wichtiges Feld bei der IC-Charakterisierung ist die Bestimmung der Position von p-n-Übergängen. Bild 69 zeigt eine unbeschädigte und eine defekte IC-Probe. Bei dem defekten Muster konnte durch Standardverfahren der HL-FA der Fehler auf einige $10 \mu\text{m}$ lokalisiert werden. Eine genauere Bestimmung war mit konventionellen Methoden nicht möglich. Mit dem SCM gelang es innerhalb einer Stunde (ohne aufwändige Probenpräparation) die Stelle zu finden und den Fehler ausfindig zu machen. Nur durch die SCM-Analyse konnte gezeigt werden, dass es sich bei dem Fehler um einen verschobenen p-n-Übergang handelt. Diese Verschiebung kann auf einen Maskenfehler zurückgeführt werden. Dieser konnte nach der SCM-Messung behoben werden. Da die Masken-Reparatur bzw. die Herstellung einer neuen Maske sehr kostenintensiv ist, müssen die Angaben zur Korrektur sehr präzise sein (auf 100nm genau) und können nicht im „Trial and Error“ Verfahren durchgeführt werden.

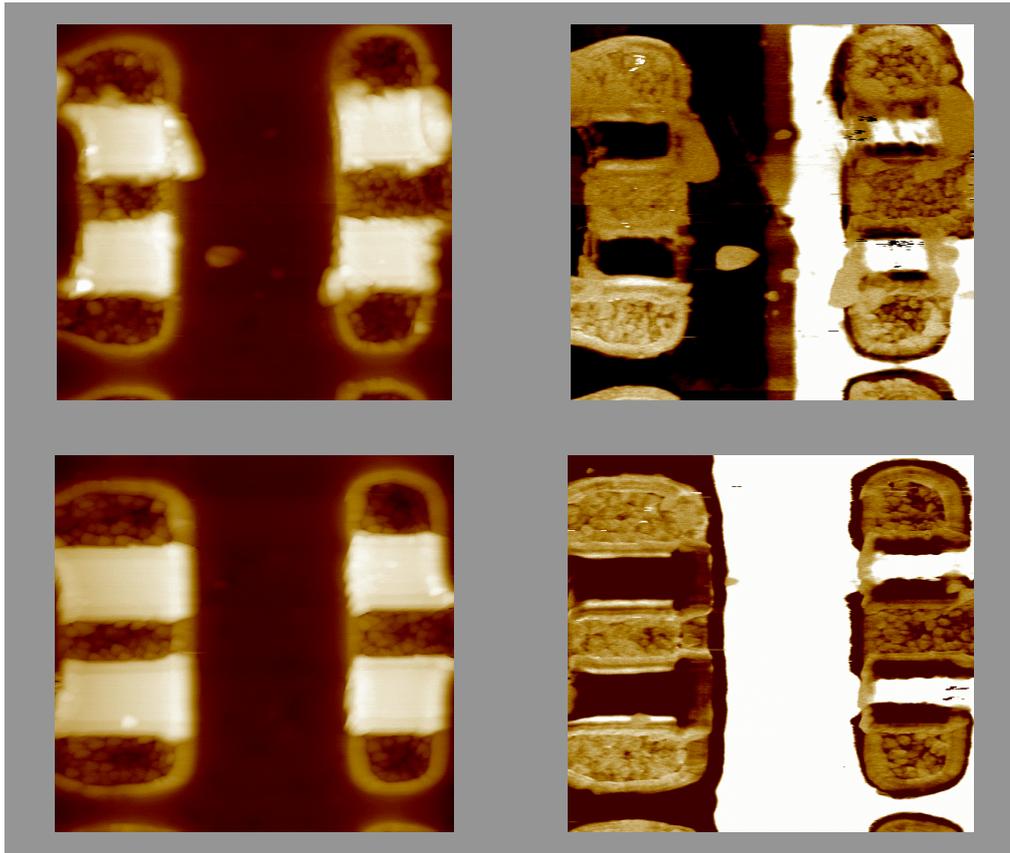


Bild 69: SCM-Bild ($5 \times 5 \mu\text{m}^2$) eines funktionstüchtigen (oben) und eines defekten ICs. Im SCM-Signal (rechts) ist deutlich eine Verschiebung des p-n-Übergangs zu erkennen.

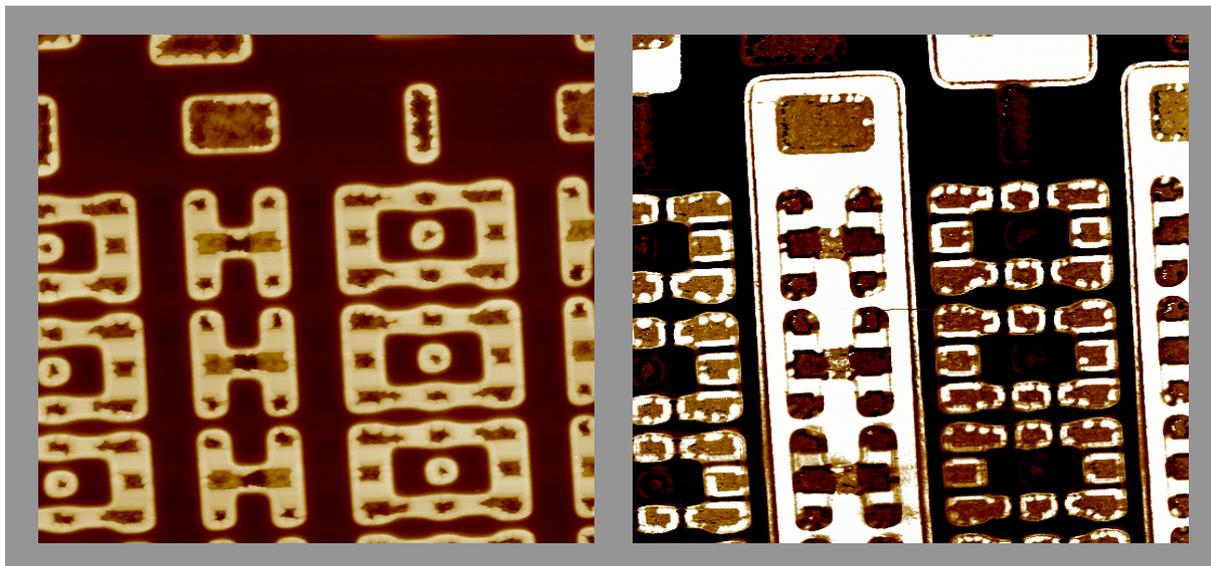


Bild 70: SCM-Bild ($20 \times 20 \mu\text{m}^2$) eines SRAMs.

Auch in der Analyse eines SRAMs (Static Random Access Memory), siehe Bild 70, ging es um die Bestimmung der Position eines p-n Übergangs bzw. der Lage der n-

Wanne (helles Rechteck im SCM-Signal). Mit dem SCM konnte die Position auf 100nm genau bestimmt werden.

Während es bei den letzten zwei Analysen um eine geforderte Auflösung von einigen 100nm ging, verlangten die folgenden zwei gezeigten Messungen eine höhere laterale Auflösung. Dies stellt natürlich höhere Anforderungen an die Probenpräparation. Die Querschnittpräparationen wurden nach dem im Abschnitt 2.2.1.3 beschriebenen Verfahren durchgeführt.

Bild 71 zeigt das Ergebnis der Vermessung eines MOSFETs. Dabei konnte die Lage der Unterdiffusion mit einer Auflösung von unter 100nm bestimmt werden.

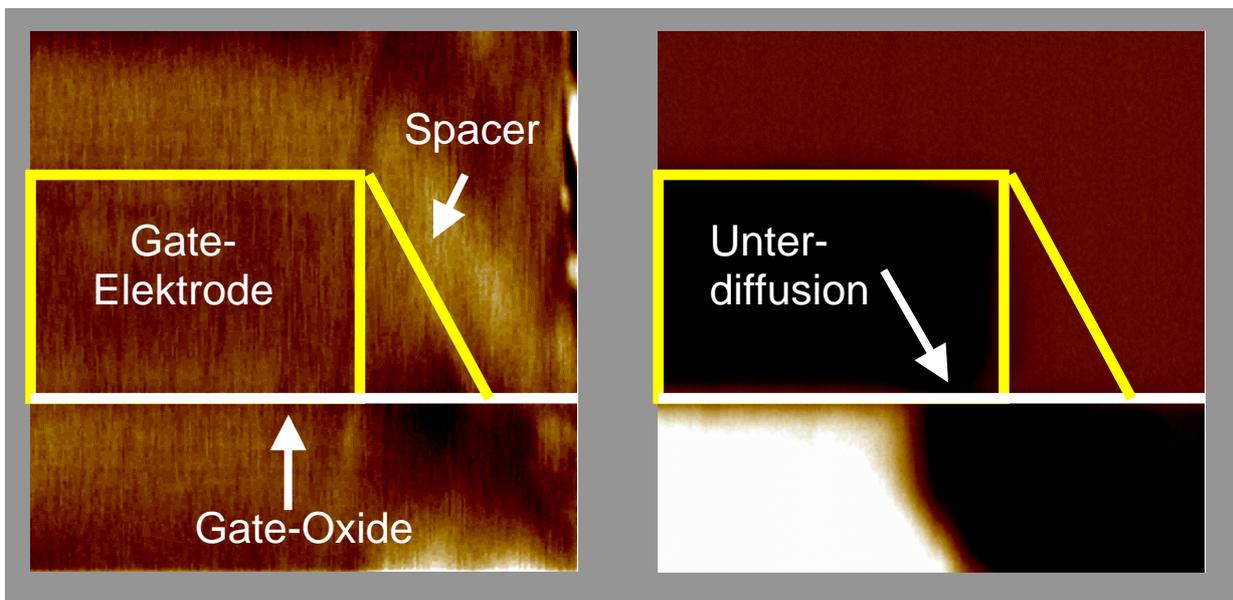


Bild 71: SCM-Bild ($1 \times 1 \mu\text{m}^2$) des Querschnitts eines MOS-FETs.

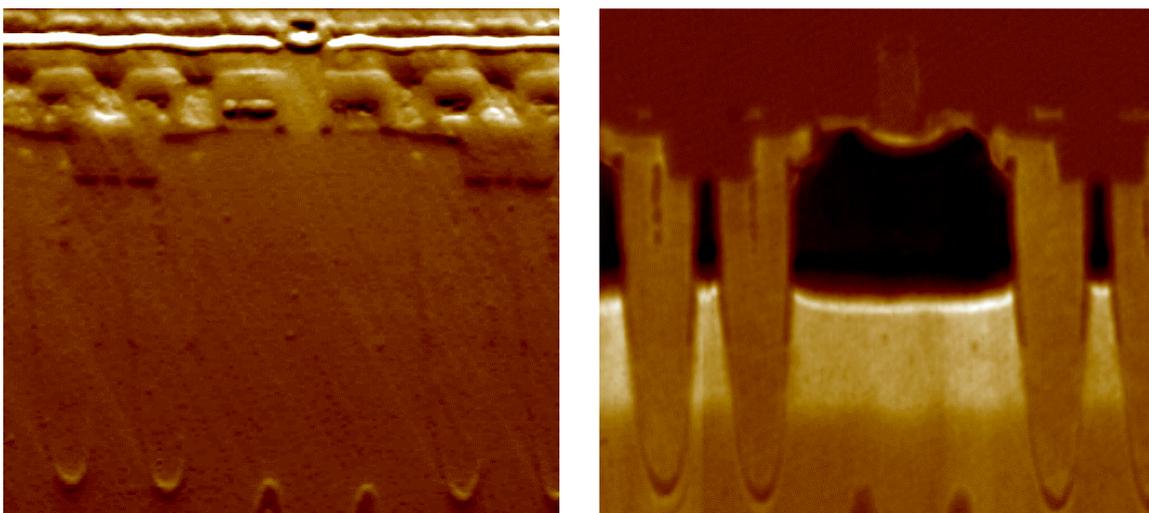


Bild 72: SCM-Bild ($3 \times 3 \mu\text{m}^2$) des Querschnitts eines 256MB-DRAMs.

Auch bei der Vermessung der Querschnittsfläche eines DRAMs (Dynamic Random Access Memory) konnte das 2D-Dotierprofil mit einer lateralen Auflösung von unter 100nm abgebildet werden. Bild 72 zeigt die erreichte hohe Güte der Probenpräparation: Im SCM-Signal sind keine Schleif- oder Polierspuren zu erkennen.

Insgesamt konnte die hervorragende Eignung des SCMs für die HL-FA gezeigt werden.

Für die Analysen von Lift-Off Proben kann mit einem geringen Arbeitsaufwand gerechnet werden (siehe die Anleitung für die Probenpräparation im Abschnitt 2.2.1.3). Die betreffende Stelle ist zudem meist leicht zu finden, so dass die entsprechende SCM-Analyse schnell durchzuführen ist.

Bei der Analyse von Querschnittspräparationen ist der Präparationsaufwand schon erheblich größer (siehe 2.2.1.3). Allerdings sollte er immer in Kauf genommen werden, da eine schlechte Probenpräparation zu falschen bzw. schwierig zu interpretierenden Analyseergebnissen führen kann.

Vor jeder Analyse sollte die Sonde charakterisiert werden (2.2.1.6) und es sollten mehrere C-V-Kurven in niedrig dotierten Bereichen der Probe aufgenommen werden. Die Kurven gestatten Aussagen bzgl. der Güte der durchgeführten Probenpräparation, der Qualität der Sonde und der Eigenschaften des Sensors. Auf die Analyse und Interpretation von C-V-Kurven wird im folgenden Abschnitt ausführlich eingegangen.

2.2.2 SCS (Scanning Capacitance Spectroscopy)

Bei der Rasterkapazitätsspektroskopie (SCS) wird die Biasspannung V_{bias} nicht wie bei der Rasterkapazitätsmikroskopie (SCM) konstant auf 0V (absolut Null) gehalten, während die Sonde über die Probe bewegt wird, sondern es werden lokale dC/dV -V-Kennlinien aufgenommen. Diese Kennlinien enthalten – im Gegensatz zu den SCM-Aufnahmen – die vollständige kapazitive Information des Nanosystems. Durch eine SCS-Analyse lassen sich folglich mehr Informationen gewinnen als aus der SCM-Analyse. Dies soll in diesem Abschnitt gezeigt werden.

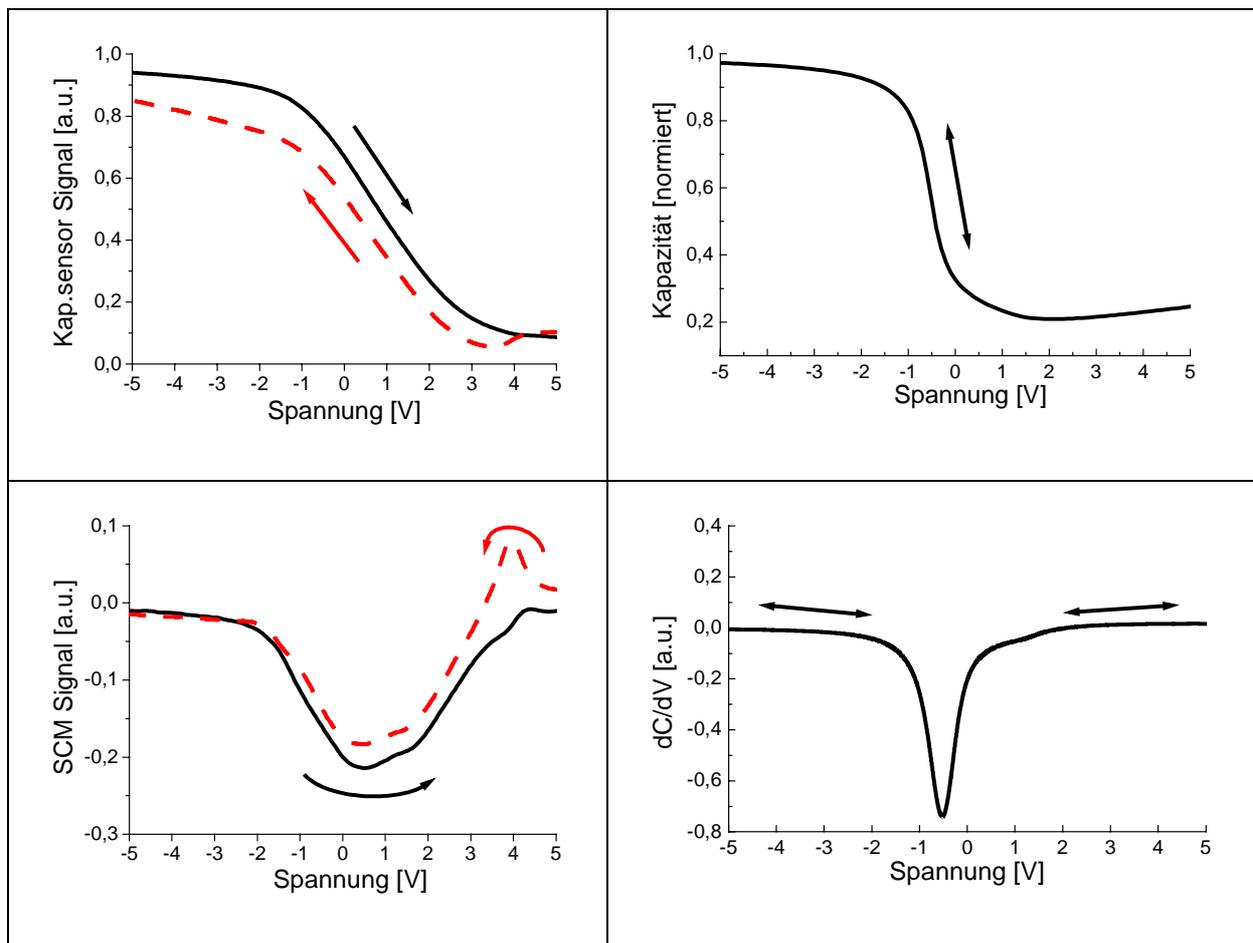


Bild 73: Experimentelle (links) und theoretische (rechts) C-V- (oben) und dC/dV -V- (unten) Kennlinie aus einem homogen p^- -dotierten Bereich.

In Bild 73 ist eine typische experimentelle und eine theoretisch zu erwartende dC/dV -V-Kennlinie (bei einer homogenen Dotierung) dargestellt. Die experimentelle Kurve ist deutlich breiter als die theoretische (aus einer 3D-Simulation mit kleinen Wechselspannungs-Amplituden). Zudem wird experimentell eine starke Hysterese zwischen den unterschiedlichen Richtungen der Spannungsrampe (einmal von negativer zu positiver Spannung, und einmal umgekehrt) beobachtet. Ferner veränderten sich die

experimentellen C-V-Kurven mit der Zeit. Zwischen der ersten und der zweiten Kurve – beide an derselben Stelle aufgenommen – konnte die größte Veränderung festgestellt werden. Nach fünf bis zehn C-V-Kurven änderte sich die Gestalt der folgenden Kurven nur noch unwesentlich. Solche Effekte können qualitative SCM-Messungen stark stören und quantitative SCM-Messungen ganz unmöglich machen. Deshalb ist eine Klärung der diesen Effekten zugrundeliegenden physikalischen Ursachen enorm wichtig für einen Fortschritt in der Rasterkapazitätsmikroskopie.

Während die Verbreiterung der C-V-Kurven gegenüber theoretischen Kurven schon im Abschnitt 2.2.1.1 auf eine große Messamplitude zurückgeführt werden konnte (durch die richtige Einstellung des Sensors müssen weitere instrumentenbedingte Effekte vermieden werden, die eine Analyse unmöglich machen; siehe 2.2.1.1), können die beobachteten Hysterese-Effekte durch Ladungen auf der Oberfläche erklärt werden [Bor98]. Ein einfaches, rein beschreibendes Modell zu dieser Annahme wurde schon in der Diplomarbeit [Bor96] entworfen. Eine Weiterentwicklung dieses Modells war eine weitere zentrale Aufgabe der vorliegenden Arbeit. Ein wichtiger Schritt in Richtung reproduzierbare Messungen war die Erarbeitung der in Abschnitt 2.2.1.3 beschriebenen Probenpräparationsvorschrift, die eine definierte saubere Oberfläche erzeugt. Dadurch konnten auch die Hysterese-Effekte deutlich reduziert werden.

Bei den Oberflächen-Ladungen kann es sich um Ionen auf der Oberfläche handeln, die durch das elektrische Feld der Sonde zu ihr hingezogen bzw. von ihr abgestoßen werden. Es können aber auch Ladungen sein, die direkt von der Sonde auf die Oberfläche aufgetragen werden. Die erste Hypothese wurde durch das zeitliche Verhalten (Wanderungsgeschwindigkeit) der Ionen auf sauberen Siliziumdioxid-Oberflächen ausgeschlossen. Bleibt die Erzeugung durch die Sonde: Auf nicht hinreichend gut präparierten Oberflächen kann die Entstehung von Ladungen auf eine Wechselwirkung von Sonde und Kontaminationen zurückgeführt werden. Da diese Oberfläche bzgl. der Kontaminationen völlig undefiniert ist, können auch keine reproduzierbaren Messungen erwartet werden.

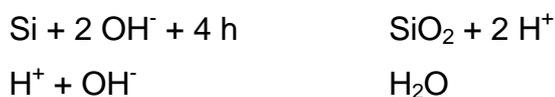
Im Folgenden wird von einer extrem sauberen SiO_2 -Oberfläche ausgegangen, die nur mit einem dünnen Wasserfilm bedeckt ist. Durch die Wechselwirkung mit Wasser ändert sich die SiO_2 -Oberfläche: Die Chemisorption von Wasser resultiert in einer hydrophilen Oberfläche, die mit Silanol-Gruppen (Si-OH) terminiert ist. An diese

lagert sich Wasser an (Physisorption), das zu einer partiellen Dissoziation führt. D.h. in wässriger Umgebung (pH=7) spalten sich hydratisierte Protonen (also Hydroniumionen, H_3O^+) ab und erzeugen dadurch eine negative Oberflächenladung. Diese unterdrückt in ihrer Umgebung die Ausbildung von Wasserstoffbrückenbindungen. Dadurch ist eine gut präparierte Siliziumdioxid-Oberfläche über mehrere Tage recht stabil gegenüber Kontaminationen aus der Umgebung. Erst nach längerer Zeit wandelt sie sich von einer hydrophilen zu einer hydrophoben Oberfläche (durch die Kontamination mit Kohlenwasserstoffen).

Die Frage, die es zu klären gilt lautet: Was für Eigenschaften haben die Oberflächenladungen und was folgt daraus für die SCS-Messung?

In der Literatur finden sich zwei Bereiche, die mit der obigen Fragestellung in Verbindung gebracht werden können: einmal das Aufbringen von Ladungen auf ein Dielektrikum (siehe z.B. [Uch96]) und einmal die spannungsinduzierte anodische Oxidation von Silizium (siehe z.B. [Avo97]).

Bei der anodischen Oxidation von Silizium wird die leitfähige Sonde eines SFMs unter Umgebungsbedingungen über eine Silizium-Oberfläche geführt. Bei einer an der Sonde anliegenden Spannung von $V_{\text{bias}} < -2\text{V}$ wird die lokale Oxidation von Silizium (oder von Metallen) beobachtet. Der Literatur nach ist diese Form der Oxidation bisher noch nicht bis ins kleinste Detail verstanden, allerdings besteht die Überzeugung, dass bei einer negativen Spannung ab -2V Hydroxidionen (OH^-) durch das vorhandene Siliziumdioxid hindurchwandern (aufgrund des elektrischen Feldes) und an der Grenzfläche zum Silizium mit Löchern zu Siliziumdioxid reagieren. Folgende zwei Reaktionsgleichungen sollen dabei gelten:



Die Entstehung der Hydroxidionen wird kaum diskutiert, allerdings konnte gezeigt werden, dass der dünne Wasserfilm auf der Probe eine wichtige Rolle spielt. Meiner Einschätzung nach entstehen die Hydroxidionen durch eine Elektrolyse des Wassers, das sich im Meniskus an der Sondenspitze befindet (darauf deutet die Spannungsabhängigkeit der Reaktion hin, denn die Zersetzungsspannung von Wasser liegt bei über -1.8V). In diesem Meniskus können die Ionen dann von der Sonde in die Probe wandern und das Silizium durch obige Reaktion oxidieren. Die Breite dieser lokalen Oxidation hängt u.a. von der relativen Luftfeuchtigkeit ab, d.h.

umso höher die Luftfeuchtigkeit ist, desto breiter ist auch der Wassermeniskus und desto breiter sind auch die erzeugten Oxidationsstrukturen. Im Vakuum, mit einer ausgeheizten Probe, wurde keine Oxidation beobachtet. Die Ausbreitung der Ionen wird durch den Meniskus und das elektrische Feld (und die Wechselwirkung zwischen beiden) bestimmt.

Bei der Untersuchung von Ladungen auf Dielektrika erwies sich die elektrostatische Rasterkraftmikroskopie (siehe Abschnitt 1.2, EFM) als die geeignete Methode. Zuerst wurde eine Spannung V_{contact} zwischen Sonde (meistens ein metallisierter Cantilever) und Probe (meistens Silizium mit einer 3-5nm dicken Oxidschicht) angelegt und beide für eine bestimmte Zeit in Kontakt gebracht. Anschließend wurde diese Stelle dann im dynamischen Modus abgebildet. Damit war es möglich, das zeitliche Verhalten dieser Ladungsscheibe (Ladungsdot) zu untersuchen. Es zeigte sich, dass ein Ladungsdot erst dann gemessen werden kann, wenn die Kontaktspannung V_{contact} , die zu seiner Entstehung benötigt wird, kleiner als -2V ist. Ab einer Spannung von $V_{\text{contact}} < -4V$ war die aufgebrachte Ladungsdichte konstant, nur die Breite des Ladungsdots nahm zu. Diese maximale Dichte liegt zwischen 10^{12} und 10^{13}cm^{-2} . Ferner konnten Dots mit gleicher und unterschiedlicher Spannungspolarität nebeneinander in einem Abstand von einigen μm deponiert werden. Die Wechselwirkung dieser Dots wurde mit dem EFM untersucht. Dabei zeigte sich, dass sich die Extrema der Dots bei gleicher Ladungspolarität aufeinander zu-, sich aber bei unterschiedlicher Polarität voneinander wegbewegen, was dem Verständnis der elektrostatischen Wechselwirkung widerspricht. In der Literatur [Uch96] wird dieses Verhalten auf eine Kombination von Feld-unabhängigen Diffusions- und Feld-abhängigen Driftprozessen zurückgeführt.

Im Rahmen dieser Arbeit konnten diese Feststellungen aus der Literatur mit einer simplen 1D-Simulation, die auf einer zufälligen Diffusion von ladungstragenden Teilchen beruht, per PC nachgeahmt werden. In dieses Random Walk Model (RWM) floss ein, dass Ladungen unterschiedlicher Polarität sich neutralisieren, und dass es eine Grenze bzgl. der Ladungsträgerdichte gibt. Ist diese Dichte erreicht, ist der Dot quasi eingefroren. D.h. es findet keine Bewegung im Dot, sondern nur an seinen Grenzen statt. Damit gelang es sogar das von Morita beobachtete 3-Phasen-Verhalten der Dots [Fuk96] per PC zu reproduzieren. Diese gute Übereinstimmung berechtigt zu der Annahme, dass mit dem RWM eine Möglichkeit gefunden wurde, die Ausbreitung von Ladungen auf einer Oberfläche zu simulieren.

Insgesamt ergibt sich folgendes Oberflächenladungs-Diffusionsmodell (ODiM): Ladungen lassen sich nur bis zu einem bestimmten Betrag (z.B. $2.2 \cdot 10^{12} \text{cm}^{-2}$) auf die Oberfläche auftragen. Große Mengen an Ladungen werden erst ab einer Spannung von mehr als $\pm 2\text{V}$ erzeugt. Sie verbreiten sich direkt um die Sonde sehr schnell – in einem durch den Wassermeniskus (und damit durch die relative Luftfeuchtigkeit) bestimmten Prozess. An den Rändern dieses Ladungsdots können sich die Ladungen dann in einem langsameren Diffusionsprozess verteilen, auf den elektrostatische Wechselwirkungen keinen Einfluss haben. Treffen negative und positive Ladungen aufeinander, so neutralisieren sie sich. Ein Beispiel für eine Simulation von ODiM mit dem oben erwähnten RWM ist in Bild 74 gegeben.

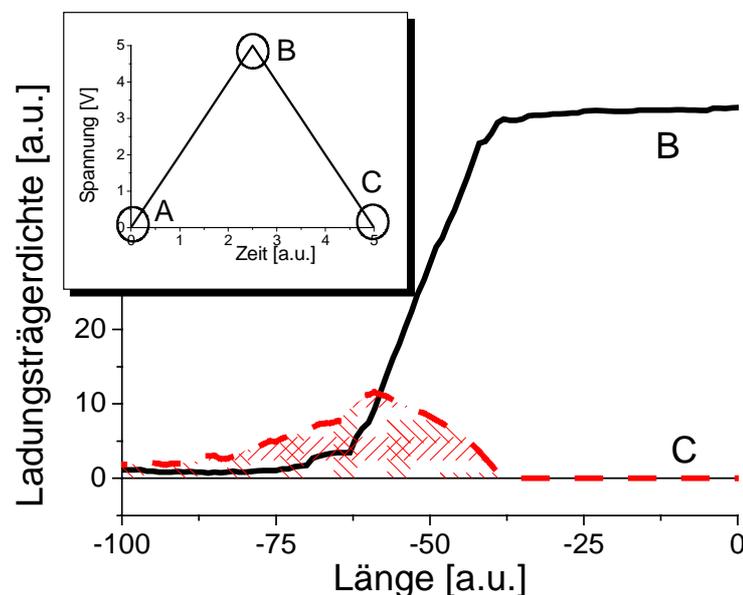


Bild 74: Simulation (ODiM & RWM) des Verhaltens von Ladungen, die von einer Sonde (bei $x=0$) aufgetragen werden. Simuliert wurde ferner eine Spannungsrampe, die von 0V (A) auf +5V (B) ansteigt und anschließend wieder zu 0V (C) zurückkehrt. In der Grafik sind augenblickliche Ladungsverteilungen am Scheitelpunkt (B) und am Ende (C) der Spannungsrampe gezeigt. Beim Startwert 0V (A) befanden sich keine Ladungen auf der Oberfläche. Mit zunehmender Spannung (bis B) wurde ein großer Bereich (bis $x=-35$) sofort mit Ladungen überschwemmt. Bei der Rückkehr zu 0V (C) ist die Oberfläche nicht mehr ungeladen. Um die Sonde herum hat sich ein Ring aus positiven Ladungen gebildet. Diese Oberflächenladungen bewirken die Hysterese in den C-V-Kurven.

ODiM soll im folgenden mit Simulationen (mit dem 3D-Device-Simulator TCAD von ISE) kombiniert werden, die die Zeitabhängigkeit der Ausbreitung der Ladungsträger auf der Oberfläche nicht berücksichtigen. Dazu wurden zuerst Simulationen durchgeführt, bei denen sich eine Scheibe mit einer definierten Oberflächenladungsdichte direkt im Anschluss an die Auflagefläche der Sonde befand [Kra00].

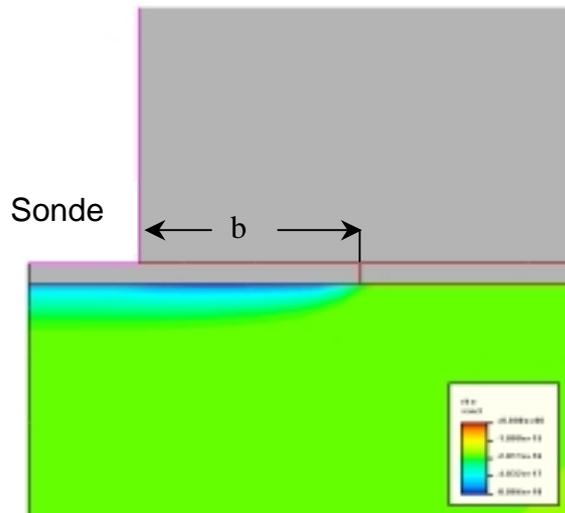


Bild 75: Simulationsmodell (in Zylindergeometrie) einer Ladungsscheibe der Breite b um die Sonde (Bildgröße ca. $250 \times 250 \text{ nm}$).

Während der Betrag der Ladungsdichte konstant gehalten wurde ($2.2 \cdot 10^{12} \text{ cm}^{-2}$), wurde die Breite b der Ladungsscheibe schrittweise variiert (Bild 75) und für jede Breite eine C-V-Kurve aufgenommen. Die Ergebnisse sind in der Abbildung 76 wiedergegeben, und zwar einmal für positive (Bild 76, rechts) und einmal für negative (Bild 76, links) Oberflächenladungen.

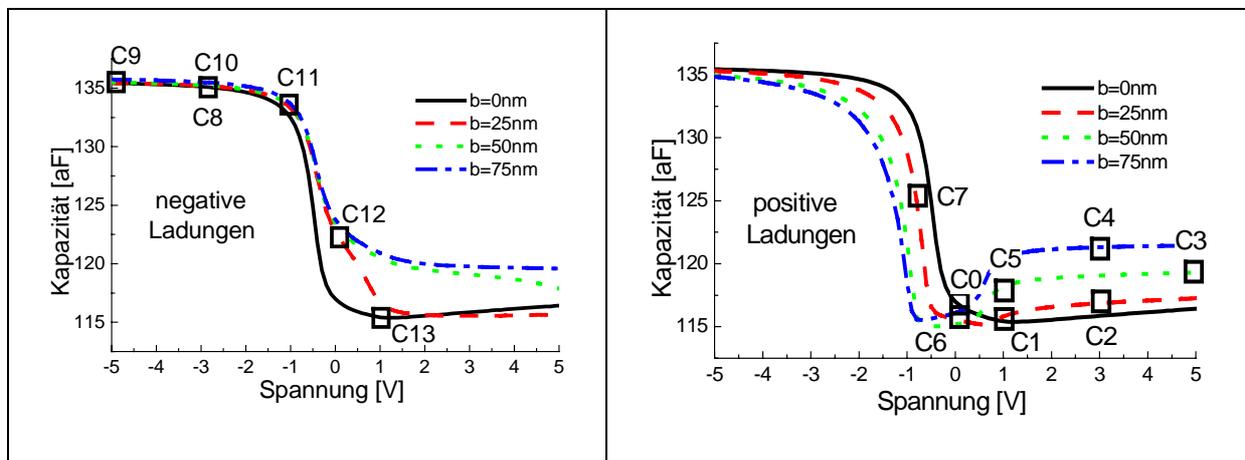


Bild 76: Simulationen des Einflusses einer Ladungsscheibe auf die C-V-Kennlinien einer homogenen p -Dotierung (10^{16} cm^{-3}). a) Links bei einer Oberflächenladungsdichte von $-2.2 \cdot 10^{12} \text{ cm}^{-2}$ und b) rechts für $+2.2 \cdot 10^{12} \text{ cm}^{-2}$ (C0, C1 usw. siehe Text).

Mit den Simulationen aus Bild 76 (mit einer Spannungs-unabhängigen Breite der Ladungsscheibe) und dem vorgestellten Diffusionsmodell ODiM sollen im Folgenden die Auswirkungen auf das SCS-Signal veranschaulicht werden. Aus den C-V-Kurven mit einer konstanten Ladungsscheibe (Bild 76) soll eine C-V-Kurve mit einer veränderlichen Ladungsscheibe konstruiert werden (Bild 77).

Das Modell zur Ausbreitung der Ladungsträger (ODiM) wurde mit dem erwähnten RWM mit einem PC simuliert. Damit konnte folgendes Verhalten der Ladungsträger festgestellt werden: Bei einem Startwert von $V_{\text{bias}}=0\text{V}$ gilt C0 (d.h. keine Ladungen auf der Oberfläche). Fahren wir die Biasspannung dann nach 1V, so tragen wir – nach ODiM – kaum Oberflächenladungen auf (C1). Erst ab 2V werden Ladungen (Ionen) aufgetragen, die sich langsam ausbreiten. Beim Zurückfahren der Spannungsrampe wird aber die Breite der Ladungsscheibe nicht abrupt verkleinert (wie unsere RWM-Simulationen gezeigt haben, siehe Bild 74), sondern die Oberflächenladungen breiten sich erst noch weiter aus, um dann dem Spannungswert mit einer gewissen Verzögerung zu folgen. Bei negativen Spannungen von kleiner -2V werden negative Ladungen aufgetragen (siehe Bild 76), die die positiven Ladungen neutralisieren und sich schließlich ebenfalls auf der Probenoberfläche ausbreiten. Damit wiederholt sich der für positive Ladungen beschriebene Vorgang mit negativen Ladungen. Fährt die Spannungsrampe schließlich wieder zurück, so erreicht die Oberflächenladungsdichte bei 1V (C13) den Wert C1. Dieser Zyklus ist in Bild 77 dargestellt.

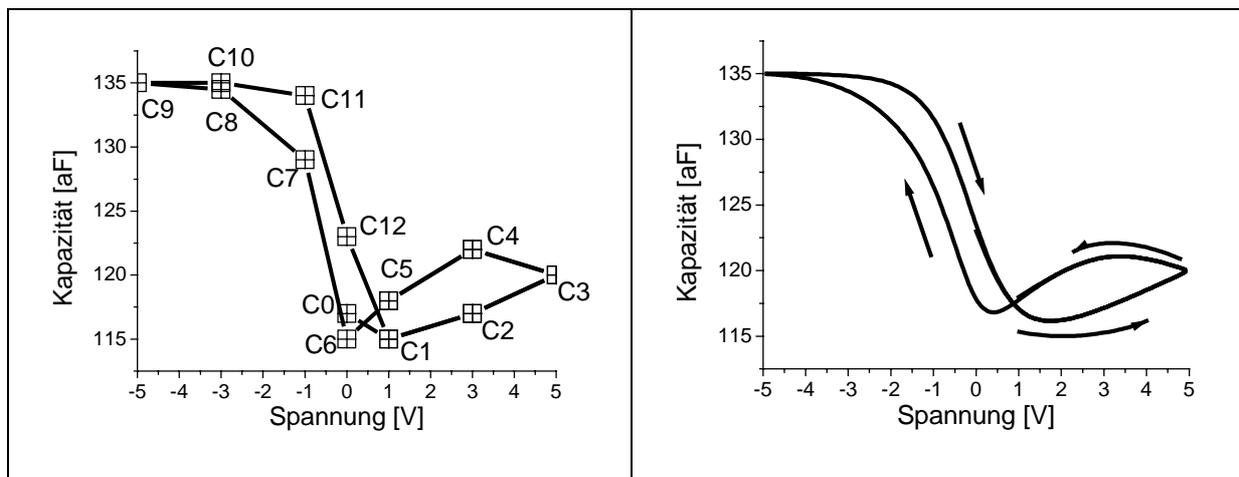


Bild 77: Einfluss von Oberflächenladungen auf die C-V-Kurve (Theorie). Links sind die konstruierten Werte und rechts ein Spline-Fit über diese Werte wiedergegeben.

Bild 78 zeigt eine experimentelle C-V-Kurve, die hervorragend durch die theoretische Kurve aus Bild 77 erklärt werden kann.

Damit ist das Ziel erreicht: die Physik hinter der Hysterese von C-V-Kurven kann modelliert und auch verstanden werden. Für die Praxis folgt daraus noch einmal die Forderung nach einer sehr guten Probenpräparation und kleinen Biasspannungen (bei $|V_{\text{bias}}|$ kleiner 2V werden nur wenige Ladungen aufgetragen).

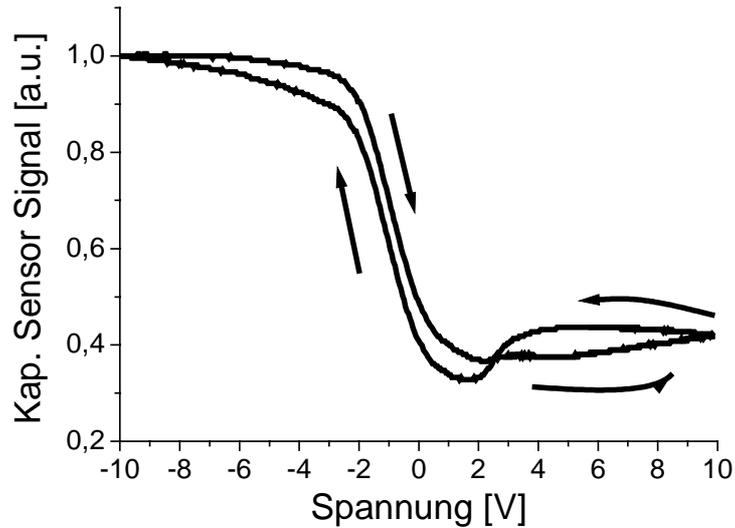


Bild 78: Einfluss von Oberflächenladungen auf die C-V-Kurve (Experiment).

Zudem sollte bei einer möglichst geringen relativen Luftfeuchtigkeit gearbeitet werden, da diese ausschlaggebend für die Breite des Meniskus und damit die Ausdehnung der Ladungsscheibe ist.

Wurde die Messspitze bei der C-V-Messung mit Druckluft angeblasen, d.h. wurde die Dicke des Wasserfilms bzw. des Meniskus reduziert, so konnte eine Reduzierung der Hysterese in der C-V-Kurve erreicht werden (Bild 79), die aber mit Abschalten der Druckluft wieder zunahm. Andere Oberflächenbehandlungen (Reinigung mit Ethanol, Isopropanol usw.) bewirkten hingegen kaum eine Verbesserung.

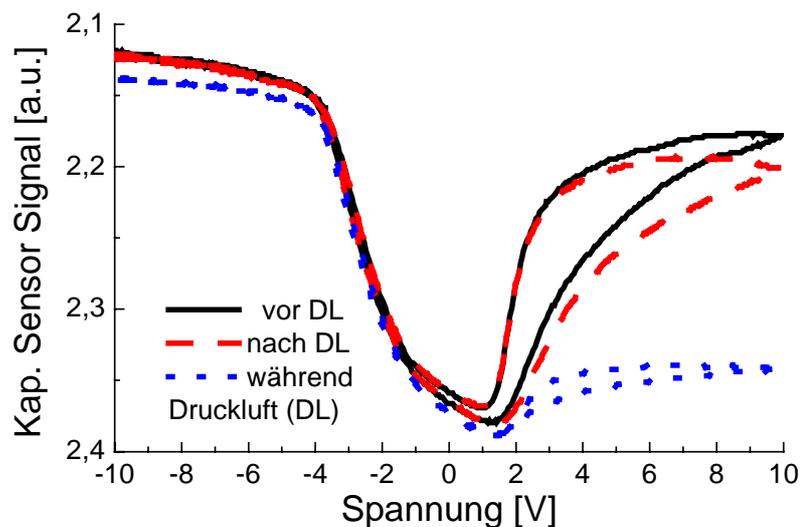


Bild 79: Drucklufteinfluss auf C-V-Messungen.

Mit SCS kann folglich die Güte der Oberflächenpräparation geprüft werden. Treten große Hysterese-Effekte auf, so sind die Messergebnisse wenig aussagekräftig, die von solch einer Probe stammen.

Die gesamte Information steckt letztendlich nicht im SCM-Bild. Sie kann nur per SCS ermittelt werden.

Auf der einen Seite ist es sinnvoll, mit $V_{\text{bias}}=-2\text{V}$ ($+2\text{V}$) im p- (n-) dotierten Bereichen zu messen, um dort die laterale Auflösung zu verbessern (siehe 2.2.1.5) und die Güte der direkten Konversion zu erhöhen (2.2.3). Auf der anderen Seite sollte mit $V_{\text{bias}}=0\text{V}$ gemessen werden um die Position eines p-n-Übergangs bestimmen zu können und eine Symmetrie zwischen p- und n-Dotierung herzustellen.

Wird z.B. ein p-n-Übergang mit unterschiedlichen Biasspannungen vermessen, so beeinflusst diese Spannung natürlich das Ladungsträgerprofil des Übergangs. Die elektrische Position, die wir mit dem SCM durch $dC/dV=0\text{V}$ bestimmen, wird aufgrund der Biasspannung verschoben, wie wir durch Simulationen ([Ise99, Kra00]) und Messungen (Bild 80) zeigen konnten.

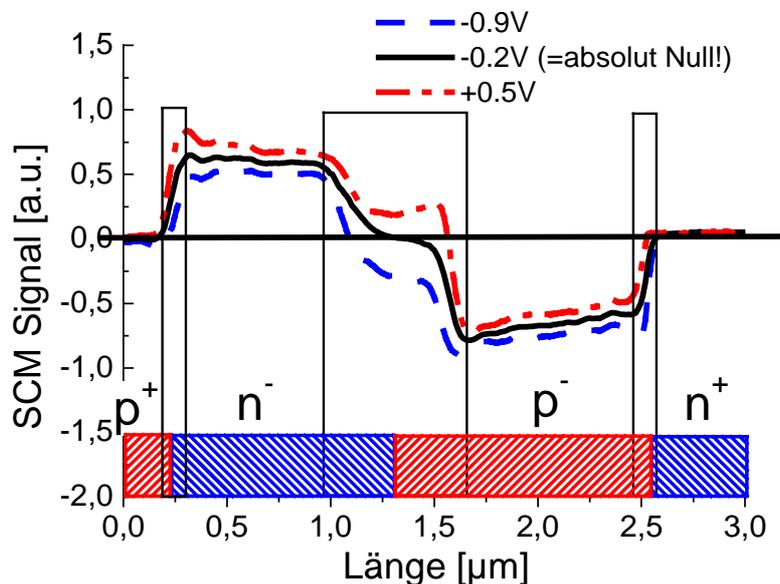


Bild 80: Experimentelles SCM-Signal eines p-n Übergangs.

Durch das Anlegen einer Biasspannung, die vom (absoluten) Nullpunkt abweicht, kann die Position um mehrere 100nm verschoben werden, wenn es sich um einen symmetrischen niedrigdotierten Übergang handelt. Zudem lässt sich nicht nur eine Spannungsabhängigkeit beobachten, sondern auch eine Abhängigkeit von der Zeit.

D.h. bei einer konstanten Biasspannung von ungleich Null, „bewegte“ sich der Übergang in den ersten Minuten. Dieses Driften ist nach unserem Modell (ODiM) verständlich, aber unerwünscht.

Aus diesem Grund wurde die Nutzung der Spektroskopie (SCS) zur Charakterisierung von p-n-Übergängen in Erwägung gezogen. Wird die Biasspannung schnell genug durchgeführt, haben die Oberflächenladungen nicht genügend Zeit, sich auf der Oberfläche auszubreiten und die Messung zu verfälschen. Da die uns zur Verfügung stehenden SCMs nicht erlaubten, automatisch ein Feld von C-V- bzw. dC/dV-V-Kurven aufzunehmen, wurden zwei Messmodi entwickelt: c-SCS (continuous) und d-SCS (discrete).

Beim c-SCS wurde die Sonde mit einer sehr langsamen Rastergeschwindigkeit kontinuierlich über die gewünschte Stelle bewegt (Linescan). Bei dieser Bewegung wurden Spannungsrampen durchgeführt und die SCS-Signale gemessen. Dieses wurde mit einem externen PC gesteuert, der mit dem eigentlichen SCM-Rechner synchronisiert war.

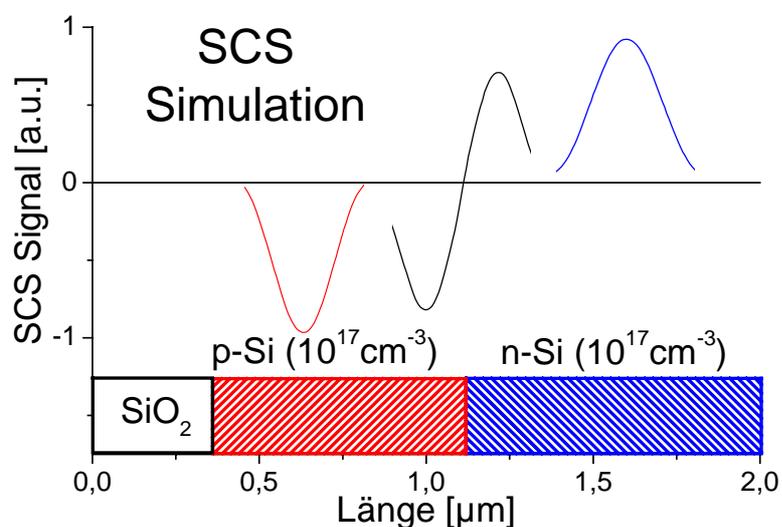


Bild 81: Simulationsergebnisse des SCS-Signals eines p-n Übergangs. Gezeigt ist je eine dC/dV-V-Kurve im homogen dotiertem Gebiet und eine dC/dV-V-Kurve direkt auf dem p-n-Übergang. Weitere Simulationen von p-n-Übergängen werden in [Ise99] und [Kra00] gegeben.

Beim d-SCS rasterete die Sonde eine Linie diskret Punkt für Punkt ab. An jedem Punkt wurden dann – mit dem externen synchronisierten PC – je zehn dC/dV-V-Kurven aufgenommen, über die automatisch gemittelt wurde. Bei der Positionierung

der Sonde wurde das SCM mit einer Makrosprache gesteuert, die für die SPM-Lithografie entwickelt wurde. Der große Nachteil des d-SCS-Modus besteht darin, dass während der d-SCS Aufnahme keine topografische Daten gewonnen werden können. Da die Steuerung auch kleinen Schwankungen (thermischer Drift) unterworfen ist, läßt die Zuordnung SCS Topografie noch Wünsche offen. Bild 82 bzw. 83 ermöglichen einen Vergleich von c- und d-SCS.

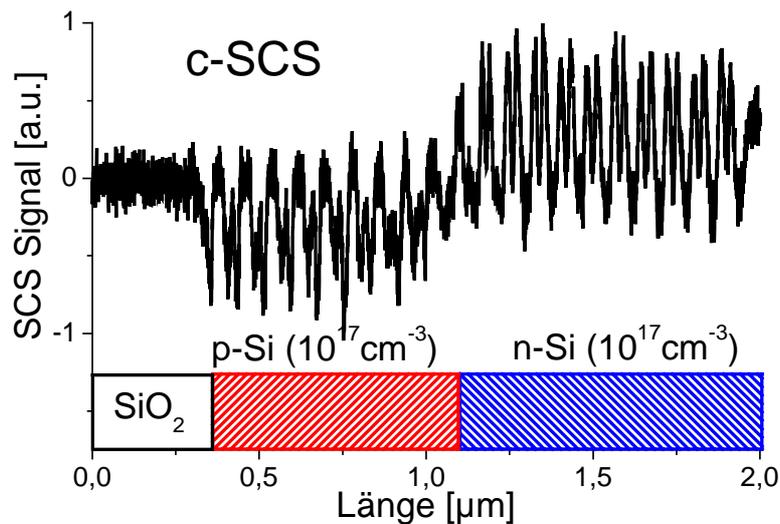


Bild 82: c-SCS-Signal eines p-n Übergangs.

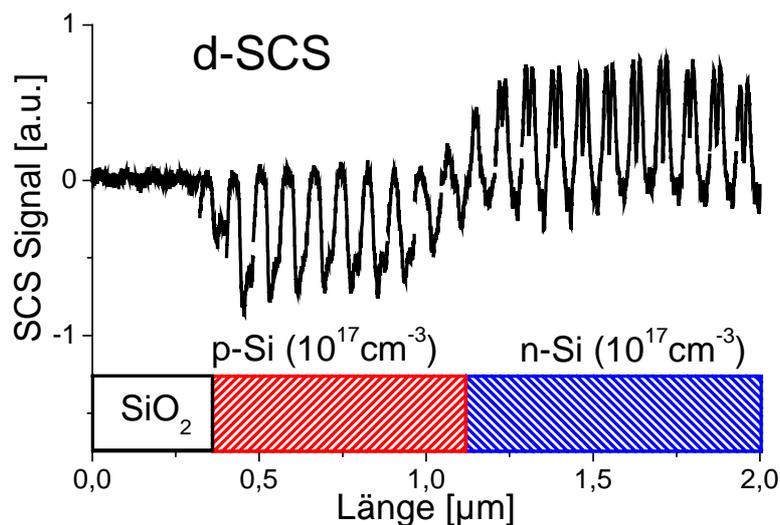


Bild 83: d-SCS-Signal eines p-n Übergangs.

Das Signal im d-SCS-Modus ist natürlich weniger verrauscht, da jede Kurve zehnfach gemittelt ist. Als problematisch beim d-SCS-Modus stellte sich die Rampengeschwindigkeit heraus. Da die externe Messkarte recht langsam war, wurde mit einer Rampengeschwindigkeit von unter 1V/s gearbeitet. Neben der thermischen Drift der Apparatur kommt es natürlich bei diesen Rampengeschwindigkeiten zu einer Ausbreitung der Oberflächenladungen. Dadurch verändern sich die dC/dV-V-Kurven mit der Zeit. Und dadurch wird das Messergebnis verfälscht.

In naher Zukunft wird die Möglichkeit gegeben sein, auf die Hardware eines SCMs direkt zuzugreifen. Dann könnte ein d-SCS-Modus realisiert werden, der neben der Topografie auch ein ganzes Feld von dC/dV-V-Kurven aufnimmt, welche mit einer hochfrequenten Spannungsrampe (10kV/s) durchgefahen werden. Dies erlaubt dann auch ein Feld von C-V-Kurven direkt aufzunehmen, denn die hohe Rampengeschwindigkeit gestattet eine Mittelung über viele Kurven bei akzeptabler Rastergeschwindigkeit. Mit diesem Modus ließe sich die volle Information über eine Probe gewinnen, ohne dass die Messkurven durch aufgebrauchte Oberflächenladungen beeinflusst würden.

2.2.3 Quantitative SCM

Ziel einer SCM-Charakterisierung eines HL-Bauelementes sind Aussagen über das zweidimensionale Dotierkonzentrations-Profil bzw. die Ladungsträgerdichte-Verteilung. Dazu muss das SCM-Signal mit der Dichte-Verteilung verknüpft werden.

$$\text{SCM-Signal} = K_{\text{sens}} \cdot dC(x,y)/dV \quad \rho(x,y)$$

Unter Verwendung dieser Ergebnisse können auch TCAD-Systeme geeicht werden (Bild 84).

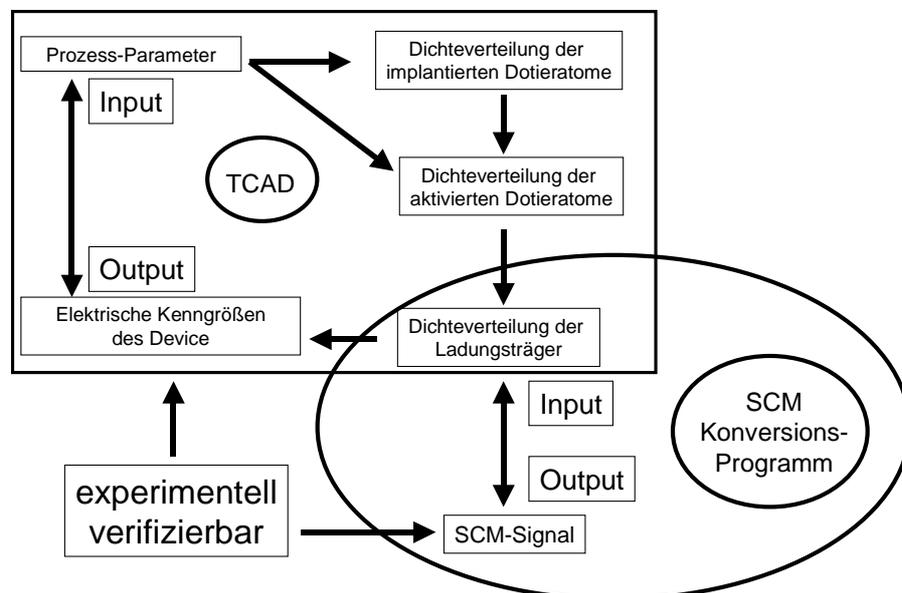


Bild 84: Einbeziehung von SCM-Daten zur Eichung eines TCAD-Systems.

In den USA (NIST: FASTC2D [Mar98, Kop98], Stanford: SUSPECT [Yu99]) wurden zu diesem Zweck große Datenbanken angelegt, die auf vielen SCM-Simulationen mit unterschiedlichen Parametern basieren. Diese sollen eine schnelle Konversion der SCM-Daten erlauben. Dabei wird wie folgt vorgegangen: An einer Eichprobe bzw. an einer hinsichtlich der Dotierkonzentration bekannten Stelle der Probe wird eine SCM-Messung durchgeführt. Diese Messung wird mit Simulations-Ergebnissen (C-V-Kurven) aus der Datenbank verglichen. Dadurch kann jedem Modell-Parameter ein konkreter Wert zugeordnet werden.

Als Parameter werden angenommen:

- 1.) die Sensitivität des SCM-Sensors K_{sens} ,
- 2.) die Oxiddicke d_{ox} ,
- 3.) die Dielektrizitätskonstante ϵ_{ox} ,
- 4.) der Spitzenradius r_{tip} ,
- 5.) der Sondenwinkel α und
- 6.) die Biasspannung V_{bias} .

Sind alle Parameter unbekannt, so gibt es natürlich viele denkbare Kombinationen (lokale Minima), um die Theorie (Datenbank) an das Experiment anzupassen. Letztendlich geschieht bei der Datenbank-Anpassung dasselbe, was in der vorliegenden Arbeit bei der Konversionskurven-Anpassung im Abschnitt 2.2.1.6 vorgeführt wurde: Mit einer Fit-Funktion mit fünf oder sechs Parametern können alle realistischen Datensätze angefitet werden. Die sich ergebenden Fit-Parameter haben aber oft nichts mit den realistischen Parametern zu tun. Eine richtige Anpassung hängt von der Güte des Fits (lokales oder globales Minimum) und der Qualität des der Fitfunktion bzw. der Datenbank zugrundeliegenden Modells ab. Z.B. modellieren die oben aufgeführten Forschergruppen die Sonde als eine kugelförmige Spitze (Kugelradius r_{tip}) und beobachten deshalb in ihren Simulationen keine große Abhängigkeit der SCM-Messung von der Sondengeometrie bzw. r_{tip} (siehe z.B. [Kop98]). Dies ist aber definitiv falsch, wie alle unsere Experimente und Simulationen mit einem realistischen Sondenmodell gezeigt haben. Zudem beruhen die Datenbanken auf (theoretischen) Kleinsignal HF-C-V-Kurven und berücksichtigen nicht die große Messamplitude.

Somit wird mit dieser Anpassung lediglich eine Konversionskurve (SCM-Signal Dotierkonzentration) bestimmt, deren Aussagekraft von der Anzahl der verwendeten Datenpunkte abhängt (siehe Spline-Fit in 2.2.1.6).

Mit dieser Konversionsfunktion (bestimmter Datensatz aus der Datenbank) kann sodann das SCM-Signal einer unbekannt Probe konvertiert werden. Hierbei ist anzumerken, dass die Probe Punkt für Punkt konvertiert wird und damit die Umgebung nicht berücksichtigt wird. Es können nur Proben mit sehr flachen Gradienten konvertiert werden (nur Stanford [Yu99] berücksichtigt ansatzweise Gradienten, konnte allerdings noch keine p-n-Übergänge konvertieren). Typischerweise wird als Maß für flache Gradienten angegeben, dass sich die Dotierung unter der Sonde nicht mehr als 5% ändern darf. Die Ergebnisse, die im Rahmen dieser Arbeit durchgeführt wurden, vermitteln ein genaueres Bild: Demnach werden richtige

Resultate erzielt, wenn die Größe der homogenen Dotiergebiete – je nach Dotierung – unterhalb der Kurven aus dem Abschnitt 2.2.1.5 (laterale Auflösung) liegt. Zudem muss zwischen reinen p- bzw. n-Dotiergebieten und p-n-Übergängen unterschieden werden.

Im Rahmen dieser Arbeit wurde keine Datenbank erstellt, da meiner Ansicht nach die Klärung der Physik hinter der Messmethode Vorrang hatte. Erst danach kann ein Modell für die Simulation und darauf aufbauend eine Datenbank angelegt werden.

Für die direkte Konversion wird, wie erwähnt, eine Konversionsfunktion benötigt. Diese wird in der vorliegenden Arbeit durch eine Eichprobe gewonnen. Dabei ist zu beachten, dass die Eichprobe und die zu untersuchende Probe (Testprobe) mit derselben Probenpräparation behandelt werden müssen, um eine Vergleichbarkeit zu garantieren. Zudem sollte die globale Geometrie von Eich- und Testprobe gleich sein, da die Streukapazität die Sensitivität des Kapazitätssensors beeinflusst. Bzgl. der Geometrie ist zwischen Lift-Off- und Querschnitts-Präparation zu unterscheiden (siehe Abschnitt 2.2.1.3). Für beide Fälle sollten Eichproben vorhanden sein. Nach der Vermessung der entsprechenden Eichprobe können dann die Werte angefitet werden (siehe Abschnitt 2.2.1.6). Als Resultat wird die entsprechende Konversionsfunktion erhalten. Damit verlangt dieser Weg gegenüber der Datenbank-Anpassung weniger Informationen über die zu untersuchende Probe, stellt aber höhere Ansprüche bzgl. einheitlicher Probenpräparation.

Im Folgenden wird, ausgehend von einer Konversionsfunktion, die Güte der direkten Konversion (Konversion der SCM-Daten mit einer Konversionsfunktion, die über Messungen von homogenen Dotiergebieten bestimmt wurde) bei einer Anwendung auf Dotierprofile mit unterschiedlichen Gradienten – einem Gaußprofil und einem p-n-Übergang – untersucht.

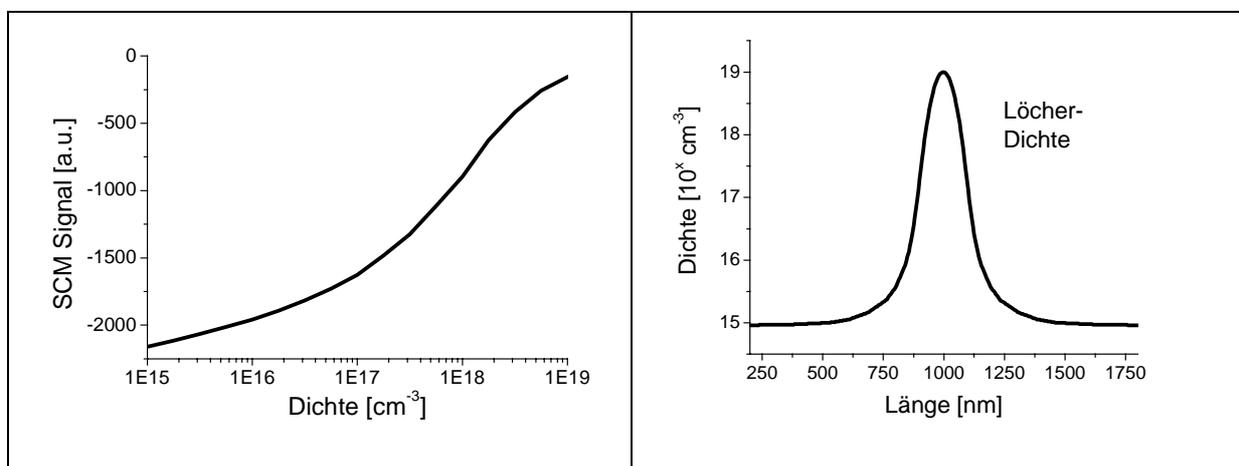


Bild 85: Konversionsfunktion (links) und Ladungsträger-Profil (rechts) des Gaußgradienten.

Dabei wird „invers vorgegangen“, d.h. dass aus einem Ladungsträger-Profil das SCM-Signal einmal mittels Simulation (Dauer: 10-20h) und einmal mittels Konversionsfunktion (Dauer: 1s) berechnet und anschließend verglichen wird.

Die verwendete Konversionsfunktion und das Ladungsträger-Profil sind in Bild 85 dargestellt. Die Breite des Profils beträgt bei einer Konzentration von 10^{18}cm^{-3} nur 125nm und stellt damit auch hohe Ansprüche an die zu erreichende laterale Auflösung.

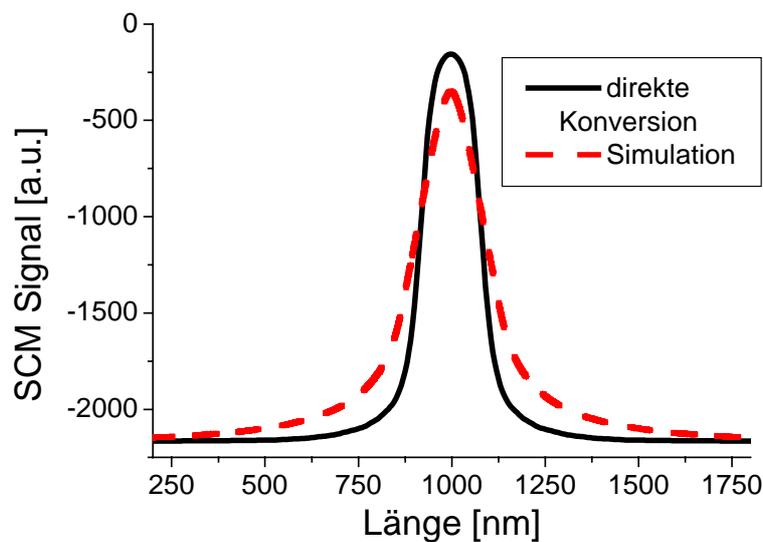


Bild 86: SCM-Signal des Gaußgradienten.

In Bild 86 können die durch die Konversionsfunktion und die Simulation erhaltenen SCM-Signale verglichen werden. Damit wird noch einmal deutlich, dass das SCM über eine bestimmte Breite mittelt. Die direkte Konversion liefert ein viel zu schmales Signal vom Gradienten. Wird aus dem Simulationssignal mittels Konversionsfunktion das Ladungsträger-Profil bestimmt, wird dieses folglich zu großen Abweichungen von der realen Ladungsträgerdichte führen.

Auf die gleiche Weise wurde die Qualität der direkten Konversion an einem p-n-Übergang geprüft (Bild 87). Dazu musste die Konversionsfunktion für beide möglichen Ladungsträgertypen (Elektronen und Löcher) erweitert werden. Der Abfall der Funktion unterhalb 10^{15}cm^{-3} ist bedingt durch die Eigenschaften des Kapazitätsensors (siehe Abschnitt 2.2.1.1).

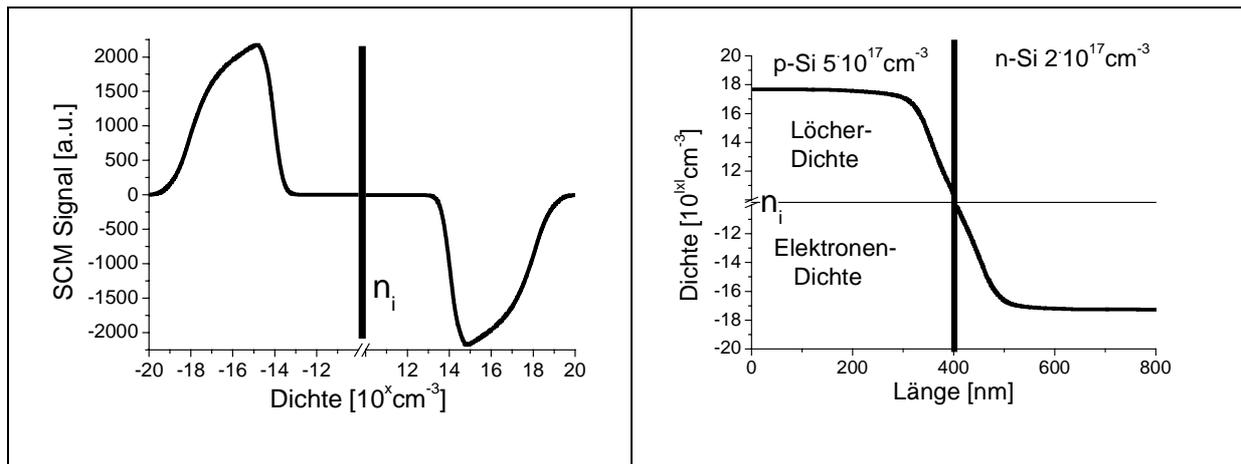


Bild 87: Konversionsfunktion (links) und Ladungsträger-Profil (rechts) des p-n Überganges.

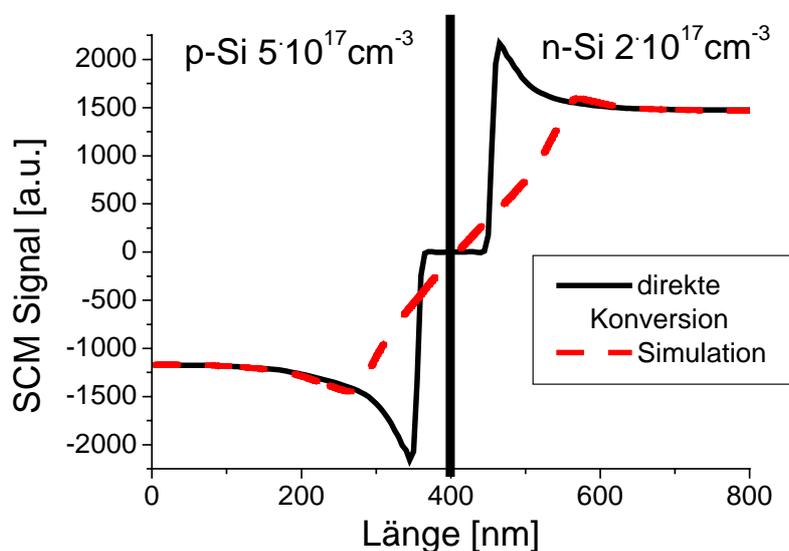


Bild 88: SCM-Signal des p-n-Übergangs.

Innerhalb des p-n-Übergangs weichen die direkte Konversion und die Simulation stark von einander ab.

Beide Teststrukturen zeigen deutlich die Grenzen der direkten Konversion auf. Sie kann somit nur einen ersten Schritt in der SCM-Analyse bilden. Die Bereiche mit einem großen Gradienten – die für die HL-Analytik interessanten Bereiche – müssen mit einer anderen Methode konvertiert werden.

Eine Möglichkeit, im Folgenden als Datenbank-Konversion (DAB-K) bezeichnet, wäre der Aufbau einer großen Datenbank, in die eine Vielzahl von Simulationen unterschiedlicher Gauß-Profile und p-n-Übergänge einfließen müsste. Da nach der direkten Konversion einer SCM-Messung schon viele Informationen vorliegen (Ausgangsdotierung, Breite und Position des p-n-Übergangs), könnte eine weitere Anpassung an das reale Profil per DAB-K relativ einfach vorgenommen werden. D.h.

das Programm würde aus der Datenbank die entsprechenden Übergänge heraussuchen und mit der tatsächlichen Messung vergleichen, um so zum realen Profil zu gelangen. Für die Konversion einer Messlinie wäre die Realisierung von DAB-K noch mit einem relativ abschätzbaren Zeitumfang zu bewältigen. Die Entwicklung eines Programms für eine vollautomatische Konversion einer 2D-SCM-Messung wäre allerdings sehr aufwändig. Aus diesem Grund wurde noch nach einer weiteren Möglichkeit gesucht, SCM-Daten in Dotierprofile zu konvertieren.

Die im Rahmen dieser Arbeit entwickelte physikalische Konversionsmethode (PHY-K) basiert auf der inversen Vorgehensweise und dem schon mehrfach erwähnten Mittelungsprozess bei einer SCM-Messung. Ausgangspunkt ist ein angenommenes Ladungsträger-Profil, aus dem mit der Konversionsfunktion das SCM-Signal ausgerechnet wird. Anschließend wird dotierungs-selektiv gemittelt. D.h. je nach lokaler Höhe der Dotierung bzw. Ladungsträgerdichte wird über einen mehr oder weniger großen Bereich in der Umgebung gemittelt. Die Funktion, welche die Dotierungskonzentration mit der Breite verknüpft, über die gemittelt wird, ist identisch mit der Funktion der lateralen Auflösung (siehe Abschnitt 2.2.1.5). Nach einer Konversion eines angenommenen Ladungsträger-Profils mit PHY-K kann dieses Resultat mit den experimentellen Ergebnissen verglichen werden. Durch eine Parametrisierung des Profils könnte eine einfache Änderung desselben und damit eine Anpassung von PHY-K und Experiment automatisch realisiert werden.

Einen Eindruck der Stärke von PHY-K liefern die Bilder 89 und 90. Hier wurde das SCM-Signal aus den Ladungsträgerdichte-Profilen des Gaußgradienten (Bild 85) und des p-n-Übergangs (Bild 87) einmal mittels PHY-K und einmal mittels Simulation berechnet. Beide Kurven stimmen (im Rahmen der Messgenauigkeit des SCMs) hervorragend überein. Eine noch bessere Übereinstimmung konnte bei lateralen MOS-Strukturen – wie sie z.B. bei einer Querschnitts-Analyse eines MOS-FETs vorkommen – erzielt werden. Bei stark asymmetrischen p-n-Übergängen war die Übereinstimmung nicht ganz so zufriedenstellend, denn neben der Mittelung durch die Messspannung müsste auch die Änderung des Ladungsträgerdichte-Profils durch die Biasspannung ins Modell einbezogen werden. PHY-K geht aber von der unbeeinflussten Ladungsträger-Verteilung aus. Die besten Ergebnisse werden dementsprechend bei einer Konversion bei der Flachbandspannung (= kein Spannungsabfall zwischen Spitze und Probe) erzielt.

Ein großer Vorteil von PHY-K ist die relativ einfache Erweiterung dieser Methode auf 2D Profile.

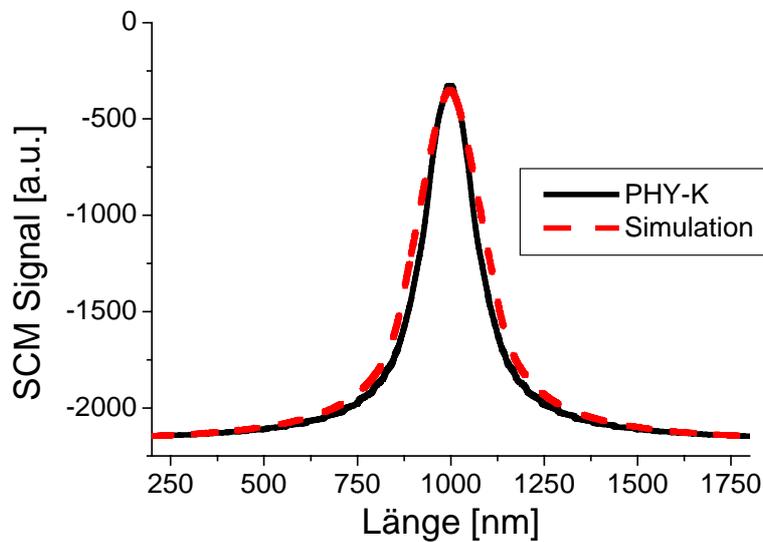


Bild 89: SCM-Signal des Gaußprofils.

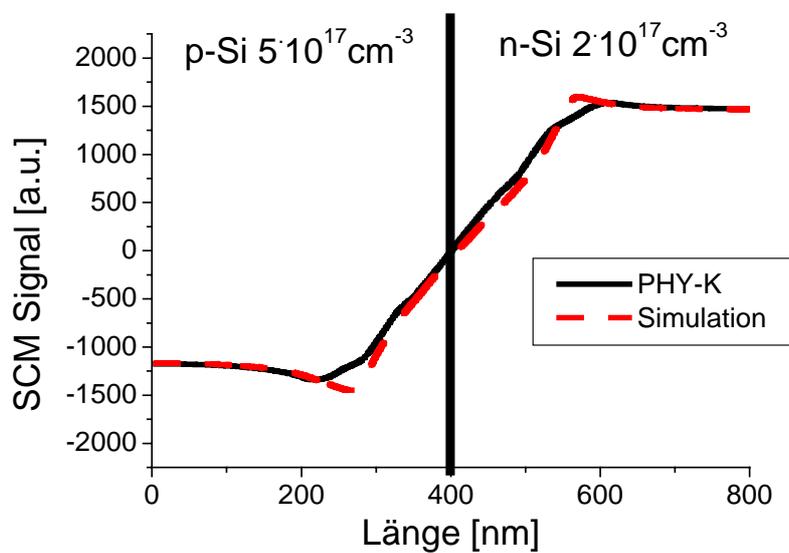


Bild 90: SCM-Signal des p-n Überganges.

Insgesamt müssen sicherlich noch weitere Untersuchungen zur Güte dieser Methode durchgeführt werden, ehe z.B. eine Datenbank auf diesem Modell aufgebaut wird.

2.3 Zusammenfassung und Ausblick

In der IC-Analyse werden dringend Analysemethoden mit einer hohen lateralen Auflösung benötigt, die schnell und einfach – d.h. unter Umgebungsbedingungen – einzusetzen sind. Dies trifft auf die Rastersondenmethoden prinzipiell zu, allerdings müssen diese Methoden noch weiterentwickelt werden. Erst wenn ein hinreichend gutes physikalisches Modell der entsprechenden SPWW existiert, können zuverlässige Aussagen über die laterale Auflösung, die Auflösung der SPWW und die Möglichkeiten einer Quantifizierung gemacht werden. Für zuverlässige Analysen muss außerdem die entsprechende Probenpräparation entwickelt werden, so dass bei jeder Messung von einer definierten Oberfläche ausgegangen werden kann. Das gleiche gilt für die Sonde: Auch diese muss eindeutig z.B. anhand einer Teststruktur charakterisiert werden können. Am Schluss solcher Untersuchungen steht die Entwicklung von Methoden, die aus der SPWW das Signal ausrechnen, das eigentlich von Interesse ist (z.B. die Stromstärke aus dem MFM-Signal von einer Leiterbahn). Wie solch eine konkrete Weiterentwicklung einer Rastersondenmethode für die IC-Analyse aussehen kann, wurde beispielhaft an der Rasterkapazitätsmikroskopie vorgeführt.

Mit der Rasterkapazitätsmikroskopie soll ein 2D-Dotierprofil eines ICs mit einer lateralen Auflösung von unter 50nm gewonnen werden. Wichtig für das Verständnis der SPWW war die Analyse des Kapazitätssensors (2.2.1.1) und die Untersuchung des Einflusses der Sondengeometrie (2.2.1.2). Als entscheidend für zuverlässige Messungen stellte sich die Probenpräparation (2.2.1.3) heraus. Da wir die Probe auf Nanometer-Skala untersuchen wollen, muss sie auch auf Nanometer-Skala definiert präparierbar sein. Die wichtigsten Präparationsschritte sind die Reduzierung der Rauigkeit (unter 5nm_{RMS}) und das Aufbringen eines hochqualitativen Oxides mit einer definierten Dicke von 3nm. Letzteres gelang mit einem speziell entwickelten UV-Ozon-Oxidationsofen.

Konkrete Aussagen konnten bzgl. der Auflösung dieses Verfahrens, einmal hinsichtlich der Dotierung (2.2.1.4) und einmal bezüglich der Lokalisierung (2.2.1.5), getroffen werden. Auch Methoden zur Eichung des Systems wurden entwickelt.

Neben diesen grundlegenden Untersuchungen wurden zahlreiche Messungen an Proben aus der HL-FA durchgeführt, von denen ein Teil in 2.2.1.7 gezeigt wird. Sie

belegen, dass das SCM eine wichtige Ergänzung zu den etablierten Analysemethoden für die HL-FA darstellt.

Letztlich konnten alle experimentellen Ergebnisse durch physikalische Modelle erfasst und mit darauf beruhenden Simulationen verifiziert werden. Sogar die Hysterese bei den Messungen konnte auf ein Modell (Ladungsträger auf der Probenoberfläche) zurückgeführt und mittels eines einfachen Programms qualitativ erklärt werden (2.2.2).

Den Abschluss dieses Abschnittes bildete die Entwicklung einer Methode zur Bestimmung eines 2D-Dotierprofils anhand des gemessenen SCM-Signals. Erstmals konnte damit sogar ein p-n-Übergang konvertiert werden (2.2.3).

Das Ziel für die Zukunft stellt der alltägliche Einsatz des SCMs in der industriellen HL-FA dar. Hierzu ist der Ausbau der vorgestellten Konversionsmethoden weiter voranzutreiben, so dass eine vollautomatische Konversion innerhalb weniger Minuten durchgeführt werden kann. D.h. einer der nächsten Schritte bildet der Aufbau einer entsprechenden Datenbank.

Um laterale und Dotierungs-Auflösung zu steigern, muss der Kapazitätssensor weiterentwickelt werden. Nur mit einer kleineren Messspannung und einer größeren Sensitivität sind die gewünschten Fortschritte zu erzielen.

Eine interessante Möglichkeit wäre der Einsatz von Koaxial-Sonden, die ebenfalls eine höhere laterale Auflösung versprechen.

Daneben wäre die Erforschung der Eigenschaften von Oberflächenladungen auf Siliziumdioxid von grundlegendem Interesse. Obwohl in diesem Oxid der Schlüssel für die gesamte MOSFET- und damit IC-Technologie liegt, ist über die atomare Struktur der Oberfläche unter Umgebungsbedingungen nur wenig bekannt. Gerade die Untersuchung des Drei-Phasensystems (Festkörper, Flüssigkeit und Dampf), das der Realität entspricht, stellt hinsichtlich seiner Komplexität eine große Herausforderung an die Grundlagenforscher dar.

Somit zeigt sich die SCM-Sonde-Probe-Wechselwirkung als ein höchst nicht-triviales Nanosystem, bei dem erst eine Kombination von Nanochemie und Nanophysik ein adäquates Verständnis des Ganzen ermöglicht.